Nonparametric and Semiparametric Methods in Medical Diagnostics

by
Eunhee Kim

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill
2009

Approved by:

Advisor, Dr. Donglin Zeng

Co-Advisor, Dr. Joseph G. Ibrahim

Reader, Dr. Jianwen Cai

Reader, Dr. David J. Couper

Reader, Dr. Gerardo Heiss

ABSTRACT

EUNHEE KIM: Nonparametric and Semiparametric Methods in Medical Diagnostics
(Under the direction of Drs. Donglin Zeng and Joseph G. Ibrahim)

In medical diagnostics, biomarkers are used as the basis for detecting or predicting disease. There has been an increased interest in using the Receiver Operating Characteristic (ROC) curve to assess the accuracy of biomarkers. In many situations, a single biomarker is not sufficient for the desired level of accuracy; furthermore, newly discovered biomarkers can provide additional information for a specific disease. Even though numerous methods have been developed to evaluate a single biomarker, few statistical methods exist to accommodate multiple biomarkers simultaneously. The first paper proposes a semiparametric transformation model for multiple biomarkers in ROC analysis to optimize classification accuracy. This model assumes that some unknown and marker-specific transformations of biomarkers follow a multivariate normal distribution; it incorporates random effects to account for within-subject correlation among biomarkers. Nonparametric maximum likelihood estimation is used for inference, and the parameter estimators are shown to be asymptotically normal and semiparametrically efficient. The proposed method is applied to analyze brain tumor imaging data and prostate cancer data.

In the second paper, we focus on assessing the accuracy of biomarkers by adjusting for covariates that can influence the performance of biomarkers. Therefore, we develop an accelerated ROC model in which the effect of covariates relates to rescaling the original ROC curve. The proposed model generalizes the usual accelerated failure time model in the survival context to the ROC analysis. An innovative method is developed

to construct estimating equations for parameter estimation. The bootstrapping method is used for inference, and the parameter estimators are shown to be asymptotically normal. We apply the proposed method to data from a prostate cancer study.

The paired-reader, paired-patient design is commonly used in reader studies when evaluating the diagnostic performance of radiological imaging systems. In this design, multiple readers interpret all test results of patients who undergo multiple diagnostic tests under study. In the third paper, we develop a method to estimate and compare accuracies of diagnostic tests in a paired-reader, paired-patient design by introducing a latent model for test results. The asymptotic property of the proposed test statistics is derived based on the theory of U-statistics. Furthermore, a method for correcting an imperfect gold standard bias and sample size formula are presented. The proposed method is applied to comparing the diagnostic performance of digital mammography and screen-film mammography in discriminating breast tumors.

ACKNOWLEDGMENTS

I would like to express the deepest appreciation to my advisor, Dr. Donglin Zeng, for his insightful guidance, ideas, and support throughout the research. I am very grateful to Dr. Lloyd Chambless, my mentor, my supervisor at the Collaborative Studies Coordinating Center for his wonderful guidance and constant support. I would also like to thank my co-advisor, Dr. Joseph Ibrahim, and committee members, Dr. Jianwen Cai, Dr. David Couper, and Dr. Gerardo Heiss for their comments. Very special thanks to Dr. Amy Herring and Dr. Fei Zou for their support and help. My heartfelt gratitude to Dr. Sinae Kim and Dr. Jungwon Mun for caring, their invaluable advice and help. I want to thank my loving parents for their support and encouragement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Literature Review

1.1 Measures of Diagnostic Accuracy

The accuracy of a diagnostic test is the ability of a test to discriminate among alternative status of health (Zweig and Campbell, 1993). The assessment of the performance of the diagnostic test is done by investigating whether test results differ for the two health states. The Receiver Operating Characteristic (ROC) analysis are widely used for the measure of performance, evaluation of diagnostic or prognostic tests and indices, and comparison of diagnostic techniques or systems. In essence, ROC analysis is an evaluation technique used in signal detection theory developed in the 1950s and 1960s (Green and Swets, 1966; Egan, 1975). It was popular in the field of radiology in the 1980s and it is increasingly used for medical and image research in recent years.

1.1.1 Sensitivity and Specificity

For binary results (such as *positive* or *negative*), the accuracy of a diagnostic test is often characterized by the true positive rate (TPR) and the false positive rate (FPR). Positive test results indicate the presence of a particular condition and negative indicates its absence. Then, the TPR is defined as the probability that the test result is positive given that the subject is truly diseased. The FPR is defined as the probability that

the test result is positive given that the subject is truly non-diseased. In biomedical research, the sensitivity and specificity are often used instead of the TPR and the FPR;

$$TPR = Pr(positive \mid disease) = sensitivity$$
$$FPR = Pr(postitive \mid nondisease) = 1 - specificity.$$

Let D denote true disease status where D=1 if the condition is present and 0 if the condition is absent and Y be the binary test result (Y=1 for a positive, 0 for a negative test result). Then,

$$TPR = Pr(Y = 1 \mid D = 1)$$

$$FPR = Pr(Y = 1 \mid D = 0).$$

1.1.2 Receiver Operating Characteristic (ROC) Curve

Many diagnostic tests results are not simply positive or negative but are measured on continuous or ordinal scales. Some tests yield qualitative results on an ordinal scale based on a subjective assessment of readers. For instance, for the mammography study of the detection of malignant lesions, readers can give a BIRAD score (American College of Radiology, 1995) where 1=normal, 2=benign, 3=probably benign, 4=suspicious, and 5=malignant. On the other hand, diagnostic tests such as temperature, serum cholesterol, and blood pressure produce continuous test results. In this thesis, we focus our attention on test results with continuous or ordinal scales and their analyses. When a diagnostic test is based on a variable measured on a continuous or ordinal scale, an assessment of the test can be made through the use of the receiver operating characteristic (ROC) curves (Hanley and McNeil, 1982; Metz, 1978).

The ROC curve is defined as a plot of TPR (sensitivity) versus FPR (1-specificity)

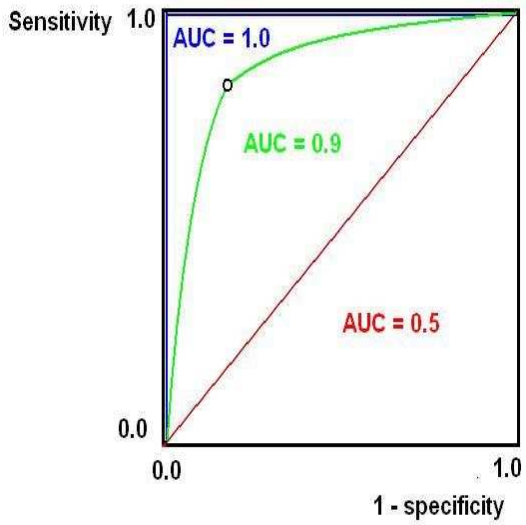

Figure 1.1: ROC curve

across all possible threshold values. Each point on the graph is created by a different threshold value (See Figure 1.1). From Figure 1.1, we can see that the FPR (1-specificity) increases as the TPR (sensitivity) increases. Thus, the ROC curve shows the range of possible tradeoffs between sensitivity and specificity.

Some basic properties of the ROC curve are:

a. The ROC curve is a monotonic increasing function mapping (0,1) onto (0,1).

b. The closer the curve follows the left and upper border of the ROC space, the more accurate the test.

c. On the other hand, the closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.

d. The ROC curve is invariant to any monotone transformations of the measurement scale. Thus, the ROC curve does not depend on the scale of the test measurements, making it useful for comparing diagnostic tests of different scales.

1.1.3 The Area Under the ROC Curve

It is often convenient to reduce an ROC curve to a single quantitative measure. A commonly used index of accuracy is the area under an ROC curve (AUC). It reflects the discriminative ability of a diagnostic procedure and can be used to make inferences for comparing ROC curves. The area can take values between 0 and 1, but typically it ranges from 0.5 to 1. The closer AUC is to 1, the better the overall diagnostic performance of the test. Conversely, an AUC of 0.5 indicates that the test is performing no better than simply guessing whether a sample is normal or abnormal, and an AUC of 1 indicates a test always classifies a sample correctly. The area under an ROC curve has several interpretations (Zou et al., 2002, p.28); (a) the average value of sensitivity for all possible values of specificity, (b) the average values of specificity for all possible values of sensitivity (Metz, 1986, 1989), and (c) the probability that a randomly selected patient with the condition has a test result indicating greater suspicion than that of a randomly chosen patient without the condition (Hanley and McNeil, 1982).

An alternative summary measure is the partial area under the ROC curve (pAUC), used to make statistical inference when only a region of the ROC space is of interest. Methods for estimating and comparing pAUCs are found in the literature (McClish, 1989; Wieand et al., 1989; Zhang et al., 2002).

1.1.4 The ROC Curve for Continuous Tests

Let D be a binary indicator of disease status with D=1 for diseased and D=0 for non-diseased subjects. Let Y denote a continuous rest result and c be a threshold that any test results greater than c are considered to be positive. For a given threshold c,

$$TPR(c) = Pr(Y \geq c \mid D = 1)$$
$$FPR(c) = Pr(Y \geq c \mid D = 0).$$

The ROC curve is the entire set of possible true and false positive rates by dichotomizing Y with different thresholds. That is,

$$ROC(\cdot) = \{(FPR(c), TPR(c)),\ c \in (-\infty, \infty)\}.$$

We also write the ROC curve as (Pepe, 2003, p.68)

$$ROC(\cdot) = \{(t, ROC(t)),\ t \in (0, 1)\},$$

where the ROC function maps t to $TPR(c)$, and c is the threshold corresponding to $FPR(c) = t$.

Let S_D and $S_{\bar{D}}$ denote the survivor functions for Y in the diseased and non-diseased populations: $S_D(y) = Pr(Y \geq y \mid D = 1)$ and $S_{\bar{D}}(y) = Pr(Y \geq y \mid D = 0)$. Then, the ROC curve can be expressed as

$$ROC(t) = S_D(S_{\bar{D}}^{-1}(t)), \quad t \in (0, 1).$$

Suppose that Y_D and $Y_{\bar{D}}$ are independent and randomly chosen test results from the diseased and non-diseased population, respectively. By the definition, $AUC = \int_0^1 ROC(t)dt = \int_0^1 S_D(S_{\bar{D}}^{-1}(t))dt$. We can easily show that

$$AUC = P(Y_D > Y_{\bar{D}}).$$

1.1.5 The Binormal ROC Curve

The most popular parametric model is the binormal ROC curve which assumes that test results are normally distributed in the diseased and non-diseased populations (Dorfman and Alf, 1969).

Let X be the test result from diseased subjects and Y be the test result from non-diseased subjects. Suppose that test results are normally distributed in the diseased and non-diseased populations. If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, then

$$ROC(t) = \Phi(a + b\Phi^{-1}(t))$$

where

$$a = \frac{\mu_X - \mu_Y}{\sigma_X},\ b = \frac{\sigma_Y}{\sigma_X}$$

and Φ denotes the standard normal cumulative distribution function.

The area under the curve for the binormal ROC curve has a following closed-form expression:

$$AUC = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right)$$

Note that the shape of the binormal ROC curve is fully characterized by two parameters; the intercept a (standardized mean difference of the distributions of test results) and the slope b (the ratio of the standard deviation of the distributions of the test results). Thus, statistical inference can be made based on the estimated parameters of a and b.

1.1.6 The ROC Curve for Ordinal Tests

Some test responses are collected on ordinal scales. For example, for the mammography data, the test results are numbered from 1=normal to 5=malignant. With ordinal test results, the unobservable latent continuous random variable model is often used and a ROC curve can be obtained by exploiting the latent variable.

Assume that there is an unobserved latent variable L corresponding to the assessor's perception of the image. The reader has decision threshold values that correspond to

his/her classification. Let Y denote the reported classification. Then,

$$Y = y \quad \Leftrightarrow \quad c_{y-1} < L < c_y, \quad y = 1, 2, \cdots, p$$

where $c_0 = -\infty$ and $c_p = \infty$. The reader classifies the image in the yth category if L falls within the interval corresponding to the reader's implicit definition for the yth category (c_{y-1}, c_y).

Our interest is in the ROC curve for L, the latent variables. Since $Y \geq y$ corresponds to $L > c_{y-1}$, we can express the true and false positive rates as $\text{TPR}(c_{y-1})$ and $\text{FPR}(c_{y-1})$ based on the threshold c_{y-1} for L. The set of $P + 1$ points from the ROC curve for L are identifiable from the distributions of the observed Y in diseased and non-diseased subject:

$$\{FPR(c_{y-1}),\ TPR(c_{y-1}), \quad y = 1, 2, \cdots, p+1\}$$

Suppose that Y_D denotes a test result from the diseased subjects and $Y_{\bar{D}}$ denotes a test result from the non-diseased subjects. For the discrete test results, the area under the ROC curve is given by

$$AUC = P(Y_D > Y_{\bar{D}}) + \frac{1}{2} Pr(Y_D = Y_{\bar{D}}).$$

1.2 The Study Design in Medical Diagnostics

1.2.1 Scale of the Test Result

The diagnostic test is used in order to classify subjects as diseased or not diseased. Test results can yield binary, ordinal, or continuous scales. The binary test result is either positive or negative: positive if a disease is present or negative if a disease is absent.

For test results on ordinal or continuous scales, the classification rule is usually set by a threshold, with results above it classified as positive for disease and results below it classified as negative, or vice versa. Tests that involve subjective assessments are often measured on ordinal scales.

1.2.2 Selection of Study Subjects

Enrollment into a study of a diagnostic test usually proceeds in one of two ways. Subjects can be selected on the basis of known true disease status. That is, a fixed number of diseased and non-diseased subjects are selected and then the diagnostic test is applied to the subjects. This design is called a case-control study. Alternatively, the diagnostic test can be applied to a set of study subjects from the population of interest and true disease status is determined for them. This design is called a cohort study because membership in the cohort is the basis for selection into the study.

1.2.3 Study Design for Comparing Tests

When multiple diagnostic tests are to be compared, a paired-reader(or patient) or unpaired-reader(or patient) design can be considered. In a paired-reader design, multiple readers interpret the results of all tests. For example, each reader interprets the results of both CT and MRA in this setting. On the other hand, in an unpaired-reader design, different readers interpret the results of different tests. Under this setting, for example, readers who interpret the CT results are not the same readers who interpret the MRA results. The paired-reader design is more powerful than the unpaired-reader design since it requires fewer patients and readers. The unpaired-reader design is used in situations that do not allow a paired-reader design. For example, a unpaired-reader design is used when different expertise is required to interpret tests and readers do not have equivalent expertise in each test. In a paired-patient design, a sample of patients

undergoes all the diagnostic tests under the study. An unpaired-patient design is the one where each patient undergoes a single test.

Paired- and unpaired- reader designs can be used with both paired- and unpaired-patient designs. Correlations between test results must be considered in evaluating a study that employs a paired design. The most commonly used design is the paired-patient, paired-reader design, in which multiple readers interpret all the test results of a sample of patients who undergoes all the diagnostic tests under the study. This design is also called traditional design. The data setup is given in Table 1.1; here, T_{kj1} and T_{kj2} denote the results of test 1 and 2 for the kth patient and are interpreted by the jth reader. This design is popular because it requires the smallest number of patient (Obuchowski and Rockette, 1995). Furthermore, it demands one of the smallest reader samples and one of the fewest number of interpretations per reader.

Table 1.1: Data setup for Paired-patient, Paired-Reader Design

	Reader 1		$\cdots$	Reader j		$\cdots$	Reader J	
	Test1	Test2		Test1	Test2		Test1	Test2
1	T_{111}	T_{112}	$\cdots$	T_{1j1}	T_{1j2}	$\cdots$	T_{1J1}	T_{1J2}
2	T_{211}	T_{212}	$\cdots$	T_{2j1}	T_{2j2}	$\cdots$	T_{2J1}	T_{2J2}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
k	T_{k11}	T_{k12}	$\cdots$	T_{kj1}	T_{kj2}	$\cdots$	T_{kJ1}	T_{kJ2}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
N	T_{N11}	T_{N12}	$\cdots$	T_{Nj1}	T_{Nj2}	$\cdots$	T_{NJ1}	T_{NJ2}

Note: This design requires N total patients and J total readers. There are $N \times I$ interpretations per reader, where I is the number of diagnostic tests under study. (Note that I=2 in the table)

In contrast, the unpaired-patient, unpaired-reader design is the most inefficient design, but situations arise when it is the only design option. The unpaired-patient, paired-reader design (See Table 1.3) and the paired-patient, unpaired-reader design(See Table 1.4) are improvements over the unpaired-patient, unpaired-reader design but inferior to the paired-patient, paired-reader design. Still, these designs may be necessary

when the tests are mutually exclusive or when the readers of the tests require different expertise (Zhou, 2002, p82).

Table 1.2: Data setup for Unpaired-patient, Unpaired-Reader Design

	Test 1						Test 2				
	Reader 1	$\cdots$	Reader j	$\cdots$	Reader J	Patient	Reader $\tilde{1}$	$\cdots$	Reader $\tilde{j}$	$\cdots$	Reader $\tilde{J}$
1	T_{111}	$\cdots$	T_{1j1}	$\cdots$	T_{1J1}	$\tilde{1}$	$T_{\tilde{1}\tilde{1}2}$	$\cdots$	$T_{\tilde{1}\tilde{j}2}$	$\cdots$	$T_{\tilde{1}\tilde{J}2}$
2	T_{211}	$\cdots$	T_{2j1}	$\cdots$	T_{2J1}	$\tilde{2}$	$T_{\tilde{2}\tilde{1}2}$	$\cdots$	$T_{\tilde{2}\tilde{j}2}$	$\cdots$	$T_{\tilde{2}\tilde{J}2}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
k	T_{k11}	$\cdots$	T_{kj1}	$\cdots$	T_{kJ1}	$\tilde{k}$	$T_{\tilde{k}\tilde{1}2}$	$\cdots$	$T_{\tilde{k}\tilde{j}2}$	$\cdots$	$T_{\tilde{k}\tilde{J}2}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
N	T_{N11}	$\cdots$	T_{Nj1}	$\cdots$	T_{NJ1}	$\tilde{N}$	$T_{\tilde{N}\tilde{1}2}$	$\cdots$	$T_{\tilde{N}\tilde{j}2}$	$\cdots$	$T_{\tilde{N}\tilde{J}2}$

Note: This design requires $I \times N$ total patients and $I \times J$ total readers, where I is the number of diagnostic tests under study. There are N interpretations per reader. (Note that I=2 in the table.)

Table 1.3: Data setup for Unpaired-patient, Paired-Reader Design

	Test 1						Test 2				
	Reader 1	$\cdots$	Reader j	$\cdots$	Reader J	Patient	Reader 1	$\cdots$	Reader j	$\cdots$	Reader J
1	T_{111}	$\cdots$	T_{1j1}	$\cdots$	T_{1J1}	$\tilde{1}$	$T_{\tilde{1}12}$	$\cdots$	$T_{\tilde{1}j2}$	$\cdots$	$T_{\tilde{1}J2}$
2	T_{211}	$\cdots$	T_{2j1}	$\cdots$	T_{2J1}	$\tilde{2}$	$T_{\tilde{2}12}$	$\cdots$	$T_{\tilde{2}j2}$	$\cdots$	$T_{\tilde{2}J2}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
k	T_{k11}	$\cdots$	T_{kj1}	$\cdots$	T_{kJ1}	$\tilde{k}$	$T_{\tilde{k}12}$	$\cdots$	$T_{\tilde{k}j2}$	$\cdots$	$T_{\tilde{k}J2}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
N	T_{N11}	$\cdots$	T_{Nj1}	$\cdots$	T_{NJ1}	$\tilde{N}$	$T_{\tilde{N}12}$	$\cdots$	$T_{\tilde{N}j2}$	$\cdots$	$T_{\tilde{N}J2}$

Note: This design requires $N \times I$ total patients and J total readers, where I is the number of diagnostic tests under study. There are $N \times I$ interpretations per reader. (Note that I=2 in the table)

Lastly, a paired-patient-per-reader, paired-reader design is the situation where the $N \times J$ total patients undergo all tests under the study, and each reader interprets the test results of the N patients. This design requires the fewest number of readers but requires many more patients than the paired-patient paired-reader design. It is an efficient design when patients can be accrued into the study quickly and inexpensively (Zhou, 2002, p83).

Table 1.4: Data setup for Paired-patient, Unpaired-Reader Design

	Test 1					Test 2				
	Reader 1	$\cdots$	Reader j	$\cdots$	Reader J	Reader $\tilde{1}$	$\cdots$	Reader $\tilde{j}$	$\cdots$	Reader $\tilde{J}$
1	T_{111}	$\cdots$	T_{1j1}	$\cdots$	T_{1J1}	$T_{1\tilde{1}2}$	$\cdots$	$T_{1\tilde{j}2}$	$\cdots$	$T_{1\tilde{J}2}$
2	T_{211}	$\cdots$	T_{2j1}	$\cdots$	T_{2J1}	$T_{2\tilde{1}2}$	$\cdots$	$T_{2\tilde{j}2}$	$\cdots$	$T_{2\tilde{J}2}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
k	T_{k11}	$\cdots$	T_{kj1}	$\cdots$	T_{kJ1}	$T_{k\tilde{1}2}$	$\cdots$	$T_{k\tilde{j}2}$	$\cdots$	$T_{k\tilde{J}2}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
N	T_{N11}	$\cdots$	T_{Nj1}	$\cdots$	T_{NJ1}	$T_{N\tilde{1}2}$	$\cdots$	$T_{N\tilde{j}2}$	$\cdots$	$T_{N\tilde{J}2}$

Note: This design requires N total patients and $J \times I$ total readers, where I is the number of diagnostic tests under study. There are N interpretations per reader. (Note that I=2 in the table)

Table 1.5: Data setup for Paired-Patient-Per-Reader, Paired-Reader Design

	Reader 1			$\cdots$	Reader j			$\cdots$	Reader J	
	Test1	Test2	Patient		Test1	Test2	Patient		Test1	Test2
1	T_{111}	T_{112}	$\tilde{1}$		$T_{\tilde{1}j1}$	$T_{\tilde{1}j2}$	$\tilde{\tilde{1}}$		$T_{\tilde{\tilde{1}}J1}$	$T_{\tilde{\tilde{1}}J2}$
2	T_{211}	T_{212}	$\tilde{2}$		$T_{\tilde{2}j1}$	$T_{\tilde{2}j2}$	$\tilde{\tilde{2}}$		$T_{\tilde{\tilde{2}}J1}$	$T_{\tilde{\tilde{2}}J2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
k	T_{k11}	T_{k12}	$\tilde{k}$		$T_{\tilde{k}j1}$	$T_{\tilde{k}j2}$	$\tilde{\tilde{k}}$		$T_{\tilde{\tilde{k}}J1}$	$T_{\tilde{\tilde{k}}J2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
N	T_{N11}	T_{N12}	$\tilde{N}$		$T_{\tilde{N}j1}$	$T_{\tilde{N}j2}$	$\tilde{\tilde{N}}$		$T_{\tilde{\tilde{N}}J1}$	$T_{\tilde{\tilde{N}}J2}$

Note: This design requires $N \times J$ total patients and J total readers. There are $N \times I$ interpretations per reader, where I is the number of diagnostic tests under study. (Note that I=2 in the table)

1.2.4 Correlated ROC Data Structures

The paired design is often used to improve the precision of the analysis. We focus our discussion on the following two types of correlated ROC data structure: (1) multiple test measurements on the same patient and (2) multireader ROC studies with multiple tests (Zou, 2002, p.274). In such studies, it is important to take into account the correlations for the estimation and the inferences on the accuracy of the diagnostic test. The statistical methodologies for the analyses from these correlated data structures are described in later sections in detail.

The first data structure arises when each patient is examined by several diagnostic tests or is repeatedly examined by the same test. Under this data structure, test responses from different patients can be assumed to be independent although test responses from the same patient are correlated. One special form from this data structure is known as clustered data. For example, in a study of the diagnostic accuracy of magnetic resonance (MRI) in lung cancer, different regions of the lung were considered for the presence of cancerous invasions (Webb et al, 1991; Beam, 1998). Beam (1998) described several approaches to the analysis of clustered data. Regression models from multiple test measurements on the same patients are discussed in section 1.4.

The second data structure arises when each patient is examined by multiple readers with multiple tests (or modalities). It is called multi-reader, multi-test study design. In this setting, the results of diagnostic tests often depend on a radiologist's subjective interpretation. For example, after conducting a mammography screening, a radiologist examines the image characteristics and gives his or her impression of the presence or absence of malignancy. Because of the variability in readers' accuracies caused by differences on interpretations, studies of such diagnostic tests usually involve several readers. The most popular design for such a study is to have multiple readers examine

the same set of patients who have undergone each of the diagnostic tests. This structure is equivalent to the notion of the paired-patient paired reader design as illustrated before. This design is most likely to occur in a radiology setting and is most efficient for a comparison of tests. We discuss several approaches (See section 1.3.4) and develop new statistical methodologies of ROC analysis arising from this data structure in Chapter 3.

1.2.5 Common Sources of Biases in Study Design

Studies of diagnostic tests are subject to an array of biases. Table 1.7 summarizes the common biases in studies of diagnostic test accuracy (Zou, 2002, p.69). Some of biases are discussed below:

a. *Selection bias*: When the sample composition has been influenced by external factors so that it does not represent the target population, we have selection bias.

b. *Spectrum bias*: Spectrum bias occurs when diseased subjects in the study are not representative of disease subjects in the population, or if controls selected for the study are different from population controls. A common mistake is to select cases that have more advanced or severe disease and select controls that are more healthier on average than non-disease subjects in the population. In this setting, test sensitivity and specificity will be higher than would be expected in the general population. Cases and controls in a diagnostic study should be randomly selected from the diseased and non-diseased target population.

c. *Imperfect gold standard bias*: A gold standard is a procedure that defines presence or absence of a condition of interest, such as a patient's disease status. Different gold standards are used for different tests and applications; Common examples are autopsy reports, surgery findings, pathology results from biopsy specimens,

and so on (Zou, 2002, p.15). If an imperfect reference test is used as a gold standard, the estimates of test accuracy usually will be biased. This phenomenon is called imperfect gold standard bias.

d. *Verification bias*: Verification bias occurs when the test being evaluated is used to determine which subjects get further evaluation, leading to a diagnostic of the disease. This problem is exceedingly common in research on diagnostic tests, particularly when the gold standard test poses some risk to subjects. If determination of disease status depends on the result of the test, then nave estimate of sensitivity and specificity based on disease-verified subjects are biased. The verification bias is also called as work-up bias, referral bias, selection bias and ascertainment bias (Pepe, 2003, p.169). For example, consider audiology screening in new borns with DPOAE. If the test suggests that a child is hearing impaired, then follow-up testing with the gold standard visual reinforcement audiometry (VRA) behavioral test is clinically indicated and is performed. However, if the DPOAE test suggests that the child's ears are responding to sound stimuli, then there are no clinical reasons to perform the VRA test.

1.3 Estimating the ROC Curve

In this section, statistical methods for estimating ROC curves, calculating the area under a ROC curve and comparing ROC curves are discussed. We introduce three approaches for estimating the ROC curve and the corresponding AUC: parametric, nonparametric, and semiparametric methods.

Table 1.6: Common Biases in Studies of Diagnostic Test Accuracy

Bias	Description
Selection bias	The composition of the sample is influenced by external factors, so the study sample is not representative of the target population
Spectrum bias	The study sample does not include the complete spectrum of patient characteristic
Imperfect gold	The reference procedure is not 100 % accurate standard bias
Workup bias	The results from the diagnostic test influence the subsequent clinical workup needed to establish the patient's diagnosis
Incorporation bias	The results from the diagnostic test under evaluation are incorporated-in full or part-into the evidence used to establish the definitive diagnosis
Verification bias	Patients with positive (or negative) test results are preferentially referred for the gold standard procedure;the bias occurs when estimates of accuracy are based only on the verified patients
Test-Review bias	The diagnostic test is evaluated without proper blinding of the results from the gold standard or competing test
Diagnostic-Review bias	The gold standard is evaluated without proper blinding of the results from the test under study
Reading-Order bias	When comparing two or more tests, the reader's interpretation is affected by his or her memory of the results from the competing test
Context bias	When the sample prevalence differs greatly from the population prevalence, the reader's interpretations may be affected, resulting is biased estimates if test accuracy

1.3.1 Parametric Method

The parametric methods model the ROC curves by assuming test results or some unknown monotonic transformation of the test results follow a certain distribution. Under the assumed distribution, parameters of ROC curves are derived and the smooth ROC curve is produced. The most commonly used distributional assumption is the binormal model in which parameters are estimated with maximum likelihood methods (Dorfman and Alf, 1969). The parameter estimates by maximum likelihood methods is fully efficient assuming that the models are correctly specified. If the distribution of scores for true-positive and true-negative test subjects are far from binormal, the parametric AUC and its corresponding standard error derived from a directly fitted binormal model may be distorted (Goddard and Hinberg, 1990). When test results are not binormal for continuous data, Zou and Hall (2000) suggested using a Box-Cox transformation to transform data to binormality and developed maximum likelihood algorithm for estimating ROC curve parameters.

The ROC curve is invariant to monotonic increasing transformations of test results. However, the parametric methods for estimating the ROC are not invariant to those transformations. This approach is not popular in practice because the fully parametric method makes strong assumptions on the distributions of test results.

1.3.2 Nonparametric Method

Alternatively, the ROC curve can be fitted empirically by using observed data without making any distributional assumptions for the test results. For continuous test results, the nonparametric ROC curve may be preferred since it passes through all observed points and provides unbiased estimates of sensitivity, specificity, and AUC in large samples (Zweig and Campbell, 1993).

Suppose the test result is measured on an ordinal or continuous scale. By convention,

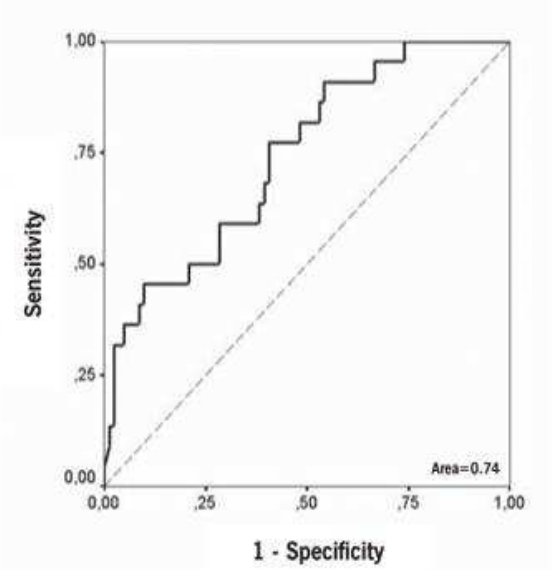

Figure 1.2: Empirical ROC curve

we assume that higher values of the test result are more indicative of disease. Let $X_i, i = 1, 2, \cdots, m$ denote test results from diseased subjects and $Y_j, j = 1, 2, \cdots, n$ denote test results from non-diseased subjects. Let $N = m + n$ be the total number of subjects under the study. Then, for each possible cutpoint c, the empirical TPR and FPR are calculated as follows.

$$\widehat{TPR}(c) = \frac{1}{m} \sum_{i=1}^{m} I(X_i \geq c)$$

$$\widehat{FPR}(c) = \frac{1}{n} \sum_{j=1}^{n} I(Y_j \geq c)$$

Then, the empirical ROC curve is a plot of $\widehat{TPR}(c)$ versus $\widehat{FPR}(c)$ for all $c \in (-\infty, \infty)$.

As shown in Figure 1.2, the empirical ROC curve is fitted by connecting the points $(\widehat{TPR}(c), \widehat{FPR}(c))$ for each c, which results in a step function. The empirical ROC curve is invariant with respect to a monotonic transformation of test results because it depends only on the ranks of observations in the combined sample (Zweig and Campbell,

1993). However, the empirical ROC curve is not smooth and the trapezoidal rule tends to underestimates the true area (Hanley and McNeil, 1982; Swet and Pickett, 1982).

The AUC can be estimated by summing the area of trapezoids formed by connecting the points of the empirical ROC curve. Nonparametric methods for estimating the area and its variance have been proposed in the literature (Bamber, 1975; Hanley and McNeil,1982; DeLong et al, 1988; Obuchowski,1997). The area under the empirical ROC curve, when the area is calculated by the trapezoidal rule, is given by

$$\hat{\theta}_{NP} = \frac{1}{mn} \sum_{j=1}^{n} \sum_{i=1}^{m} \phi(X_i, Y_j),$$

where

$$\phi(X_i, Y_j) = \begin{cases} 1 & \text{if } Y < X, \\ 1/2 & \text{if } Y = X, \\ 0 & \text{if } Y > X. \end{cases}$$

Note that $\hat{\theta}_{NP}$ provides an unbiased estimate of $\theta_{NP} = Pr(Y < X) + \frac{1}{2}Pr(X = Y)$. $Pr(X = Y) = 0$ for continuous test results.

$\hat{\theta}_{NP}$ is equivalent to the Mann-Whitney U-statistic (Bamber, 1975; Hanley and McNeil, 1982). Thus, the statistical properties of the Wilcoxon statistic can be applied to predict the properties of the area under an ROC curve. The AUC is a measure based on pair-wise comparisons of scores from diseased versus non-diseased subjects, not depending on the actual values of test result. It is the probability that test results from a randomly selected pair of diseased and non-diseased subjects are correctly ordered. The variance of the Mann-Whitney statistic can be derived from theory developed for generalized U-statistics by Hoeffing (1948). Define

$$\delta_{10} = E[\phi(X_i, Y_j)\phi(X_i, Y_k)] - \theta^2, \quad j \neq k$$
$$\delta_{01} = E[\phi(X_i, Y_j)\phi(X_k, Y_j)] - \theta^2, \quad i \neq k$$
$$\delta_{11} = E[\phi(X_i, Y_j)\phi(X_i, Y_j)] - \theta^2,$$

Then,

$$Var(\hat{\theta}_{NP}) = \frac{(n-1)\delta_{10} + (m-1)\delta_{01}}{mn} + \frac{\delta_{11}}{mn}.$$

Hanley and McNeil (1983) proposed a nonparametric method for the estimation of the AUC but used the assumption of Gaussian distribution to estimate variances of the areas for non-continuous test results. On the other hand, DeLong et al. (1988) developed the completely nonparametric covariance estimation by using the theory on generalized U-statistics.

The variance estimator proposed by DeLong et al. (1988) is as follows:

a) Define the X-components for the ith subject, $V_{10}(X_i)$ and the Y-components for the jth subject, $V_{01}(Y_j)$.

$$V_{10}(X_i) = \frac{1}{n}\sum_{j=1}^{n} \phi(X_i, Y_j), \quad i = 1, 2, \cdots, m$$

$$V_{01}(Y_j) = \frac{1}{m}\sum_{i=1}^{m} \phi(X_i, Y_j), \quad j = 1, 2, \cdots, n.$$

b) The AUC can be estimated using either the X or Y components.

$$\hat{\theta}_{NP} = \sum_{i=1}^{m} \frac{V_{10}(X_i)}{m} = \sum_{j=1}^{n} \frac{V_{01}(Y_j)}{n}.$$

c) Let S_{10} and S_{01} be covariance estimates for the X and Y components (S_{10} and

S_{01} are estimates of δ_{10} and δ_{01}, respectively). Then,

$$S_{10} = \frac{1}{m-1} = \sum_{i=1}^{m}(V_{10}(X_i) - \hat{\theta}_{NP})^2$$

and

$$S_{01} = \frac{1}{n-1} = \sum_{j=1}^{n}(V_{01}(Y_j) - \hat{\theta}_{NP})^2.$$

d) Then, the variance of the area under the curve is estimated by

$$V(\hat{\theta}_{NP}) = \frac{1}{m}S_{10} + \frac{1}{n}S_{01}.$$

An asymptotic (1-α) percent confidence interval for the ROC area is given by

$$\left(\hat{\theta}_{NP} - z_{1-\alpha/2}\sqrt{\hat{Var}(\hat{\theta}_{NP})}, \quad \hat{\theta}_{NP} + z_{1-\alpha/2}\sqrt{\hat{Var}(\hat{\theta}_{NP})}\right).$$

Hoeffding's theory extends to a vector of U statistics. Let $\hat{\boldsymbol{\theta}}$=$(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)$ be the vector of the AUC estimators. Let $\{X_i^r\}$ and $\{Y_j^r\}$ $(i = 1, \cdots, m; j = 1, \cdots, n; 1 \leq r \leq k)$ be the test results of rth diagnostic measure. Define

$$\begin{aligned}
\delta_{10}^{rs} &= E[\phi(X_i^r, Y_j^r)\phi(X_i^s, Y_k^s)] - \theta^r\theta^s, \quad j \neq k \\
\delta_{01}^{rs} &= E[\phi(X_i^r, Y_j^r)\phi(X_k^s, Y_j^s)] - \theta^r\theta^s, \quad i \neq k \\
\delta_{11}^{rs} &= E[\phi(X_i^r, Y_j^r)\phi(X_i^s, Y_j^s)] - \theta^r\theta^s,
\end{aligned}$$

Then, the covariance of the rth and sth statistic is $Cov(\hat{\theta}^r, \hat{\theta}^s) = \frac{(n-1)\delta_{10}^{rs} + (m-1)\delta_{01}^{rs}}{mn} + \frac{\delta_{11}^{rs}}{mn}$.

The estimated covariance matrix for $\hat{\boldsymbol{\theta}}$ can be computed as follows:

a) For the rth statistic, $\hat{\theta}^r$, the X-components for the ith subject, $V_{10}^r(X_i)$ and the

Y-components for the$j$th subject, $V_{01}^r(Y_j)$ are defined respectively, as

$$V_{10}^r(X_i) = \frac{1}{n}\sum_{j=1}^{n}\phi(X_i^r, Y_j^r), \quad i = 1, 2, \cdots, m$$

$$V_{01}^r(Y_j) = \frac{1}{m}\sum_{i=1}^{m}\phi(X_i^r, Y_j^r), \quad j = 1, 2, \cdots, n.$$

b) Compute $k \times k$ matrix S_{10} such that the (r,s)th element is

$$S_{10}^{rs} = \frac{1}{m-1} = \sum_{i=1}^{m}[V_{10}^r(X_i) - \hat{\theta^r}][V_{10}^s(X_i) - \hat{\theta^s}],$$

and similarly,

$$S_{01}^{rs} = \frac{1}{n-1} = \sum_{j=1}^{n}[V_{10}^r(Y_j) - \hat{\theta^r}][V_{01}^s(Y_j) - \hat{\theta^s}].$$

c) The estimated covariance matrix for the $\hat{\boldsymbol{\theta}}$=$(\hat{\theta_1}, \hat{\theta_2}, \cdots, \hat{\theta_k})$ is $\mathbf{S} = \frac{1}{m}\mathbf{S_{10}} + \frac{1}{n}\mathbf{S_{01}}$.

Let g be a linear function of $\hat{\boldsymbol{\theta}}$. If $\lim_{N\to\infty} m/n$ is bounded and nonzero, for any contrast $\boldsymbol{L\theta'}$, where $\boldsymbol{L}$ is a row vector of coefficients,

$$\frac{\boldsymbol{L\hat{\theta}' - L\theta'}}{\left[\boldsymbol{L}\left(\frac{1}{m}\boldsymbol{S_{10}} + \frac{1}{n}\boldsymbol{S_{10}}\right)\boldsymbol{L}'\right]^{1/2}}$$

has a standard normal distribution.

If there are multiple test results from the same subjects, estimation and inference of the accuracy of diagnostic tests must account for intracluster correlation. Obuchowski (1997) proposed a method for estimating the area under the ROC curve in the presence of clustered data by applying Delong et al. (1988)'s structural components approach to ROC curve estimation but extended their method to clustered ROC data using the ideas of Rao and Scott (1992). Note that Delong et al. (1988)'s method assumes that the test results are independent observations. In contrast, under the clustered data

structure, test results from the same subjects are correlated though test results from different subjects are assumed to be independent. The methodology for the analysis of clustered data in ROC studies was reviewed by Beam (1998).

1.3.3 Semiparametric Method

Under the semiparametric approach to estimating the ROC curve, the data are transformed to follow a binormal distribution using an unspecified monotonic transformation. A common approach to the semiparametric estimation of the ROC curve is to model the ROC curve parametrically without making additional assumptions on the distribution of test results. It is also called the parametric distribution free approach (Pepe, 2000; Alonzo and Pepe, 2002).

The maximum likelihood estimation algorithm for fitting binormal ROC curves to ordinal data has long been available (Swet and Pickett, 1982; Dorfman and Alf, 1969; Grey and Morgan, 1972, Metz et al. 1998). Dorfman and Alf (1969) proposed an iterative method for obtaining the maximum likelihood estimates of the parameters of a binormal ROC curve to ordinal data. The methods for fitting ROC curves to continuous data are less established than for ordinal data. Metz et al. (1998) developed a semiparametric method for continuous data to estimate the ROC curve. They assumed that the data come from a distribution of a latent variable and created ordinal data by categorizing the original continuous data, thereby using an ML curve-fitting algorithm for the ordinal data. The computer algorithm LABROC4 (a true ML algorithm) and LABROC5 (a quasi-ML algorithm) are developed for this method. However, this method is less sensitive to non-normality than the direct parametric method. On the other hand, Zou and Hall (2000) proposed the maximum likelihood rank-based estimator of the ROC curve for continuous data. They derived the probability distribution of ranks of test results using the theorem of Hoeffding (1951) and estimated binormal

parameters from the likelihood which depends only on the rank order statistics of the data by a Monte Carlo procedure. This method is efficient but computationally intensive. Pepe (2000a) and Alonzo and Pepe (2002) developed similar semiparametric methods to estimate ROC curves by applying procedures for fitting generalized linear models to binary data. Cai et al. (2004) proposed two semiparametric methods for estimating location and scale parameters in the binormal ROC model; (a) maximum profile likelihood approach, (b) Pseduo maximum likelihood approach.

1.3.4 Analysis of Multi-Reader, Multi-Test Studies

For the studies with multiple test measurements on the same patients, we discussed DeLong et al. (1988) in section 1.3.2. In this section, we focus on the studies in which each patient is examined by multiple readers with multiple tests.

Obuchowski and Rockette (1995) proposed a Mixed-Effects ANOVA model on the accuracy indices (e.g. the ROC curve area), where the tests were considered fixed and the readers were considered random. They modified the usual ANOVA F-tests to correct for the correlations between and within readers. Note that their method makes strong assumptions as follows: First, it assumes that the complex correlation structure from having the same patient sample evaluated by several readers in a set of tests can be described by only three correlations; correlation of error terms in diagnostic accuracies of the same reader in different tests, the correlation of error terms in diagnostic accuracies of different readers in the same test, and the correlation of error terms in diagnostic accuracies of different readers in different tests on the same patients. Second, it is not clear how well the modified F statistic follows a F distribution, especially in small samples. Furthermore, this approach does not provide variance estimates, due to readers and due to the interaction of readers and test, and it cannot handle covariates on either the patient or reader level (Zhou, 2002, p290).

Dorfman et al. (1992) proposed a Mixed-Effects ANOVA model on Jacknife pseudovalues for the test statistic. A mixed effect linear model for the jackknife pseudovalues is fitted in which readers and patients are random factors and tests are a fixed factor. They assumed that the random effects and error term in the model are normally and independently distributed. This method has been widely used in practice but has some weaknesses. In this framework, jackknife pseudovalues are treated as observed data and considered independent, which is, in fact, correlated.

Ishwaran and Gatsonis (2000) developed a Bayesian hierarchical ordinal regression models to deal with multi-reader ROC data and other types of multilevel cluster data. The models include covariates reflecting characteristics of the units at various levels of aggregation, such as individual patients, radiologists, and hospitals. We further examine the multi-reader studies with regression framework in section 1.4.

1.4 Regression Analysis for ROC Data

The performance of a diagnostic test can be influenced by risk factors beyond disease status. Thus, it is important to identify such factors to determine optimal conditions for test performance. The regression analysis can be used to evaluate or control for the possible covariate effects. There are three main approaches to incorporating covariate effects into ROC analysis. The first approach is to specify a model for the test result as a function of disease status and covariates. A second approach directly models covariate effects on ROC curve (Pepe 1997, 2000; Alonzo and Pepe, 2001). It is called parametric distribution free approach for the reason that it assumes a parametric model for the ROC curve but is distribution-free regarding the distribution of the test results. The third approach is to model ROC curve summary indices as a function of covariates. Note that the first two approaches can be applied to both discrete and continuous covariates. However, the third approach can be used only when the covariates are

discrete and there are enough observations in each covariate combination to permit calculation of the summary accuracy measure. Because of limitations in the third approach, we focus on the first two approaches.

Let Y denote a diagnostic test result and X denote a set of covariates of interest. The true disease status for the unit being tested is a known binary random variable, denoted by D, with D=1 if the unit is diseased and 0 if it is not. Without loss of generality, we assume that larger values of Y are more indicative of disease. The ROC curve associated with covariate vector X is denoted by $ROC_X(t)$, where t is the false positive rate. ROC curve can be written as $ROC(t) = S_D(S_{\bar{D}}^{-1}(t))$, where S_D and $S_{\bar{D}}$ are the survivor functions for Y given X in the diseased and non-diseased populations, respectively.

1.4.1 Modeling Covariate Effects on Test Results

Tosteson and Begg (1988) proposed the location-scale-type ordinal regression model for ordinal test results. They postulate that

$$P[Y \geq y|Z, D] = S_0\left(\frac{c_y - \mu(D, Z)}{\sigma(D, Z)}\right)$$

with particular specification for μ, σ, and S_0. For identifiability of c_y, it is assumed that $\mu(0, Z_0) = 0$ and $\sigma(0, Z_0) = 1$ at a baseline covariate value Z_0.

Under the above ordinal regression model, the induced ROC function at Z is

$$ROC_z(t) = S_0(-a(Z) + b(Z)S_{\bar{D}}^{-1}(t))$$

for $t \in T(Z) = S_0([c_y - \mu(0, Z)]/\sigma(0, Z)), y = 1, \cdots, P-1$, where

$$a(Z) = (\mu(1, Z) - \mu(0, Z))/\sigma(1, Z) \quad \text{and} \quad b(Z) = \sigma(0, Z)/\sigma(1, Z).$$

Toledano and Gatsonis (1995) extended the regression model to include the situation of correlated data arising from a combination of multiple modalities and/or multiple readers and developed the the technique of generalized estimating equations (GEEs) to account for the correlation between the observations. Tosteson and Begg (1988)'s method has been extended to random effects models (Beam, 1995; Gatsonis, 1995) and Bayesian methods (Peng and Hall, 1996; Hellmich et al. 1998; Ishwaran and Gatsonsis, 2000) when test results are ordinal.

On the other hand, the linear regression model for continuous test result Y conditional on disease status D and covariates Z is as follows:

$$Y = \mu(D, Z) + \sigma(D, Z)\varepsilon,$$

where ε is the residual term with mean 0 and variance 1 but with an unknown survivor function S_0. Then, the corresponding covariate-specific ROC curve is

$$ROC_z(t) = S_0(-a(Z) + b(Z)S_{\bar{D}}^{-1}(t)),$$

where $a(Z) = (\mu(1, Z) - \mu(0, Z))/\sigma(1, Z)$ and $b(Z) = \sigma(0, Z)/\sigma(1, Z)$.

With regard to techniques for estimation, parameters in fully parameterized models can be estimated using the usual likelihood or GEE techniques. The delta method yields standard errors for induced ROC parameters and confidence bands for induced ROC curves. Bootstrapping techniques may also be applied which are often easier to implement (Pepe, 2003, p.146).

1.4.2 Modeling Covariate Effects on ROC Curves

There are several advantages of modelling covariate effects directly on ROC curves (Pepe, 2003, p.151, p.165). First, the interpretation of model parameters pertains

directly to the ROC curves. The second advantage is that multiple tests can be evaluated and compared with each other within the regression framework even if the test results are measured in different unit or on different scales which cannot be achieved by modeling test results.

Pepe (1997, 2000) developed the ROC-GLM regression model,

$$g(ROC_X(t)) = h_0(t) + \beta X$$

where $h_0(\cdot)$ and $g(\cdot)$ denote monotone increasing(or decreasing) functions on (0,1) and $t \in T_Z \subset (0,1)$. The link function g is specified as part of the model. Examples are probit with $g(t) = \Phi^{-1}(t)$, logistic with $g(t)$=logit(t)=log$(t/(1-t))$ or logarithmic with $g(t)$=log(t). $h_0(t)$ is a baseline function specified up to some real parameters. The baseline function h defines the location and shape of the ROC curve, and β quantifies covariate effects. Pepe (1997, 2000) put forth an interpretation for each point on the ROC curve as being a conditional probability of a test result from a random diseased subject exceeding that from a random nondiseased subject. Then, they noted that generalized linear model methods applied to binary indicator variables can be used to model ROC curves and estimate parameters in the model. The key limitation of the ROC-GLM regression model by Pepe (1997) was that the parameter estimation required special programming that made the approach difficult to implement. Pepe (2000) and Alonzo and Pepe (2000) simplified the implementation process to perform parameter estimation.

Recently, Cai and Pepe (2002) extended the parametric ROC regression model by allowing an arbitrary nonparametric baseline function for h_0. Suppose that the data for analysis are organized as N_D data records for n_D subjects with disease, $\{(Y_{ik}, \mathbf{Z}_{ik}, \mathbf{Z}_{Dik}), k = 1, \cdots, K_i, i = 1, \cdots, n_D\}$, and $N_{\bar{D}}$ data records for $n_{\bar{D}}$ subjects

without disease, $\{(Y_{jl}, \mathbf{Z}_{jl}), k = 1, \cdots, K_j, i = n_D + 1, \cdots, n_D + n_{\bar{D}}\}$, where each subject may have more than one data record, $N_D = \sum_{i=1}^{n_D} K_i$ and $N_{\bar{D}} = \sum_{j=n_D+1}^{n_D+n_{\bar{D}}} K_j$. The covariates denoted by $\mathbf{Z}$ are relevant to both diseased and nondiseased subjects.

Then, the ROC curve is modeled as

$$R0C_{Z,Z_D}(u) = g\{h_0(u) + \boldsymbol{\beta}'\mathbf{Z} + \boldsymbol{\beta}'_D\mathbf{Z}_D\},$$

where g is a monotone increasing function mapping $(-\infty, \infty)$ to $(0, 1)$ and h_0 is an unspecified increasing function from $(0, 1)$ to $(-\infty, \infty)$. Using the fact that conditional on the covariates $\{\mathbf{Z}, \mathbf{Z}_D\}$ and $D = 1$, the expected value of $I(Y \geq S^{-1}_{\bar{D},\mathbf{Z}}(u))$ is $R0C_{Z,Z_D}(u) = g\{h_0(u) + \boldsymbol{\beta}'\mathbf{Z} + \boldsymbol{\beta}'_D\mathbf{Z}_D\}$, they constructed the following class of estimating equations for $\boldsymbol{\theta}_0 = (\beta, \beta_D)$ based on binary indicator variables:

$$\sum_{i=1}^{n_D}\sum_{k=1}^{K_i}\int_a^b w(X_{ik}, u)\mathbf{X}_{ik}[I\{Y_{ik} \geq S^{-1}_{\bar{D},\mathbf{Z}}(u)\} - g\{\boldsymbol{\theta}'\mathbf{X}_{ik} + h(u)\}]d\hat{v}(u) = 0,$$

where the prespecified constants (a, b) are chosen such that $P\{Y_{11} < S^{-1}_{\bar{D},\mathbf{Z}_{11}}(a)\}$ and $P\{Y_{11} > S^{-1}_{\bar{D},\mathbf{Z}_{11}}(b)\}$ are positive; $\mathbf{X} = [\mathbf{Z}', \mathbf{Z}'_d]'$, $\boldsymbol{\theta}'\mathbf{X}_{ik} = \boldsymbol{\beta}'\mathbf{Z} + \boldsymbol{\beta}'_D\mathbf{Z}_D$, w is a positive bounded uniformly continuous weight function, $\hat{v}(\cdot)$ is a known increasing but possibly data dependent function. They used a semiparametric location model (Pepe 1998; Heagerty and Pepe 1999), $S_{\bar{D},\mathbf{Z}}(c) = S_0(c - \gamma'_0\mathbf{Z})$, and estimated the parameter $\boldsymbol{\gamma}_0$ as the solution to

$$\sum_{j=n_D+1}^{n_D+n_{\bar{D}}}\sum_{l=1}^{K_j}\mathbf{Z}_{jl}(Y_{jl} - \boldsymbol{\gamma}'\mathbf{Z}_{jl}) = 0,$$

which is denoted by $\hat{\boldsymbol{\gamma}}$, and the survivor function S_0 with the empirical distribution of the residuals

$$\hat{S}_0(c) = \frac{1}{N_{\bar{D}}}\sum_{j=n_D+1}^{n_D+n_{\bar{D}}}\sum_{l=1}^{K_j} I(Y_{jl} - \boldsymbol{\gamma}'\mathbf{Z}_{jl} \geq c).$$

They then estimated the baseline ROC function h_0 and the parameter $\boldsymbol{\theta}_0$ simultaneously

as solution to

$$\sum_{i=1}^{n_D}\sum_{k=1}^{K_i}[I\{Y_{ik} \geq S^{-1}_{\bar{D},\mathbf{Z}}(u)\} - g\{\boldsymbol{\theta}'\mathbf{X}_{ik} + h(u)\}] = 0,$$

for $u \in [a, b]$ and

$$\sum_{i=1}^{n_D}\sum_{k=1}^{K_i}\int_a^b w(\mathbf{X}_{ik}, u)\mathbf{X}_{ik}[I\{Y_{ik} \geq S^{-1}_{\bar{D},\mathbf{Z}}(u)\} - g\{\boldsymbol{\theta}'\mathbf{X}_{ik} + h(u)\}]d\hat{v}(u) = 0,$$

where $\hat{S}^{-1}_{\bar{D},\mathbf{Z}}(u) = \hat{S}^{-1}_0(u) + \hat{\boldsymbol{\gamma}}'\mathbf{Z}$.

Cai and Pepe (2002) showed that their semiparametric methods fit the model with efficiency comparable to that of the fully parametric approach. However, their method requires to construct the high dimensional estimating equations for β and h, which makes the implementation process demanding.

1.4.3 Modeling ROC Summary Indices

The third approach to ROC regression is to model some summary index of the ROC curve as a function of covariates. This approach is feasible if covariates are discrete and there are sufficient numbers of disease and non-disease observations at each distinct covariate level in order to estimate the summary index. If we denote the estimated summary index by $\hat{\theta}_Z$, the idea is to fit the model

$$E\{g(\hat{\theta}_Z)\} = \beta_0 + \beta_1 X$$

using standard linear regression methods. Dorfman et al. (1992) and Obuchowski (1995a) suggested modeling the AUC, while Thompson and Zucchini (1989) recommended modelling the partial area under the curve, pAUC.

Chapter 2

Combining Continuous Biomarkers Using Semiparametric Transformation Models in the ROC Analysis

2.1 Introduction

Recent technological advances continue to provide non-invasive and more accurate biomarkers for evaluating disease status. Thus, the assessment of accuracy needs to keep pace with developments. One standard tool for assessing the accuracy of diagnostic tests is the Receiver Operating Characteristic (ROC) curve (Swets and Pickett, 1982; Hanley, 1989). Many methods have been developed to evaluate a single continuous-scale biomarker in the framework of ROC analysis.

A common approach to the semiparametric estimation of ROC curves is to model the ROC curve parametrically without making assumptions about the distribution of test results (Pepe, 2000; Alonzo and Pepe, 2002). The most popular semiparametric model is the binormal ROC model. Let X denote test results from diseased subjects

and Y denote test results from non-diseased subjects. A binormal ROC curve for X and Y assumes that for some (unknown) strictly increasing transformation h, $h(X)$ and $h(Y)$ have normal distributions. The binormal ROC model is then written as

$$ROC(u) = \Phi\{a + b\Phi^{-1}(u)\},$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Metz et al. (1998) categorized continuous data into ordinal-scale categorical data and used an ML curve-fitting algorithm to estimate binormal ROC curve. On the other hand, Zou and Hall (2000) proposed the maximum likelihood rank-based estimator of the ROC curve by ranking original continuous data and numerically solving the score equations derived from the likelihood function of the order statistics using a Monte Carlo procedure. Cai and Moskowitz (2004) proposed a maximum profile likelihood and a pseudo-maximum likelihood approaches for estimating the binormal ROC model.

Even though numerous methods have been developed for single biomarkers, few statistical methods exist to accommodate multiple biomarkers that can be used simultaneously for disease detection. In many situations, single biomarkers are not sufficient to achieve the desired level of accuracy and newly discovered biomarkers can provide additional information. We consider a linear combination of biomarkers in order to optimize diagnostic accuracy. One possible objective function to be optimized is the area under an ROC curve (AUC) of a biomarker combination. Using this framework, Su and Liu (1993) assumed that markers follow a multivariate normal distribution in both disease and non-disease groups, and showed that Fisher's linear discriminant function provides an optimal linear combination of biomarkers. In contrast, Pepe and Thompson (2000) proposed a nonparametric rank-based method for AUC maximization. A second approach to combining biomarkers uses the likelihood function as an objective function to be optimized such as the logistic regression model. Pepe and

Thompson (2000) showed that the nonparametric rank-based method performs as well as the logistic likelihood-based method when the logistic model holds, and better than the logistic regression model when it does not. However, finding optimal coefficients for a nonparametric AUC is computationally demanding when the number of biomarkers exceeds 2. Furthermore, Pepe and Thompson (2000)'s method does not account for correlated biomarker structure, and relying on a linear combination of biomarkers can be misleading for prediction because test results can be affected by original biomarker scales.

In this chapter, we propose joint transformation models for analyzing multiple biomarkers. Subject-specific random effects are included in the models. Our model generalizes the usual binormal ROC model and naturally accounts for the dependence among biomarkers. The derived diagnostic rule does not depend on any monotone transformation of biomarkers and is not sensitive to extreme biomarker values. In Section 2.2, we introduce our models and give the best rule for combining biomarkers. In Section 2.3, we propose an inference procedure to estimate model parameters based on the nonparametric maximum likelihood estimation. The asymptotic properties of the estimators are provided in Section 2.4. Section 2.5 presents the results from simulation studies. The applications of the model analyzing vessel attributes in magnetic resonance angiography as well as measures for serum prostate-specific antigen (PSA) for prostate cancer disease are given in Section 2.6. A discussion is given in Section 2.7. All technical proofs are given in the Appendix.

2.2 Model

Suppose we observe K biomarker measurements from random sample of size n (n_1 diseased, n_0 non-diseased, $n = n_1 + n_0$) subjects. Then, the observed data for subject i can be represented as (X_{ik}, D_i), where X_{ik} is the measurement of kth continuous

biomarker from ith subject ($i = 1, \cdots, n; k = 1, \cdots, K$) and D_i is the binary disease outcome ($D_i = 1$ for diseased; 0 for non-diseased). In a multivariate normal ROC model, we assume that there exist K non-decreasing transformations $H_1, \cdots, H_K$ such that $(H_1(X_{i1}), \cdots, H_K(X_{iK}))$ follows a multivariate normal distribution in each of diseased and non-diseased groups;

$$[H_1(X_{i1}), \cdots, H_K(X_{iK})|Z_i, D_i = 0] = Z_i' a_i + (\epsilon_{i1}, \cdots, \epsilon_{iK}),\ i = 1, \cdots, n_0 \qquad (2.1)$$

$$[H_1(X_{i1}), \cdots, H_K(X_{iK})|Z_i, D_i = 1] = Z_i' a_i + (\epsilon_{i1}', \cdots, \epsilon_{iK}'),\ i = 1, \cdots, n_1, \qquad (2.2)$$

where $a_i = (a_{i1}, \cdots, a_{ip})'$ is a $p \times 1$ vector of subject-specific random effect and $Z_i = (z_{i1}, \cdots, z_{ip})'$ is a $p \times 1$ vector of covariates for ith subject with $z_{i1} = 1$. $\epsilon_i = (\epsilon_{i1}, \cdots, \epsilon_{iK})'$ and $\epsilon_i' = (\epsilon_{i1}', \cdots, \epsilon_{iK}')'$ are $K \times 1$ vectors of random errors. Here a_i and ϵ_i (or ϵ_i') are independent and are normally distributed with

$$a_i \sim N_p(\mathbf{0}, \Sigma_a), \quad \epsilon_i \sim N_K(\mu_0 Z_i, \Sigma_0), \quad \text{and} \quad \epsilon_i' \sim N_K(\mu_1 Z_i, \Sigma_1),$$

where $\mu_0 = (\mu_{01}', \cdots \mu_{0K}')'$ with $\mu_{0k} = (\mu_{0k1}, \mu_{0k2}, \cdots, \mu_{0kp})'$, and $\mu_1 = (\mu_{11}', \cdots \mu_{1K}')'$ with $\mu_{1k} = (\mu_{1k1}, \mu_{1k2}, \cdots, \mu_{1kp})'$. Thus, $[H_1(X_{i1}), \cdots, H_K(X_{iK})|D_i = 0] \sim N(\mu_0 Z_i,$ $Z_i' \Sigma_a Z_i \mathbf{1}\mathbf{1}' + \Sigma_0)$ and $[H_1(X_{i1}), \cdots, H_K(X_{iK})|D_i = 1] \sim N(\mu_1 Z_i, Z_i' \Sigma_a Z_i \mathbf{1}\mathbf{1}' + \Sigma_1)$, $\mathbf{1} = (1, 1, \cdots, 1)'_{K \times 1}$. We give the following restrictions for the identifiability of the model.

(i) $H_k(0) = 0,\ k = 1, \cdots, K.$

(ii) $\Sigma_0 = \text{diag}(1, 1, \cdots, 1)$ and $\Sigma_1 = \text{diag}(\sigma_{11}^2, \sigma_{12}^2, \cdots, \sigma_{1K}^2)$.

The restrictions (i) and (ii) are imposed not to allow location and scale shifts of the transformations and parameters to result in same parameters, respectively.

Under the above assumptions, we propose to use a linear combination of the transformed biomarkers $H_1(X_1), H_2(X_2), \cdots, H_K(X_K)$ for describing the performance of such biomarkers. Consider a linear combination $\beta_1 H_1(X_1)+\beta_2 H_2(X_2)+\cdots+\beta_K H_K(X_K)$. Since $(H_1(X_1), H_2(X_2), \cdots, H_K(X_K))$ has a multivariate normal distribution for diseased and non-diseased populations given covariates $Z = (z_1, \cdots, z_p)'$, the linear combination is normally distributed with

$$\beta_1 H_1(X_1) + \beta_2 H_2(X_2) + \cdots + \beta_K H_K(X_K)|D = 0 \ \sim N(\beta'\mu_0 Z, \beta'(Z'\Sigma_a Z\mathbf{1}\mathbf{1}' + \Sigma_0)\beta),$$

$$\beta_1 H_1(X_1) + \beta_2 H_2(X_2) + \cdots + \beta_K H_K(X_K)|D = 1 \ \sim N(\beta'\mu_1 Z, \beta'(Z'\Sigma_a Z\mathbf{1}\mathbf{1}' + \Sigma_1)\beta),$$

where $\beta = (\beta_1, \cdots, \beta_K)'_{K\times 1}$.

The AUC_Z of such a linear combination is given by

$$\frac{\beta'(\mu_1 - \mu_0)Z}{\sqrt{\beta'(\Sigma_1 + \Sigma_0 + 2Z'\Sigma_a Z\mathbf{1}\mathbf{1}')\beta}}.$$

As shown by Su and Liu (1993), the coefficients for the best linear combination are

$$\beta_{opt,Z} \propto (\Sigma_1 + \Sigma_0 + 2Z'\Sigma_a Z\mathbf{1}\mathbf{1}')^{-1}(\mu_1 - \mu_0)Z.$$

The AUC of the optimal linear combination (optimal AUC) is given by

$$AUC_{opt,Z} = \Phi\sqrt{\left(Z^T(\mu_1 - \mu_0)^T(\Sigma_1 + \Sigma_0 + 2Z'\Sigma_a Z\mathbf{1}\mathbf{1}')^{-1}(\mu_1 - \mu_0)Z\right)}. \qquad (2.3)$$

2.3 Inference Procedures

We propose to use the nonparametric maximum likelihood estimation (NPMLE) to estimate parameters μ_0, μ_1, Σ_1, and Σ_a and all the transformations $H_1, H_2, \cdots, H_K$. In

the NPMLE, $H_k(\cdot)$ will be assumed to be a non-decreasing step function with jumps at the observed data for kth biomarker and $H_k^{'}(t)$ will be replaced by its jump size at t in the likelihood function. We assume the kth biomarker measurement is censored if it is larger or smaller than fixed threshold values m_k and M_k. Our model includes uncensored cases of biomarkers in which $m_k = -\infty$ and $M_k = \infty$. The indicator variables are introduced to denote the censored status; $\delta_{ik}^O = I(m_k < X_{ik} < M_k)$ and $\delta_{ik}^R = I(X_{ik} > M_k)$. Define $\mathbf{X}_i = (X_{i1}, \cdots, X_{iK})'$, $\mathbf{M} = (M_1, \cdots, M_K)'$ and $\mathbf{m} = (m_1, \cdots, m_K)'$. Under models (2.1) and (2.2), the observed likelihood function concerning parameters of interest is given by

$$\prod_{i=1}^{n} \int_{a_i} \left[\prod_{k=1}^{K} \{f_k(X_{ik}|a_i, Z_i, D_i)\}^{\delta_i^O} \{1 - F_k(M_k - |a_i, Z_i, D_i)\}^{\delta_i^R} F_k(m_k|a_i, Z_i, D_i)^{1-\delta_i^O-\delta_i^R} \right] \times f(a_i) da_i,$$

where f_k is the conditional density of X_k given covariates, random effects and disease status, F_k is the corresponding cumulative distribution function, and $f(a_i)$ is the density of random effects. Particularly,

$$f_k(X_k|a, Z, D = d) = H_k'(X_k)(2\pi\sigma_{dk}^2)^{-1/2} \exp\left\{ -\frac{1}{2\sigma_{dk}^2} (H_k(X_k) - \widetilde{Z}^T a - \mu_{dk}^T Z)^2 \right\}$$

and $f(a) = (2\pi)^{-q/2} \exp\{-a^T \Sigma_a^{-1} a/2\}$.

In addition, we define

$$H_k^+(X) = H_k(X) \text{ if } X > 0, \quad H_k^-(-X) = -H_k(X) \text{ if } X < 0.$$

Then, both H_k^+ and H_k^- are increasing functions from $(-\infty, \infty)$ to $(0, \infty)$ and $H_k(X_{ik}) = H_k^+(X_{ik}) I(X_{ik} > 0) - H_k^-(\widetilde{X}_{ik}) I(\widetilde{X}_{ik} > 0)$ with $\widetilde{X}_{ik} = -X_{ik}$. Let $\boldsymbol{\theta} \equiv (\mu_0, \mu_1, \Sigma_1)$ and

$\boldsymbol{\theta}^* \equiv (\boldsymbol{\theta}, \Sigma_a)$. The complete data log likelihood function for $(\boldsymbol{\theta}^*, H^+, H^-)$ is given by

$$
\begin{aligned}
l_c(\boldsymbol{\theta}^*, H^+, H^-) \;=\; & \sum_{X_{ik}>0}\sum_{k=1}^{2}\delta_{ik}^{O}\log(h_k^+(X_{ik})) \\
& + \sum_{X_{ik}>0}\sum_{k=1}^{2}\Bigg[\Big\{ -\frac{\delta_{ik}^{O}}{2}(H_k^+(X_{ik}) - \mu_{0k}'Z_i - Z_i'a_i)^2 \\
& \qquad +\delta_{ik}^{R}\log P(X_{ik} > M_k|a_i, D_i)\Big\}(1-D_i) \\
& \qquad +\Big\{\delta_{ik}^{O}\Big(-\frac{1}{2}\log\sigma_{1k}^2 - \frac{1}{2\sigma_{1k}^2}(H_k^+(X_{ik}) - \mu_{1k}'Z_i - Z_i'a_i)^2\Big) \\
& \qquad +\delta_{ik}^{R}\log P(X_{ik} > M_k|a_i, D_i)\Big\}D_i\Bigg] \\
& + \sum_{\widetilde{X}_{ik}>0}\sum_{k=1}^{2}\delta_{ik}^{O}\log(h_k^-(\widetilde{X}_{ik})) \\
& + \sum_{\widetilde{X}_{ik}>0}\sum_{k=1}^{2}\Bigg[\Big\{ -\frac{\delta_{ik}^{O}}{2}(H_k^-(\widetilde{X}_{ik}) + \mu_{0k}'Z_i + Z_i'a_i)^2 \\
& \qquad +(1-\delta_{ik}^{O}-\delta_{ik}^{R})\log P(X_{ik} < m_k|a_i, D_i)\Big\}(1-D_i) \\
& \qquad +\Big\{\delta_{ik}^{O}\Big(-\frac{1}{2}\log\sigma_{1k}^2 - \frac{1}{2\sigma_{1k}^2}(H_k^-(\widetilde{X}_{ik}) + \mu_{1k}'Z_i + Z_i'a_i)^2\Big) \\
& \qquad +(1-\delta_{ik}^{O}-\delta_{ik}^{R})\log P(X_{ik} < m_k|a_i, D_i)\Big\}D_i\Bigg] \\
& + \sum_{i=1}^{n}\log P(a_i;\Sigma_a), \qquad (2.4)
\end{aligned}
$$

where $h_{ik}^+(\equiv h_k^+(X_{ik}) \equiv H_k^+\{X_{ik}\})$ is the jump size of $H_k^+(\cdot)$ at X_{ik} and $h_{ik}^-(\equiv h_k^-(\widetilde{X}_{ik}) \equiv H_k^-\{\widetilde{X}_{ik})\})$ is the jump size of $H_k^-(\cdot)$ at $\widetilde{X}_{ik}$. Note that $H_{ik}^+ = H_k^+(X_{ik}) = \sum_{j=1}^{n_k^+} h_{jk}^+ I(X_{jk} \le X_{ik})$ and $H_{ik}^- = H_k^-(\widetilde{X}_{ik}) = \sum_{j=1}^{n_k^-} h_{jk}^- I(\widetilde{X}_{jk} \le \widetilde{X}_{ik})$ where $n_k^+(n_k^-)$ is the number of positive (negative) observed values of the kth biomarker.

We apply an the expectation-maximization (EM) algorithm (Dempster et al., 1977) to calculate NPMLEs and variances of $\boldsymbol{\theta}^*$ and all jump sizes, $\mu_0 = (\mu_{0kr}), \mu_1 = (\mu_{1kr}), \Sigma_1 = diag(\sigma_{11}^2, \cdots, \sigma_{1K}^2), \Sigma_a = (\sigma_{arr'}^2), h_{ik}^+$, and h_{jk}^- $(k = 1, \cdots, K; r, r' = 1, \cdots, p;$

$i = 1, \cdots, n_k^+; j = 1, \cdots, n_k^-$). The maximization for the jump sizes of H_k^+ or H_k^- can be done separately for each k and the likelihood is strictly concave in these jump sizes so that most of the optimization algorithm converges in finding the maximum. In the EM framework, the random effect a_i is treated as missing. Let $\hat{E}[\cdot]$ denote the conditional expectation given the observed data and the current parameter estimates. The conditional expectations of any functions of a_i is computed in the E-step and the conditional expectation of (2.4) given observed data is maximized in the M-step. Since σ^2 can be updated by maximizing $\sum_{i=1}^n \hat{E}[\log P(a_i; \Sigma_a)]$ directly, $\boldsymbol{\theta}$ should be updated in the M-step. Thus, the objective function becomes

$$
\begin{aligned}
\hat{E}[l_c(\boldsymbol{\theta}, H^+, H^-)] \;=\; & \sum_{X_{ik}>0} \sum_{k=1}^{K} \delta_{ik}^O \log(h_k^+(X_{ik})) + \sum_{X_{ik}>0} \sum_{k=1}^{K} \Bigg[\Bigg\{ -\frac{\delta_{ik}^O}{2} \hat{E}[(H_k^+(X_{ik}) - \mu_{0k}' Z_i \\
& - Z_i' a_i)^2] + \delta_{ik}^R \hat{E}[\log P(X_{ik} > M_k | a_i, D_i)] \Bigg\} (1 - D_i) \\
& + \Bigg\{ \delta_{ik}^O \Big(-\frac{1}{2} \log \sigma_{1k}^2 - \frac{1}{2\sigma_{1k}^2} \hat{E}[(H_k^+(X_{ik}) - \mu_{1k}' Z_i - Z_i' a_i)]^2 \Big) \\
& + \delta_{ik}^R \hat{E}[\log P(X_{ik} > M_k | a_i, D_i)] \Bigg\} D_i \Bigg] + \sum_{\widetilde{X}_{ik}>0} \sum_{k=1}^{K} \delta_{ik}^O \log(h_k^-(X_{ik})) \\
& + \sum_{\widetilde{X}_{ik}>0} \sum_{k=1}^{K} \Bigg[\Bigg\{ -\frac{\delta_{ik}^O}{2} \hat{E}[(H_k^-(\widetilde{X}_{ik}) + \mu_{0k}' Z_i + Z_i' a_i)^2] \\
& + (1 - \delta_{ik}^O - \delta_{ik}^R) \hat{E}[\log P(X_{ik} < m_k | a_i, D_i)] \Bigg\} (1 - D_i) \\
& + \Bigg\{ \delta_{ik}^O \Big(-\frac{1}{2} \log \sigma_{1k}^2 - \frac{1}{2} \frac{1}{\sigma_{1k}^2} \hat{E}[(H_k^-(\widetilde{X}_{ik}) + \mu_{1k}' Z_i + Z_i' a_i)^2] \Big) \\
& + (1 - \delta_{ik}^O - \delta_{ik}^R) \hat{E}[\log P(X_{ik} < m_k | a_i, D_i)] \Bigg\} D_i \Bigg]. \qquad (2.5)
\end{aligned}
$$

Let $\hat{\boldsymbol{\theta}}^* \equiv (\hat{\mu}_0, \hat{\mu}_1, \hat{\Sigma}_1, \hat{\Sigma}_a)$, $\hat{H}^+$ and $\hat{H}^-$ denote the nonparametric maximum likelihood estimates for $\boldsymbol{\theta}^*$, H^+, and H^-, respectively. We estimate the variance of $(\hat{\boldsymbol{\theta}}^*, \hat{H}^+, \hat{H}^-)$ using the observed information matrix of $\hat{\boldsymbol{\theta}}^*$ and jump sizes $\hat{H}_k^+\{X_{ik}\}$

and $\hat{H}_k^-\{\widetilde{X}_{ik}\}$ by Louis's formula (1982). The optimal AUC given in (2.3) is estimated at $\hat{\boldsymbol{\theta}}^*$ and its variance is computed by the delta method.

2.4 Asymptotic Theory

In this section, we derive the asymptotic properties of the NPMLEs under finite censorship condition. Particularly, we cast our transformation models into the transformation models for multivariate failure times as in Zeng and Lin (2007), also described in Zeng and Lin (2009). To this end, biomarkers are treated as survival times and threshold values as censoring time. Specifically, each biomarker is divided by two groups: (G1) X_{ik} (> 0), positive biomarker; (G2) $-X_{ik}$ (> 0), biomarker transformed to be positive by switching a sign. We have a total of $2K$ different groups based on K types of biomarkers. Define T_l $(l = 1, \cdots, 2K)$ is a biomarker in the lth group where $T_l, l = 1, \cdots, K$ is the kth type of biomarker in G1; and $T_l, l = K+1, \cdots, 2K$ is the $(l-K)$th type of biomarker in G2. Let C_l denote censoring time (M_l for G1 and $|m_l|$ for G2), Z_l covariates, and τ the duration of the study ($\tau = C_l$). Define $\mu_{0k} = (\mu_{0k1}, \mu_{0k2}, \cdots, \mu_{0kp})'$, $\mu_{1k} = (\mu_{1k1}, \mu_{1k2}, \cdots, \mu_{1kp})'$ with $\mu_0 = (\mu_{01}', \cdots \mu_{0K}')'$ and $\mu_1 = (\mu_{11}', \cdots \mu_{1K}')'$, $\Sigma_1 = \text{diag}(\sigma_{11}^2, \cdots, \sigma_{1K}^2)$, and $\Sigma_a = (\sigma_{arr'})$, $k = 1, \cdots, K; r, r' = 1, \cdots, p$. Then, the linear transformation model has a following form.

$$\widetilde{H}(T_{il}) \;=\; \mu_{1l}' Z_{il} D_i + \mu_{0l}' Z_{il}(1 - D_i) + Z_{il}' a_i + \epsilon_{il},\ i = 1, \cdots, n_l; l = 1, \cdots, L.$$

where $\widetilde{H}(\cdot)$ is an unspecified increasing function, ϵ_{il} is normally distributed with mean zero and variance $\sigma_{1l}^2 D_i + (1 - D_i)$. $\mu_{0l} = \mu_{0(l+K)}, \mu_{1l} = \mu_{1(l+K)}, \sigma_{0l}^2 = \sigma_{0(l+K)}^2, \sigma_{1l}^2 = \sigma_{1(l+K)}^2$, $l < K$ and $L = 2K$. Note that $H(X) = \widetilde{H}(X)$ if $X > 0$ and $H(X) = -\widetilde{H}(-X)$ if $X < 0$.

Write $\Lambda(T_{il}) = \exp\{\widetilde{H}(T_{il})\}$. Let $N_{il}(t)$ denote the number of lth group of biomarkers that are smaller or equal to t. Our model implied that the cumulative intensity for $N_{il}(t)$ takes the form

$$\Lambda_l(t|Z_{il}; a_i) \;=\; G_l\left[\int_0^t R_{il}(s)\exp\{\mu_{1l}'Z_{il}D_i + \mu_{0l}'Z_{il}(1-D_i) + Z_{il}'a_i\}d\Lambda_l(s)\right] \quad (2.6)$$

where G_l is a continuously differentiable and strictly increasing function; $G_l(x) = -\log\{1-\Phi(\log x)\}$ if $D_i = 0$ and $G_l(x) = -\log\{1-\Phi\left(\frac{\log x}{\sigma_{1l}}\right)\}$ if $D_i = 1$. $R_{il}(t) = I(C_{il} \geq t)$ is an indicator process, μ_{1l} and μ_{0l} are vectors of unknown regression parameters, and $\Lambda_l(\cdot)$ is an unspecified increasing function. Define $\boldsymbol{\theta}_{d\times 1} = (\mu_{0kr}, \mu_{1kr}, \sigma_{1k}^2)'$, $k = 1, \cdots, K; r = 1, \cdots, p,\ d = 2Kp + K$. Model (2.6) has the same form as the transformation models with random effects for dependent failure times in Zeng and Lin (2007, 2009). The likelihood for $\boldsymbol{\theta}$ and Λ_l is given by

$$\begin{aligned}\prod_{i=1}^n \int \prod_{l=1}^L \prod_{t\leq\tau} \Big[& R_{il}(t)\lambda_l(t)\exp\Big(\mu_{1l}'Z_{il}D_i + \mu_{0l}'Z_{il}(1-D_i) + Z_{il}'a_i\Big) \\ & G_l'\Big\{\textstyle\int_0^t R_{il}(s)\exp\Big(\mu_{1l}'Z_{il}D_i + \mu_{0l}'Z_{il}(1-D_i) + Z_{il}'a_i\Big)d\Lambda_l(s)\Big\}\Big]^{dN_{il}(t)} \\ & \exp\Big[-G_l\Big\{\textstyle\int_0^\tau R_{il}(t)\exp\Big(\mu_{1l}'Z_{il}D_i + \mu_{0l}'Z_{il}(1-D_i) + Z_{il}'a_i\Big)d\Lambda_l(s)\Big\}\Big] \\ & f(a;\Sigma_a)da,\end{aligned} \quad (2.7)$$

with $\lambda_l(t) = \Lambda_l'(t)$ $(l = 1, \cdots, 2K)$. This can be written in the following form

$$\prod_{i=1}^n \prod_{l=1}^L \prod_{t\leq\tau} \lambda_l(t)^{R_{il}(t)dN_{il}(t)}\Psi(O_i; \boldsymbol{\theta}, \mathcal{A}),$$

where $\mathcal{A} = (\Lambda_1, \cdots, \Lambda_L)$, $\mathcal{O}_i$ pertains to the observation on the ith subject, and Ψ is a function of $\mathcal{O}_i$, $\boldsymbol{\theta}$, and $\mathcal{A}$. Note that $H(X)$ is same as $log\Lambda(X)I(X > 0) - log\Lambda(-X)I(X < 0)$. For the nonparametric maximum likelihood estimation, we allow $\hat{H}$ to be discontinuous with jumps at the observed biomarkers and maximize the

modified likelihood function

$$\prod_{i=1}^{n}\prod_{l=1}^{L}\prod_{t\leq\tau}\lambda_l\{t\}^{R_{il}(t)dN_{il}(t)}\Psi(O_i;\boldsymbol{\theta},\mathcal{A}),$$

where $\lambda_l\{t\}$ is the jump size of the monotone function $\exp(H(T))$ (for group G1) and $\exp(-H(T))$ (for group G2) at t.

We establish the asymptotic properties of the NPMLEs, $\widehat{\boldsymbol{\theta}}$ and $\widehat{H}_k, k = 1, ..., K$, under finite detection limits condition, i.e., $-\infty < m_k < M_k < \infty$. We impose the following regularity conditions.

(C1) The parameter values $(\mu_0, \mu_1, \sigma_{1k}, \Sigma_a)^T_{d\times 1}$ belongs to the interior of a compact set Θ and $H_l^{'}(t) > 0$ for all $t \in [m_k, M_k]$. Moreover, for $k = 1, ..., K$, the true transformation for H_k, denoted by H_{k0}, is twice-continuously differentiable and $H'_{k0}(t) > 0$ for all $t \in [m_k, M_k]$.

(C2) With probability 1, $\lim_{n\to\infty} n_1/n = q$ with $0 < q < 1$.

(C3) (Identifiability Condition) If there exist a vector ν and a symmetric matrix M such that $\nu^T Z = \mathbf{0}$ and $Z^T M Z = O$, then $\nu = \mathbf{0}$ and $M = O$. This condition is equivalent to the linear independence of covariates.

Remark 2.1. (C1) and (C2) are standard conditions for this type of problem.

The following Theorems 2.1 and 2.2 state the consistency and efficiency of the NPMLEs.

Theorem 2.1. *Under Conditions (C1)-(C3),*

$$|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| + \sum_{k=1}^{K} sup_{t\in[m_k,M_k]}|\widehat{H}_k(t) - \widehat{H}_{k0}(t)| \to_{a.s.} 0.$$

To derive the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0, \widehat{H}_k - H_{k0}, k = 1, ..., K)$, we

first define the metric space for which such a random element is defined. We introduce $\mathcal{V} = \{\boldsymbol{\nu} \in R^d, |\boldsymbol{\nu}| \leq 1\}$ and $\mathcal{Q}_k = \{h_k(t) : \|h_k(t)\|_{V[m_k,M_k]} \leq 1\}$, where $V[m_k, M_k]$ denotes the space of functions with bounded total variations in $[m_k, M_k]$. We then identify $\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0, \widehat{H}_k - H_{k0}, k = 1, ..., K)$ as a random element in $l^\infty(\mathcal{V}\times\mathcal{Q}_1\times\cdots\times\mathcal{Q}_K)$ by letting

$$\begin{aligned}\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0, \widehat{H}_k - H_{k0}, k = 1, ..., K)[\boldsymbol{\nu}, h_1, ..., h_K] \;\equiv\; & \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\boldsymbol{\nu} \\ & + \sum_{k=1}^{K} \sqrt{n} \int_0^\tau h_k(s) d(\widehat{H}_k - \widehat{H}_{0k})(s).\end{aligned}$$

Thus, the weak convergence of $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0, \widehat{H}_k - H_{k0}, k = 1, ..., K)$ is with respect to the same metric space. The following theorem holds.

Theorem 2.2. *Under Conditions (C1)-(C3),*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0, \widehat{H}_k - H_{k0}, k = 1, ..., K) \longrightarrow_d \mathcal{G} \text{ in } l^\infty(\mathcal{V} \times \mathcal{Q}_1 \times \cdots \times \mathcal{Q}_K),$$

where $\mathcal{G}$ is a mean-zero and tight Gaussian process. Furthermore, the limiting covariance matrix of $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ attains the semiparametric efficiency bound (Bickel et al. 1993).

Theorems 2.1 and 2.2 are proved in Section 2.8.

2.5 Simulation Studies

Simulation studies were conducted to examine the small-sample performance of the proposed approach, and to compare it with the performances based on the nonparametric and logistic regression methods (Pepe and Thompson 2000). We used an equal number of diseased and non-diseased subjects but varied the total sample size n from

200 to 400. We considered two biomarker measurements for each subject (K=2). Moreover, two transformation functions were set to $H_1(X) = \tan(X_1)$ for biomarker 1 and $H_2(X) = 2\log(X_2+1)$ for biomarker 2, very non-linear transformations. No covariates, Z, were used in the simulations. Specifically, biomarker data were generated from the following models: among non-diseased subjects,

$$\begin{pmatrix} H_1(X_1) \\ H_2(X_2) \end{pmatrix} = a + \begin{pmatrix} \mu_{01} \\ \mu_{02} \end{pmatrix} + \begin{pmatrix} \epsilon_{01} \\ \epsilon_{02} \end{pmatrix}, \quad \begin{pmatrix} \epsilon_{01} \\ \epsilon_{02} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right),$$

and among diseased subjects,

$$\begin{pmatrix} H_1(X_1) \\ H_2(X_2) \end{pmatrix} = a + \begin{pmatrix} \mu_{11} \\ \mu_{12} \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \end{pmatrix}, \quad \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^2 & 0 \\ 0 & \sigma_{12}^2 \end{pmatrix} \right),$$

where a was a random effect generated from a normal distribution with mean zero and variance σ_a^2. In the simulation studies, we selected the true parameters as $\mu_{01} = -2.5$, $\mu_{02} = 1.5$, $\mu_{11} = 1$, $\mu_{12} = 2.5$, $\sigma_{11}^2 = 2.3$, $\sigma_{12}^2 = 2.7$ and $\sigma_a^2 = 4.5$. We also considered the situation without detection limits as well as the one with detection limits. For the latter, we let $m_1 = -1.4, M_1 = 1.3$ and $m_2 = -0.7, M_2 = 25$, which resulted in about 4% left censoring and 8% right censoring for biomarker 1, and about 4% left and right censoring for biomarker 2.

For each simulated data, we applied the EM algorithm to calculate the NPMLEs. In the E-step, the conditional expectations for the functions of random effect a_i were evaluated by the Gaussian-Hermite approximation, where 20 quadratures were used. The maximization in the M-step was carried out by the Matlab optimization toolbox. Specifically, the gradient and Hessian matrix of the conditional expectation of the complete log-likelihood function were provided by us for the optimization; a "fminunc" function was used, where each search entails the use of a subspace trust region method

based on the interior-reflective Newton method (Coleman and Li 1994, 1996). For each M-step, we set the maximal number of the quasi-Newton search to be five. When the EM algorithm converged to fixed points, we calculated the observed information matrix using Louis' formula (see Appendix A.2). The inverse of this matrix is considered as an estimation of the asymptotic covariance of the parameter estimators. Finally, we computed the optimal AUC using formula (2.3), and estimated its variance using the Delta method.

Table 2.1 and Table 2.2 summarize the respective simulation results from 1,000 replicates for the situation without detection limits and the one with detection limits, respectively. Column "Est" is the average value of the estimates from 1,000 replicates; column "ASE" is the average of the estimated standard errors; column "SE" is the standard deviation of the estimates; column "CP" gives the ($100\times$) coverage proportion of the 95% confidence intervals based on the asymptotic normality. Overall, the estimated parameters and transformations are very close to the actual values across sample sizes, and the estimated standard errors using the observed information matrix approximate the empirical standard errors well. In addition, the coverage proportions of 95% CIs approach the nominal level of 95% as the sample size increases. Next, we present the true and estimated transformation values at six fixed points at each scenario of sample sizes in Table 2.1. Similarly, the true and estimated transformation values are shown in Table 2.2, but the transformations at the minimum and maximum points in Table 2.1 are excluded by observing detection limits. This shows that the estimation of the unknown transformation is pretty accurate in small samples, even at the tails of biomarker distributions. As shown in figures 2.1 and 2.2, the empirical transformation functions are almost identical to the true transformation functions for both biomarkers even with the small sample size, no matter whether biomarkers are censored or not.

Two biomarkers generated in the simulation studies are highly correlated. For instance, from the parameter estimators in Table 2.1, it can be shown that the estimated correlations between two biomarkers is 0.64 for diseased and 0.83 for non-diseased subjects when a sample size is 200. At the bottom of Tables 2.1 and 2.2, the proposed optimal AUC (AUC_{opt}), with the nonparametric AUC (AUC_{emp}) and the AUC based on the logistic model (AUC_{lgit}), are presented. We can see that for highly correlated biomarkers, the proposed optimal AUC estimate is larger than optimal AUC estimates by the nonparametric and logistic regression based methods, and its empirical standard error is smaller than those by the other two methods. Next, we generated two biomarkers having low correlation and estimated the optimal AUCs using the three approaches mentioned above. We found that the proposed optimal AUC is similar to those by the nonparametric and logistic model-based methods when biomarkers have a weak association.

2.6 Applications

2.6.1 Brain Tumor Data

We apply our method to analyzing brain tumor imaging data from magnetic resonance angiograms. The American Cancer Society estimates that approximately 20,000 Americans were diagnosed with brain tumors in 2008, so there is clinical need for a reliable, noninvasive method of assessing tumor malignancy and evaluating treatment response. One critical stage in tumor growth is the establishment of blood supply, and magnetic resonance imaging is useful for detecting malignancy because it can search for the foci of neoangiogenesis of abnormal vascular permeability. Bullitt et al. (2003) extracted the quantitative measurements of vessel shapes, as visualized by magnetic resonance

Table 2.1: Simulation results with complete biomarkers

		$n = 200$				$n = 400$			
Par.	True	Est	ASE	SE	CP	Est	ASE	SE	CP
μ_{01}	-2.5	-2.567	0.348	0.397	91.4	-2.551	0.242	0.256	93.5
μ_{02}	1.5	1.564	0.310	0.355	92.5	1.531	0.214	0.232	93.4
μ_{11}	1	1.047	0.351	0.372	94.4	1.015	0.243	0.245	94.9
μ_{12}	2.5	2.576	0.379	0.418	93.6	2.542	0.263	0.269	95.0
σ^2_{11}	2.3	2.526	1.170	1.316	92.0	2.396	0.789	0.765	95.0
σ^2_{12}	2.7	2.791	1.028	1.111	94.1	2.823	0.711	0.769	94.2
σ^2_a	4.5	4.859	0.998	1.519	89.6	4.711	0.676	0.737	93.2
$H_1(-4\pi/10)$	-3.078	-3.143	0.348	0.425	91.7	-3.135	0.240	0.255	93.6
$H_1(-2\pi/10)$	-0.727	-0.743	0.173	0.183	94.0	-0.737	0.121	0.123	95.5
$H_1(-\pi/10)$	-0.325	-0.331	0.116	0.120	94.5	-0.331	0.082	0.081	94.6
$H_1(\pi/10)$	0.325	0.336	0.119	0.127	95.3	3.313	0.083	0.086	94.9
$H_1(2\pi/10)$	0.727	0.744	0.183	0.198	93.2	0.733	0.127	0.132	94.9
$H_1(4\pi/10)$	3.078	3.180	0.491	0.564	91.8	3.113	0.334	0.333	95.4
$H_2(-0.5)$	-1.386	-1.415	0.263	0.289	92.5	-1.410	0.183	0.188	94.4
$H_2(0.5)$	0.811	0.829	0.172	0.176	94.5	0.819	0.120	0.119	95.2
$H_2(4)$	3.219	3.297	0.339	0.393	91.2	3.265	0.232	0.249	93.2
$H_2(8)$	4.394	4.506	0.430	0.510	91.0	4.460	0.294	0.311	92.8
$H_2(12)$	5.130	5.264	0.498	0.599	91.4	5.217	0.341	0.360	93.8
$H_2(20)$	6.089	6.257	0.607	0.735	91.0	6.198	0.415	0.446	93.3
AUC_{opt}	0.8812	0.8841	0.020	0.025	88.9	0.8826	0.014	0.017	90.7
AUC_{emp}		0.8486		0.027		0.8457		0.019	
AUC_{lgit}		0.8473		0.028		0.8451		0.020	

Table 2.2: Simulation results with censored biomarkers

		$n = 200$				$n = 400$			
Par.	True	Est	ASE	SE	CP	Est	ASE	SE	CP
μ_{01}	-2.5	-2.559	0.342	0.378	92.3	-2.546	0.241	0.256	94.5
μ_{02}	1.5	1.554	0.304	0.324	94.5	1.530	0.212	0.230	93.2
μ_{11}	1	1.041	0.348	0.363	95.0	1.011	0.243	0.245	94.1
μ_{12}	2.5	2.565	0.373	0.393	94.3	2.537	0.261	0.270	95.1
σ^2_{11}	2.3	2.492	1.170	1.287	92.1	2.393	0.811	0.798	96.1
σ^2_{12}	2.7	2.784	1.052	1.113	93.7	2.820	0.745	0.799	94.1
σ^2_a	4.5	4.750	0.983	1.139	92.0	4.683	0.672	0.733	93.0
$H_1(-2\pi/10)$	-0.727	-0.740	0.172	0.177	93.9	-0.736	0.120	0.124	95.4
$H_1(-\pi/10)$	-0.325	-0.329	0.116	0.118	95.3	-0.330	0.082	0.081	95.4
$H_1(\pi/10)$	0.325	0.335	0.118	0.123	95.7	0.331	0.083	0.086	94.8
$H_1(2\pi/10)$	0.727	0.743	0.182	0.190	94.4	0.732	0.126	0.133	94.6
$H_2(0.5)$	0.811	0.825	0.171	0.169	94.5	0.818	0.120	0.119	94.8
$H_2(4)$	3.219	3.281	0.334	0.342	94.0	3.261	0.231	0.250	93.5
$H_2(8)$	4.394	4.484	0.424	0.436	93.4	4.454	0.293	0.314	92.5
$H_2(12)$	5.130	5.242	0.492	0.512	94.1	5.211	0.340	0.363	93.4
AUC_{opt}	0.8812	0.8842	0.020	0.024	90.0	0.8823	0.014	0.017	90.5
AUC_{emp}		0.8556		0.027		0.8529		0.019	
AUC_{lgit}		0.8488		0.027		0.8469		0.019	

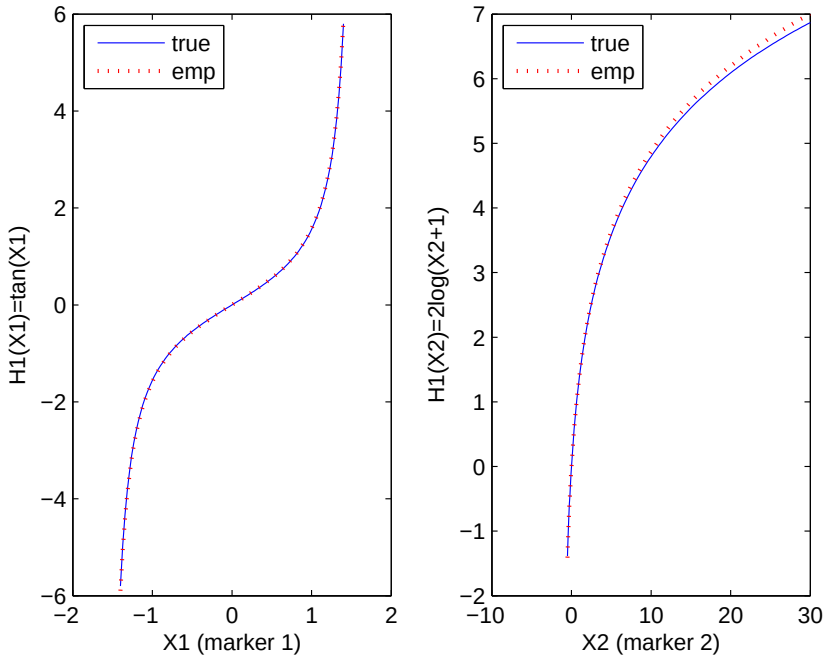

Figure 2.1: Plots of the true and estimated transformation functions without detection limits ($n = 400$): the solid curve is the true transformation and the dashed curve is the average of the estimated curves from 1,000 replicates.

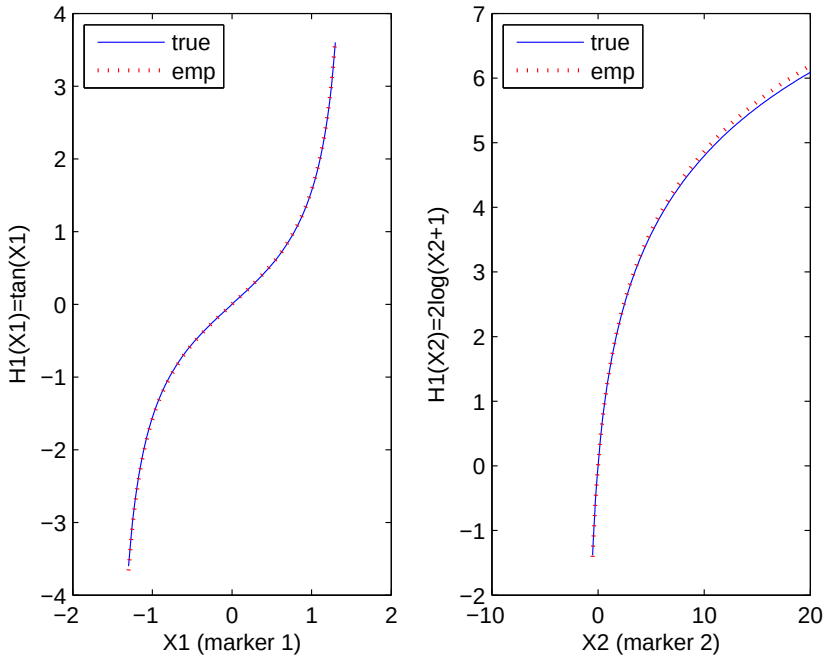

Figure 2.2: Plots of the true and estimated transformation functions with detection limits ($n = 400$): see Figure 2.1.

angiography, to assess tumor malignancy in a cross-sectional study and evaluate treatment response among patients undergoing anti-VEGF and cytotoxic therapy. The two most important biomarkers of tortuosity measures are (a) the sum of angles metric (SOAM), which is the sum of the angles between consecutive trios of points along the space curve represented by the vessel skeleton and then normalized by path length to measure the vessel curvature; (b) the inflection count metric (ICM), which counts inflection points along each space curve, multiplies this number (plus one) times the total path length, and then divides the product by the distance between endpoints. The data we used contain both SOAM and ICM biomarker measurements from 45 regions of interest, where 11 tumors were identified as benign and 34 as malignant. To apply our approach, we assume models (2.1) and (2.2) hold for these two biomarkers; that is, some non-decreasing unknown transformations of SOAM and ICM follow a bivariate normal distribution in both benign and malignant groups. After using the EM algorithm for inference, we report the parameter estimates in Table 2.3. Table 2.3 shows that the biomarkers tend to have larger means and smaller variances for malignant tumors than for benign tumors. The dependence between the two transformed biomarkers is low, with $\sigma_a^2 = 0.0217$. The transformations for SOAM and ICM, plotted in Figure 2.3, suggest that the quadratic transformations may be appropriate for obtaining the normality. Figure 2.4 further demonstrates that our transformations yield the normality and can actually downsize some extreme observations in the original biomarkers.

With our approach, the optimal linear combination of these two biomarkers is $1.455\widehat{H}_1(SOAM) + 0.676\widehat{H}_2(ICM)$, where $\widehat{H}_1$ and $\widehat{H}_2$ are the two estimated transformations. The optimal AUC estimate is 0.9823 with a standard error of 0.016, while the logistic regression yields a combination of 2.8270(SOAM) + 0.1376(ICM) with an optimal AUC estimate of 0.9893, which is very close to the optimal AUC estimate based

Table 2.3: Analysis of brain tumor imaging data

Parameter	Estimate	SE	P-value
μ_{01}	-0.094	0.3232	0.7716
μ_{02}	-0.323	0.3358	0.3360
μ_{11}	2.449	0.9773	0.0122
μ_{12}	0.751	0.2611	0.0040
σ^2_{11}	0.685	0.9839	0.4866
σ^2_{12}	0.452	0.2457	0.0659
σ^2_a	0.022	0.1213	0.8580

on the nonparametric method. There appears to be little difference among three approaches, possibly because of the high discrimination power from each single biomarker (the AUC for SOAM only is 0.9759, and the AUC for ICM only is 0.8262) and the low dependence between the two biomarkers. However, in this tumor data, the advantage of normalizing biomarkers in our approach is evident even with such a small sample size.

2.6.2 Prostate Cancer Data

Next, we illustrate our approach with a prostate cancer study. Serum prostate-specific antigen (PSA) is the most widely used biomarker to detect prostate cancer. We used a dataset of 71 prostate cancer subjects and 71 controls who participated in the Beta-Carotene and Retinol Efficacy Trial (CARET) which is a randomized lung cancer prevention study that began in 1985 and terminated in 1994. Subjects had serum samples drawn at baseline and at two-year intervals thereafter (Goodman et al., 1993; Etzioni et al. 1999). We considered the single serum sample per subject measured at the time closest to the diagnosis among multiple samples accumulated before a participant was diagnosed as having prostate cancer or not, leaving 139 men (71 cases and 68 controls) in the dataset.

Two measures of PSA, total PSA and the ratio of free to total PSA, were used for

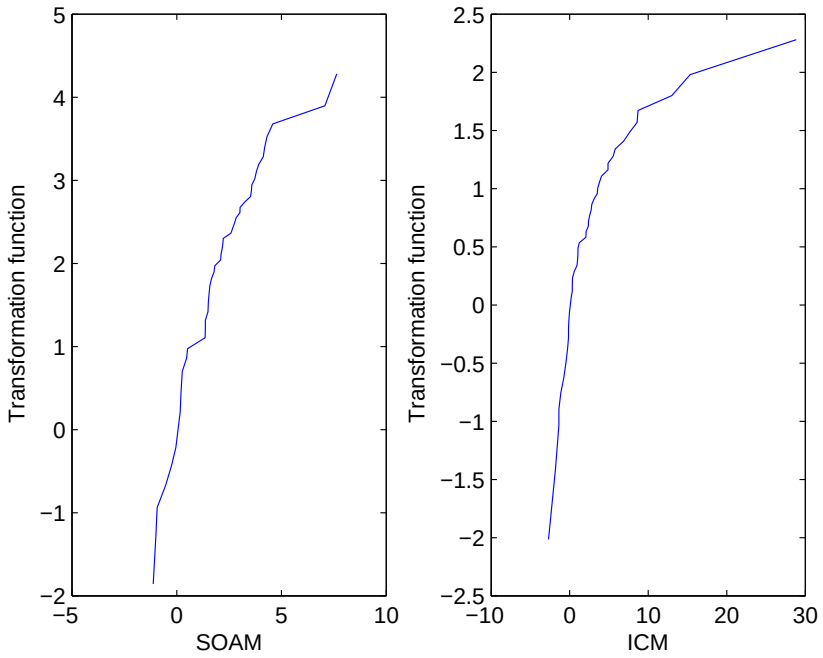

Figure 2.3: Estimated transformations for biomarkers in brain tumor imaging data

prostate cancer. We standardized the total serum PSA and the ratio of free to total PSA measurements by subtracting their mean values, and defined the standardized PSA values as Y_1=log(total PSA) and Y_2=-log(free PSA/total PSA). The correlation between these two biomarkers is fairly high with a correlation coefficient of 0.56521.

Parameter estimates from models (2.1) and (2.2) are given in Table 2.4, showing that Y_1 and Y_2 have higher means and variances in cases than in controls. The estimated correlations between two biomarkers is about 0.41 for diseased and about 0.57 for non-diseased subjects, suggesting the high dependence of the two biomarkers. The transformations for Y_1 and Y_2 given in Figure 2.5 appear to be linear and Figure 2.6 shows that the histograms after transformations are closer to a normal shape than those before transformations.

We found that the optimal linear combination is $0.5562\hat{H}_1(Y_1) + 0.0213\hat{H}_2(Y_2)$, and its corresponding AUC is 0.9199 (SE 0.0210). On the other hand, the optimal linear

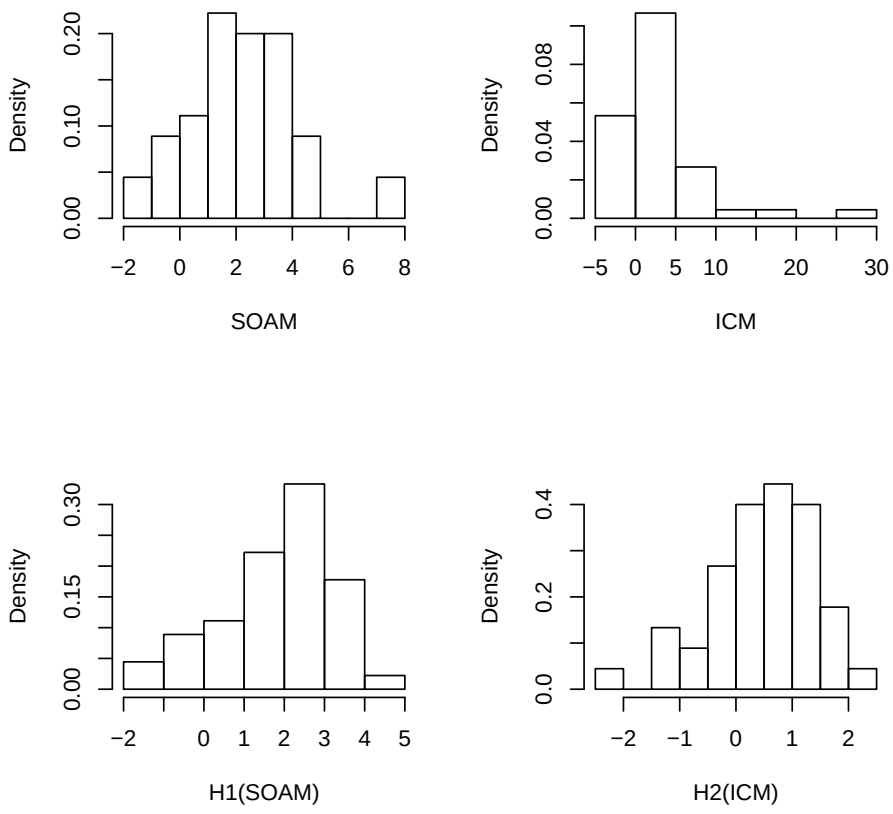

Figure 2.4: Histogram of biomarkers before transformation and after transformation in brain tumor imaging data

Table 2.4: Analysis of prostate cancer data

Parameter	Estimate	SE	P-value
μ_{01}	-2.032	0.2737	<.0001
μ_{02}	-1.148	0.2313	<.0001
μ_{11}	1.456	0.3825	1.9999
μ_{12}	0.433	0.2437	1.9244
σ^2_{11}	2.520	1.1182	1.9758
σ^2_{12}	1.405	0.6208	1.9764
σ^2_a	1.324	0.3496	1.9998

combination based on the nonparametric method is $Y_1 + 0.191Y_2$, giving an AUC estimate of 0.9095 (SE 0.0252 based on 1000 bootstrap replicates). In addition, the logistic regression yields $2.0494Y_1 + 0.4886Y_2$ and a corresponding AUC estimate of 0.9089 (SE 0.0256 based on 1000 bootstrap replicates). The coefficients of the optimal linear combinations from all three methods suggest that total PSA performed better than the ratio of free to total PSA measured closest to the diagnosis. However, the proposed method gives a larger value of optimal AUC than those from the nonparametric and logistic regression methods, demonstrating that our method performs better than the other two methods for highly correlated biomarkers.

2.7 Discussion

The proposed transformation model has several advantages over Pepe and Thompson's (2000) nonparametric method. First, we allow a completely unknown transformation, so the diagnostic rule is less sensitive to the extreme values of biomarkers. Second, since our method is applied to the biomarkers characterized by left or right censoring and accounts for the correlated structure of biomarkers by introducing random effects, it describes the data properly. Third, in our method, finding optimal coefficients for the linear combination of biomarkers does not require extensive computation, as the

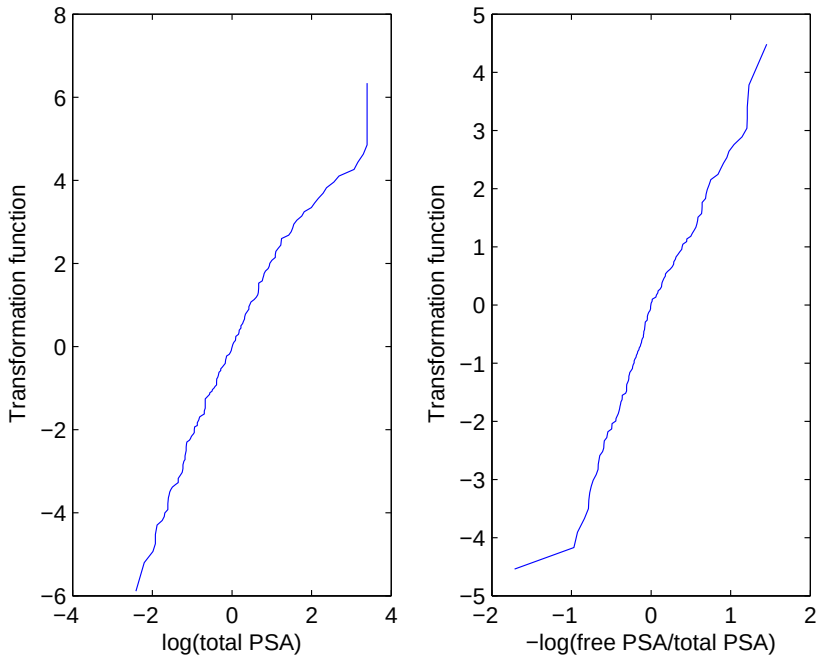

Figure 2.5: Estimated transformations for biomarkers in prostate cancer data

nonparametric method does.

Our method combines multiple biomarkers but does not address selecting biomarkers. We will generalize our method to high dimensional data such as microarray data and will propose a method for biomarker selection and classification in an ROC framework. In addition, we will develop a method for longitudinal biomarkers and methods accounting for the situations where disease status is not binary but in continuous scale and the imperfect disease status is presented.

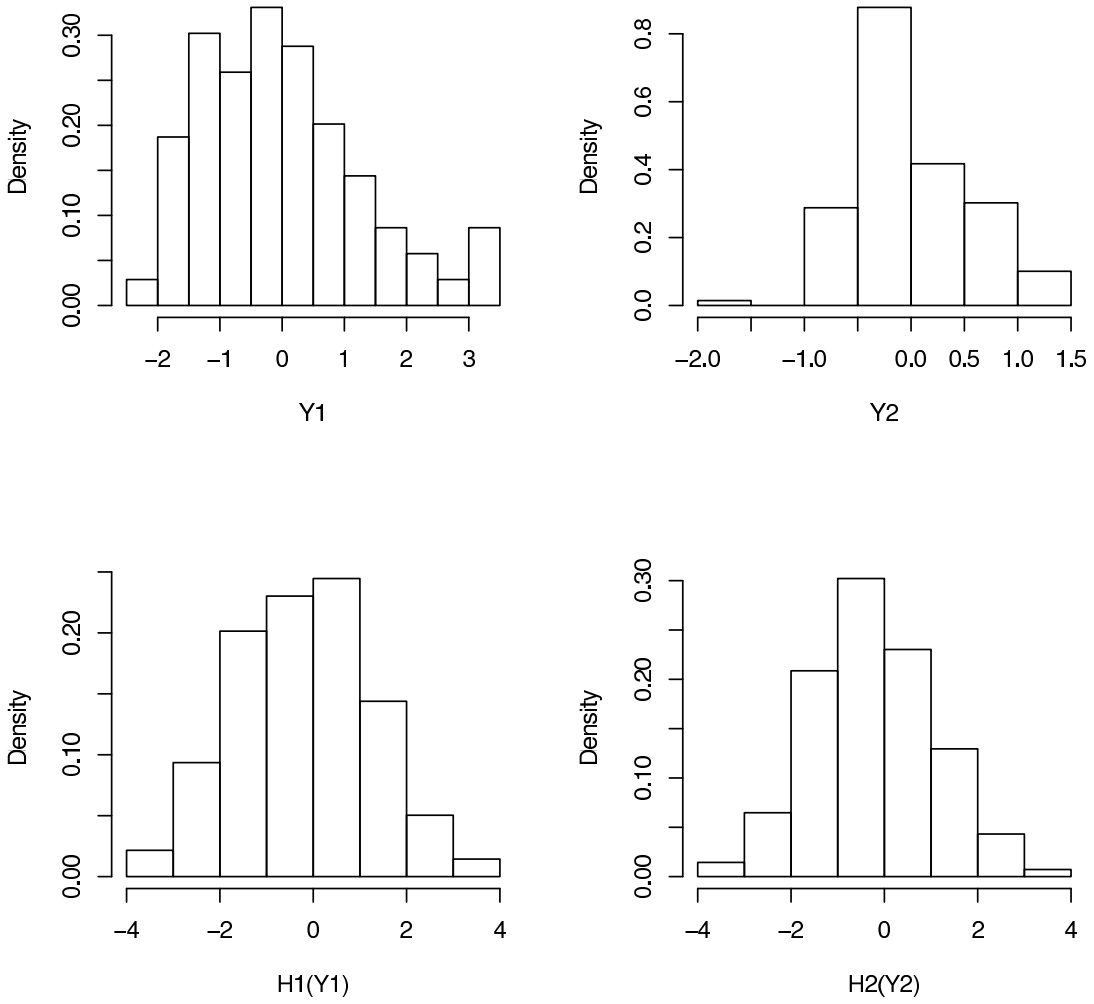

Figure 2.6: Histogram of biomarkers before transformation and after transformation in prostate cancer data

2.8 Appendix A

2.8.1 Proofs of Theorems 2.1 and 2.2

The likelihood (2.6) in section 2.4 has a same form as that of the transformation models with random effects for dependent failure times in Zeng and Lin (2007, 2009) and the proofs of Theorems 2.1 and 2.2 are equivalent to the asymptotic proof for the multiple type of events in Zeng and Lin (2009). Therefore, we only need to verify their conditions (D1)-(D8).

First of all, Zeng and Lin (2009)'s condition (D1) is equivalent to our condition (C1). The condition (D2) naturally holds for time independent covariates Z, (D3) is always true because censoring time is same as the duration of the study ($C_l = \tau$), and (D4) is equivalent to our condition (C2). Note that function G_l in (2.6) is equal to $-\log\{1 - \Phi(\log x)\}$ for $D_i = 0$ and is $-\log\{1 - \Phi(\log x/\sigma_{1l})\}$ for $D_i = 1$. Zeng and Lin (2007, 2009) verified the linear transformation model $G_l(x) = -\log\{1 - \Phi(\log x)\}$ satisfies the following condition (D5). Even though $G_l(x)$ for the diseased group depends on σ_{1l}, the condition (D5) can be proven by exactly same arguments as Zeng and Lin (2009) assuming that $\sigma_{1l} > 0$. Next, (D6) holds naturally for the normally distributed random effect a. (D7) and (D8) are conditions to ensure parameter identifiability and non-singular information matrix, and these conditions are satisfied by the following Propositions A1 and A2.

Proposition A1. Suppose two sets of parameters $(\boldsymbol{\theta}, H_1, ..., H_K)$ and $(\widetilde{\boldsymbol{\theta}}, \widetilde{H}_1, ..., \widetilde{H}_k)$ give the same observed likelihood function with probability one. Then $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}$, $\Sigma_a = \widetilde{\Sigma}_a$ and $H_k = \widetilde{H}_k$ for $k = 1, ..., K$.

proof.

Let $\theta_k = (\mu_{0k}, \mu_{1k}, \sigma^2_{0k}, \sigma^2_{1k})$ and $\widetilde{\theta}_k = (\widetilde{\mu}_{0k}, \widetilde{\mu}_{1k}, \widetilde{\sigma}^2_{0k}, \widetilde{\sigma}^2_{1k})$. It suffices to show that

$l_k(H_k, \theta_k, \Sigma_a|X, Z) = l_k(\widetilde{H}_k, \widetilde{\theta}_k, \widetilde{\Sigma}_a|X, Z)$ implies $H_k = \widetilde{H}_k, \theta_k = \widetilde{\theta}_k$ and $\Sigma_a = \widetilde{\Sigma}_a$ for a fixed k $(k = 1, \cdots, K)$.

The marginal densities for kth biomarker given covariates and disease status are

$$H_k(X)|Z, D = 0 \sim N(\mu'_{0k}Z, Z'\Sigma_a Z + \sigma^2_{0k}),\ H_k(X)|Z, D = 1 \sim N(\mu'_{1k}Z, Z'\Sigma_a Z + \sigma^2_{1k}).$$

Similarly, for $(\widetilde{H}_k, \widetilde{\theta}_k, \widetilde{\Sigma}_a)$

$$\widetilde{H}_k(X)|Z, D = 0 \sim N(\widetilde{\mu}'_{0k}Z, Z'\widetilde{\Sigma}_a Z + \widetilde{\sigma}^2_{0k}),\ \widetilde{H}_k(X)|Z, D = 1 \sim N(\widetilde{\mu}'_{1k}Z, Z'\widetilde{\Sigma}_a Z + \widetilde{\sigma}^2_{1k}).$$

First, assume $l_k(H_k, \theta_k, \Sigma_a|Z, D = 0) = l_k(\widetilde{H}_k, \widetilde{\theta}_k, \widetilde{\Sigma}_a|Z, D = 0)$. Then,

$$\frac{(H_k(X) - \mu'_{0k}Z)^2}{Z'\Sigma_a Z + \sigma^2_{0k}} = \frac{(\widetilde{H}_k(X) - \widetilde{\mu}'_{0k}Z)^2}{Z'\widetilde{\Sigma}_a Z + \widetilde{\sigma}^2_{0k}}. \tag{2.8}$$

If $(H_k(X) - \mu'_{0k}Z) = \gamma_{0k}(\widetilde{H}_k(X) - \widetilde{\mu}'_{0k}Z)$, $\gamma_{0k} = \sqrt{\dfrac{Z'\Sigma_a Z + \sigma^2_{0k}}{Z'\widetilde{\Sigma}_a Z + \widetilde{\sigma}^2_{0k}}}$ and

$$Cov(H_k(X), H_l(X)|D = 0) = Z'\Sigma_a Z = \gamma_{0k}\gamma_{0l}Z'\widetilde{\Sigma}_a Z. \tag{2.9}$$

Under the restriction $\sigma^2_{0k} = \widetilde{\sigma}^2_{0k} = 1$, $\gamma_{0k} = \sqrt{\dfrac{Z'\Sigma_a Z + 1}{Z'\widetilde{\Sigma}_a Z + 1}}$ and from (2.9)

$$Z'\Sigma_a Z = \Big(\frac{Z'\Sigma_a Z + 1}{Z'\widetilde{\Sigma}_a Z + 1}\Big) Z'\widetilde{\Sigma}_a Z \iff Z'\Sigma_a Z = Z'\widetilde{\Sigma}_a Z.$$

Therefore, $\Sigma_a = \widetilde{\Sigma}_a$ and $\gamma_{0k} = 1$. Since $H_k(0) = 0$, it can be shown $\mu_{0k} = \widetilde{\mu}_{0k}$ and $H_k = \widetilde{H}_k$ from (2.8).

Likewise, if we assume $l_k(H_k, \theta_k, \Sigma_a|Z, D = 1) = l_k(\widetilde{H}_k, \widetilde{\theta}_k, \widetilde{\Sigma}_a|Z, D = 1)$ for the diseased group,

$$\frac{(H_k(X) - \mu'_{1k}Z)^2}{Z'\Sigma_a Z + \sigma^2_{1k}} = \frac{(H_k(X) - \widetilde{\mu}'_{1k}Z)^2}{Z'\widetilde{\Sigma}_a Z + \widetilde{\sigma}^2_{1k}}. \tag{2.10}$$

If

$$(H_k(X) - \mu'_{1k}Z) = \gamma_{1k}(H_k(X) - \widetilde{\mu}'_{1k}Z), \tag{2.11}$$

$Cov(H_k(X), H_l(X)|D = 1) = Z'\Sigma_a Z = \gamma_{1k}\gamma_{1l}Z'\widetilde{\Sigma}_a Z = \gamma_{1k}\gamma_{1l}Z'\Sigma_a Z$, which gives $\gamma_{1k} = \gamma_{1l} = 1$. Then, $\mu_{1k} = \widetilde{\mu}_{1k}$ from (2.11) and $\sigma^2_{1k} = \widetilde{\sigma}^2_{1k}$ from (2.10).

Proposition A2. If $H_k = H_{k0} + \epsilon \int h_k dH_{k0}$ $(k = 1, \cdots, K)$, $\mu_0 = \mu_{00} + \epsilon\nu_0$, $\mu_1 = \mu_{10} + \epsilon\nu_1$, $\Sigma_0 = \Sigma_{00} + \epsilon M_0$, $\Sigma_1 = \Sigma_{10} + \epsilon M_1$, and $\Sigma_a = \Sigma_{a0} + \epsilon M_a$, then $\int h_k dH_{k0} = 0$ and ν_0, ν_1, M_0 and M_a are zero matrices.

proof.

Define $H(\mathbf{X}_i) = (H_1(X_{i1}), \cdots, H_K(X_{iK}))'$. Note that conditional on disease status,

$$H(\mathbf{X}_i)|D_i = 0 \sim N(\mu_0 Z_i, Z_i'\Sigma_a Z_i \mathbf{1}\mathbf{1}' + \Sigma_0),\ H(\mathbf{X}_i)|D_i = 1 \sim N(\mu_1 Z_i, Z_i'\Sigma_a Z_i \mathbf{1}\mathbf{1}' + \Sigma_1).$$

First of all, the non-diseased group log likelihood function can be expressed as

$$\begin{aligned} \log l_{0i} = &\sum_{i=1}^{K} \log H'_k(X_{ik}) - \frac{1}{2}\log|Z_i'\Sigma_a Z_i \mathbf{1}\mathbf{1}' + \Sigma_0| \\ &-\frac{1}{2}(H(\mathbf{X}_i) - \mu_0 Z_i)'(Z_i'\Sigma_a Z_i \mathbf{1}\mathbf{1}' + \Sigma_0)^{-1}(H(\mathbf{X}_i) - \mu_0 Z_i). \end{aligned}$$

Define $H_k = H_{k0} + \epsilon \int h_k dH_{k0}$ $(k = 1, \cdots, K)$ with $H_0(\mathbf{X}_i) = (H_{10}(X_{i1}), \cdots, H_{K0}(X_{iK}))'$ and $h^* = (\int h_1 dH_{10}, \cdots, \int h_K dH_{K0})'$, $\mu_0 = \mu_{00} + \epsilon\nu_0$, $\mu_1 = \mu_{10} + \epsilon\nu_1$, $\Sigma_0 = \Sigma_{00} + \epsilon M_0$, $\Sigma_1 = \Sigma_{10} + \epsilon M_1$, and $\Sigma_a = \Sigma_{a0} + \epsilon M_a$. Then,

$$\begin{aligned} \log l_{0i} = &\sum_{i=1}^{K} \log(H'_{k0}(X_{ik}) + \epsilon h_k(X_{ik}) H'_{k0}(X_{ik}))) - \frac{1}{2}\log|Z_i'(\Sigma_{a0} + \epsilon M_a) Z_i \mathbf{1}\mathbf{1}' + \Sigma_0| \\ &-\frac{1}{2}\Big(H_0(\mathbf{X}_i) + \epsilon h^* - (\mu_{00} + \epsilon\nu_0)Z_i\Big)'\Big(Z_i'(\Sigma_{a0} + \epsilon M_a) Z_i \mathbf{1}\mathbf{1}' + (\mu_{00} + \epsilon\nu_0)\Big)^{-1}\Big(H_0(\mathbf{X}_i) \\ &+\epsilon h^* - (\mu_{00} + \epsilon\nu_0)Z_i\Big). \end{aligned} \tag{2.12}$$

First, M_0 is a zero matrix by the definition of Σ_0. We will show all elements of h^*, ν_0, ν_1, and M_a are equal to zeros. $\frac{\partial log l_{0i}}{\partial \epsilon}\Big|_{\epsilon=0}$ is given by

$$\begin{aligned} l_0^* &= \sum_{k=1}^{K} h_k(X_{ik}) - \frac{1}{2}\Big\{(Z_i'M_aZ_i)tr((Z_i'\Sigma_{a0}Z_i+\Sigma_{00})^{-1}\mathbf{1}\mathbf{1}') + tr((Z_i'\Sigma_{a0}Z_i\mathbf{1}\mathbf{1}'+\Sigma_{00})^{-1}M_0)\Big\} \\ &\quad -(h^*(\mathbf{X}_i)-\nu_0 Z_i)'(Z_i'\Sigma_{a0}Z_i\mathbf{1}\mathbf{1}'+\Sigma_{00})^{-1}(H_0(\mathbf{X}_i)-\mu_{00}Z_i) + \frac{1}{2}(H_0(\mathbf{X}_i)-\mu_{00}Z_i)' \\ &\quad (Z_i'\Sigma_{a0}Z_i\mathbf{1}\mathbf{1}'+\Sigma_{00})^{-1}(Z_i'M_aZ_i\mathbf{1}\mathbf{1}'+M_0)(Z_i'\Sigma_{a0}Z_i+\Sigma_{00})^{-1}(H_0(\mathbf{X}_i)-\mu_{00}Z_i). \quad (2.13) \end{aligned}$$

Let $A_0 = (a_{ij}) = (Z_i'\Sigma_{a0}Z_i\mathbf{1}\mathbf{1}'+\Sigma_{00})^{-1}$ and $B_0 = (b_{ij}) = (Z_i'\Sigma_{a0}Z_i\mathbf{1}\mathbf{1}'+\Sigma_{00})^{-1}(Z_i'M_aZ_i\mathbf{1}\mathbf{1}'+M_0)(Z_i'\Sigma_{a0}Z_i+\Sigma_{00})^{-1}$ $i,j=1,\cdots,K$.

$$\frac{\partial l_0^*}{\partial X_{ik}X_{il}} = -a_{kl}h_kH_{k0}'H_{l0}' - a_{lk}h_lH_{k0}'H_{l0}' + b_{kl}H_{k0}'H_{l0}' = 0 \quad (k\neq l),$$

which gives $a_{kl}(h_k+h_l)-b_{kl}=0$ so that h_k should be constant.

Define $h_k = c_k$, $C = diag(c_1,\cdots,c_K)$, $\mathbf{e} = (H_0(\mathbf{X}_i)-\mu_{00}Z_i)$, and $\mathbf{e}' = (H_0(\mathbf{X}_i)-\mu_{10}Z_i)$. Then l_0^* in (2.13) can be rewritten as

$$\begin{aligned} &\textstyle\sum_{k=1}^{K} c_k - \frac{1}{2}\Big\{(Z_i'M_aZ_i)tr((Z_i'\Sigma_{a0}Z_i\mathbf{1}\mathbf{1}'+\Sigma_{00})^{-1}\mathbf{1}\mathbf{1}') + tr((Z_i'\Sigma_{a0}Z_i\mathbf{1}\mathbf{1}'+\Sigma_{00})^{-1}M_0)\Big\} \\ &\quad +\mathbf{e}'\Big\{\frac{1}{2}(Z_i'\Sigma_{a0}Z_i\mathbf{1}\mathbf{1}'+\Sigma_{00})^{-1}(Z_i'M_aZ_i\mathbf{1}\mathbf{1}'+M_0)(Z_i'\Sigma_{a0}Z_i+\Sigma_{00})^{-1} \\ &\quad -C'(Z_i'\Sigma_{a0}Z_i\mathbf{1}\mathbf{1}'+\Sigma_{00})^{-1}\Big\}\mathbf{e} - Z_i'(C\mu_{00}-\nu_0)'(Z_i'\Sigma_{a0}Z_i\mathbf{1}\mathbf{1}'+\Sigma_{00})^{-1}\mathbf{e} \\ &= 0. \end{aligned}$$

Therefore, the following two conditions hold.

$$(A1)\ Z_i'M_aZ_i\mathbf{1}\mathbf{1}' = 2(Z_i'\Sigma_{a0}Z_i\mathbf{1}\mathbf{1}'+\Sigma_{00})C \quad (A2)\ C\mu_{00}=\nu_0$$

Condition (A1) implies $Z_i'M_aZ_i = c_k(Z_i'\Sigma_{a0}Z_i+1)$ and $Z_i'M_aZ_i = c_kZ_i'\Sigma_{a0}Z_i$. Thus, it can be shown that $c_k=0$, $\int h_k dH_{k0}=0$ and $\nu_0=O$. $Z_i'M_aZ_i=O$ implies $M_a=O$.

Similarly, for the diseased group, $l_1^* = \left.\frac{\partial log l_{1i}}{\partial \epsilon}\right|_{\epsilon=0}$ is same as

$$\sum_{k=1}^{K} c_k - \frac{1}{2}\Big\{(Z_i' M_a Z_i) tr((Z_i'\Sigma_{a0} Z_i \mathbf{1}\mathbf{1}' + \Sigma_{10})^{-1}\mathbf{1}\mathbf{1}') + tr((Z_i'\Sigma_{a0} Z_i \mathbf{1}\mathbf{1}' + \Sigma_{10})^{-1} M_1)\Big\}$$
$$+\mathbf{e}'\Big\{\frac{1}{2}(Z_i'\Sigma_{a0} Z_i \mathbf{1}\mathbf{1}' + \Sigma_{10})^{-1}(Z_i' M_a Z_i \mathbf{1}\mathbf{1}' + M_1)(Z_i'\Sigma_{a0} Z_i + \Sigma_{10})^{-1}$$
$$-C'(Z_i'\Sigma_{a0} Z_i \mathbf{1}\mathbf{1}' + \Sigma_{10})^{-1}\Big\}\mathbf{e} - Z_i'(C\mu_{10} - \nu_1)'(Z_i'\Sigma_{a0} Z_i \mathbf{1}\mathbf{1}' + \Sigma_{10})^{-1}\mathbf{e}$$
$$= 0.$$

Using the same argument as before, we obtain

$$(A3)\ Z_i' M_a Z_i \mathbf{1}\mathbf{1}' + M_1 = 2(Z_i'\Sigma_{a0} Z_i \mathbf{1}\mathbf{1}' + \Sigma_{10})C \quad (A4)\ C\mu_{10} = \nu_1$$

Since M_a and C are zero matrices, M_1 and ν_1 are also zero matrices from the condition (A3) and (A4), respectively.

2.9 Appendix B

2.9.1 EM Algorithm

Define (X_{i1}, X_{i2}) are two continuous biomarkers from subject i $(i = 1, \cdots, n)$. It is assumed that there exist non-decreasing transformations H_1 and H_2 such that $(H_1(X_1), H_2(X_2))$ follows a multivariate normal distribution in each of diseased and non-diseased groups;

$$[H_1(X_1), H_2(X_2)|D = 0] = a_i + (\epsilon_{i1}, \epsilon_{i2}),\ (\epsilon_{i1}, \epsilon_{i2}) \sim N(\mu_0, \text{diag}(1, 1)) \text{ and}$$

$$[H_1(X_1), H_2(X_2)|D = 1] = a_i + (\epsilon'_{i1}, \epsilon'_{i2}),\ (\epsilon'_{i1}, \epsilon'_{i2}) \sim N(\mu_1, \text{diag}(\sigma^2_{11}, \sigma^2_{12})),$$

with random effect $a_i \sim N(0, \sigma^2)$. The parameters to be estimated are $\mu_0(\mu_{01}, \mu_{02})$, $\mu_1(\mu_{11}, \mu_{12})$, $\Sigma_1(\sigma^2_{11}, \sigma^2_{12})$, jump sizes $(h^+_{ik}, h^-_{jk})_{k=1,2;i=1,\cdots,n^+_k;j=1,\cdots,n^-_k}$ and σ^2. The complete data likelihood for $\boldsymbol{\theta^*}$ and (H^+, H^-) is given by

$$\begin{aligned}
L_c(\boldsymbol{\theta^*}, H^+, H^-) \;=\; & \prod_{X_{ik}>0} \prod_{k=1}^{2} \Big[\big(h^+_k(X_{ik})\phi(H^+_k(X_{ik}) - \mu_{0k} - a_i)\big)^{\delta^O_{ik}} \\
& \qquad (1 - \Phi(H^+_k(M_k) - \mu_{0k} - a_i))^{\delta^R_{ik}} \Big]^{D_i=0} \\
& \qquad \Big[\big(h^+_k(X_{ik})\phi((H^+_K(X_{ik}) - \mu_{1k} - a_i)/\sigma_{1k})\big)\big)^{\delta^O_{ik}} \\
& \qquad (1 - \Phi((H^+_k(M_k) - \mu_{1k} - a_i)/\sigma_{1k}))^{\delta^R_{ik}} \Big]^{D_i=1} \phi(a_i; \sigma^2) \\
& \times \prod_{\widetilde{X}_{ik}>0} \prod_{k=1}^{2} \Big[\big(h^-_k(\widetilde{X}_{ik})\phi((H^-_k(\widetilde{X}_{ik}) + \mu_{0k} + a_i)\big)^{\delta^O_{ik}} \\
& \qquad (1 - \Phi(H^-_k(\widetilde{m}_k) + \mu_{0k} + a_i))^{1-\delta^O_{ik}-\delta^R_{ik}} \Big]^{D_i=0} \\
& \qquad \Big[\big(h^-_k(\widetilde{X}_{ik})\phi((H^-_k(\widetilde{X}_{ik}) + \mu_{1k} + a_i)/\sigma_{1k})\big)^{\delta^O_{ik}} \\
& \qquad (1 - \Phi((H^-_k(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k}))^{1-\delta^O_{ik}-\delta^R_{ik}} \Big]^{D_i=1} \phi(a_i; \sigma^2),
\end{aligned}$$

where $\phi(\cdot)$ is a standard normal density. Thus,

$$\hat{E}[l_c(\boldsymbol{\theta}^*, H^+, H^-) = \hat{E}[l_c(\boldsymbol{\theta}, H^+, H^-)] + \sum_{i=1}^{n} \hat{E}[\log\phi(a_i; \sigma^2)], \tag{2.14}$$

with

$$\begin{aligned}
\hat{E}[l_c(\boldsymbol{\theta}, H^+, H^-)] \;=\; & \sum_{X_{ik}>0} \sum_{k=1}^{2} \delta_{ik}^{O} \log(h_k^+(X_{ik})) + \sum_{X_{ik}>0} \sum_{k=1}^{2} \Big[\Big\{ -\frac{\delta_{ik}^{O}}{2} \Big((H_k^+(X_{ik}) - \mu_{0k})^2 \\
& -2\hat{E}(a_i)(H_k^+(X_{ik}) - \mu_{0k}) + \hat{E}[a_i^2] \Big) \\
& +\delta_{ik}^{R} \hat{E}[\log(1 - \Phi(H_k^+(M_k) - \mu_{0k} - a_i))] \Big\} (1 - D_i) \\
& +\Big\{ \delta_{ik}^{O} \Big(-\frac{1}{2}\log\sigma_{1k}^2 - \frac{1}{2\sigma_{1k}^2} [(H_k^+(X_{ik}) - \mu_{1k})^2 \\
& -2\hat{E}(a_i)(H_k^+(X_{ik}) - \mu_{1k}) + \hat{E}(a_i^2)] \Big) \\
& +\delta_{ik}^{R} \hat{E}[\log(1 - \Phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k}))] \Big\} D_i \Big] \\
& + \sum_{\widetilde{X}_{ik}>0} \sum_{k=1}^{2} \delta_{ik}^{O} \log(h_k^-(\widetilde{X}_{ik})) + \sum_{\widetilde{X}_{ik}>0} \sum_{k=1}^{2} \Big[\Big\{ -\frac{\delta_{ik}^{O}}{2} \Big((H_k^-(\widetilde{X}_{ik}) + \mu_{0k})^2 \\
& +2\hat{E}(a_i)(H_k^-(\widetilde{X}_{ik}) + \mu_{0k}) + \hat{E}[a_i^2] \Big) \\
& +(1 - \delta_{ik}^{O} - \delta_{ik}^{R}) \hat{E}[\log(1 - \Phi(H_k^-(\widetilde{m}_k) + \mu_{0k} + a_i))] \Big\} (1 - D_i) \\
& +\Big\{ \delta_{ik}^{O} \Big(-\frac{1}{2}\log\sigma_{1k}^2 - \frac{1}{2\sigma_{1k}^2} [(H_k^-(\widetilde{X}_{ik}) + \mu_{1k})^2 \\
& +2\hat{E}(a_i)(H_k^-(\widetilde{X}_{ik}) + \mu_{1k}) + \hat{E}(a_i^2)] \Big) \\
& +(1 - \delta_{ik}^{O} - \delta_{ik}^{R}) \hat{E}[\log(1 - \Phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k}))] \Big\} D_i \Big]
\end{aligned} \tag{2.15}$$

where $\hat{E}[\cdot]$ denotes the conditional expectation given the observed data and the current parameter estimates.

In the E-step, the conditional expectations of any functions of a_i are computed via Gaussian quadrature approximations since the density of $a_i|X_i$ does not have a closed

form. In detail, $\hat{E}[g(a_i)]$ is computed in the following manner.

$$\hat{E}[g(a_i)] = \int g(a_i) f(a_i|X_i) da_i = \int g(a_i) \frac{f(X_i|a_i)\phi(a_i;\sigma^2)}{f(X_i)} da_i = \frac{1}{f(X_i)} E_{a_i}[g(a_i) f(X_i|a_i)]$$

and $f(X_i) = \int f(X_i|a_i)\phi(a_i;\sigma^2)da_i = E_{a_i}[g(a_i)f(X_i|a_i)]$, where

$$\begin{aligned} f(X_i|a_i) &= \prod_{k=1}^{2} f(X_{ik}|a_i)^{\delta_{ik}^O} P(X_{ik} > M_k|a_i)^{\delta_{ik}^R} P(X_{ik} < m_k|a_i)^{1-\delta_{ik}^O-\delta_{ik}^R} \\ &= \prod_{k=1}^{2} \Big[\big(H_k'(X_{ik})\phi(H_k(X_{ik}) - \mu_{0k} - a_i)\big)^{\delta_{ik}^O} (1 - \Phi(H_k(M_k) - \mu_{0k} - a_i))^{\delta_{ik}^R} \\ &\qquad \Phi(H_k(m_k) - \mu_{0k} - a_i)^{1-\delta_{ik}^O-\delta_{ik}^R} \Big]^{D_i=0} \\ &\qquad \Big[\big(H_k'(X_{ik})\phi((H_K(X_{ik}) - \mu_{1k} - a_i)/\sigma_{1k})\big)^{\delta_{ik}^O} (1 - \Phi((H_k(M_k) \\ &\qquad -\mu_{1k} - a_i)/\sigma_{1k}))^{\delta_{ik}^R} \Phi((H_k(m_k) - \mu_{1k} - a_i)/\sigma_{1k})^{1-\delta_{ik}^O-\delta_{ik}^R} \Big]^{D_i=1}. \end{aligned}$$

Specifically, $\hat{E}[a_i]$ and $\hat{E}[a_i^2]$, conditional expectations of functions of a_i defined in (B.1) and (B.2) are computed via Gaussian quadrature approximations. Next, in the M-step, we maximize $\hat{E}[l_c(\boldsymbol{\theta}^*, H^+, H^-)$ from (2.14). Since σ^2 can be updated directly by maximizing $\sum_{i=1}^{n} \hat{E}[\log\phi(a_i;\sigma^2)]$, the rest of parameters $\mu_{0k}, \mu_{1k}, \sigma_{1k}^2, h_{ik}^+$ and h_{jk}^- $(k = 1, 2; i = 1, \cdots, n_k^+; j = 1, \cdots, n_k^-)$ will be updated in the M-step.

First of all, the score equations are obtained by differentiating $L \equiv \hat{E}[l_c(\boldsymbol{\theta}, H^+, H^-)]$ from (2.15) with respect to $\theta_k \equiv (\mu_{0k}, \mu_{1k}, \sigma_{1k}^2, h_{ik}^+ h_{jk}^-)$ $(k = 1, 2; i = 1, \cdots, n_k^+; j = 1, \cdots, n_k^-)$.

$$\begin{aligned} \frac{\partial L}{\partial \mu_{0k}} &= \sum_{X_{ik}>0} \Big\{ \delta_{ik}^O (H_k^+(X_{ik}) - \mu_{0k} - \hat{E}(a_i)) + \delta_{ik}^R \hat{E}\left[\frac{\phi(H_k^+(M_k) - \mu_{0k} - a_i)}{1 - \Phi(H_k^+(M_k) - \mu_{0k} - a_i)} \right] \Big\} (1 - D_i) \\ &\quad - \sum_{\widetilde{X}_{ik}>0} \Big\{ \delta_{ik}^O (H_k^-(\widetilde{X}_{ik}) + \mu_{0k} + \hat{E}(a_i)) \end{aligned}$$

$$+(1-\delta_{ik}^O-\delta_{ik}^R)\hat{E}\left[\frac{\phi(H_k^-(\widetilde{m}_k)+\mu_{0k}+a_i)}{1-\Phi(H_k^-(\widetilde{m}_k)+\mu_{0k}+a_i)}\right]\Big\}(1-D_i)$$

$$\frac{\partial L}{\partial \mu_{1k}} = \sum_{X_{ik}>0}\Big\{\frac{\delta_{ik}^O}{\sigma_{1k}^2}(H_k^+(X_{ik})-\mu_{1k}-\hat{E}(a_i))+\frac{\delta_{ik}^R}{\sigma_{1k}}\hat{E}\left[\frac{\phi((H_k^+(M_k)-\mu_{1k}-a_i)/\sigma_{1k})}{1-\Phi((H_k^+(M_k)-\mu_{1k}-a_i)/\sigma_{1k})}\right]\Big\}D_i$$

$$-\sum_{\widetilde{X}_{ik}>0}\Big\{\frac{\delta_{ik}^O}{\sigma_{1k}^2}(H_k^-(\widetilde{X}_{ik})+\mu_{1k}+\hat{E}(a_i))$$

$$+\frac{(1-\delta_{ik}^O-\delta_{ik}^R)}{\sigma_{1k}}\hat{E}\left[\frac{\phi((H_k^-(\widetilde{m}_k)+\mu_{1k}+a_i)/\sigma_{1k})}{1-\Phi((H_k^-(\widetilde{m}_k)+\mu_{1k}+a_i)/\sigma_{1k})}\right]\Big\}D_i$$

$$\frac{\partial L}{\partial \sigma_{1k}^2} = \frac{1}{2}\sum_{X_{ik}>0}\Big\{\frac{\delta_{ik}^O}{\sigma_{1k}^4}(-\sigma_{1k}^2+(H_k^+(X_{ik})-\mu_{1k})^2-2\hat{E}(a_i)(H_k^+(X_{ik})-\mu_{1k})+\hat{E}[a_i^2])$$

$$+\delta_{ik}^R(\sigma_{1k}^2)^{-3/2}\hat{E}\left[\frac{(H_k^+(M_k)-\mu_{1k}-a_i)\phi((H_k^+(M_k)-\mu_{1k}-a_i)/\sigma_{1k})}{1-\Phi(H_k^+(M_k)-\mu_{1k}-a_i)/\sigma_{1k})}\right]\Big\}D_i$$

$$+\frac{1}{2}\sum_{\widetilde{X}_{ik}>0}\Big\{\frac{\delta_{ik}^O}{\sigma_{1k}^4}(-\sigma_{1k}^2+(H_k^-(\widetilde{X}_{ik})+\mu_{1k})^2+2\hat{E}(a_i)(H_k^-(\widetilde{X}_{ik})+\mu_{1k})+\hat{E}[a_i^2])$$

$$+(1-\delta_{ik}^O-\delta_{ik}^R)(\sigma_{1k}^2)^{-3/2}\hat{E}\left[\frac{(H_k^-(\widetilde{m}_k)+\mu_{1k}+a_i)\phi((H_k^-(\widetilde{m}_k)+\mu_{1k}+a_i)/\sigma_{1k})}{1-\Phi((H_k^-(\widetilde{m}_k)+\mu_{1k}+a_i)/\sigma_{1k})}\right]\Big\}D_i$$

$$\frac{\partial L}{\partial h_{jk}^+} = \frac{\delta_{jk}^O}{h_{jk}^+}-\sum_{X_{ik}>0}\Big\{\delta_{ik}^O(H_k^+(X_{ik})-\mu_{0k}-\hat{E}(a_i))I(X_{jk}\le X_{ik})$$

$$+\delta_{ik}^R\hat{E}\left[\frac{\phi(H_k^+(M_k)-\mu_{0k}-a_i)}{1-\Phi(H_k^+(M_k)-\mu_{0k}-a_i)}\right]I(X_{jk}\le M_k)\Big\}(1-D_i)$$

$$+\Big\{\frac{\delta_{ik}^O}{\sigma_{1k}^2}(H_k^+(X_{ik})-\mu_{1k}-\hat{E}(a_i))I(X_{jk}\le X_{ik})$$

$$+\delta_{ik}^R(\sigma_{1k}^2)^{-1/2}\hat{E}\left[\frac{\phi((H_k^+(M_k)-\mu_{1k}-a_i)/\sigma_{1k})}{1-\Phi((H_k^+(M_k)-\mu_{1k}-a_i)/\sigma_{1k})}\right]I(X_{jk}\le M_k)\Big\}D_i\ j=1,\cdots,n_k^+$$

$$\frac{\partial L}{\partial h_{jk}^-} = \frac{\delta_{jk}^O}{h_{jk}^-}-\sum_{\widetilde{X}_{ik}>0}\Big\{\delta_{ik}^O(H_k^-(\widetilde{X}_{ik})+\mu_{0k}+\hat{E}(a_i))I(\widetilde{X}_{jk}\le\widetilde{X}_{ik})$$

$$+(1-\delta_{ik}^O-\delta_{ik}^R)\hat{E}\left[\frac{\phi(H_k^-(\widetilde{m}_k)+\mu_{0k}+a_i)}{1-\Phi(H_k^-(\widetilde{m}_k)+\mu_{0k}+a_i)}\right]I(\widetilde{X}_{jk}\le\widetilde{m}_k)\Big\}(1-D_i)$$

$$+\Big\{\frac{\delta_{ik}^O}{\sigma_{1k}^2}(H_k^-(\widetilde{X}_{ik})+\mu_{1k}+\hat{E}(a_i))I(\widetilde{X}_{jk}\le\widetilde{X}_{ik})$$

$$+(1-\delta_{ik}^O-\delta_{ik}^R)(\sigma_{1k}^2)^{-1/2}\hat{E}\left[\frac{\phi((H_k^-(\widetilde{m}_k)+\mu_{1k}+a_i)/\sigma_{1k})}{1-\Phi((H_k^-(\widetilde{m}_k)+\mu_{1k}+a_i)/\sigma_{1k})}\right]I(\widetilde{X}_{jk}\le\widetilde{m}_k)\Big\}D_i$$

$$j=1,\cdots,n_k^-$$

First, define equations O_{jk} $(j = 1, \cdots, 6;\ k = 1, 2)$:

$$
\begin{aligned}
O_{1k} &= \frac{\phi(H_k^+(M_k) - \mu_{0k} - a_i)}{1 - \Phi(H_k^+(M_k) - \mu_{0k} - a_i)} \\
O_{2k} &= \frac{\phi(H_k^-(\widetilde{m}_k) + \mu_{0k} + a_i)}{1 - \Phi(H_k^-(\widetilde{m}_k) + \mu_{0k} + a_i)} \\
O_{3k} &= \frac{\phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})}{1 - \Phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})} \\
O_{4k} &= \frac{\phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})}{1 - \Phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})} \\
O_{5k} &= \frac{(H_k^+(M_k) - \mu_{1k} - a_i)\phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})}{1 - \Phi(H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})} \\
O_{6k} &= \frac{(H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)\phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})}{1 - \Phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})}
\end{aligned}
$$

B.1. We need to compute following conditional expectations $\hat{E}_{jk}$ $(j = 1, \cdots, 6; k = 1, 2)$ from the score equations.

$$
\begin{aligned}
\hat{E}_{1k} &= \hat{E}(O_{1k}) = \hat{E}\left[\frac{\phi(H_k^+(M_k) - \mu_{0k} - a_i)}{1 - \Phi(H_k^+(M_k) - \mu_{0k} - a_i)}\right] \\
\hat{E}_{2k} &= \hat{E}(O_{2k}) = \hat{E}\left[\frac{\phi(H_k^-(\widetilde{m}_k) + \mu_{0k} + a_i)}{1 - \Phi(H_k^-(\widetilde{m}_k) + \mu_{0k} + a_i)}\right] \\
\hat{E}_{3k} &= \hat{E}(O_{3k}) = \hat{E}\left[\frac{\phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})}{1 - \Phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})}\right] \\
\hat{E}_{4k} &= \hat{E}(O_{4k}) = \hat{E}\left[\frac{\phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})}{1 - \Phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})}\right] \\
\hat{E}_{5k} &= \hat{E}(O_{5k}) = \hat{E}\left[\frac{(H_k^+(M_k) - \mu_{1k} - a_i)\phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})}{1 - \Phi(H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})}\right] \\
\hat{E}_{6k} &= \hat{E}(O_{6k}) = \hat{E}\left[\frac{(H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)\phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})}{1 - \Phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})}\right]
\end{aligned}
$$

Next, the elements of the Hessian matrix are as follows.

$$
\begin{aligned}
\frac{\partial^2 L}{\partial \mu_{0k}^2} &= \sum_{X_{ik}>0}\left[-\delta_{ik}^O+\delta_{ik}^R\frac{\partial \hat{E}_{1k}}{\partial \mu_{0k}}\right](1-D_i)-\sum_{\widetilde{X}_{ik}>0}\left[\delta_{ik}^O+(1-\delta_{ik}^O-\delta_{ik}^R)\frac{\partial \hat{E}_{2k}}{\partial \mu_{0k}}\right](1-D_i)\\
\frac{\partial^2 L}{\partial \mu_{0k}\partial \mu_{1k}} &= \frac{\partial^2 L}{\partial \mu_{1k}\partial \mu_{0k}}=0\\
\frac{\partial^2 L}{\partial \mu_{0k}\partial \sigma_{1k}^2} &= \frac{\partial^2 L}{\partial \sigma_{1k}^2\partial \mu_{0k}}=0\\
\frac{\partial^2 L}{\partial \mu_{0k}\partial h_{jk}^+} &= \frac{\partial^2 L}{\partial h_{jk}^+\partial \mu_{0k}}=\sum_{X_{ik}>0}\left[\delta_{ik}^O I(X_{jk}\le X_{ik})+\delta_{ik}^R\frac{\partial \hat{E}_{1k}}{\partial h_{jk}^+}\right](1-D_i)\\
&\quad j=1,\cdots,n_k^+\\
\frac{\partial^2 L}{\partial \mu_{0k}\partial h_{jk}^-} &= \frac{\partial^2 L}{\partial h_{jk}^-\partial \mu_{0k}}=-\sum_{\widetilde{X}_{ik}>0}\left[\delta_{ik}^O I(\widetilde{X}_{jk}\le \widetilde{X}_{ik})+(1-\delta_{ik}^O-\delta_{ik}^R)\frac{\partial \hat{E}_{2k}}{\partial h_{jk}^-}\right](1-D_i)\\
&\quad j=1,\cdots,n_k^-\\
\frac{\partial^2 L}{\partial \mu_{1k}^2} &= \sum_{X_{ik}>0}\left[-\frac{\delta_{ik}^O}{\sigma_{1k}^2}+\frac{\delta_{ik}^R}{\sigma_{1k}}\frac{\partial \hat{E}_{3k}}{\partial \mu_{1k}}\right]D_i-\sum_{\widetilde{X}_{ik}>0}\left[\frac{\delta_{ik}^O}{\sigma_{1k}^2}+\frac{(1-\delta_{ik}^O-\delta_{ik}^R)}{\sigma_{1k}}\frac{\partial \hat{E}_{4k}}{\partial \mu_{1k}}\right]D_i\\
\frac{\partial^2 L}{\partial \mu_{1k}\partial \sigma_{1k}^2} &= \frac{\partial^2 L}{\partial \sigma_{1k}^2\partial \mu_{1k}}=\sum_{X_{ik}>0}\Bigg[-\frac{\delta_{ik}^O}{\sigma_{1k}^4}(H_k^+(X_{ik})-\mu_{1k}-\hat{E}(a_i))\\
&\quad +\delta_{ik}^R\Big\{-\frac{1}{2}(\sigma_{1k}^2)^{-3/2}\hat{E}_{3k}+(\sigma_{1k}^2)^{-1/2}\frac{\partial \hat{E}_{3k}}{\partial \sigma_{1k}^2}\Big\}\Bigg]D_i+\sum_{\widetilde{X}_{ik}>0}\Bigg[\frac{\delta_{ik}^O}{\sigma_{1k}^4}(H_k^-(\widetilde{X}_{ik})\\
&\quad +\mu_{1k}+\hat{E}(a_i))-(1-\delta_{ik}^O-\delta_{ik}^R)\Big\{-\frac{1}{2}(\sigma_{1k}^2)^{-3/2}\hat{E}_{4k}+(\sigma_{1k}^2)^{-1/2}\frac{\partial \hat{E}_{4k}}{\partial \sigma_{1k}^2}\Big\}\Bigg]D_i\\
\frac{\partial^2 L}{\partial \mu_{1k}\partial h_{jk}^+} &= \frac{\partial^2 L}{\partial h_{jk}^+\partial \mu_{1k}}=\sum_{X_{ik}>0}\left[\frac{\delta_{ik}^O}{\sigma_{1k}^2}I(X_{jk}\le X_{ik})+\delta_{ik}^R(\sigma_{1k}^2)^{-1/2}\frac{\partial \hat{E}_{3k}}{\partial h_{jk}^+}\right]D_i,\ j=1,\cdots,n_k^+\\
\frac{\partial^2 L}{\partial \mu_{1k}\partial h_{jk}^-} &= \frac{\partial^2 L}{\partial h_{jk}^-\partial \mu_{1k}}=-\sum_{\widetilde{X}_{ik}>0}\left[\frac{\delta_{ik}^O}{\sigma_{1k}^2}I(\widetilde{X}_{jk}\le \widetilde{X}_{ik})+(1-\delta_{ik}^O-\delta_{ik}^R)(\sigma_{1k}^2)^{-1/2}\frac{\partial \hat{E}_{4k}}{\partial h_{jk}^-}\right]D_i,\\
&\quad j=1,\cdots,n_k^-\\
\frac{\partial^2 L}{\partial \sigma_{1k}^4} &= \frac{1}{2}\sum_{X_{ik}>0}\Bigg[\delta_{ik}^O\Big(\frac{1}{\sigma_{1k}^4}-\frac{2}{\sigma_{1k}^6}[(H_k^+(X_{ik})-\mu_{1k})^2-2\hat{E}(a_i)(H_k^+(X_{ik})-\mu_{1k})\\
&\quad +\hat{E}(a_i^2)]\Big)+\delta_{ik}^R\Big\{-\frac{3}{2}(\sigma_{1k}^2)^{-5/2}\hat{E}_{5k}+(\sigma_{1k}^2)^{-3/2}\frac{\partial \hat{E}_{5k}}{\partial \sigma_{1k}^2}\Big\}\Bigg]D_i
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{2}\sum_{\widetilde{X}_{ik}>0}\Bigg[\delta_{ik}^{O}\Big(\frac{1}{\sigma_{1k}^{4}}-\frac{2}{\sigma_{1k}^{6}}[(H_k^{-}(\widetilde{X}_{ik})+\mu_{1k})^2+2\hat{E}(a_i)(H_k^{-}(\widetilde{X}_{ik})+\mu_{1k})\\
&+\hat{E}(a_i^2)]\Big)+(1-\delta_{ik}^{O}-\delta_{ik}^{R})\Big\{-\frac{3}{2}(\sigma_{1k}^{2})^{-5/2}\hat{E}_{6k}+(\sigma_{1k}^{2})^{-3/2}\frac{\partial\hat{E}_{6k}}{\partial\sigma_{1k}^{2}}\Big\}\Bigg]D_i\\
\frac{\partial^2 L}{\partial\sigma_{1k}^{2}\partial h_{jk}^{+}} &= \frac{\partial^2 L}{\partial h_{jk}^{+}\partial\sigma_{1k}^{2}}=\sum_{X_{ik}>0}\Bigg[\frac{\delta_{ik}^{O}}{\sigma_{1k}^{4}}(H_k^{+}(X_{ik})-\mu_{1k}-\hat{E}(a_i))I(X_{jk}\le X_{ik})\\
&-\delta_{ik}^{R}\Big\{-\frac{1}{2}(\sigma_{1k}^{2})^{-3/2}\hat{E}_{3k}+(\sigma_{1k}^{2})^{-1/2}\frac{\partial\hat{E}_{3k}}{\partial\sigma_{1k}^{2}}\Big\}I(X_{jk}\le M_k)\Bigg]D_i,\ j=1,\cdots,n_k^{+}\\
\frac{\partial^2 L}{\partial\sigma_{1k}^{2}\partial h_{jk}^{-}} &= \frac{\partial^2 L}{\partial h_{jk}^{-}\partial\sigma_{1k}^{2}}=\sum_{\widetilde{X}_{ik}>0}\Bigg[\frac{\delta_{ik}^{O}}{\sigma_{1k}^{4}}(H_k^{-}(\widetilde{X}_{ik})+\mu_{1k}+\hat{E}(a_i))I(\widetilde{X}_{jk}\le\widetilde{X}_{ik})\\
&-(1-\delta_{ik}^{O}-\delta_{ik}^{R})\Big\{-\frac{1}{2}(\sigma_{1k}^{2})^{-3/2}\hat{E}_{4k}+(\sigma_{1k}^{2})^{-1/2}\frac{\partial\hat{E}_{4k}}{\partial\sigma_{1k}^{2}}\Big\}I(\widetilde{X}_{jk}\le\widetilde{m}_k)\Bigg]D_i,\\
&j=1,\cdots,n_k^{-}\\
\frac{\partial^2 L}{\partial h_{jk}^{+}\partial h_{j'k}^{+}} &= -\frac{\delta_{jk}^{O}I(j=j')}{(h_k^{+}(X_{jk}))^2}-\sum_{X_{ik}>0}\Big\{\delta_{ik}^{O}I(X_{j'k}\le X_{ik})I(X_{jk}\le X_{ik})\\
&+\delta_{ik}^{R}\frac{\partial\hat{E}_{1k}}{\partial h_{j'k}^{+}}I(X_{jk}\le M_k)\Big\}(1-D_i)+\Big\{\frac{\delta_{ik}^{O}}{\sigma_{1k}^{2}}I(X_{j'k}\le X_{ik})I(X_{jk}\le X_{ik})\\
&+\delta_{ik}^{R}(\sigma_{1k}^{2})^{-1/2}\frac{\partial\hat{E}_{3k}}{\partial h_{j'k}^{+}}I(X_{jk}\le M_k)\Big\}D_i,\ j,j'=1,\cdots,n_k^{+}\\
\frac{\partial^2 L}{\partial h_{jk}^{+}\partial h_{j'k}^{-}} &= \frac{\partial^2 L}{\partial h_{j'k}^{-}\partial h_{jk}^{+}}=0,\ j=1,\cdots,n_k^{+};j'=1,\cdots,n_k^{-}\\
\frac{\partial^2 L}{\partial h_{jk}^{-}\partial h_{j'k}^{-}} &= -\frac{\delta_{jk}^{O}I(j=j')}{(h_k^{-}(\widetilde{X}_{jk}))^2}-\sum_{\widetilde{X}_{ik}>0}\Big\{\delta_{ik}^{O}I(\widetilde{X}_{j'k}\le\widetilde{X}_{ik})I(\widetilde{X}_{jk}\le\widetilde{X}_{ik})+(1-\delta_{ik}^{O}-\delta_{ik}^{R})\\
&\frac{\partial\hat{E}_{2k}}{\partial h_{j'k}^{-}}I(\widetilde{X}_{jk}\le\widetilde{m}_k)\Big\}(1-D_i)+\Big\{\frac{\delta_{ik}^{O}}{\sigma_{1k}^{2}}I(\widetilde{X}_{j'k}\le\widetilde{X}_{ik})I(\widetilde{X}_{jk}\le\widetilde{X}_{ik})\\
&+(1-\delta_{ik}^{O}-\delta_{ik}^{R})(\sigma_{1k}^{2})^{-1/2}\frac{\partial\hat{E}_{4k}}{\partial h_{j'k}^{-}}I(\widetilde{X}_{jk}\le\widetilde{m}_k)\Big\}D_i,\ j,j'=1,\cdots,n_k^{-}
\end{aligned}
$$

B.2. $\dfrac{\partial\hat{E}_{jk}}{\partial\theta_k}$ $(j=1,\cdots,6;k=1,2)$ from the above elements of the Hessian matrix are derived below.

$$\frac{\partial \hat{E}_{1k}}{\partial \mu_{0k}} = \hat{E}\left[\frac{\partial O_{1k}}{\partial \mu_{0k}}\right] = \hat{E}\Big[\Big\{(H_k^+(M_k) - \mu_{0k} - a_i)\phi(H_k^+(M_k) - \mu_{0k} - a_i)(1 - \Phi(H_k^+(M_k) - \mu_{0k} - a_i)) - \phi^2(H_k^+(M_k) - \mu_{0k} - a_i)\Big\}/(1 - \Phi(H_k^+(M_k) - \mu_{0k} - a_i))^2\Big]$$

$$\frac{\partial \hat{E}_{2k}}{\partial \mu_{0k}} = \hat{E}\left[\frac{\partial O_{2k}}{\partial \mu_{0k}}\right] = \hat{E}\Big[\Big\{ - (H_k^-(\widetilde{m}_k) + \mu_{0k} + a_i)\phi(H_k^-(\widetilde{m}_k) + \mu_{0k} + a_i)(1 - \Phi(H_k^-(\widetilde{m}_k) + \mu_{0k} + a_i)) + \phi^2(H_k^-(\widetilde{m}_k) + \mu_{0k} + a_i)\Big\}/(1 - \Phi(H_k^-(\widetilde{m}_k) + \mu_{0k} + a_i))^2\Big]$$

$$\frac{\partial \hat{E}_{3k}}{\partial \mu_{1k}} = \hat{E}\left[\frac{\partial O_{3k}}{\partial \mu_{1k}}\right] = \hat{E}\Big[\Big\{\frac{1}{\sigma_{1k}^2}(H_k^+(M_k) - \mu_{1k} - a_i)\phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k}) (1 - \Phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})) - \frac{1}{\sigma_{1k}}\phi^2((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})\Big\}/ (1 - \Phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k}))^2\Big]$$

$$\frac{\partial \hat{E}_{4k}}{\partial \mu_{1k}} = \hat{E}\left[\frac{\partial O_{4k}}{\partial \mu_{1k}}\right] = \hat{E}\Big[\Big\{ - \frac{1}{\sigma_{1k}^2}(H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)\phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k}) (1 - \Phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})) + \frac{1}{\sigma_{1k}}\phi^2(H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})\Big\}/ (1 - \Phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k}))^2\Big]$$

$$\frac{\partial \hat{E}_{3k}}{\partial \sigma_{1k}^2} = \hat{E}\left[\frac{\partial O_{3k}}{\partial \sigma_{1k}^2}\right] = \hat{E}\Big[\Big\{\frac{1}{2\sigma_{1k}^4}(H_k^+(M_k) - \mu_{1k} - a_i)^2\phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k}) (1 - \Phi(H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k}) - \frac{1}{2}(\sigma_{1k}^2)^{-3/2}(H_k^+(M_k) - \mu_{1k} - a_i) \phi^2((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})\Big\}/(1 - \Phi(H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k}))^2\Big]$$

$$\frac{\partial \hat{E}_{4k}}{\partial \sigma_{1k}^2} = \hat{E}\left[\frac{\partial O_{4k}}{\partial \sigma_{1k}^2}\right] = \hat{E}\Big[\Big\{\frac{1}{2\sigma_{1k}^4}(H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)^2\phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k}) (1 - \Phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})) - \frac{1}{2}(\sigma_{1k}^2)^{-3/2}(H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i) \phi^2((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})\Big\}/(1 - \Phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k}))^2\Big]$$

$$\frac{\partial \hat{E}_{5k}}{\partial \sigma_{1k}^2} = \hat{E}\left[\frac{\partial O_{5k}}{\partial \sigma_{1k}^2}\right] = \hat{E}\Big[\Big\{\frac{1}{2\sigma_{1k}^4}(H_k^+(M_k) - \mu_{1k} - a_i)^3\phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k}) (1 - \Phi(H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k}) - \frac{1}{2}(\sigma_{1k}^2)^{-3/2}(H_k^+(M_k) - \mu_{1k} - a_i)^2 \phi^2((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})\Big\}/(1 - \Phi(H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k}))^2\Big]$$

$$\frac{\partial \hat{E}_{6k}}{\partial \sigma_{1k}^2} = \hat{E}\left[\frac{\partial O_{6k}}{\partial \sigma_{1k}^2}\right] = \hat{E}\left[\Big\{\frac{1}{2\sigma_{1k}^4}(H_k^-(\widetilde{m}_k)+\mu_{1k}+a_i)^3\phi((H_k^-(\widetilde{m}_k)+\mu_{1k}+a_i)/\sigma_{1k})\right.$$
$$(1-\Phi((H_k^-(\widetilde{m}_k)+\mu_{1k}+a_i)/\sigma_{1k})) - \frac{1}{2}(\sigma_{1k}^2)^{-3/2}(H_k^-(\widetilde{m}_k)+\mu_{1k}+a_i)^2$$
$$\left.\phi^2((H_k^-(\widetilde{m}_k)+\mu_{1k}+a_i)/\sigma_{1k})\Big\}/(1-\Phi((H_k^-(\widetilde{m}_k)+\mu_{1k}+a_i)/\sigma_{1k}))^2\right]$$

$$\frac{\partial \hat{E}_{1k}}{\partial h_{jk}^+} = \hat{E}\left[\frac{\partial O_{1k}}{\partial h_{jk}^+}\right] = \hat{E}\left[\Big\{-(H_k^+(M_k)-\mu_{0k}-a_i)\phi(H_k^+(M_k)-\mu_{0k}-a_i)\right.$$
$$(1-\Phi(H_k^+(M_k)-\mu_{0k}-a_i))\ I(X_{jk}\le M_k) + \phi^2(H_k^+(M_k)-\mu_{0k}-a_i)$$
$$\left.I(X_{jk}\le M_k)\Big\}/(1-\Phi(H_k^+(M_k)-\mu_{0k}-a_i))^2\right],\ X_{jk}>0, j=1,\cdots,n_k^+$$

$$\frac{\partial \hat{E}_{3k}}{\partial h_{jk}^+} = \hat{E}\left[\frac{\partial O_{3k}}{\partial h_{jk}^+}\right] = \hat{E}\left[\Big\{-\frac{1}{\sigma_{1k}^2}(H_k^+(M_k)-\mu_{1k}-a_i)\phi((H_k^+(M_k)-\mu_{1k}-a_i)/\sigma_{1k})\right.$$
$$(1-\Phi((H_k^+(M_k)-\mu_{1k}-a_i)/\sigma_{1k}))I(X_{jk}\le M_k) + (\sigma_{1k}^2)^{-1/2}\phi^2((H_k^+(M_k)$$
$$\left.-\mu_{1k}-a_i)/\sigma_{1k})I(X_{jk}\le M_k)\Big\}/(1-\Phi((H_k^+(M_k)-\mu_{1k}-a_i)/\sigma_{1k}))^2\right],$$
$$X_{jk}>0, j=1,\cdots,n_k^+$$

$$\frac{\partial \hat{E}_{2k}}{\partial h_{jk}^-} = \hat{E}\left[\frac{\partial O_{2k}}{\partial h_{jk}^-}\right] = \hat{E}\left[\Big\{-(H_k^-(\widetilde{m}_k)+\mu_{0k}+a_i)\phi(H_k^-(\widetilde{m}_k)+\mu_{0k}+a_i)\right.$$
$$(1-\Phi(H_k^-(\widetilde{m}_k)+\mu_{0k}+a_i))I(\widetilde{X}_{jk}\le\widetilde{m}_k) + \phi^2(H_k^-(\widetilde{m}_k)+\mu_{0k}+a_i)$$
$$\left.I(\widetilde{X}_{jk}\le\widetilde{m}_k)\Big\}/(1-\Phi(H_k^-(\widetilde{m}_k)+\mu_{0k}+a_i))^2\right],\ \widetilde{X}_{jk}>0, j=1,\cdots,n_k^-$$

$$\frac{\partial \hat{E}_{4k}}{\partial h_{jk}^-} = \hat{E}\left[\frac{\partial O_{4k}}{\partial h_{jk}^-}\right] = \hat{E}\left[\Big\{-\frac{1}{\sigma_{1k}^2}(H_k^-(\widetilde{m}_k)+\mu_{1k}+a_i)\phi((H_k^-(\widetilde{m}_k)+\mu_{1k}+a_i)/\sigma_{1k})\right.$$
$$(1-\Phi((H_k^-(\widetilde{m}_k)+\mu_{1k}+a_i)/\sigma_{1k}))I(\widetilde{X}_{jk}\le\widetilde{m}_k) + (\sigma_{1k}^2)^{-1/2}\phi^2((H_k^-(\widetilde{m}_k)$$
$$\left.+\mu_{1k}+a_i)/\sigma_{1k})I(\widetilde{X}_{jk}\le\widetilde{m}_k)\Big\}/(1-\Phi((H_k^-(\widetilde{m}_k)+\mu_{1k}+a_i)/\sigma_{1k}))^2\right],$$
$$\widetilde{X}_{jk}>0, j=1,\cdots,n_k^-$$

2.9.2 Variance Estimation

Let $\hat{\boldsymbol{\theta}}^* \equiv (\hat{\boldsymbol{\theta}}, \hat{\Sigma}_a)$ be the estimates of $\boldsymbol{\theta} = (\mu_{01}, \mu_{02}, \mu_{11}, \mu_{12}, \sigma_{11}^2, \sigma_{12}^2)$ and Σ_a at convergence. We estimate the variance of $(\hat{\boldsymbol{\theta}}^*, \hat{H}^+, \hat{H}^-)$ with observed information matrix

of $\hat{\boldsymbol{\theta}}^*$ and jump sizes $\hat{H}_k^+\{X_{ik}\}$ and $\hat{H}_k^-\{\widetilde{X}_{ik}\}$ by Louis's formula (1982). Note that $\hat{H}_k^+(X_{ik}) = \sum_{X_{jk}\le X_{ik}} \hat{H}_k^+\{X_{jk}\}$ and $\hat{H}_k^-(\widetilde{X}_{ik}) = \sum_{\widetilde{X}_{jk}\le \widetilde{X}_{ik}} \hat{H}_k^-\{\widetilde{X}_{jk}\}$. Let $l_i(\boldsymbol{\theta}^*, H^+, H^-)$ be the complete data log likelihood function for ith subject and $\boldsymbol{\delta} \equiv (\boldsymbol{\theta}^*, H^+, H^-)$. Define $U_i(\hat{\boldsymbol{\delta}}) = \dfrac{\partial}{\partial \boldsymbol{\delta}} l_i(\hat{\boldsymbol{\delta}})\Big|_{\boldsymbol{\delta}=\hat{\boldsymbol{\delta}}} = (U_i'(\hat{\boldsymbol{\theta}}^*), U_i'(\hat{H}^+), U_i'(\hat{H}^-))'$ and $\dot{U}_i(\boldsymbol{\delta}) = \dfrac{\partial}{\partial \boldsymbol{\delta}} U_i(\hat{\boldsymbol{\delta}})$. Then, the observed information matrix is estimated by

$$
\begin{aligned}
&-\textstyle\sum_{i=1}^n E[\dot{U}_i(\hat{\boldsymbol{\delta}})|X_i, D_i, \hat{\boldsymbol{\delta}}] - \sum_{i=1}^n \{E[U_i(\hat{\boldsymbol{\delta}})U_i'(\hat{\boldsymbol{\delta}})|X_i, D_i, \hat{\boldsymbol{\delta}}] \\
&\qquad + E[U_i(\hat{\boldsymbol{\delta}})|X_i, D_i, \hat{\boldsymbol{\delta}}]E'[U_i(\hat{\boldsymbol{\delta}})|X_i, D_i, \hat{\boldsymbol{\delta}}]\}.
\end{aligned}
$$

Complete data log likelihood function for ith subject is expressed as

$$
\begin{aligned}
l_i(\boldsymbol{\theta}^*, H^+, H^-) \;=\; & \sum_{k=1}^{2} \delta_{ik}^O \log H_k^+\{X_{ik}\} I(X_{ik} > 0) + \sum_{k=1}^{2} \Bigg[\Big\{ -\frac{\delta_{ik}^O}{2}(H_k^+(X_{ik}) - \mu_{0k} - a_i)^2 \\
& + \delta_{ik}^R \log(1 - \Phi(H_k^+(M_k) - \mu_{0k} - a_i)) \Big\}(1 - D_i) \\
& + \sum_{k=1}^{2} \Big\{ -\frac{\delta_{ik}^O}{2}\Big(\log\sigma_{1k}^2 + \frac{1}{\sigma_{1k}^2}(H_k^+(X_{ik}) - \mu_{1k} - a_i)^2\Big) \\
& + \delta_{ik}^R \log(1 - \Phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})) \Big\} D_i \Bigg] I(X_{ik} > 0) \\
& + \sum_{k=1}^{2} \delta_{ik}^O \log H_k^-\{\widetilde{X}_{ik}\} I(\widetilde{X}_{ik} > 0) + \sum_{k=1}^{2} \Bigg[\Big\{ -\frac{\delta_{ik}^O}{2}(H_k^-(\widetilde{X}_{ik}) + \mu_{0k} + a_i)^2 \\
& + (1 - \delta_{ik}^O - \delta_{ik}^R) \log(1 - \Phi(H_k^-(\widetilde{m}_k) + \mu_{0k} + a_i)) \Big\}(1 - D_i) \\
& + \sum_{k=1}^{2} \Big\{ -\frac{\delta_{ik}^O}{2}\Big(\log\sigma_{1k}^2 + \frac{1}{\sigma_{1k}^2}(H_k^-(\widetilde{X}_{ik}) + \mu_{1k} + a_i)^2\Big) \\
& + (1 - \delta_{ik}^O - \delta_{ik}^R) \log(1 - \Phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})) \Big\} D_i \Bigg] \\
& I(\widetilde{X}_{ik} > 0) + \log\phi(a_i; \sigma^2) \qquad (2.16)
\end{aligned}
$$

The elements of $U_i(\boldsymbol{\delta})$ and $\dot{U}_i(\boldsymbol{\delta})$ for fixed i $(i = 1, \cdots, n)$ are computed below.

B.3. $U_i(\boldsymbol{\delta}) = (U_i(\boldsymbol{\theta^*}),\ U_i(H^+),\ U_i(H^-))'$

$$
\begin{aligned}
U_i(\mu_{0k}) &= \frac{\partial l_i}{\partial \mu_{0k}} = \left[\delta_{ik}^O(H_k^+(X_{ik}) - \mu_{0k} - a_i) + \delta_{ik}^R \frac{\phi(H_k^+(M_k) - \mu_{0k} - a_i)}{1 - \Phi(H_k^+(M_k) - \mu_{0k} - a_i)}\right] \\
&\quad I(X_{ik} > 0)(1 - D_i) - \left[\delta_{ik}^O(H_k^-(\widetilde{X}_{ik}) + \mu_{0k} + a_i)\right. \\
&\quad \left. + (1 - \delta_{ik}^O - \delta_{ik}^R)\frac{\phi(H_k^-(\widetilde{m}_k) + \mu_{0k} + a_i)}{1 - \Phi(H_k^-(\widetilde{m}_k) + \mu_{0k} + a_i)}\right] I(\widetilde{X}_{ik} > 0)(1 - D_i) \\
U_i(\mu_{1k}) &= \frac{\partial l_i}{\partial \mu_{1k}} = \left[\frac{\delta_{ik}^O}{\sigma_{1k}^2}(H_k^+(X_{ik}) - \mu_{1k} - a_i)\right. \\
&\quad \left. + \delta_{ik}^R(\sigma_{1k}^2)^{-1/2}\frac{\phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})}{1 - \Phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})}\right] I(X_{ik} > 0)D_i \\
&\quad - \left[\frac{\delta_{ik}^O}{\sigma_{1k}^2}(H_k^-(\widetilde{X}_{ik}) + \mu_{1k} + a_i)\right. \\
&\quad \left. + (1 - \delta_{ik}^O - \delta_{ik}^R)(\sigma_{1k}^2)^{-1/2}\frac{\phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})}{1 - \Phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})}\right] I(\widetilde{X}_{ik} > 0)D_i \\
U_i(\sigma_{1k}^2) &= \frac{\partial l_i}{\partial \sigma_{1k}^2} = \frac{1}{2}\left[\frac{\delta_{ik}^O}{\sigma_{1k}^4}(-\sigma_{1k}^2 + (H_k^+(X_{ik}) - \mu_{1k} - a_i)^2 + \delta_{ik}^R(\sigma_{1k}^2)^{-3/2}\right. \\
&\quad \left. (H_k^+(M_k) - \mu_{1k} - a_i)\frac{\phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})}{1 - \Phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})}\right] I(X_{ik} > 0)D_i \\
&\quad + \frac{1}{2}\left[\frac{\delta_{ik}^O}{\sigma_{1k}^4}(-\sigma_{1k}^2 + (H_k^-(\widetilde{X}_{ik}) + \mu_{1k} + a_i))^2 + (1 - \delta_{ik}^O - \delta_{ik}^R)\right. \\
&\quad \left. (\sigma_{1k}^2)^{-3/2}(H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)\frac{\phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})}{1 - \Phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})}\right] \\
&\quad I(\widetilde{X}_{ik} > 0)D_i \\
U_i(\sigma^2) &= \frac{\partial l_i}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{a_i^2}{2\sigma^4} \\
U_i(H_k^+\{X_{jk}\}) &= \frac{\partial l_i}{\partial H_k^+\{X_{jk}\}} = \frac{\delta_{jk}^O I(i = j)}{H_k^+\{X_{jk}\}} - \left[\delta_{ik}^O(H_k^+(X_{ik}) - \mu_{0k} - a_i)I(X_{jk} \le X_{ik})\right. \\
&\quad \left. + \delta_{ik}^R\frac{\phi(H_k^+(M_k) - \mu_{0k} - a_i)}{1 - \Phi(H_k^+(M_k) - \mu_{0k} - a_i)}I(X_{jk} \le M_k)\right] I(X_{ik} > 0)(1 - D_i) \\
&\quad - \left[\frac{\delta_{ik}^O}{\sigma_{1k}^2}(H_k^+(X_{ik}) - \mu_{1k} - a_i)I(X_{jk} \le X_{ik})\right. \\
&\quad \left. + \delta_{ik}^R(\sigma_{1k}^2)^{-1/2}\frac{\phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})}{1 - \Phi((H_k^+(M_k) - \mu_{1k} - a_i)/\sigma_{1k})}I(X_{jk} \le M_k)\right] \\
&\quad I(X_{ik} > 0)D_i,\ X_{jk} > 0,\ j = 1, \cdots, n_k^+
\end{aligned}
$$

$$
\begin{aligned}
U_i(H_k^-\{\widetilde{X}_{jk}\}) &= \frac{\partial l_i}{\partial H_k^-\{\widetilde{X}_{jk}\}} = \frac{\delta_{jk}^O I(i=j)}{H_k^-\{\widetilde{X}_{jk}\}} - \Bigg[\delta_{ik}^O (H_k^-(\widetilde{X}_{ik}) + \mu_{0k} + a_i) I(\widetilde{X}_{jk} \le \widetilde{X}_{ik}) \\
&\quad + (1 - \delta_{ik}^O - \delta_{ik}^R) \frac{\phi(H_k^-(\widetilde{m}_k) + \mu_{0k} + a_i)}{1 - \Phi(H_k^-(\widetilde{m}_k) + \mu_{0k} + a_i)} I(\widetilde{X}_{jk} \le \widetilde{m}_k) \Bigg] \\
&\quad I(\widetilde{X}_{ik} > 0)(1 - D_i) - \Bigg[\frac{\delta_{ik}^O}{\sigma_{1k}^2} (H_k^-(\widetilde{X}_{ik}) + \mu_{1k} + a_i) I(\widetilde{X}_{jk} \le \widetilde{X}_{ik}) \\
&\quad + (1 - \delta_{ik}^O - \delta_{ik}^R)(\sigma_{1k}^2)^{-1/2} \frac{\phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})}{1 - \Phi((H_k^-(\widetilde{m}_k) + \mu_{1k} + a_i)/\sigma_{1k})} \\
&\quad I(\widetilde{X}_{jk} \le \widetilde{m}_k) \Bigg] I(\widetilde{X}_{ik} > 0) D_i,\ \widetilde{X}_{jk} > 0\ j = 1, \cdots, n_k^-
\end{aligned}
$$

B.4. $\dot{U}_i(\boldsymbol{\delta}) = \dfrac{\partial}{\partial \boldsymbol{\delta}} U_i(\boldsymbol{\delta})$

$$
\begin{aligned}
\frac{\partial U_i(\mu_{0k})}{\partial \mu_{0k}} &= \left[-\delta_{ik}^O + \delta_{ik}^R \frac{\partial O_{1k}}{\partial \mu_{0k}} \right] I(X_{ik} > 0)(1 - D_i) \\
&\quad - \left[\delta_{ik}^O + (1 - \delta_{ik}^O - \delta_{ik}^R) \frac{\partial O_{2k}}{\partial \mu_{0k}} \right] I(\widetilde{X}_{ik} > 0)(1 - D_i) \\
\frac{\partial U_i(\mu_{0k})}{\partial \mu_{1k}} &= \frac{\partial U_i(\mu_{1k})}{\partial \mu_{0k}} = 0 \\
\frac{\partial U_i(\mu_{0k})}{\partial \sigma_{1k}^2} &= \frac{\partial U_i(\sigma_{1k}^2)}{\partial \mu_{0k}} = 0 \\
\frac{\partial U_i(\mu_{0k})}{\partial \sigma^2} &= \frac{\partial U_i(\sigma^2)}{\partial \mu_{0k}} = 0 \\
\frac{\partial U_i(\mu_{0k})}{\partial H_k^+\{X_{jk}\}} &= \frac{\partial U_i(H_k^+\{X_{jk}\})}{\partial \mu_{0k}} = \left[\delta_{ik}^O I(X_{jk} \le X_{ik}) + \delta_{ik}^R \frac{\partial O_{1k}}{\partial H_k^+\{X_{jk}\}} \right] I(X_{ik} > 0) \\
&\quad (1 - D_i),\ X_{jk} > 0, j = 1, \cdots, n_k^+ \\
\frac{\partial U_i(\mu_{0k})}{\partial H_k^-\{\widetilde{X}_{jk}\}} &= \frac{\partial U_i(H_k^-\{\widetilde{X}_{jk}\})}{\partial \mu_{0k}} = - \left[\delta_{ik}^O I(\widetilde{X}_{jk} \le \widetilde{X}_{ik}) + (1 - \delta_{ik}^O - \delta_{ik}^R) \frac{\partial O_{2k}}{\partial H_k^-\{\widetilde{X}_{jk}\}} \right] \\
&\quad I(\widetilde{X}_{ik} > 0)(1 - D_i),\ \widetilde{X}_{jk} > 0, j = 1, \cdots, n_k^- \\
\frac{\partial U_i(\mu_{1k})}{\partial \mu_{1k}} &= \left[-\frac{\delta_{ik}^O}{\sigma_{1k}^2} + \frac{\delta_{ik}^R}{\sigma_{1k}} \frac{\partial O_{3k}}{\partial \mu_{1k}} \right] I(X_{ik} > 0) D_i \\
&\quad - \left[\frac{\delta_{ik}^O}{\sigma_{1k}^2} + \frac{(1 - \delta_{ik}^O - \delta_{ik}^R)}{\sigma_{1k}} \frac{\partial O_{4k}}{\partial \mu_{1k}} \right] I(\widetilde{X}_{ik} > 0) D_i
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial U_i(\mu_{1k})}{\partial \sigma_{1k}^2} &= \frac{\partial U_i(\sigma_{1k}^2)}{\partial \mu_{1k}} = \Big[-\frac{\delta_{ik}^O}{\sigma_{1k}^4}(H_k^+(X_{ik}) - \mu_{1k} - a_i) + \delta_{ik}^R\Big\{ -\frac{1}{2}(\sigma_{1k}^2)^{-3/2}O_{3k} \\
&\quad +(\sigma_{1k}^2)^{-1/2}\frac{\partial O_{3k}}{\partial \sigma_{1k}^2}\Big\}\Big] I(X_{ik} > 0)D_i - \Big[-\frac{\delta_{ik}^O}{\sigma_{1k}^4}(H_k^-(\widetilde{X}_{ik}) + \mu_{1k} + a_i) \\
&\quad +(1 - \delta_{ik}^O - \delta_{ik}^R)\Big\{ -\frac{1}{2}(\sigma_{1k}^2)^{-3/2}O_{4k} + (\sigma_{1k}^2)^{-1/2}\frac{\partial O_{4k}}{\partial \sigma_{1k}^2}\Big\}\Big] I(\widetilde{X}_{ik} > 0)D_i \\
\frac{\partial U_i(\mu_{1k})}{\partial \sigma^2} &= \frac{\partial U_i(\sigma^2)}{\partial \mu_{1k}} = 0 \\
\frac{\partial U_i(\mu_{1k})}{\partial H_k^+\{X_{jk}\}} &= \frac{\partial U_i(H_k^+\{X_{jk}\})}{\partial \mu_{1k}} = \Big[\frac{\delta_{ik}^O}{\sigma_{1k}^2}I(X_{jk} \le X_{ik}) + \delta_{ik}^R(\sigma_{1k}^2)^{-1/2}\frac{\partial O_{3k}}{\partial H_k^+\{X_{jk}\}}\Big] \\
&\quad I(X_{ik} > 0)D_i,\ X_{jk} > 0, j = 1, \cdots, n_k^+ \\
\frac{\partial U_i(\mu_{1k})}{\partial H_k^-\{\widetilde{X}_{jk}\}} &= \frac{\partial U_i(H_k^-\{X_{jk}\})}{\partial \mu_{1k}} = -\Big[\frac{\delta_{ik}^O}{\sigma_{1k}^2}I(\widetilde{X}_{jk} \le \widetilde{X}_{ik}) + (1 - \delta_{ik}^O - \delta_{ik}^R) \\
&\quad (\sigma_{1k}^2)^{-1/2}\frac{\partial O_{4k}}{\partial H_k^-\{\widetilde{X}_{jk}\}}\Big] I(\widetilde{X}_{ik} > 0)D_i,\ \widetilde{X}_{jk} > 0, j = 1, \cdots, n_k^- \\
\frac{\partial U_i(\sigma_{1k}^2)}{\partial \sigma_{1k}^2} &= \frac{1}{2}\Big[\delta_{ik}^O\Big\{\frac{1}{\sigma_{1k}^4} - \frac{2}{\sigma_{1k}^6}(H_k^+(X_{ik}) - \mu_{1k} - a_i)^2\Big\} \\
&\quad +\delta_{ik}^R\Big\{ -\frac{3}{2}(\sigma_{1k}^2)^{-5/2}O_{5k} + (\sigma_{1k}^2)^{-3/2}\frac{\partial O_{5k}}{\partial \sigma_{1k}^2}\Big\}\Big] I(X_{ik} > 0)D_i \\
&\quad +\frac{1}{2}\Big[\delta_{ik}^O\Big\{\frac{1}{\sigma_{1k}^4} - \frac{2}{\sigma_{1k}^6}(H_k^-(\widetilde{X}_{ik}) + \mu_{1k} + a_i)^2\Big\} + (1 - \delta_{ik}^O - \delta_{ik}^R) \\
&\quad \Big\{ -\frac{3}{2}(\sigma_{1k}^2)^{-5/2}O_{6k} + (\sigma_{1k}^2)^{-3/2}\frac{\partial O_{6k}}{\partial \sigma_{1k}^2}\Big\}\Big] I(\widetilde{X}_{ik} > 0)D_i \\
\frac{\partial U_i(\sigma_{1k}^2)}{\partial \sigma^2} &= \frac{\partial U_i(\sigma^2)}{\partial \sigma_{1k}^2} = 0 \\
\frac{\partial U_i(\sigma_{1k}^2)}{\partial H_k^+\{X_{jk}\}} &= \frac{\partial U_i(H_k^+\{X_{jk}\})}{\partial \sigma_{1k}^2} = \Big[\frac{\delta_{ik}^O}{\sigma_{1k}^4}(H_k^+(X_{ik}) - \mu_{1k} - a_i)I(X_{jk} \le X_{ik}) \\
&\quad -\delta_{ik}^R\Big(-\frac{1}{2}(\sigma_{1k}^2)^{-3/2}O_{3k} + (\sigma_{1k}^2)^{-1/2}\frac{\partial O_{3k}}{\partial \sigma_{1k}^2}\Big)I(X_{jk} \le M_k)\Big] I(X_{ik} > 0)D_i \\
&\quad X_{jk} > 0, j = 1, \cdots, n_k^+ \\
\frac{\partial U_i(\sigma_{1k}^2)}{\partial H_k^-\{\widetilde{X}_{jk}\}} &= \frac{\partial U_i(H_k^-\{\widetilde{X}_{jk}\})}{\partial \sigma_{1k}^2} = \Big[\frac{\delta_{ik}^O}{\sigma_{1k}^4}(H_k^-(\widetilde{X}_{ik}) + \mu_{1k} + a_i)I(\widetilde{X}_{jk} \le \widetilde{X}_{ik}) \\
&\quad -(1 - \delta_{ik}^O - \delta_{ik}^R)\Big(-\frac{1}{2}(\sigma_{1k}^2)^{-3/2}O_{4k} + (\sigma_{1k}^2)^{-1/2}\frac{\partial O_{4k}}{\partial \sigma_{1k}^2}\Big)I(\widetilde{X}_{jk} \le \widetilde{m}_k)\Big] \\
&\quad I(\widetilde{X}_{ik} > 0)D_i,\ \widetilde{X}_{jk} > 0, j = 1, \cdots, n_k^-
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial U_i(\sigma^2)}{\partial \sigma^2} &= \frac{1}{2\sigma^4} - \frac{a_i^2}{\sigma^6} \\
\frac{\partial U_i(\sigma^2)}{\partial H_k^+\{X_{jk}\}} &= \frac{\partial U_i(H_k^+\{X_{jk}\})}{\partial \sigma^2} = 0,\ j = 1, \cdots, n_k^+ \\
\frac{\partial U_i(\sigma^2)}{\partial H_k^-\{\widetilde{X}_{jk}\}} &= \frac{\partial U_i(H_k^-\{\widetilde{X}_{jk}\})}{\partial \sigma^2} = 0, j = 1, \cdots, n_k^- \\
\frac{\partial U_i(H_k^+\{X_{jk}\})}{\partial H_k^+\{X_{j'k}\}} &= -\frac{\delta_{jk}^O I(i=j=j')}{(H_k^+\{X_{jk}\})^2} I(X_{jk}>0) - \Bigg[\delta_{ik}^O I(X_{j'k} \le X_{ik}) I(X_{jk} \le X_{ik}) \\
&\quad + \delta_{ik}^R \frac{\partial O_{1k}}{\partial H_k^+\{X_{j'k}\}} I(X_{jk} \le M_k) \Bigg] I(X_{ik}>0)(1-D_i) \\
&\quad - \Bigg[\frac{\delta_{ik}^O}{\sigma_{1k}^2} I(X_{j'k} \le X_{ik}) I(X_{jk} \le X_{ik}) + \delta_{ik}^R (\sigma_{1k}^2)^{-1/2} \frac{\partial O_{3k}}{\partial H_k^+\{X_{j'k}\}} \\
&\quad I(X_{jk} \le M_k) \Bigg] I(X_{ik}>0) D_i,\ X_{jk}>0, X_{j'k}>0, j, j' = 1, \cdots, n_k^+ \\
\frac{\partial U_i(H_k^+\{X_{jk}\})}{\partial H_k^-\{\widetilde{X}_{j'k}\}} &= \frac{\partial U_i(H_k^-\{\widetilde{X}_{j'k}\})}{\partial H_k^+\{X_{jk}\}} = 0,\ X_{jk}>0, \widetilde{X}_{j'k}>0, j = 1, \cdots, n_k^+; j' = 1, \cdots, n_k^- \\
\frac{\partial U_i(H_k^-\{\widetilde{X}_{jk}\})}{\partial H_k^-\{\widetilde{X}_{j'k}\}} &= -\frac{\delta_{jk}^O I(i=j=j')}{(H_k^-\{\widetilde{X}_{jk}\})^2} I(\widetilde{X}_{jk}>0) - \Bigg[\delta_{ik}^O I(\widetilde{X}_{j'k} \le \widetilde{X}_{ik}) I(\widetilde{X}_{jk} \le \widetilde{X}_{ik}) \\
&\quad + (1-\delta_{ik}^O - \delta_{ik}^R) \frac{\partial O_{2k}}{\partial H_k^-\{\widetilde{X}_{j'k}\}} I(\widetilde{X}_{jk} \le \widetilde{m}_k) \Bigg] I(\widetilde{X}_{ik}>0)(1-D_i) \\
&\quad - \Bigg[\frac{\delta_{ik}^O}{\sigma_{1k}^2} I(\widetilde{X}_{j'k} \le \widetilde{X}_{ik}) I(\widetilde{X}_{jk} \le \widetilde{X}_{ik}) + (1-\delta_{ik}^O - \delta_{ik}^R)(\sigma_{1k}^2)^{-1/2} \frac{\partial O_{4k}}{\partial H_k^-\{\widetilde{X}_{j'k}\}} \\
&\quad I(\widetilde{X}_{jk} \le \widetilde{m}_k) \Bigg] I(\widetilde{X}_{ik}>0) D_i,\ \widetilde{X}_{jk}>0, \widetilde{X}_{j'k}>0, j, j' = 1, \cdots, n_k^-
\end{aligned}
$$

2.9.3 The Optimal AUC and its Variance

The linear combination of $H_1(X_1)$ and $H_2(X_2)$, $\beta_1 H_1(X_1) + \beta_2 H_2(X_2)$ $(\beta = (\beta_1, \beta_2)')$ follows a multivariate normal distribution with

$$
\beta_1 H_1(X_1) + \beta_2 H_2(X_2) | D = 0 \ \sim N(\beta' \mu_0, \beta'(\sigma^2 \mathbf{1}\mathbf{1}' + \Sigma_0)\beta)
$$

$$\beta_1 H_1(X_1) + \beta_2 H_2(X_2)|D = 1 \ \sim N(\beta'\mu_1, \beta'(\sigma^2 \mathbf{1}\mathbf{1}' + \Sigma_1)\beta).$$

The coefficients for the best linear combination are $\beta_{opt,Z} = (\Sigma_1 + \Sigma_0 + 2\sigma^2 \mathbf{1}\mathbf{1}')^{-1}(\mu_1 - \mu_0)$ and the area under the ROC curve of the optimal linear combination is

$$AUC_{opt,Z} = \Phi\sqrt{(\mu_1 - \mu_0)^T(\Sigma_1 + \Sigma_0 + 2\sigma^2 \mathbf{1}\mathbf{1}')^{-1}(\mu_1 - \mu_0)}.$$

Define $\eta = \sqrt{(\mu_1 - \mu_0)^T(\Sigma_1 + \Sigma_0 + 2\sigma^2 \mathbf{1}\mathbf{1}')^{-1}(\mu_1 - \mu_0)}$ and $\boldsymbol{\theta^*} \equiv(\boldsymbol{\theta}, \sigma^2)$ with $\boldsymbol{\theta} \equiv (\mu_{01}, \mu_{02}, \mu_{11}, \mu_{12}, \sigma_{11}^2, \sigma_{12}^2)$. Then, using the multivariate delta method, the variance of $AUC_{opt,Z}$ given by

$$(\phi(\eta))^2 \left\{ \frac{g_2}{4g_1^3} Var(g_1) + \frac{1}{4g_1 g_2} Var(g_2) - \frac{1}{2g_1^2} Cov(g_1, g_2) \right\},$$

where

$$g_1(\boldsymbol{\theta^*}) = (\sigma_{11}^2 + 2\sigma^2 + 1)(\sigma_{12}^2 + 2\sigma^2 + 1) - 4\sigma^2 \text{ and}$$

$$g_2(\boldsymbol{\theta^*}) = (\sigma_{12}^2 + 2\sigma^2 + 1)(\mu_{11} - \mu_{01})^2 + (\sigma_{11}^2 + 2\sigma^2 + 1)(\mu_{12} - \mu_{02})^2 - 4\sigma^2(\mu_{11} - \mu_{01})(\mu_{12} - \mu_{02}).$$

Let $\Sigma_{\boldsymbol{\theta^*}}$ denote the covariance of $\boldsymbol{\theta^*}$. The covariance of g_1 and g_2 is $\Sigma_{g_1,g_2} = \nabla g(\boldsymbol{\theta^*})^T \Sigma_{\boldsymbol{\theta^*}} \nabla g(\boldsymbol{\theta^*})$, where

$$\nabla g(\boldsymbol{\theta^*}) = \begin{pmatrix} 0 & -2(\mu_{11} - \mu_{01})(\sigma_{12}^2 + 2\sigma_a^2 + 1) + 4\sigma_a^2(\mu_{12} - \mu_{02}) \\ 0 & -2(\mu_{12} - \mu_{02})(\sigma_{11}^2 + 2\sigma_a^2 + 1) + 4\sigma_a^2(\mu_{11} - \mu_{01}) \\ 0 & 2(\mu_{11} - \mu_{01})(\sigma_{12}^2 + 2\sigma_a^2 + 1) - 4\sigma_a^2(\mu_{12} - \mu_{02}) \\ 0 & 2(\mu_{12} - \mu_{02})(\sigma_{11}^2 + 2\sigma_a^2 + 1) - 4\sigma_a^2(\mu_{11} - \mu_{01}) \\ (\sigma_{12}^2 + 2\sigma_a^2 + 1) & (\mu_{12} - \mu_{02})^2 \\ (\sigma_{11}^2 + 2\sigma_a^2 + 1) & (\mu_{11} - \mu_{01})^2 \\ 2(\sigma_{11}^2 + \sigma_{12}^2 + 2) & 2(\mu_{11} - \mu_{01})^2 + 2(\mu_{12} - \mu_{02})^2 - 4(\mu_{11} - \mu_{01})(\mu_{12} - \mu_{02}) \end{pmatrix}$$

Chapter 3

Accelerated Regression Model in the ROC Analysis

3.1 Introduction

Recent technological advances continue to provide non-invasive and more accurate biomarkers for evaluating disease status and patients' treatment response. Examples include the use of Prostate-Specific Antigen and CA-125 to detect the presence of prostate cancer and ovarian cancer, respectively. The Receiver-Operating Characteristic (ROC) curve is a useful tool to assess the accuracy of biomarkers for diagnosis and prognosis of disease (Swets and Pickett, 1982; Hanley, 1989; Pepe 2000b). Let Y_1 denote the biomarker for diseased subjects and Y_0 denote the biomarker for nondiseased subjects. Let c be a threshold value that any test results greater than c are considered to be positive. Without loss of generality, we assume that higher values of test results are more indicative of disease. Then, for a given threshold c, the true positive and false positive rates are respectively

$$S_1(c) = P(Y_1 \geq c) \text{ and } S_0(c) = P(Y_0 \geq c).$$

The ROC curve is a plot of the true positive rates versus the false positive rates, $ROC(\cdot) = \{(S_0(c), S_1(c)),\ c \in (-\infty, \infty)\}$. Equivalently, the ROC function can be expressed as $ROC(\cdot) = \{(t, ROC(t)),\ t \in (0,1)\} = \{(t, S_1(S_0^{-1}(t)),\ t \in (0,1)\}$.

The performance of a diagnostic test can be influenced by risk factors. For example, subject characteristics such as age and gender, the experience and expertise of persons performing the test, and the environment in which and the time when a test is performed can affect test results. Thus, it is important to identify such factors to understand and determine the optimal conditions for the best performance.

In the existing literature, three approaches to incorporating covariate effects into ROC analysis have been suggested (Pepe, 1998). The first approach is to model the ROC curve summary indices as a function of covariates. Dorfman et al. (1992) and Obuchowski (1995) suggested modeling the area under the curve (AUC), while Thompson and Zucchini (1989) recommended modelling the partial area under the curve (pAUC). This approach is feasible only when covariates are discrete and there are enough patients in each covariate combination to permit the reliable calculation of the summary accuracy measure. The second approach is to model the distributions of test results as a function of disease status and covariates. Tosteson and Begg (1988) described the use of ordinal regression model to induce the regression models for the ROC curve for tests with ordinal outcomes. Their method has been extended to random effects models (Beam, 1995; Gatsonis, 1995) and Bayesian methods (Peng and Hall, 1996; Hellmich et al. 1998; Ishwaran and Gatsonsis, 2000). However, in this approach, the parameter estimates do not reflect the covariate effects on the ROC curve. The third approach directly models covariate effects on the ROC curve (Pepe 1997, 2000a; Alonzo and Pepe, 2001). It is called the parametric distribution free approach since it only assumes a parametric model for the ROC curve but is distribution-free regarding the distribution of the test results. The most important advantage of this approach is

that the interpretation of model parameters pertains directly to the ROC curves.

Specifically, in the the third approach mentioned above, Pepe (1997, 2000a) proposed the parametric ROC regression models of the generalized linear model (GLM) form,

$$ROC_X(t) = g(h(t) + \beta^T X), \quad t \in (0,1),$$

where $ROC_X(t)$ denotes the ROC curve at a false positive rate t associated with covariates X, $g(.)$ is a known link function, and $h(.)$ is a baseline function specified up to some real parameters. The baseline function h defines the location and shape of the ROC curve, and β quantifies covariate effects. Recently, Cai and Pepe (2002) extended the parametric ROC regression model to a semi-parametric approach by allowing an arbitrary nonparametric baseline function for h. They assumed a semiparametric location model for $S_0(y|X)$ (Pepe 1998; Heagerty and Pepe 1999) and constructed high-dimensional estimating equations for β and h. They showed that their semiparametric methods fit the model with efficiency comparable to that of the fully parametric approach. Note that the last two models assume that the effects of covariates are related to the location shift of the same ROC curve. This may not be true in some practice.

In this chapter, we develop an alternative regression model, namely, the accelerated ROC model by adjusting for covariates that can influence the performance of a biomarker. We consider modeling covariates directly on the ROC curve and our model generalizes the usual accelerated failure time model in the survival context to the ROC analysis. In Section 3.2, we describe an accelerated ROC model as well as the procedures for estimating parameters of covariates β and the ROC function. The asymptotic properties of β and the ROC function are given in Section 3.3 and the simulation studies are followed in Section 3.4. As an example, we apply our method to a prostate cancer dataset in Section 3.5. A discussion and all technical proofs are given in Sections 3.6 and 3.7, respectively.

3.2 Model and Inference Procedure

Suppose we observe n_1 biomarker measurements from diseased subjects and n_0 biomarker measurements from nondiseased subjects. Y_{i1} $(i = 1, \cdots, n_1)$ denotes the biomarker measurement for diseased subject i and Y_{j0} denotes biomarker measurements for nondiseased subject j. We assume that each subject may have more than one type of covariates and denote them as X_{i1} and X_{j0} for diseased subject i and nondiseased subject j, respectively. In most practice, the biomarker measurement is subject to an upper detection limit, denoted by τ. Thus, the observed data consist of $\{(\min(Y_{i1} \wedge \tau), X_{i1}, \Delta_{i1}),\ i = 1, \cdots, n_1\}$ for diseased subjects and $\{(\min(Y_{j0} \wedge \tau), X_{j0}, \Delta_{j0}),\ j = 1, \cdots, n_0\}$ for nondiseased subjects, where $\Delta_{i1} = I(Y_{i1} \leq \tau)$ and $\Delta_{j0} = I(Y_{j0} \leq \tau)$.

To model the covariate effects on the ROC curve, we propose the following accelerated ROC model,

$$ROC_X(t) = G(e^{\beta^T X} \log t), \quad t \in (0, 1), \tag{3.1}$$

where $G(.)$ is an unknown and increasing function satisfying $G(0) = 1$ and $G(-\infty) = 0$. The latter is because $ROC(1) = 1$ and $ROC(0) = 0$. Note that the effect of X in the ROC model relates to rescaling the original ROC curve. To see how this is different from the model in Pepe (1997, 2000a), we plot the ROC curves based on these two models in Figure 3.1, where the the ROC curve on the left side is based on the parametric ROC regression model by Pepe (1997, 2000a) and the ROC curve on the right side is based on our accelerated ROC model (3.1). It is clear that our model implies the covariate affects sensitivity dramatically for the low false positive rates compared to the model in Pepe (1997, 2000a).

The model (3.1) can be rewritten as

$$S_1(Y|X) = G(e^{\beta^T X} \log S_0(Y|X)). \tag{3.2}$$

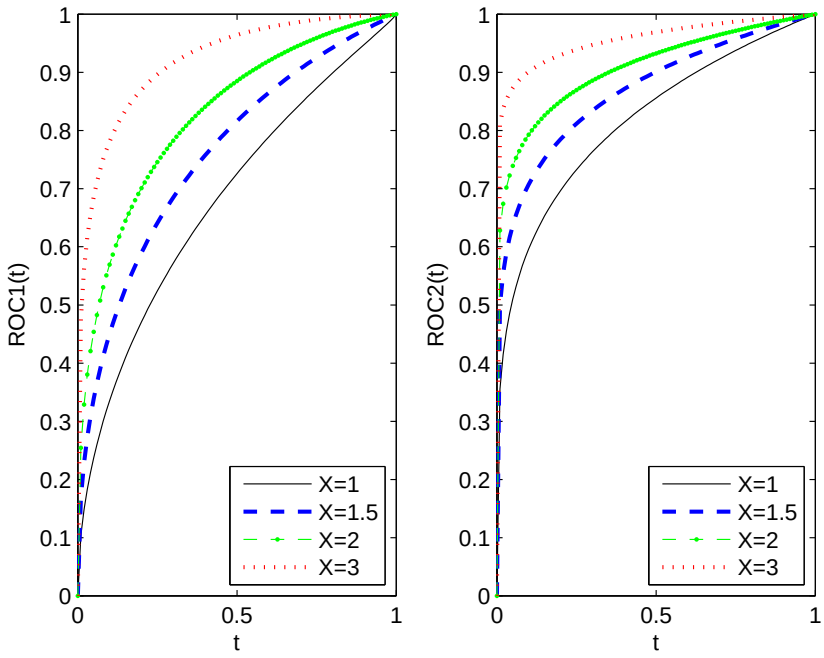

Figure 3.1: $ROC_1(t) = \Phi(0.6X + 0.8\Phi^{-1}(t))$, $ROC_2(t) = \exp(0.5\mathrm{e}^{-0.8X}\log(t))$

Note that we do not make any assumptions on the model for $S_0(Y|X)$. To estimate β, we define $Z_{i1} = -logS_0(Y_{i1}|X_{i1})$. Using equation (3.2), it can be shown that

$$\begin{aligned} P(Z_{i1} \leq z|X_{i1}) &= 1 - P(-\log S_0(Y_{i1}) > z|X_{i1}) = 1 - G(e^{\beta^T X_{i1}}\log e^{-z}) \\ &= 1 - G(-ze^{\beta^T X_{i1}}) \equiv F(ze^{\beta^T X_{i1}}), \end{aligned}$$

with $F(x) = 1 - G(-x)$. Hence, Z_{i1} satisfies the accelerated failure time model, so the inference for β can be conducted by solving the log-rank estimating equation, which is commonly used for the estimation in the accelerated failure time model. Specifically, the log-rank estimating equation is given by

$$\sum_{i=1}^{n_1} \Delta_{i1}\left\{X_{i1} - \frac{\sum_j I(\log Z_{j1} + \beta^T X_{j1} \geq \log Z_{i1} + \beta^T X_{i1})X_{j1}}{\sum_j I(\log Z_{j1} + \beta^T X_{j1} \geq \log Z_{i1} + \beta^T X_{i1})}\right\} = 0. \qquad (3.3)$$

Since S_0 is unknown, we estimate S_0 nonparametrically using the smoothed Breslow estimator as follows:

$$\hat{S}_0(y|x) = \exp\left\{ -\sum_{j=1}^{n_0} I(Y_{j0} \le y) \frac{\triangle_{j0} K_{a_n}(X_{j0} - x)}{\sum_{k=1}^{n_0} I(Y_{k0} \ge Y_{j0}) K_{a_n}(X_{k0} - x)} \right\}, \tag{3.4}$$

where $K_{a_n}(x) = K(x/a_n)/a_n^d$ with a_n being the bandwidth and d the dimension of X. Thus, Z_{i1} is estimated by $\hat{Z}_{i1} = -\log \hat{S}_0(Y_{i1}|X_{i1})$. After plugging $\hat{Z}_{i1}$ into (3.3), $\hat{\beta}$ is obtained by solving

$$\sum_{i=1}^{n_1} \Delta_{i1} \left\{ X_{i1} - \frac{\sum_j I(\log \hat{Z}_{j1} + \hat{\beta}^T X_{j1} \ge \log \hat{Z}_{i1} + \hat{\beta}^T X_{i1}) X_{j1}}{\sum_j I(\log \hat{Z}_{j1} + \hat{\beta}^T X_{j1} \ge \log \hat{Z}_{i1} + \hat{\beta}^T X_{i1})} \right\} = 0.$$

Remark 3.1. When X is discrete, the estimator for S_0, $\hat{S}_0(y|x)$ in (3.4) can be replaced by the Breslow estimator using the data with $X_{j0} = x$. i.e.,

$$\hat{S}_0(y|x) = \exp\left\{ -\sum_{j=1}^{n_0} \frac{I(Y_{j0} \le y) \triangle_{j0} I(X_{j0} = x)}{\sum_{k=1}^{n_0} I(Y_{k0} \ge Y_{j0}) I(X_{k0} = x)} \right\}.$$

Remark 3.2. When X has more than one continuous covariate, the kernel estimate $\hat{S}_0$ may not perform well with a moderate sample size. In this case, we suggest estimating $S_0(y|x)$ based on the Cox regression model using the nondiseased data. That is,

$$\hat{S}_0(y|x) = \exp\Big[-\hat{\Lambda}(y) \exp(\hat{\gamma}^T x) \Big],$$

where $\hat{\Lambda}(y)$ is the estimated cumulative baseline cumulative function and $\hat{\gamma}$ is the regression parameter estimate.

We next describe the procedures for estimating the function G and the ROC function

given in (3.1). Clearly, $P(Z_{i1}e^{\beta^T X_{i1}} \leq z|X_{i1}) = 1 - G(-z)$. Therefore, $Z_{i1}e^{\beta^T X_{i1}}$ is independent of X_{i1} and has distribution function $1 - G(-z)$. This implies that we can estimate G consistently by using the empirical distribution of $W_{i1} \equiv Z_{i1}e^{\beta^T X_{i1}}$. Since Z_{i1} is subject to right-censoring, so is W_{i1}. We use the Kaplan-Meier estimator to estimate the survival function of W_{i1}. After replacing W_{i1} with its estimate

$$\hat{W}_{i1} = -e^{\hat{\beta}^T X_{i1}} \log \hat{S}_0(Y_{i1}; X_{i1}), \quad i = 1, ..., n_1,$$

we estimate $G(\cdot)$ using

$$\hat{G}(t) = \prod_{i=1}^{n_1} \left[1 - \frac{\Delta_{i1} I(\hat{W}_{i1} \leq -t)}{\sum_{j=1}^{n_1} I(\hat{W}_{j1} \geq \hat{W}_{i1})} \right]. \tag{3.5}$$

Finally, the ROC curve for any covariate value X is estimated by

$$\widehat{ROC}_X(t) = \widehat{G}(e^{\hat{\beta}^T X} \log t), \quad t \in (0, 1). \tag{3.6}$$

To make inference, we estimate the variances of $\hat{\beta}$ and $\hat{G}$ using the bootstrap method; Bootstrap samples are drawn repeatedly with replacement from the dataset, and for each bootstrap sample, β and G are estimated. We then use the variances of these $\hat{\beta}$'s and $\hat{G}$'s as our estimates. Alternatively, the variances can be estimated by a different resampling method, which is described in Section 3.7.2.

Remark 3.3. The proposed approach can be generalized to handle the situation when each subject may have multiple or repeated biomarkers. In this case, the estimating equation for β is replaced by

$$\sum_{i=1}^{n_1} \sum_{k=1}^{n_{i1}} \Delta_{ik1} \left\{ X_{ik1} - \frac{\sum_j \sum_{l=1}^{n_{j1}} I(\log \hat{Z}_{jl1} + \hat{\beta}^T X_{jl1} \geq \log \hat{Z}_{ik1} + \hat{\beta}^T X_{ik1}) X_{jl1}}{\sum_j \sum_{l=1}^{n_{j1}} I(\log \hat{Z}_{jl1} + \hat{\beta}^T X_{jl1} \geq \log \hat{Z}_{ik1} + \hat{\beta}^T X_{ik1})} \right\} = 0,$$

where Δ_{ij1}, $\hat{Z}_{ij1}$ and X_{ij1} are the observations of jth measurement for subject i in the diseased group, and $\hat{Z}_{ij1}$ can be estimated similarly as $\hat{Z}_{i1}$. The bootstrapping method can still be used for inference by randomly selecting subjects for each bootstrap sample.

3.3 Asymptotic Properties

In this section, we derive the asymptotic properties of $\hat{\beta}$ and $\hat{G}$. First, we assume the following conditions hold.

(C.1) The true parameter value, β_0, belongs to a compact set $\mathcal{B}$.

(C.2) The true densities with respective to a dominating measure for (Y_1, C_1, X_1) and (Y_0, C_0, X_0) are $(\chi + 1)$-continuously differentiable, where $\chi > d/2$. Additionally, X_1 and X_0 have bounded support.

(C.3) The matrix $[1, X_1]$ is linearly independent with positively.

(C.4) The kernel function $K(\cdot)$ is differentiable with bounded symmetric support and first $(\chi - 1)$ moments begin zero. Moreover, $na_n^d \to \infty$ and $na_n^{2\chi} \to 0$.

(C.5) $n_0/n \to \nu \in (0, 1)$, where $n = n_0 + n_1$.

(C.1) and (C.5) are standard conditions for this type of problem. In (C.2), both C_1 and C_2 are same as τ based on our model. (C.3) ensures the identifiability of the regression parameters, and (C.4) states the restrictions on the choice of possible kernel functions. Examples of the kernel function include the Gaussian kernel and the Epanechnikov kernel for $\chi = 2$. Both (C.2) and (C.4) are necessary conditions to prove the asymptotic distribution of $\hat{\beta}$. Obviously, if S_0 is estimated using the Breslow method with discrete X_1 or from the Cox regression method, (C.4) is not needed.

Under these conditions, the following consistency theorem holds.

Theorem 3.1. *Under Conditions (C1)-(C5),* $|\widehat{\beta} - \beta_0| \rightarrow_{a.s.} 0$.

The following two theorems state the asymptotic normality of $\hat{\beta}$ and $\hat{G}$.

Theorem 3.2. *Under Conditions (C1)-(C5),* $\sqrt{n}(\hat{\beta} - \beta_0)$ *converges in distribution to a mean zero normal random vector as* $n \rightarrow \infty$.

Theorem 3.3. *Under Conditions (C1)-(C5),* $\sqrt{n}(\hat{G}(\log t) - G_0(\log t))$ *converges weakly to a zero mean Gaussian process in* $l^{\infty}([0,1])$.

The proofs of Theorems 3.1-3.3 are provided in Section 3.7. For the proof of Theorem 3.1, we use the fact that $\hat{S}_0(y;x)$ converges uniformly in (y,x) to $S_0(y;x)$ as n goes to ∞, which is given in Zeng (2004). We then apply Theorems 2.10.3 and 5.9 of van der Vaart (1998). The proofs of Theorem 3.2 and 3.3 follow the same arguments as in Zeng (2004), and we use the central limit theorems for the empirical process indexed by classes depending on samples (Theorem 2.11.23, van der Vaart and Wellner, 1996).

3.4 Simulation Studies

Simulation studies were conducted to examine the performance of the proposed method. First, we defined the true function of G as $G(x) = \exp(\alpha x)$. Then, the ROC function given in (3.1) becomes $ROC_X(t) = \exp[\alpha e^{\beta^T X} \log(t)]$. The biomarker values for diseased and non-diseased subjects, y_1 and y_0, were generated by

$$y_0 = -\ln(U_0)/(\exp(\gamma^T X)) \text{ and } y_1 = -\ln(U_1)/(\alpha \exp(\gamma^T X + \beta^T X)), \qquad (3.7)$$

where U_0 and U_1 are uniform random variables from $U[0,1]$. It is easy to check such (y_0, y_1) gives the above ROC function. We used an equal number of diseased and nondiseased subjects but varied the total sample size n from 200 to 400. Additionally, we set

the upper detection limit τ as the 95th percentile of the biomarker in the nondiseased group.

We conducted three different simulations with different types of covariates. For the first simulation, a binary covariate X was generated from a Bernoulli distribution with probability 0.5 and true parameters in (3.7) were set to $\beta = 0.5$, $\gamma = -0.5$ and $\alpha = 0.5$. Because X was discrete, we estimated $S_0(y|x)$ using the Breslow estimator given in Remark 3.1. In the second simulation, we used a continuous covariate generated from uniform (0,1) distribution, and true parameters were set to $\beta = -1$, $\gamma = -0.5$ and $\alpha = 1.5$. In this simulation, $S_0(y|x)$ was estimated using the smoothed Breslow estimator given in (3.4), where the Gaussian kernel function $K(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$ with $a_n = {n_1}^{-1/3}$ were applied. For the last simulation, two continuous covariates, both generated from uniform (0,1), were used with $\beta = (-1.2, -2)^T$, $\gamma = (-2, 2)^T$ and $\alpha = 5$. We then fitted the Cox model to estimate $\hat{S}_0$ as described in Remark 3.2. In all the simulation studies, we obtained $\hat{\beta}$ by solving the log-rank estimating equation (3.3) through bisection search.

Table 3.1 summarizes the simulation results based on 1000 replicates. Column "Est" is the average value of the estimates from 1,000 replicates; column "ASE" is the average of the estimated standard errors by the bootstrap method with 1000 replicates; column "SE" is the standard deviation of the estimates; column "CP" gives the ($100\times$) coverage proportion of the 95% confidence intervals based on the asymptotic normality. Overall, the estimates for β are very close to the actual values across sample sizes, and the estimated standard errors using the bootstrap method approximate the empirical standard errors well. In addition, the coverage proportions of 95% CIs are close to the nominal level of 95% across sample sizes. In the same table, we present the true and estimated function G at three fixed points, where the three points were chosen to be the quartiles of the true distribution of $-W_1$ ($= -e^{\beta^T X_1} \log S_0(Y_1; X_1)$). The coverage

Table 3.1: Simulation results with $G(t) = \exp(\alpha t)$

Par.	True	$n_1 = n_0 = 100$				$n_1 = n_0 = 200$			
		Est	ASE	SE	CP	Est	ASE	SE	CP
Simulation Study 1. X: 0 or 1									
β	0.5	0.479	0.309	0.302	95.7	0.517	0.219	0.210	94.9
$G(-2.7)$	0.259	0.255	0.081	0.080	95.6	0.262	0.060	0.060	93.0
$G(-1.3)$	0.522	0.519	0.084	0.082	94.8	0.524	0.060	0.059	94.0
$G(-0.5)$	0.779	0.774	0.063	0.060	96.0	0.778	0.044	0.042	94.3
Simulation Study 2. X $\sim$ Uniform(0,1)									
β	-1	-1.120	0.510	0.494	95.4	-1.123	0.364	0.346	94.7
$G(-0.8)$	0.301	0.282	0.103	0.101	96.9	0.279	0.076	0.076	94.9
$G(-0.4)$	0.549	0.525	0.109	0.106	94.3	0.526	0.080	0.077	94.3
$G(-0.15)$	0.799	0.780	0.080	0.076	95.1	0.782	0.056	0.053	95.0
Simulation Study 3. $X_1 \sim$ Uniform(0,1), $X_2 \sim$ Uniform(0,1)									
β_1	-1.2	-1.196	0.582	0.564	95.2	-1.198	0.393	0.376	95.4
β_2	-2	-2.064	0.578	0.575	95.6	-2.029	0.392	0.384	95.0
$G(-0.26)$	0.273	0.272	0.135	0.141	92.3	0.275	0.101	0.101	95.0
$G(-0.14)$	0.497	0.479	0.148	0.153	94.1	0.492	0.108	0.106	94.5
$G(-0.06)$	0.741	0.715	0.121	0.195	95.3	0.732	0.083	0.077	95.2

proportions of 95% CIs for G were calculated based on the log transformation of G. For all simulations, the estimated values of G are very close to the actual values at all three points. Figures 3.2-3.4 display the true and estimated ROC curves from the three simulations. Apparently, all three figures show that estimated ROC curves are extremely close to the true ROC curves.

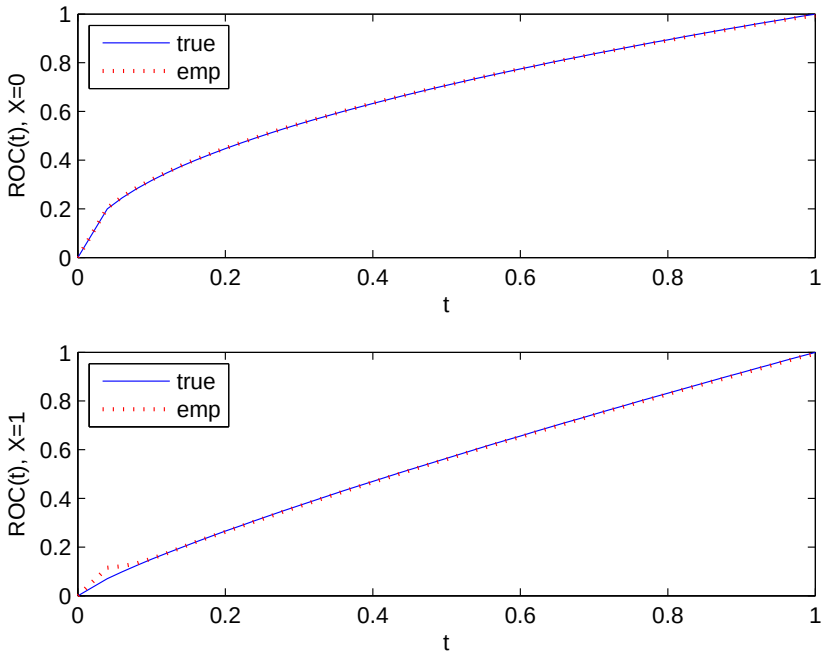

Figure 3.2: Plots of the true and estimated function G (n=200) from Simulation Study 1 (X: 0 or 1): the solid curve is the true G and the dashed curve is the average of the estimated curves from 1,000 replicates.

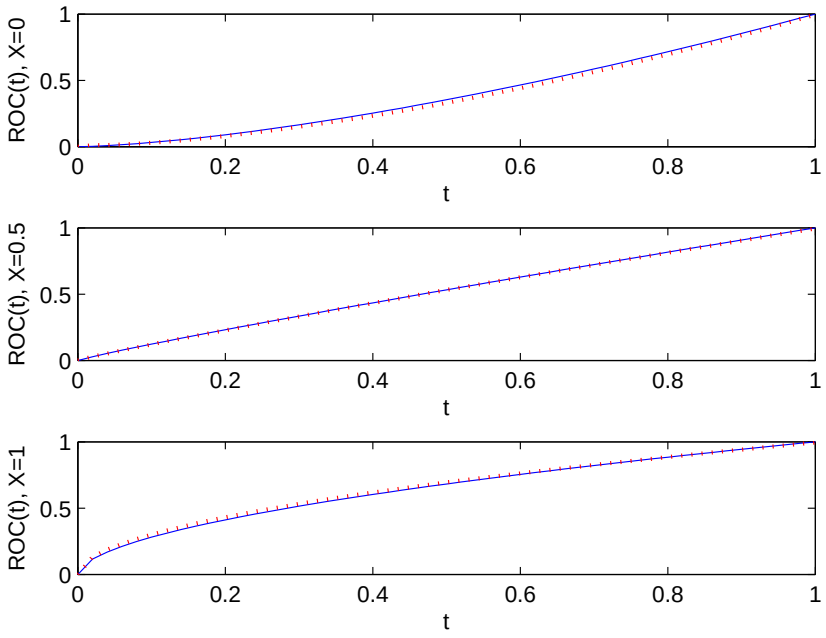

Figure 3.3: Plots of the true and estimated function G (n=200) from Simulation Study 2 (X$\sim$ Uniform(0,1)): the solid curve is the true G and the dashed curve is the average of the estimated curves from 1,000 replicates.

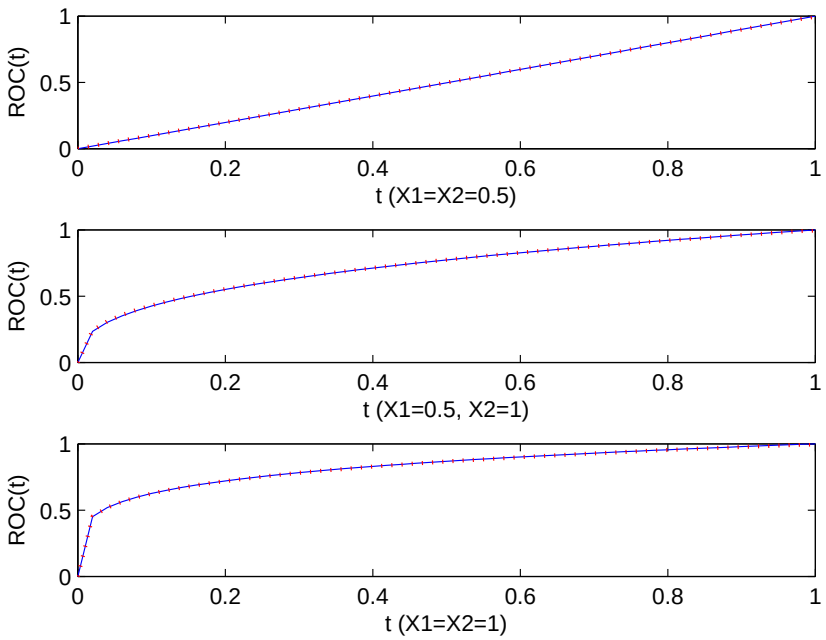

Figure 3.4: Plots of the true and estimated function G (n=200) from Simulation Study 3 (X1, X2$\sim$ Uniform(0,1)): the solid curve is the true G and the dashed curve is the average of the estimated curves from 1,000 replicates.

3.5 Application

We illustrate our approach with a prostate cancer dataset. Prostate-specific antigen (PSA) is a protein produced by the prostate gland, and the PSA test measures the level of PSA in the blood. Most healthy men have PSA levels under 4 nanograms per milliliter (ng/mL) of blood, and the chance of having prostate cancer goes up as the PSA level increases. PSA occurs in 2 major forms in the blood. One form is attached to blood proteins while the other circulates free (unattached). The free PSA (fPSA) is the ratio of how much PSA circulates free compared to the total PSA level. Low free PSA may indicate prostate cancer and most men with prostate cancer have a free PSA below 15%. If free PSA is below 7%, prostrate cancer is most likely. According to American Cancer Society and National Cancer Institute, men with free PSA at 7% or lower should undergo biopsy. We used a dataset of 71 prostate cancer subjects and 68 controls who participated in the Beta-Carotene and Retinol Efficacy Trial (CARET).

The objective of this analysis was to evaluate the capacity of free PSA for discriminating men with prostate cancer from those without before the onset of clinical symptoms. This trial enrolled 12,025 men at high risk of lung cancer as a result of smoking or asbestos exposure and evaluated the efficacy of beta-carotene and retinol in preventing lung cancer, which began in 1985 and terminated in 1994. Subjects who participated in CARET had serum samples drawn at baseline and at two-year intervals thereafter (Goodman et al., 1993; Etzioni et al., 1999). Blood samples drawn after diagnosis of prostate cancer were excluded from this analysis, leaving on average 1 to 7 blood samples per subject (average 3.2 samples per case and 3.5 samples per control). The average age was 63.7 (range from 46.7 to 80.8) and the average time was -3.06 years (range from -9.008 to -0.003 yrs). Previous studies have suggested that age and the time PSA measured may affect the discrimination of prostate cancer. Let X be the age when PSA was measured and T be the time between the onset of symptoms and

the time at which the serum sample was drawn, so that time is negative and increase to 0 as measurements are closer to the time of clinical diagnosis. We then fitted the following accelerated ROC model adjusting for age X and time T:

$$ROC_{T,X}(u) = G(\mathrm{e}^{\beta_x X + \beta_t T} \log u).$$

We found $\hat{\beta}_x = 0.0485$ with SE 0.0248 (p-value 0.0505) and $\hat{\beta}_t = -0.0587$ with SE 0.0442 (p-value 0.1841). The positive coefficient for age suggests that discrimination is better in younger men than in old men, and the negative coefficient for time implies that discrimination improves when PSA is measured closer to diagnosis although the time T is not significant. Figure 3.5 displays the estimated ROC curves at age=57, 63, and 68 when T=-2.82, which is the median of time T. The AUCs are 0.8579, 0.8103, and 0.7623 at age=57, 63, and 68, respectively. Again, it appears that free PSA performes better for younger men. To assess the fit of our model, in Figure 3.6, we plotted the empirical ROC curves for free PSA for each of three age groups where the groups were based on the categorization of age X. The figure shows a similar pattern as Figure 3.5. The empirical AUCs of each age group are 0.8575, 0.8062, and 0.7527 for age $\leq$ 61, 61$<$ age $\leq$ 65, and age $\geq$ 65. The AUCs based on the proposed method agree well with the empirical AUCs for these groups, demonstrating the good fit of the accelerated regression model.

3.6 Discussion

In this chapter, we have focused on assessing the accuracy of biomarkers by adjusting for covariates that could influence the performance of biomarkers. We developed an accelerated ROC model by employing the properties of the accelerated failure time model. Based on Pepe (1997, 2000a)'s method, the covariate effect is related to the

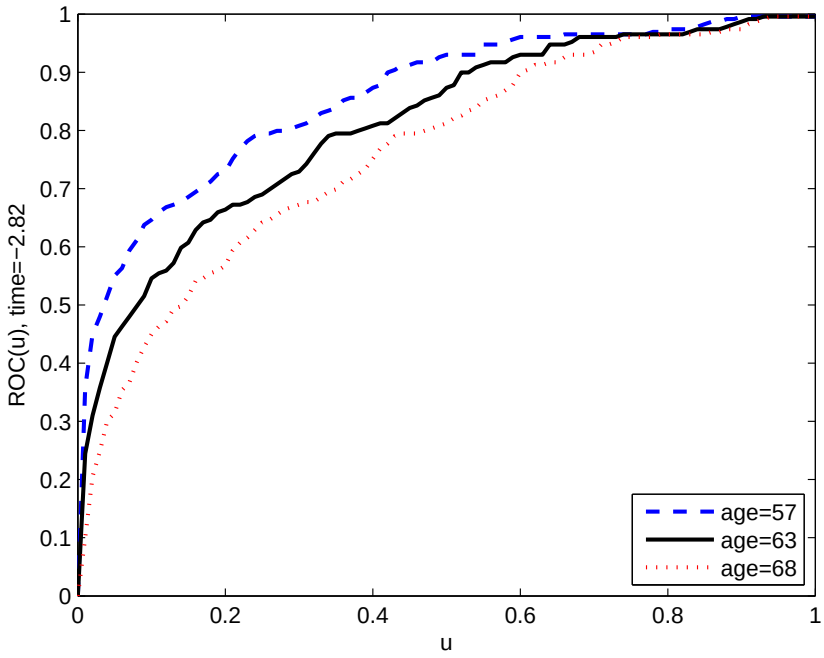

Figure 3.5: Estimated ROC curve for PSA adjusted for age and time T

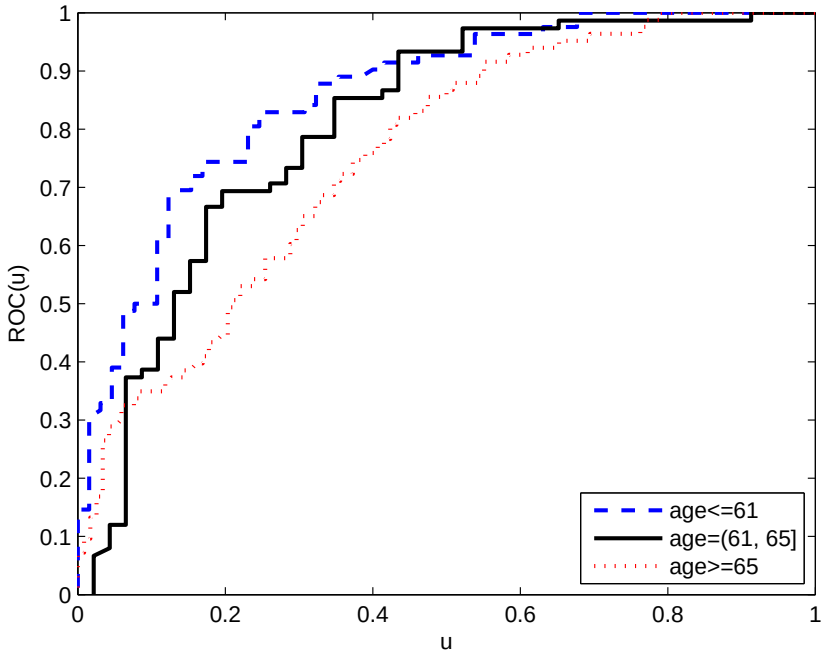

Figure 3.6: Empirical ROC curve for PSA by time

location-shift of the original ROC curves. On the other hand, in our proposed method, the effect of covariates in the ROC model relates to rescaling the original ROC curve. Therefore, our model provides a useful alternative to the traditional method. Note that the parameter estimates of covariates based on the log-rank estimating equation may not be efficient. Regarding this issue, we will explore other methods which attain the semiparametric efficiency.

3.7 Appendix

3.7.1 Proof of Theorem 3.1

With direct calculations, condition (C.3) implies that matrix

$$\sigma_1 \equiv -\frac{\partial}{\partial\beta} E\left[\Delta_1 \frac{Q_1(\log Z_1 + \beta^T X_1)}{Q_0(\log Z_1 + \beta^T X_1)}\right]$$

is positive for $\beta \in \mathcal{B}$, where $Q_1(x) = E[X_1 I(\log Z_1 + \beta^T X_1 \geq x)]$ and $Q_0 = E[I(\log Z_1 + \beta^T X_1 \geq x)]$. Therefore, β_0 must be the unique solution to the following equation

$$E\left[\Delta\left(X_1 - \frac{Q_1(\log Z_1 + \beta^T X_1)}{Q_0(\log Z_1 + \beta^T X_1)}\right)\right] = 0,$$

We introduce the following notations. We use $\mathbf{P}_{n1}$ and $\mathbf{P}_1$ to denote the empirical measure and expectation base on i.i.d observations in the diseased group, i.e., $(Y_{i1}, X_{i1}, \Delta_{i1}), i = 1, ..., n_1$. Similarly, we use $\mathbf{P}_{n0}$ and $\mathbf{P}_0$ to denote the empirical measure and expectation based on i.i.d observations in the non-diseased group, i.e., $(Y_{j0}, X_{j0}, \Delta_{j0}), j = 1, ..., n_0$. Moreover, we use $\mathbf{G}_{n1}$ and $\mathbf{G}_{n0}$ denote empirical processes $\sqrt{n_1}(\mathbf{P}_{n1} - \mathbf{P}_1)$ and $\sqrt{n_0}(\mathbf{P}_{n0} - \mathbf{P}_0)$ respectively. Thus, by definition, $\hat{\beta}$ should solve

$$0 = \mathbf{P}_{n1}\left[\Delta_1\left\{X_1 - \frac{\sum_{i=1}^{n_1} I(\log \hat{Z}_{i1} + \hat{\beta}^T X_{i1} \geq \log \hat{Z}_1 + \hat{\beta}^T X_1) X_{i1}}{\sum_{i=1}^{n_1} I(\log \hat{Z}_{i1} + \hat{\beta}^T X_{i1} \geq \log \hat{Z}_1 + \hat{\beta}^T X_1)}\right\}\right].$$

We start to show the consistency of $\hat{\beta}$. First, conditional on non-diseased data, $(\hat{Z}_{i1}, X_{i1}, \Delta_{i1})$ are i.i.d. Therefore, the class

$$\mathcal{F} \equiv \left\{I(x \geq \log \hat{Z}_1 + \beta^T X_1) : x \in (-\infty, \infty), \beta \in \mathcal{B}\right\}$$

is the VC-class, so is Donsker. Note that the following random functions

$$n^{-1}\sum_{i=1}^{n_1} I(\log \hat{Z}_{i1}+\beta^T X_{i1} \geq \log \hat{Z}_1+\beta^T X_1)X_{i1},\ \ n^{-1}\sum_{i=1}^{n_1} I(\log \hat{Z}_{i1}+\beta^T X_{i1} \geq \log \hat{Z}_1+\beta^T X_1),$$

and

$$E^*\left[\frac{I(\log \hat{Z}_1+\beta^T X_1 \geq \log \hat{Z}_1^*+\beta^T X_1^*)X_1}{n^{-1}\sum_{i=1}^{n_1} I(\log \hat{Z}_{i1}+\beta^T X_{i1} \geq \log \hat{Z}_1^*+\beta^T X_1^*)}\right],$$

where here and in the squeal, E^* and E^{**} denote the expectation with respect to those random variables with asterisk and double asterisk respectively, can be expressed as the limit of the convex combinations of $\mathcal{F}$ and are bounded from above. Thus, they belong to $sconv\mathcal{F}$, which is a Donsker class from Theorem 2.10.3 of van der Vaart and Wellner (1996). Therefore, by the Glivenko-Cantelli theorem, it is easy to see

$$\sup_{\beta}\Bigg|\mathbf{P}_{n1}\left[\Delta_1\left\{X_1-\frac{\sum_{i=1}^{n_1} I(\log \hat{Z}_{i1}+\beta^T X_{i1} \geq \log \hat{Z}_1+\beta^T X_1)X_{i1}}{\sum_{i=1}^{n_1} I(\log \hat{Z}_{i1}+\beta^T X_{i1} \geq \log \hat{Z}_1+\beta^T X_1)}\right\}\right]$$

$$-E\left[\Delta_1\left\{X_1-\frac{E^*[I(\log \hat{Z}_1^*+\beta^T X_1^* \geq \log \hat{Z}_1+\beta^T X_1)X_1^*]}{E^*[I(\log \hat{Z}_1^*+\beta^T X_1^* \geq \log \hat{Z}_1+\beta^T X_1)]}\right\}\right]\Bigg| \rightarrow_{a.s.} 0.$$

Furthermore, as n goes to ∞, $\hat{S}_0(y;x)$ converges uniformly in (y,x) to $S_0(y;x)$ as shown in Zeng (2004). Thus, the limit function

$$E\left[\Delta_1\left\{X_1-\frac{E^*[I(\log \hat{Z}_1^*+\beta^T X_1^* \geq \log \hat{Z}_1+\beta^T X_1)X_1^*]}{E^*[I(\log \hat{Z}_1^*+\beta^T X_1^* \geq \log \hat{Z}_1+\beta^T X_1)]}\right\}\right]$$

converges uniformly in β to

$$E\left[\Delta_1\left\{X_1-\frac{Q_1(\log Z_1+\beta^T X_1)}{Q_0(\log Z_1+\beta^T X_1)}\right\}\right].$$

The latter has a unique minimum zero at β_0 by condition (C.1). Additionally, it satisfies the separability at β_0 by condition (C.3). Therefore, from Theorem 5.9 of van der Vaart

(1998), $\hat{\beta}$ converges almost surely to β_0.

3.7.2 Proof of Theorem 3.2

Next, we derive the asymptotic distribution of $\hat{\beta}$. From equation

$$\mathbf{P}_{n1}\left[\Delta_1\left\{X_1 - \frac{\sum_{i=1}^{n_1} I(\log \hat{Z}_{i1} + \hat{\beta}^T X_{i1} \geq \log \hat{Z}_1 + \beta^T X_1) X_{i1}}{\sum_{i=1}^{n_1} I(\log \hat{Z}_{i1} + \hat{\beta}^T X_{i1} \geq \log \hat{Z}_1 + \hat{\beta}^T X_1)}\right\}\right] = 0. \tag{3.8}$$

if we define

$$\hat{Q}_1(x) = E[X_1 I(\log \hat{Z}_1 + \hat{\beta}^T X_1 \geq x)], \quad \hat{Q}_0(x) = E[I(\log \hat{Z}_1 + \hat{\beta}^T X_1 \geq x)],$$

then we obtain that

$$\begin{aligned}
& \mathbf{G}_{n1}\left[\Delta_1\left\{X_1 - \frac{\sum_{i=1}^{n_1} I(\log \hat{Z}_{i1} + \hat{\beta}^T X_{i1} \geq \log \hat{Z}_1 + \hat{\beta}^T X_1) X_{i1}}{\sum_{i=1}^{n_1} I(\log \hat{Z}_{i1} + \hat{\beta}^T X_{i1} \geq \log \hat{Z}_1 + \hat{\beta}^T X_1)}\right\}\right] \\
& -\mathbf{G}_{n1} E^*\left[\frac{\Delta_1^* I(\log \hat{Z}_1 + \hat{\beta}^T X_1 \geq \log \hat{Z}_1^* + \hat{\beta}^T X_1^*) X_1}{n^{-1}\sum_{i=1}^{n_1} I(\log \hat{Z}_{i1} + \hat{\beta}^T X_{i1} \geq \log \hat{Z}_1^* + \hat{\beta}^T X_1^*)}\right] \\
& +\mathbf{G}_{n1} E^*\left[\frac{\Delta_1^* I(\log \hat{Z}_1 + \hat{\beta}^T X_1 \geq \log \hat{Z}_1^* + \hat{\beta}^T X_1^*)\hat{Q}_1(\log \hat{Z}_1^* + \hat{\beta}^T X_1^*)}{n^{-1}\sum_{i=1}^{n_1} I(\log \hat{Z}_{i1} + \hat{\beta}^T X_{i1} \geq \log \hat{Z}_1^* + \hat{\beta}^T X_1^*)\hat{Q}_0(\log \hat{Z}_1^* + \hat{\beta}^T X_1^*)]}\right] \\
= & -\sqrt{n_1} E\left[\Delta_1\left\{X_1 - \frac{E^*[I(\log \hat{Z}_1^* + \hat{\beta}^T X_1^* \geq \log \hat{Z}_1 + \hat{\beta}^T X_1) X_1^*]}{E^*[I(\log \hat{Z}_1^* + \hat{\beta}^T X_1^* \geq \log \hat{Z}_1 + \hat{\beta}^T X_1)]}\right\}\right].
\end{aligned}$$

From the Donsker theorem, we have

$$-\sqrt{n_1} E\left[\Delta_1\left\{X_1 - \frac{\hat{Q}_1(\log \hat{Z}_1 + \hat{\beta}^T X_1)}{\hat{Q}_0(\log \hat{Z}_1 + \hat{\beta}^T X_1)}\right\}\right] = \mathbf{G}_{n1} g(\Delta_1, X_1, Z_1; \beta_0) + o_p(1), \tag{3.9}$$

where

$$g(\Delta_1, X_1, Z_1; \beta_0) = \Delta_1\left\{X_1 - \frac{Q_1(\log Z_1 + \beta_0^T X_1)}{Q_0(\log Z_1 + \beta_0^T X_1)}\right\}$$

$$-E^*\left[\frac{\Delta_1^* I(\log Z_1+\beta^T X_1 \geq \log Z_1^*+\beta^T X_1^*)X_1}{Q_0(\log Z_1^*+\beta_0^T X_1^*)}\right]$$
$$+E^*\left[\frac{\Delta_1^* I(\log Z_1+\beta^T X_1 \geq \log Z_1^*+\beta^T X_1^*)Q_1(\log Z_1^*+\beta_0^T X_1^*)}{Q_0(\log Z_1^*+\beta_0^T X_1^*)^2}\right].$$

On the other hand, from condition (C.2),

$$\hat{Q}_0(x)=E\left[P\left(Y_1\geq \hat{H}_0^{-1}(e^{x-\hat{\beta}^T X_1};X_1)\Big|X_1\right)\right],$$

where $\hat{H}_0^{-1}(y;x)$ denotes the inverse of $H_0(y;x)\equiv -\log S_0(y;x)$ for given x. Thus, if letting $f_1(y|x)$ be the conditional density of Y_1 given X_1, then

$$\hat{Q}_0(x)=-E\left[f_1\left(H_0^{-1}(e^{x-\hat{\beta}^T X_1};X_1)\Big|X_1\right)\left(\hat{H}_0^{-1}(e^{x-\hat{\beta}^T X_1};X_1)-H_0^{-1}(e^{x-\hat{\beta}^T X_1};X_1)\right)\right]$$
$$+E\left[P\left(Y_1\leq H_0^{-1}(e^{x-\hat{\beta}^T X_1});X_1)\Big|X_1\right)\right]+o(1).$$

By slightly modifying the inverse map lemma (Lemma 3.9.20, van der Vaart and Wellner 1996), we can show

$$\hat{H}_0^{-1}(e^{x-\hat{\beta}^T X_1};X_1)-H_0^{-1}(e^{x-\hat{\beta}^T X_1};X_1)$$
$$=-\frac{\hat{H}_0(H_0^{-1}(e^{x-\hat{\beta}^T X_1};X_1);X_1)-H_0(H_0^{-1}(e^{x-\hat{\beta}^T X_1};X_1);X_1)}{H_0'(H_0^{-1}(e^{x-\beta_0^T X_1};X_1);X_0=X_1)}+o(1),$$

and it holds uniformly in $x,\hat{\beta}$ and X_1. Moreover, since $\hat{H}_0(\cdot;x)$ converges to $H_0(\cdot;x)$ in $D[0,\tau]$ uniformly in x, we obtain

$$\begin{aligned}
&\hat{Q}_0(\log\hat{Z}_1^*+\beta_1^T X_1^*)\\
=\;& E\left[\frac{f_1\left(H_0^{-1}(Z_1^* e^{\beta_0^T X_1^*-\beta_0^T X_1};X_1)|X_1\right)}{H_0'(H_0^{-1}(Z_1^* e^{\beta_0^T X_1^*-\beta_0^T X_1};X_1);X_0=X_1)}\right.\\
&\left.\times\left(\hat{H}_0(H_0^{-1}(Z_1^* e^{\beta_0^T X_1^*-\beta_0^T X_1};X_1);X_1)-H_0(H_0^{-1}(Z_1^* e^{\beta_0^T X_1^*-\beta_0^T X_1};X_1);X_1)\right)\right]\\
&+E\left[P\left(Y_1\geq H_0^{-1}(\hat{Z}_1^* e^{\hat{\beta}_0^T X_1^*-\hat{\beta}^T X_1};X_1)\Big|X_1\right)\right]+o(1).
\end{aligned}$$

The last term on the right-hand side can be further approximated by

$$E\left[P\left(Y_1 \geq H_0^{-1}(Z_1^* e^{\beta_0^T X_1^* - \beta_0^T X_1}; X_1)\Big| X_1\right)\right]$$

$$E\left[\frac{f_1(H_0^{-1}(Z_1^* e^{\beta_0^T X_1^* - \beta_0^T X_1}; X_1)|X_1)}{H_0'(H_0^{-1}(Z_1^* e^{\beta_0^T X_1^* - \beta_0^T X_1}; X_1); X_0 = X_1)} Z_1^* e^{\beta_0^T X_1^* - \beta_0^T X_1}\right.$$

$$\left.\times\left\{\frac{\hat{H}_0(Y_1^*; X_1^*) - H_0(Y_1^*; X_1^*)}{H_0(Y_1^*; X_1^*)} + (\hat{\beta} - \beta_0)(X_1^* - X_1)\right\}\right].$$

Similarly, we can expand the numerator term in the left-hand side of (3.11), i.e., $\hat{Q}_1(\log \hat{Z}_1 + \hat{\beta}^T X_1)$. We eventually obtain that (3.11) is equivalent to

$$\begin{aligned}
\mathbf{G}_{n1} g(\Delta_1, X_1, Z_1; \beta_0) + o_p(1) = \sqrt{n_1}\sigma_1(\hat{\beta} - \beta_0)
\\
+\sqrt{n_1}E\left[\sigma_2(Z_1, X_1, X_1^*)\left(\hat{H}_0(H_0^{-1}(Z_1 e^{\beta_0^T X_1 - \beta_0^T X_1^*}; X_1^*); X_1^*)\right.\right.
\\
\left.\left.-H_0(H_0^{-1}(Z_1 e^{\beta_0^T X_1 - \beta_0^T X_1^*}; X_1^*)\right)\right]
\\
+\sqrt{n_1}E\left[\sigma_3(Y_1, X_1)\left(\hat{H}_0(Y_1; X_1) - H_0(Y_1; X_1)\right)\right], \qquad (3.12)
\end{aligned}$$

for some differentiable functions σ_2 and σ_3. Particularly, σ_1 has the same expression as given in condition (C.3) with $\beta = \beta_0$ so σ_1 is non-singular.

Using the same arguments as in Zeng (2004) and condition (C.4), we can show that uniformly in x and $y \in [0, \tau]$,

$$\begin{aligned}
&(\hat{H}_0(y; x) - H_0(y, x))
\\
= &\left\{(\mathbf{P}_{n0} - \mathbf{P}_0)\left[\frac{\Delta I(Y_0 \leq y)(n_0 a_n^d)^{-1} K_{a_n}(X_0 - x)}{(n_0 a_n^d)^{-1}\sum_{k=1}^{n_0} I(Y_{k0} \geq Y_0) K_{a_n}(X_{k0} - x)}\right]\right.
\\
&-(\mathbf{P}_{n0} - \mathbf{P}_0)
\\
&\left.E^*\left[\frac{\Delta^* I(Y_0 \geq Y_0^*)(n_0 a_n)^{-1} K_{a_n}(X_0^* - x)}{(n_0 a_n^d)^{-1}\sum_{k=1}^{n_0} I(Y_{k0} \geq Y_0^*) K_{a_n}(X_{k0} - x) E^{**}[I(Y_0^{**} \geq Y_0^*) K_{a_n}(X_0^{**} - x)]}\right]\right\}
\end{aligned}$$

$$
\begin{aligned}
& +O(a_n^{\chi}) \\
\equiv\ & (\mathbf{P}_{n0} - \mathbf{P}_0) q_n(y, x, Y_0, X_0) + o_p(1).
\end{aligned}
$$

We plug the above expression into equation (3.13) then (3.11). From condition (C.4), we obtain

$$
\begin{aligned}
& \mathbf{G}_{n1} g(\Delta_1, X_1, Z_1; \beta_0) + o_p(1) \\
=\ & \sqrt{n_1}\sigma_1(\hat{\beta} - \beta_0) \\
& + \sqrt{n_1}(\mathbf{P}_{n0} - \mathbf{P}_0) E\left[\sigma_2(Z_1, X_1, X_1^*) q_n(H_0^{-1}(Z_1^* e^{-\beta_0^T X_1 + \beta_0^T X_1^*}; X_1^*), X_1^*, Y_0, X_0)\right] \\
& + \sqrt{n_1}(\mathbf{P}_{n0} - \mathbf{P}_0) E\left[\sigma_3(Y_1, X_1) q_n(Y_1, X_1, Y_0, X_0)\right] \qquad (3.13)
\end{aligned}
$$

Finally, we apply Theorem 2.11.23 in van der Vaart and Wellner (1996) to the last two terms in the right-hand side of equation (3.13). Particularly, their conditions are satisfies by observing that after integration by parts, both

$$
E\left[\sigma_2(Z_1, X_1, X_1^*) q_n(H_0^{-1}(Z_1 e^{\beta_0^T X_1 - \beta_0^T X_1^*}; X_1^*), X_1^*, Y_0, X_0)\right]
$$

and

$$
E\left[\sigma_3(Y_1, X_1) q_n(Y_1, X_1, Y_0, X_0)\right]
$$

converges uniformly in (Y_0, X_0) and they have bounded total variation in Y_0 uniformly in X_0 and are Lipschitz continuous in X_0. The latter implies the entropy condition in Theorem 2.11.23. Therefore, combined the above results and the non-singularity of σ_1 in (3.13), we obtain the asymptotic normality of $\hat{\beta}$.

Remark A.1. When X's take discrete values, the proof can be much simplified. Particulary, we can set $a_n = 1/n$ and $K_{a_n}(x) = I(x = 0)$ in the above arguments.

Remark A.2. From the proof, it is clear that the asymptotic variance for $\hat{\beta}$ is the summation of two sources of variations: one is that the variability, denoted by V_1, in the log-rank estimation conditional on the non-diseased data; the other one, denote by V_2, is due to estimating $S_0(y;x)$ using the non-diseased data. Therefore, to estimate the asymptotic variance of $\hat{\beta}$, we can estimate these two source of variation separately. Specifically, we estimate V_1 using the log-rank estimation procedure as given in Zeng and Lin (2008). While, we estimate V_2 using the following resampling method: according to the expansion in Zeng (2004), we generate $G_1, ..., G_{n_0}$ from $N(0,1)$ and define

$$\tilde{H}_0(y,x) = \hat{H}_0(y,x) + n_0^{-1} \sum_{j=1}^{n_0} G_j \left[\frac{\Delta_j I(Y_{j0} \le y) K_{a_n}(X_{j0} - x)}{\sum_{k=1}^{n_0} I(Y_{k0} \ge Y_{j0}) K_{a_n}(X_{k0} - x)} \right]$$

$$- n_0^{-1} \sum_{j=1}^{n_0} G_j \sum_{s=1}^{n_0} \left[\frac{\Delta_{s0} I(Y_{j0} \ge Y_{s0}) K_{a_n}(X_{s0} - x)}{(\sum_{k=1}^{n_0} I(Y_{k0} \ge Y_{s0}) K_{a_n}(X_{k0} - x))^2} \right].$$

We then replace $\hat{Z}_{1i}$ by $\tilde{H}_0(Y_{1i}; X_{1i})$ and re-estimate β. Then V_2 can be estimated as the sample variance of β's by repeating this procedure for a number of times. The validity of such a resampling method can be justified using the procedure as in the previous proof.

Remark A.3. When $S_0(y|x)$ is estimated by the Cox model, the only difference is in the expressions of $\hat{H}_0(y;x) - H_0(y,x)$; the influence function $q_n(y,x,Y_0,X_0)$ is given by the influence function of $\exp[-\hat{\Lambda}(y)\exp(\hat{\gamma}^T x)]$, where $(\hat{\Lambda}, \hat{\gamma})$ is the nonparametric maximum likelihood estimator in the Cox model.

3.7.3 Proof of Theorem 3.3

The asymptotic property of $\hat{G}(t)$ follows the same expansion as the proof of Theorem 3.2 but we utilize the differentiability of the product-limit function. Let S_W denote the survival function for W_1 and H_W denote the cumulative hazard function of W_1. We have

$$\hat{G}(t)-G_0(t) = -G_0(t)(\mathbf{P}_{n1}-\mathbf{P}_1)\left[\frac{\Delta_1 I(W_1 \leq -t)}{E^*[I(W_1^* \geq W_1)]} - E^*\left\{\frac{I(W_1 \geq W_1^*)\Delta_1^* I(W_1^* \leq -t)}{S_W(W_1^*)^2}\right\}\right]$$

$$-G_0(t)\left\{E\left[\frac{\Delta_1 I(\hat{W}_1 \leq -t)}{E^*[I(\hat{W}_1^* \geq \hat{W}_1)]}\right] - H_W(t)\right\} + o_p(n^{-1/2}).$$

We further expand the second term in the right-hand side as in the previous section to obtain

$$\tilde{\sigma}_1(\hat{\beta}-\beta_0) + E\left[\tilde{\sigma}_2(Z_1, X_1, X_1^*)\left(\hat{H}_0(H_0^{-1}(Z_1 e^{\beta_0^T X_1 - \beta_0^T X_1^*}; X_1^*); X_1^*)\right.\right.$$

$$\left.\left. -H_0(H_0^{-1}(-Z_1 e^{\beta_0^T X_1 - \beta_0^T X_1^*}; X_1^*)\right)\right]$$

$$+\sqrt{n_1}E\left[\tilde{\sigma}_3(Y_1, X_1)\left(\hat{H}_0(Y_1; X_1) - H_0(Y_1; X_1)\right)\right] + o_p(n^{-1/2}).$$

Hence, the asymptotic distribution of $\hat{G}(t)$ follows from the same arguments as in Theorem 3.2.

Chapter 4

Comparing Areas under ROC Curves in a Paired-Patient, Paired-Reader Design

4.1 Introduction

The Receiver Operating Characteristic (ROC) curve is widely used to evaluate the performance of a diagnostic test when test results are based on a continuous or ordinal variable (Metz, 1978; Hanly and McNeil, 1982; Swets and Pickett, 1982). In an ROC curve, the true positive rate is plotted in function of the false positive rate across all possible cutpoints. The area under the receiver operating characteristic curve (AUC) is a commonly used summary measure to evaluate the accuracy of a diagnostic test and relative accuracies of diagnostic tests can be compared by their corresponding areas under the ROC curves.

Diagnostic test results often depend on a reader's subjective interpretation, expertise, or experience. Because of the variability in readers' accuracy, studies of such diagnostic tests usually involve several readers. The most popular design for such a

multi-reader study is the paired-patient, paired-reader design (Obuchowski and Rockette, 1995), in which multiple readers interpret all test results of a sample of patients who undergoes multiple diagnostic tests. This design is most likely to occur in a radiology setting and is most efficient for comparing tests because it requires the smallest number of subjects. In addition, it demands one of the smallest reader samples and one of the fewest number of interpretations per reader compared with other study design (Zhou *et al.*, 2002).

The AUCs of diagnostic tests are correlated since diagnostic tests are based on the same patients or from the same readers. Therefore, the correlated structure of data must be taken into account in the analysis to avoid the inflation of testing powers. For the correlated ROC curves, nonparametric approaches have been proposed to estimate the AUCs. Especially, DeLong *et al.* (1988) developed a fully nonparametric approach to the comparison of correlated ROC curves by using the theory on U-statistics. This method, however, is applicable to the cases when each patient is examined by multiple tests or repeatedly examined using a single test. Obuchowski (1997) extended Delong *et al.* (1988)'s method and proposed a method for comparing the AUCs when there are multiple test results per subject. Under this data structure, test results from different patients are assumed to be independent although test results from the same patients are correlated. This method is not appropriate for the paired-patient, paired-reader design since, in this design, test results from different patients are correlated if they are read by a same reader. In addition, these two nonparametric methods do not consider the reader-variability caused by differences in interpretations.

On the other hand, several methods have been developed using the mixed-effects analysis of variance (ANOVA) models for the analysis of multi-reader ROC studies with multiple tests (Dorfman *et al.*, 1992; Obuchowski and Rockette, 1995; Beiden *et al.*, 2000). Dorfman *et al.* (1992) proposed a mixed-effects ANOVA model on the jackknife

pseudovalues of the summary measures of ROC curve. Obuchowski and Rockette (1995) applied a mixed-effects ANOVA model to the estimated summary measures of the ROC curve for each combination of patients, readers and tests. They developed the adjusted the ANOVA F-test in order to test differences in diagnostic accuracies. However, the validity of these methods depends heavily on assumptions on the underlying distribution of the random variables.

In this chapter, we introduce a latent model to estimate and compare correlated AUCs in a paired-patient, paired-reader design. We assume diagnostic test results come from some unknown and monotone functions of continuous latent variables and further assume reader variability is characterized by random effects due to a specific reader from a given diagnostic test. In Section 4.2, we estimate the AUCs nonparametrically based on the ranks of test results and suggest inference procedure to account for complicated correlated structures of test results. We also provide the asymptotic normality of the AUC estimates in Section 4.2. Since, it is common that disease status may be misclassified, in Section 4.3, we present a method for comparing correlated AUCs when an imperfect gold standard bias is presented. A sample size formula is presented in Section 4.4. In Section 4.5, we conduct simulation studies to investigate the performance of the correlated AUC differences for two diagnostic tests and to evaluate power by varying the number of patients and readers. In Section 4.6, we present an example from a breast cancer study. A discussion is followed in Section 4.7 and all proofs are provided in Section 4.8.

4.2 Inference for Correlated AUCs

Suppose a total of h tests are performed on a sample of N patients (m diseased, n non-diseased, $N = m + n$) where r readers independently examine the test results from N patients. Let X_{ik}^{l} be the test result of diseased subject i obtained from reader k with

diagnostic test l ($i = 1, \cdots, m; k = 1, \cdots, r; l = 1, \cdots, h$). Likewise, let Y_{jk}^l be the test result of non-diseased subject j from reader k with diagnostic test l ($j = 1, \cdots, n; k = 1, \cdots, r; l = 1, \cdots, h$). The test results can be based on either continuous or ordinal random variables. Without loss of generality, we assume that higher values of test results are more indicative of disease. Suppose X_{ik}^l and Y_{jk}^l come from distributions of $f(W_{ik}^l)$ and $f(W_{jk}^l)$ for an unknown increasing function f. W_{ik}^l and W_{jk}^l denote unobserved continuous latent variables of X_{ik}^l and Y_{jk}^l, respectively.

$$\begin{aligned} W_{ik}^l &= b_k^l + \epsilon_{ik}^l \\ W_{jk}^l &= b_k^l + \tilde{\epsilon}_{jk}^l, \end{aligned} \tag{4.1}$$

where b_k^l is a random effect due to reader k of diagnostic test l, $\epsilon_i = (\epsilon_{ik}^l)_{k=1,\cdots,r;\ l=1,\cdots,h}$ is a random error of diseased subject i, and $\tilde{\epsilon}_j = (\epsilon_{jk}^l)_{k=1,\cdots,r;\ l=1,\cdots,h}$ is a random error of non-diseased subject j. It is assumed that ϵ_i^l's (or $\tilde{\epsilon}_j^l$'s) are iid for a fixed l, and b_k^l, ϵ_{ik}^l, and $\tilde{\epsilon}_{jk}^l$ are independent.

Let θ_k^l denote the AUC of diagnostic test l by reader k. When the AUC is calculated by the trapezoidal rule, θ_k^l is equal to $Pr(X_k^l > Y_k^l) + \frac{1}{2} Pr(X_k^l = Y_k^l)$ by Bamber(1975)'s formula. The empirical AUC of θ_k^l applying the Mann-Whitney U statistic is given by

$$\hat{\theta}_k^l = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \phi(X_{ik}^l, Y_{jk}^l),$$

with

$$\phi(X, Y) = \begin{cases} 1 & \text{if } X > Y, \\ 1/2 & \text{if } X = Y, \\ 0 & \text{if } X < Y. \end{cases}$$

$E(\hat{\theta}_k^l) = Pr(X_k^l > Y_k^l) + \frac{1}{2} Pr(X_k^l = Y_k^l) = \theta_k^l$ with $Pr(X_k^l = Y_k^l) = 0$ for continuous

test results. Since X_{ik}^l and Y_{jk}^l are assumed to follow underlying distributions of W_{ik}^l and W_{jk}^l from (4.1), $\phi(X_{ik}^l, Y_{jk}^l)$ is equal to $\phi(W_{ik}^l, W_{jk}^l)$ and $\phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l)$. Let $\hat{\theta}^l$ be the empirical AUC of diagnostic test l. The accuracy of diagnostic test l is described by the average of the reader specific AUCs from diagnostic test l.

$$\begin{aligned}\hat{\theta}^l &= \frac{1}{r}\sum_{k=1}^{r}\hat{\theta}_k^l = \frac{1}{mnr}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{r}\phi(X_{ik}^l, Y_{jk}^l) \\ &= \frac{1}{mnr}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{r}\phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l),\end{aligned}$$

where $E(\hat{\theta}^l) = \frac{1}{r}\sum_{k=1}^{r}\theta_k^l (= \text{let } \theta^l)$.

Let $\hat{\boldsymbol{\theta}} = (\hat{\theta}_k^l)_{k=1,\cdots,r;l=1,\cdots,h}$. $\hat{\theta}_k^l$'s are correlated each other because of having same patients or same readers. We consider three different types of correlations among $\phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l)$ values.

$$\begin{aligned}\rho_{1((k,l),(k',l'))} &= Corr[\phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l),\ \phi(\epsilon_{ik'}^{l'}, \tilde{\epsilon}_{j'k'}^{l'})],\ j \neq j' \\ \rho_{2((k,l),(k',l'))} &= Corr[\phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l),\ \phi(\epsilon_{i'k'}^{l'}, \tilde{\epsilon}_{jk'}^{l'})],\ i \neq i' \\ \rho_{3((k,l),(k',l'))} &= Corr[\phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l),\ \phi(\epsilon_{ik'}^{l'}, \tilde{\epsilon}_{jk'}^{l'})].\end{aligned} \tag{4.2}$$

$\rho_{1((k,l),(k',l'))}$ denotes the correlation due to same diseased subjects with two different non-diseased subjects; $\rho_{2((k,l),(k',l'))}$, the correlation due to same non-diseased subjects with two different diseased subjects; and $\rho_{3((k,l),(k',l'))}$, the correlation due to same diseased and non-diseased subjects.

The asymptotic normality is derived from the theory of U-statistics by Hoeffding (1948).

Theorem 4.1. *Let* $\hat{\boldsymbol{\theta}} = (\hat{\theta}_k^l)_{k=1,\cdots,r;l=1,\cdots,h}$ *and* $\boldsymbol{\theta} = (\theta_k^l)_{k=1,\cdots,r;l=1,\cdots,h}$. *If* $\lim_{N\to\infty} m/N = \lambda$ *and* $\lim_{N\to\infty} n/N = 1-\lambda$ *with* $0 < \lambda < 1$, *and if* $E[\phi^2(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l)] < \infty$, *then* $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ *is asymptotically normal with zero mean vector and covariance matrix* $\boldsymbol{\Sigma} = (\sigma_{((k,l),(k',l'))})$

where

$$\sigma_{((k,l),(k',l'))} = \left[\frac{1}{\lambda}\xi_{10}^{((k,l),(k',l'))} + \frac{1}{(1-\lambda)}\xi_{01}^{((k,l),(k',l'))}\right]$$

and

$$\xi_{10}^{((k,l),(k',l'))} = Cov[\phi(\epsilon_{ik}^{l}, \tilde{\epsilon}_{jk}^{l}), \phi(\epsilon_{ik'}^{l'}, \tilde{\epsilon}_{j'k'}^{l'})|D_i = 1, D_j = 0, D_{j'} = 0],\ j \neq j'$$

$$\xi_{01}^{((k,l),(k',l'))} = Cov[\phi(\epsilon_{ik}^{l}, \tilde{\epsilon}_{jk}^{l}), \phi(\epsilon_{i'k'}^{l'}, \tilde{\epsilon}_{jk'}^{l'})|D_i = 1, D_j = 0, D_{i'} = 1],\ i \neq i'.$$

To assess statistical significance, we use the test statistic $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \rightarrow N(\mathbf{0}, \hat{\boldsymbol{\Sigma}})$, with $\hat{\boldsymbol{\Sigma}} = (\hat{\sigma}_{((k,l),(k',l'))}) = \left[\frac{1}{\lambda}\hat{\xi}_{10}^{((k,l),(k',l'))} + \frac{1}{(1-\lambda)}\hat{\xi}_{01}^{((k,l),(k',l'))}\right]$. Σ can be estimated by the method of structural components developed by Sen (1960).

Corollary 4.2. *Under the above assumptions and if the number of readers r is bounded, for any two correlated empirical ROC areas $\hat{\theta}^l$ and $\hat{\theta}^{l'}$, $\sqrt{N}((\hat{\theta}^l - \hat{\theta}^{l'}) - (\theta^l - \theta^{l'}))$ is asymptotically normal with mean zero and variance $\sigma_{l,l'}$, where*

$$\begin{aligned}\sigma_{l,l'} = {} & \frac{1}{r^2}\sum_{t=l,l'}\left[\sum_{k=1}^{r}\left(\frac{1}{\lambda}\xi_{10}^{((k,t),(k,t))} + \frac{1}{(1-\lambda)}\xi_{01}^{((k,t),(k,t))}\right) + \sum_{k\neq k'}\left(\frac{1}{\lambda}\xi_{10}^{((k,t),(k',t))}\right.\right. \\ & \left.\left. + \frac{1}{(1-\lambda)}\xi_{01}^{((k,t),(k',t))}\right)\right] - \frac{2}{r^2}\left[\sum_{k=1}^{r}\left(\frac{1}{\lambda}\xi_{10}^{((k,l),(k,l'))} + \frac{1}{(1-\lambda)}\xi_{01}^{((k,l),(k,l'))}\right)\right. \\ & \left. + \sum_{k\neq k'}\left(\frac{1}{\lambda}\xi_{10}^{((k,l),(k',l'))} + \frac{1}{(1-\lambda)}\xi_{01}^{((k,l),(k',l'))}\right)\right].\end{aligned}$$

4.3 Inference with an Imperfect Gold Standard Bias

The imperfect gold standard bias refers to the misclassification of disease status in which disease status based on a gold standard and true disease status are not consistent. A perfect gold standard is often not available or too expensive to implement for many

diseases, which leads investigators to use an imperfect gold standard. If an imperfect gold standard is used, the estimated accuracy of the tests would be biased. In this section, we present a method for comparing correlated AUCs when an imperfect gold standard bias is presented. Let Z_{ik}^l be the test result of subject i by reader k of diagnostic test l ($i = 1, \cdots, m+n; k = 1, \cdots, r; l = 1, \cdots, h$). D_i denotes disease status of subject i based on a gold standard and D_i^0 denotes true disease status of subject i in which D_i and D_i^0 are assumed to be independent. Both of D_i and D_i^0 have 1 for diseased and 0 for non-diseased. m and n are the number of diseased and non-diseased subjects based on a gold standard. $Z_{ik}^l|D_i = 1$ and $Z_{jk}^l|D_j = 0$ are assumed to follow underlying distributions of W_{ik}^l and W_{jk}^l, respectively, where $W_{ik}^l = b_k^l + \epsilon_{ik}^l$ and $W_{jk}^l = b_k^l + \tilde{\epsilon}_{jk}^l$. The assumptions for b_k^l, ϵ_{ik}^l, and $\tilde{\epsilon}_{jk}^l$ are described in section 4.2. Suppose the probability of having disease based on a gold standard given that a subject is truly non-diseased, $Pr(D_i = 1|D_i^0 = 0)$, equals p, the probability of having non-diseased based on a gold standard given that a subject is truly diseased, $Pr(D_i = 0|D_i^0 = 1)$, equals q, and the prevalence of truly diseased population, $Pr(D_i^0 = 1)$, equals ω. In addition, we assume that test results do not depend on observed disease status given true disease status.

The area under the empirical ROC curve of diagnostic test l by reader k in the presence of misclassification of disease status is

$$\hat{\theta}_k^l = \frac{1}{\sum_{i=1}^{m+n} D_i \sum_{j=1}^{m+n}(1-D_j)} \sum_{i=1}^{m+n}\sum_{j=1}^{m+n} \phi(Z_{ik}^l, Z_{jk}^l) D_i(1-D_j) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} \phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l),$$

where

$$\phi(Z_{ik}^l, Z_{jk}^l) = \begin{cases} 1 & \text{if } Z_{ik}^l > Z_{jk}^l, \\ 1/2 & \text{if } Z_{ik}^l = Z_{jk}^l, \\ 0 & \text{if } Z_{ik}^l < Z_{jk}^l. \end{cases}$$

Similar to (4.2), three different types of correlations are considered.

$$\begin{aligned}
\rho^*_{1((k,l),(k',l'))} &= Corr[\phi(\epsilon^l_{ik}, \tilde{\epsilon}^l_{jk}),\ \phi(\epsilon^{l'}_{ik'}, \tilde{\epsilon}^{l'}_{j'k'})|D_i = 1, D_j = 0, D_{j'} = 0],\ j \neq j' \\
\rho^*_{2((k,l),(k',l'))} &= Corr[\phi(\epsilon^l_{ik}, \tilde{\epsilon}^l_{jk}),\ \phi(\epsilon^{l'}_{i'k'}, \tilde{\epsilon}^{l'}_{jk'})|D_i = 1, D_j = 0, D_{i'} = 1],\ i \neq i' \\
\rho^*_{3((k,l),(k',l'))} &= Corr[\phi(\epsilon^l_{ik}, \tilde{\epsilon}^l_{jk}),\ \phi(\epsilon^{l'}_{ik'}, \tilde{\epsilon}^{l'}_{jk'})|D_i = 1, D_j = 0].
\end{aligned} \tag{4.3}$$

Theorem 4.3. *The expectation and variance of an empirical AUC of diagnostic test l by reader k.*

i. $E(\hat{\theta}^l_k) = a\theta^l_k + b$ *with*

$$a = \frac{(1-p)(1-q)\omega(1-\omega) - pq\omega(1-\omega)}{(1-p)(1-q)\omega(1-\omega) + p(1-p)(1-\omega)^2 + pq\omega(1-\omega) + q(1-q)\omega^2}$$

$$b = \frac{\frac{1}{2}p(1-p)(1-\omega)^2 + pq\omega(1-\omega) + \frac{1}{2}q(1-q)\omega^2}{(1-p)(1-q)\omega(1-\omega) + p(1-p)(1-\omega)^2 + pq\omega(1-\omega) + q(1-q)\omega^2}.$$

ii. $Var(\hat{\theta}^l_k) = \frac{V^{l*}_k}{mn}[1 + (n-1)\rho^*_{1((k,l),(k,l))} + (m-1)\rho^*_{2((k,l),(k,l))}]$, *where*

$$\begin{aligned}
V^{l*}_k &= Var(\phi(\epsilon^l_{ik}, \tilde{\epsilon}^l_{jk})|D_i = 1, D_j = 0) \\
&= (a\theta^l_k + b) - \frac{1}{4}Pr(\epsilon^l_{ik} = \tilde{\epsilon}^l_{jk}|D_i = 1, D_j = 0) - (a\theta^l_k + b)^2.
\end{aligned}$$

It is noted that the empirical AUC $\hat{\theta}^l_k$ is biased if imperfect gold standard bias is present. Define $\hat{\theta}^{l*}_k = (\hat{\theta}^l_k - b)/a$ is the bias corrected AUC estimate of $\hat{\theta}^l_k$. Then, the bias corrected AUC estimate of diagnostic test l is defined by the average of $\hat{\theta}^{l*}_k$'s $(k = 1, \cdots, r)$.

$$\hat{\theta}^{l*} = \frac{1}{r}\sum_{k=1}^{r}\hat{\theta}^{l*}_k = \frac{1}{r}\sum_{k=1}^{r}\frac{\hat{\theta}^l_k - b}{a},$$

and

$$
\begin{aligned}
Var(\hat{\theta}^{l*}) &= Var\left[\frac{1}{r}\sum_{k=1}^{r}\left(\frac{\hat{\theta}_k^l - b}{a}\right)\right] = \frac{1}{a^2r^2}\left[\sum_{k=1}^{r} Var(\hat{\theta}_k^l) + \sum_{k\neq k'} Cov(\hat{\theta}_k^l, \hat{\theta}_{k'}^l)\right] \\
&= \frac{1}{mna^2r^2}\left[\sum_{k=1}^{r} V_k^{l*}(1 + (n-1)\rho^*_{1((k,l),(k,l))} + (m-1)\rho^*_{2((k,l),(k,l))}) \right. \\
&\quad \left. + \sum_{k\neq k'} \sqrt{V_k^{l*}}\sqrt{V_{k'}^{l*}}((n-1)\rho^*_{1((k,l),(k',l))} + (m-1)\rho^*_{2((k,l),(k',l))} + \rho^*_{3((k,l),(k',l))})\right].
\end{aligned}
$$

Theorem 4.4. *Let* $\hat{\boldsymbol{\theta}}^* = (\hat{\theta}_k^{l*})_{k=1,\cdots,r;l=1,\cdots,h}$ *and* $\boldsymbol{\theta} = (\theta_k^l)_{k=1,\cdots,r;l=1,\cdots,h}$. *If* $\lim_{N\to\infty} m/N = \lambda$ *and* $\lim_{N\to\infty} n/N = 1-\lambda$ *with* $0 < \lambda < 1$, *and if* $E[\phi^2(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l)] < \infty$, *then* $\sqrt{N}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta})$ *is asymptotically normal with zero mean vector and covariance matrix* $\boldsymbol{\Sigma}^* = (\sigma^*_{((k,l),(k',l'))})$ *with*

$$
\sigma^*_{((k,l),(k',l'))} = \frac{1}{a^2}\left[\frac{1}{\lambda}\xi_{10}^{((k,l),(k',l'))} + \frac{1}{(1-\lambda)}\xi_{01}^{((k,l),(k',l'))}\right].
$$

Corollary 4.5. *If the conditions of Theorem 4.4 hold and the number of readers* r *is bounded, for any two correlated empirical ROC areas* $\hat{\theta}^{l*}$ *and* $\hat{\theta}^{l'*}$, $\sqrt{N}((\hat{\theta}^{l*} - \hat{\theta}^{l'*}) - (\theta^l - \theta^{l'}))$ *is asymptotically normal with mean zero and variance* $\sigma^*_{l,l'}$ *with*

$$
\begin{aligned}
\sigma^*_{l,l'} &= \frac{1}{a^2r^2}\sum_{t=l,l'}\left[\sum_{k=1}^{r}\left(\frac{1}{\lambda}\xi_{10}^{((k,t),(k,t))} + \frac{1}{(1-\lambda)}\xi_{01}^{((k,t),(k,t))}\right) + \sum_{k\neq k'}\left(\frac{1}{\lambda}\xi_{10}^{((k,t),(k',t))}\right.\right. \\
&\quad \left.\left. + \frac{1}{(1-\lambda)}\xi_{01}^{((k,t),(k',t))}\right)\right] - \frac{2}{a^2r^2}\left[\sum_{k=1}^{r}\left(\frac{1}{\lambda}\xi_{10}^{((k,l),(k,l'))} + \frac{1}{(1-\lambda)}\xi_{01}^{((k,l),(k,l'))}\right)\right. \\
&\quad \left. + \sum_{k\neq k'}^{r}\left(\frac{1}{\lambda}\xi_{10}^{((k,l),(k',l'))} + \frac{1}{(1-\lambda)}\xi_{01}^{((k,l),(k',l'))}\right)\right].
\end{aligned}
$$

4.4 Sample Size Calculation

There exists only one method for determining sample sizes for multi-reader studies in the literature (Obuchowski, 1995a, 1995b). The method is based on a mixed-effects ANOVA model for the summary measure of the ROC curve. Let $\hat{\theta}_{ijq}$ be the estimated summary measure of the ROC curve for ith test by jth reader at the qth occasion. The mixed effect linear model is defined by $\hat{\theta}_{ijq} = \mu + \mu_i + r_j + (\mu r)_{ij} + \epsilon_{ijq},\ 1 \leq i \leq I, 1 \leq j \leq J,$ and $1 \leq q \leq Q$. Here μ is overall mean, μ_i is a fixed effect corresponding to the ith test, r_j is a random effect due to the jth reader, $(\mu r)_{ij}$ is a random effect due to the interaction between the ith test and jth reader, and ϵ_{ijq} is a random error. Each of r_j, $(\mu r)_{ij}$, and ϵ_{ijq} is assumed to follow a normal distribution. For sample size calculations, the diagnostic accuracies of the J readers of the two tests are assumed to follow a multivariate normal distribution and an approximated F statistic with 1 and $(J-1)$ degrees of freedom is used for testing the null hypothesis that the mean diagnostic accuracies of the tests are equal (Zhou *et al.*, 2002). However, sample sizes under this approach are sensitive to the assumptions on variance components of the mixed model. In addition, the response of the mixed model is fitted nonparametrically first and then this nonparametric estimate is fitted again based on the model assumptions. This two-stage sample size determination can be misleading because the nonparametric estimates of the response are not considered in the second step of fitting a mixed model.

Instead, we propose to make an inference under the asymptotic normality of the empirical area under the ROC curves. Based on the test statistic $((\hat{\theta}^{l*} - \hat{\theta}^{l'*}) - (\theta^l - \theta^{l'}))/\sqrt{Var(\hat{\theta}^{l*} - \hat{\theta}^{l'*})}$, which possesses approximately a standard normal distribution, we determine the asymptotic power for a hypothesis test of the difference in accuracy:

$$\mathrm{H}_0 :\ \delta_0 = \theta^l - \theta^{l'} = 0 \quad \mathrm{H}_1 :\ \delta_1 = \theta^l - \theta^{l'} \neq 0$$

Corollary 4.6. *The variance of difference between two correlated, bias corrected empirical AUCs*

$$
\begin{aligned}
& Var(\hat{\theta}^{l*} - \hat{\theta}^{l'*}) \\
= \; & \frac{1}{mna^2r^2} \sum_{t=l,l'} \left[\sum_{k=1}^{r} V_k^{t*}((1+(n-1)\rho^*_{1((k,t),(k,t))} + (m-1)\rho^*_{2((k,t),(k,t))}) \right. \\
& \left. + \sum_{k \neq k'} \sqrt{V_k^{t*}} \sqrt{V_{k'}^{t*}} ((n-1)\rho^*_{1((k,t),(k',t))} + (m-1)\rho^*_{2((k,t),(k',t))} + \rho^*_{3((k,t),(k',t))}) \right] \\
& - \frac{2}{mna^2r^2} \left[\sum_{k=1}^{r} \sqrt{V_k^{l*}} \sqrt{V_k^{l'*}} ((n-1)\rho^*_{1((k,l),(k,l'))} + (m-1)\rho^*_{2((k,l),(k,l'))} + \rho^*_{3((k,l),(k,l'))}) \right. \\
& \left. + \sum_{k \neq k'}^{r} \sqrt{V_k^{l*}} \sqrt{V_{k'}^{l'*}} ((n-1)\rho^*_{1((k,l),(k',l'))} + (m-1)\rho^*_{2((k,l),(k',l'))} + \rho^*_{3((k,l),(k',l'))}) \right],
\end{aligned}
$$

where $V_k^{l*} = Var(\phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l | D_i = 1, D_j = 0)) = (a\theta_k^l + b) - \frac{1}{4} Pr(\epsilon_{ik}^l = \tilde{\epsilon}_{jk}^l | D_i = 1, D_j = 0) - (a\theta_k^l + b)^2$ *from Theorem 4.3(ii).*

For sample size determination, the followings are assumed:

a. Some of correlations defined in (4.3) are simplified as

$$
\begin{aligned}
\rho^*_{1((k,l),(k,l'))} &= \rho^*_{2((k,l),(k,l'))} = \rho_a, \\
\rho^*_{1((k,l),(k',l))} &= \rho^*_{2((k,l),(k',l))} = \rho_b, \\
\rho^*_{1((k,l),(k',l'))} &= \rho^*_{2((k,l),(k',l'))} = \rho_c.
\end{aligned} \tag{4.4}
$$

b. $V_k^{l*} = Var(\phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l | D_i = 1, D_j = 0)) = (a\theta_k^l + b) - \frac{1}{4} Pr(\epsilon_{ik}^l = \tilde{\epsilon}_{jk}^l | D_i = 1, D_j = 0) - (a\theta_k^l + b)^2$ from Corollary 4.6 is simplified as $V^* = (a\bar{\theta} + b) - (a\bar{\theta} + b)^2$ assuming that the variance of $\phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l | D_i = 1, D_j = 0)$ is same across readers and modalities. Here $\bar{\theta}$ denotes average of two comparing AUCs and $Pr(\epsilon_{ik}^l = \tilde{\epsilon}_{jk}^l | D_i = 1, D_j = 0)$ is assumed to be zero.

Then, the variance of the difference between two AUCs from corollary 4.6 is simplified as

$$Var(\hat{\theta}^{l*} - \hat{\theta}^{l'*}) \approx \frac{2(m+n)V^*}{mna^2r}[(1-\rho_a) + (r-1)(\rho_b - \rho_c)],\ V^* = (a\bar{\theta}+b) - (a\bar{\theta}+b)^2.$$

Under the above assumptions, the power at level α is

$$\begin{aligned} Power &= \Phi\left(\frac{-z_{1-\alpha/2}\sqrt{2(m+n)V^*((1-\rho_a)+(r-1)(\rho_b-\rho_c))/mna^2r} + \delta_1}{\sqrt{2(m+n)V^*((1-\rho_a)+(r-1)(\rho_b-\rho_c))/mna^2r}}\right) \\ &\quad +\Phi\left(\frac{-z_{1-\alpha/2}\sqrt{2(m+n)V^*((1-\rho_a)+(r-1)(\rho_b-\rho_c))/mna^2r} - \delta_1}{\sqrt{2(m+n)V^*((1-\rho_a)+(r-1)(\rho_b-\rho_c))/mna^2r}}\right) \end{aligned} \tag{4.5}$$

with $V^* = (a\bar{\theta}+b) - (a\bar{\theta}+b)^2$.

4.5 Simulation Studies

4.5.1 Data Generation

Let m and n denote the number of diseased and non-diseased subjects and r denote the number of readers. $X = (X_i)_{i=1,\cdots,m}$ denotes diseased results and $Y = (Y_j)_{j=1,\cdots,n}$ denotes non-diseased results with $X_i = ((X_{ik}^1)_{k=1,\cdots,r}, (X_{ik}^2)_{k=1,\cdots,r})'$ and $Y_j = ((Y_{jk}^1)_{k=1,\cdots,r}, (Y_{jk}^2)_{k=1,\cdots,r})'$. X_{ik}^l is the ith diseased subject's test result by the kth reader of the lth test and Y_{jk}^l is the jth non-diseased subject's test result by the kth reader of the lth test. For the purpose of illustration, X and Y are assumed to be independent. Regarding correlations among X_{ik}^l's (Y_{jk}^l's) for the ith (jth) subject with varying k and l, we assume that the correlation when same subjects are evaluated by different readers using a same test is 0.3, the correlation when same subjects are evaluated by a same reader using different tests 0.8, and the correlation when same subjects

are evaluated by different readers using different tests is 0.25 for both continuous and ordinal data.

Continuous data

X_i's and Y_j's ($i = 1, \cdots, m;\ j = 1, \cdots, n$) are generated from multivariate normal distributions; $X_i \sim N((\mu_{d1}\mathbf{1}, \mu_{d2}\mathbf{1})', \mathbf{\Sigma})$ and $Y_j \sim N(\mathbf{0}, \mathbf{\Sigma})$ where $\mathbf{1} = (1, \cdots, 1)'_{r\times 1}$ and $\mathbf{\Sigma} = \begin{pmatrix} \Sigma_{11(r\times r)} & \Sigma_{12(r\times r)} \\ \Sigma_{12(r\times r)} & \Sigma_{11(r\times r)} \end{pmatrix}$ with $\Sigma_{11}(i,i) = 1$, $\Sigma_{11}(i,j) = 0.3$, $\Sigma_{12}(i,i) = 0.8$, and $\Sigma_{12}(i,j) = 0.25$ $(i \neq j)$. The true AUC for diagnostic test l is given by

$$\theta^l = \frac{1}{r}\sum_{k=1}^{r}\theta_k^l = \frac{1}{r}\sum_{k=1}^{r} Pr(X_k^l > Y_k^l). \tag{4.6}$$

μ_{d1} and μ_{d2} values are set to 1.465738 and 1.190232 since $X_{ik}^1 \sim N(1.465738, 1)$, $Y_{jk}^1 \sim N(0,1)$, $X_{ik}^2 \sim N(1.190232, 1)$ and $Y_{jk}^2 \sim N(0,1)$ give $\theta_k^1 = 0.85$ and $\theta_k^2 = 0.8$ for a fixed k $(k = 1, \cdots, r)$. It follows that $\theta_k^1 = 0.85$ and $\theta_k^1 = 0.8$ by equation (4.6). Similarly, we set $\mu_{d1} = \mu_{d2} = 1.190232$ to make $\theta^1 = \theta^2 = 0.8$.

Ordinal data

We generated ordinal data by making true AUCs as close as 0.85 or 0.8. We found the following setup makes $\theta_1 = 0.8502313$ and $\theta_2 = 0.8005894$ which is used for the true AUCs; First we generated continuous values from

$$X^1 \sim N(2.325216, 1.75) \quad Y^1 \sim N(0, 0.75)$$
$$X^2 \sim N(1.81, 1.75) \quad Y^2 \sim N(0, 0.75)$$

and categorized them as

-2 $(X_{ik}^l \leq -2)$, -1 $(-2 < X_{ik}^l \leq -1)$, 0 $(-1 < X_{ik}^l \leq 0)$, 1 $(0 < X_{ik}^l \leq 1)$, 2 $(X_{ik}^l > 1)$;

-2 ($Y_{jk}^l \leq -2$), -1 ($-2 < Y_{jk}^l \leq -1$), 0 ($-1 < Y_{jk}^l \leq 0$), 1 ($0 < Y_{jk}^l \leq 1$), 2 ($Y_{jk}^l > 1$).

4.5.2 Simulation Results

We set the number of diseased and non-diseased subjects at 50, 100, and 200 ($m = n = 50, 100, 200$) and the number of readers at 2, 5, and 10 ($r = 2, 5, 10$). In each combination of subject and reader sample sizes, we used 5000 stimulations and accounted for an imperfect gold standard bias ($p = q = 0, 0.02$, or 0.05). Correlations defined in (4.4) are assumed to be $\rho_a = 0.7$ and $\rho_b = \rho_c = 0$. Tables 4.1 (continuous) and 4.3 (ordinal) contain the results when there is no difference between the two AUCs. In each scenario of sample sizes, the empirical power (when α=0.05) is very close to 0.05, and the coverage proportions of 90% and 95% confidence intervals based on the asymptotic normal approximation are close to 0.9 or 0.95, respectively. Next, we tested to see if there is a statistically significant difference between the two AUCs when the true difference is 0.05. Results are given in Table 4.2 (continuous) and Table 4.4 (ordinal). In Table 4.2, when there is no imperfect gold standard bias ($p = q = 0$), empirical powers are greater than 80% in all sample sizes except for $m = n = 50$ and $k = 2$. When $p = q = 0.02$ or 0.05, the empirical powers increase as the number of subjects or readers increase. Similar results are shown in Table 4.4.

4.6 Application to Breast Cancer Data

We used mammographic images of 201 women with dense breast (Cole *et al.*, 2005). The goal of this study was to compare the diagnostic accuracy of digital mammography with that of screen-film mammography in distinguishing breast cancer status (benign, malignant).

There were a total of nine readers who participated in the reader study. The readers

Table 4.1: Continuous data, $\theta_1 = \theta_2 = 0.8$

m, n	k	p, q	$CP_1^{(1)}$	$CP_2^{(2)}$	$emp.power^{(3)}$
50	2	0	0.911	0.954	0.046
		0.02	0.907	0.95	0.047
		0.05	0.906	0.951	0.049
	5	0	0.901	0.950	0.050
		0.02	0.903	0.950	0.050
		0.05	0.905	0.955	0.045
	10	0	0.895	0.947	0.053
		0.02	0.902	0.954	0.046
		0.05	0.900	0.953	0.047
100	2	0	0.899	0.953	0.047
		0.02	0.896	0.949	0.051
		0.05	0.892	0.947	0.053
	5	0	0.899	0.952	0.048
		0.02	0.896	0.948	0.052
		0.05	0.899	0.950	0.050
	10	0	0.902	0.954	0.046
		0.02	0.894	0.945	0.055
		0.05	0.897	0.953	0.047
200	2	0	0.891	0.947	0.053
		0.02	0.897	0.949	0.051
		0.05	0.896	0.948	0.052
	5	0	0.900	0.949	0.051
		0.02	0.899	0.951	0.050
		0.05	0.906	0.954	0.046
	10	0	0.900	0.949	0.051
		0.02	0.892	0.947	0.053
		0.05	0.899	0.949	0.051

(1) CP_1: coverage proportion of 90% confidence interval based on normal approximation

(2) CP_2: coverage proportion of 95% confidence interval based on normal approximation

(3) *emp.power*: Empirical power based on the significance level 0.05

Table 4.2: Continuous data, $\theta_1 = 0.85$ and $\theta_2 = 0.8$

m, n	k	p, q	$CP_1^{(1)}$	$CP_2^{(2)}$	$emp.power^{(3)}$
50	2	0	0.894	0.943	0.590
		0.02	0.906	0.957	0.365
		0.05	0.902	0.955	0.203
	5	0	0.896	0.953	0.837
		0.02	0.897	0.946	0.551
		0.05	0.905	0.949	0.298
	10	0	0.898	0.949	0.921
		0.02	0.896	0.947	0.652
		0.05	0.897	0.950	0.385
100	2	0	0.899	0.950	0.892
		0.02	0.905	0.957	0.641
		0.05	0.891	0.944	0.377
	5	0	0.914	0.956	0.991
		0.02	0.892	0.951	0.851
		0.05	0.894	0.948	0.542
	10	0	0.892	0.944	0.997
		0.02	0.890	0.947	0.918
		0.05	0.897	0.949	0.646
200	2	0	0.905	0.953	0.996
		0.02	0.900	0.954	0.920
		0.05	0.887	0.944	0.646
	5	0	0.901	0.952	1.000
		0.02	0.895	0.950	0.989
		0.05	0.896	0.947	0.843
	10	0	0.899	0.954	1.000
		0.02	0.892	0.947	0.997
		0.05	0.903	0.952	0.915

(1) CP_1: coverage proportion of 90% confidence interval based on normal approximation

(2) CP_2: coverage proportion of 95% confidence interval based on normal approximation

(3) *emp.power*: Empirical power based on the significance level 0.05

Table 4.3: Ordinal data, $\theta_1 = \theta_2 = 0.8$

m, n	k	p, q	$CP_1^{(1)}$	$CP_2^{(2)}$	$emp.power^{(3)}$
50	2	0	0.901	0.952	0.048
		0.02	0.907	0.951	0.049
		0.05	0.891	0.948	0.053
	5	0	0.892	0.943	0.057
		0.02	0.899	0.948	0.052
		0.05	0.896	0.947	0.053
	10	0	0.901	0.948	0.052
		0.02	0.902	0.951	0.050
		0.05	0.898	0.953	0.047
100	2	0	0.898	0.946	0.054
		0.02	0.897	0.949	0.051
		0.05	0.894	0.942	0.058
	5	0	0.900	0.953	0.047
		0.02	0.903	0.953	0.047
		0.05	0.889	0.947	0.054
	10	0	0.899	0.952	0.048
		0.02	0.887	0.946	0.054
		0.05	0.903	0.945	0.055
200	2	0	0.894	0.947	0.053
		0.02	0.894	0.947	0.053
		0.05	0.893	0.943	0.057
	5	0	0.892	0.947	0.053
		0.02	0.901	0.953	0.047
		0.05	0.904	0.949	0.051
	10	0	0.905	0.953	0.047
		0.02	0.902	0.951	0.049
		0.05	0.890	0.944	0.056

(1) CP_1: coverage proportion of 90% confidence interval based on normal approximation

(2) CP_2: coverage proportion of 95% confidence interval based on normal approximation

(3) *emp.power*: Empirical power based on the significance level 0.05

Table 4.4: Ordinal data, $\theta_1 = 0.85$ and $\theta_2 = 0.8$

m, n	k	p, q	$CP_1^{(1)}$	$CP_2^{(2)}$	$emp.power^{(3)}$
50	2	0	0.892	0.947	0.857
		0.02	0.896	0.950	0.543
		0.05	0.896	0.946	0.280
	5	0	0.871	0.937	0.988
		0.02	0.877	0.934	0.780
		0.05	0.880	0.942	0.457
	10	0	0.857	0.924	0.999
		0.02	0.869	0.932	0.900
		0.05	0.887	0.939	0.556
100	2	0	0.870	0.934	0.991
		0.02	0.876	0.937	0.835
		0.05	0.884	0.942	0.501
	5	0	0.828	0.900	1.000
		0.02	0.862	0.926	0.975
		0.05	0.882	0.936	0.744
	10	0	0.795	0.882	1.000
		0.02	0.845	0.918	0.995
		0.05	0.872	0.934	0.854
200	2	0	0.808	0.892	1.000
		0.02	0.851	0.917	0.983
		0.05	0.872	0.932	0.784
	5	0	0.732	0.836	1.000
		0.02	0.815	0.894	1.000
		0.05	0.863	0.923	0.955
	10	0	0.683	0.790	1.000
		0.02	0.797	0.883	1.000
		0.05	0.851	0.924	0.985

(1) CP_1: coverage proportion of 90% confidence interval based on normal approximation

(2) CP_2: coverage proportion of 95% confidence interval based on normal approximation

(3) *emp.power*: Empirical power based on the significance level 0.05

reported a probability of malignancy based on 5-point scale, with the scale defined as 1-No Findings or definitely not, 2-Probably not, 3-Probably, 4- Possibly and 5-Definitely. Due to a large number of missing test results across readers, we created three imputed readers for each patient. The scores of the three readers were created by taking average scores of the actual readers 1-3, 4-6, and 7-9, respectively, where the decimal points of mean scores were rounded up. A total of 137 patients (benign 81, malignant 56) out of 210 were used for the analysis after excluding patients with missing mean scores (64 patients) and with invalid cancer status (9 patients).

The reader averaged AUCs were 0.770 (SE 0.028) for the screen-film mammography and 0.683 (SE 0.033) for the digital mammography. The estimated AUC difference between the two modalities was 0.087 (SE 0.03) and the estimated AUC for the screen-film was significantly larger than that of the digital (p-value 0.004). Therefore, we concluded that screen-film mammography perform better than digital mammography in discriminating benign and malignant breast tumors.

4.7 Discussion

The paired-patient, paired-reader design is most widely used in radiological studies to compare different diagnostic techniques because it requires the smallest number of subjects. Our method can be applied to compare correlated ROC curves of this design and to a situation in which there is an imperfect gold standard bias. We propose to make an inference under the asymptotic normality of the empirical AUCs and determine the asymptotic power for a hypothesis test of the AUC differences. In seeking a formula for sample size, we found that the theoretical powers are very conservative compared to empirical powers, especially when the imperfect gold standard bias is present. A more accurate sample size formula is needed.

4.8 Appendix

Proof of Theorem 4.1 and Corollary 4.2. Since there exists no imperfect gold standard bias, Theorem 4.1 and Corollary 4.2 hold when $p = Pr(D = 1|D^0 = 0) = 0$ and $q = Pr(D = 0|D^0 = 1) = 0$ from Theorem 4.4 and Corollary 4.5, respectively.

Proof of Theorem 4.3. The expectation and variance of the empirical AUC of diagnostic test l by reader k are given as follows.

$$\begin{aligned} E(\hat{\theta}_k^l) &= Pr(Z_{ik}^l > Z_{jk}^l|D_i = 1, D_j = 0) + \frac{1}{2}Pr(Z_{ik}^l = Z_{jk}^l|D_i = 1, D_j = 0) \\ &= \frac{[Pr(Z_{ik}^l > Z_{jk}^l, D_i = 1, D_j = 0) + \frac{1}{2}Pr(Z_{ik}^l = Z_{jk}^l, D_i = 1, D_j = 0)]}{Pr(D_i = 1, D_j = 0)}. \end{aligned}$$

We define $Pr(D_i = 1|D_i^0 = 0) = p$, $Pr(D_i = 0|D_i^0 = 1) = q$, and $Pr(D_i^0 = 1) = \omega$. Furthermore, we assume that test results do not depend on observed disease status given true disease status. Then, the numerator is computed as

$$\begin{aligned} & Pr(Z_{ik}^l > Z_{jk}^l, D_i = 1, D_j = 0) + \frac{1}{2}Pr(Z_{ik}^l = Z_{jk}^l, D_i = 1, D_j = 0) \\ = & \sum_{r=0}^{1}\sum_{s=0}^{1}\Big\{Pr(Z_{ik}^l > Z_{jk}^l|D_i = 1, D_j = 0, D_i^0 = r, D_j^0 = s)Pr(D_i = 1, D_j = 0|D_i^0 \\ & = r, D_j^0 = s)Pr(D_i^0 = r, D_j^0 = s) + \frac{1}{2}Pr(Z_{ik}^l = Z_{jk}^l|D_i = 1, D_j = 0, D_i^0 = r, \\ & D_j^0 = s)Pr(D_i = 1, D_j = 0|D_i^0 = r, D_j^0 = s)Pr(D_i^0 = r, D_j^0 = s)\Big\} \\ = & \sum_{r=0}^{1}\sum_{s=0}^{1}\Big\{Pr(Z_{ik}^l > Z_{jk}^l|D_i^0 = r, D_j^0 = s)Pr(D_i = 1, D_j = 0|D_i^0 = r, D_j^0 = s) \\ & Pr(D_i^0 = r, D_j^0 = s) + \frac{1}{2}Pr(Z_{ik}^l = Z_{jk}^l|D_i^0 = r, D_j^0 = s)Pr(D_i = 1, D_j = 0| \\ & D_i^0 = r, D_j^0 = s)Pr(D_i^0 = r, D_j^0 = s)\Big\} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{r=0}^{1}\sum_{s=0}^{1}\Big\{(Pr(Z_{ik}^l > Z_{jk}^l|D_i^0=r,D_j^0=s)+\frac{1}{2}Pr(Z_{ik}^l=Z_{jk}^l|D_i^0=r,D_j^0=s)\Big\}\\
&\qquad Pr(D_i=1,D_j=0|D_i^0=r,D_j^0=s)Pr(D_i^0=r,D_j^0=s)\\
&= \frac{1}{2}p(1-p)(1-\omega)^2+(1-\theta_k^l)pq\omega(1-\omega)+\theta_k^l(1-p)(1-q)\omega(1-\omega)+\frac{1}{2}q(1-q)\omega^2\\
&= \Big\{\frac{1}{2}p(1-p)(1-\omega)^2+pq\omega(1-\omega)+\frac{1}{2}q(1-q)\omega^2\Big\}\\
&\quad +\Big\{(1-p)(1-q)\omega(1-\omega)-pq\omega(1-\omega)\Big\}\theta_k^l
\end{aligned}
$$

where

$$
\begin{aligned}
&Pr(Z_{ik}^l > Z_{jk}^l|D_i^0=0,D_j^0=0)+\frac{1}{2}Pr(Z_{ik}^l=Z_{jk}^l|D_i^0=0,D_j^0=0)=1/2\\
&Pr(Z_{ik}^l > Z_{jk}^l|D_i^0=1,D_j^0=1)+\frac{1}{2}Pr(Z_{ik}^l=Z_{jk}^l|D_i^0=1,D_j^0=1)=1/2\\
&Pr(Z_{ik}^l > Z_{jk}^l|D_i^0=1,D_j^0=0)+\frac{1}{2}Pr(Z_{ik}^l=Z_{jk}^l|D_i^0=1,D_j^0=0)=\theta_k^l\\
&Pr(Z_{ik}^l > Z_{jk}^l|D_i^0=0,D_j^0=1)+\frac{1}{2}Pr(Z_{ik}^l=Z_{jk}^l|D_i^0=0,D_j^0=1)=1-\theta_k^l
\end{aligned}
$$

and the denominator $Pr(D_i=1,D_j=0)$ is given by

$$
\begin{aligned}
&Pr(D_i=1,D_j=0)\\
&= \sum_{r=0}^{1}\sum_{s=0}^{1}Pr(D_i=1,D_j=0|D_i^0=r,D_j^0=s)Pr(D_i^0=r,D_j^0=s)\\
&= (1-p)(1-q)\omega(1-\omega)+p(1-p)(1-\omega)^2+pq\omega(1-\omega)+q(1-q)\omega^2.
\end{aligned}
$$

Thus

$$
\begin{aligned}
E(\hat{\theta}_k^l) &= \frac{\big[\frac{1}{2}p(1-p)(1-\omega)^2+pq\omega(1-\omega)+\frac{1}{2}q(1-q)\omega^2\big]}{(1-p)(1-q)\omega(1-\omega)+p(1-p)(1-\omega)^2+pq\omega(1-\omega)+q(1-q)\omega^2}\\
&\quad +\frac{\big[(1-p)(1-q)\omega(1-\omega)-pq\omega(1-\omega)\big]\theta_k^l}{(1-p)(1-q)\omega(1-\omega)+p(1-p)(1-\omega)^2+pq\omega(1-\omega)+q(1-q)\omega^2}\\
&= a\theta_k^l+b. \qquad (4.7)
\end{aligned}
$$

where

$$a = \frac{(1-p)(1-q)\omega(1-\omega) - pq\omega(1-\omega)}{(1-p)(1-q)\omega(1-\omega) + p(1-p)(1-\omega)^2 + pq\omega(1-\omega) + q(1-q)\omega^2}, \quad (4.8)$$

$$b = \frac{\frac{1}{2}p(1-p)(1-\omega)^2 + pq\omega(1-\omega) + \frac{1}{2}q(1-q)\omega^2}{(1-p)(1-q)\omega(1-\omega) + p(1-p)(1-\omega)^2 + pq\omega(1-\omega) + q(1-q)\omega^2}. \quad (4.9)$$

Let V_k^{l*} denote $Var\big(\phi(Z_{ik}^l, Z_{jk}^l)|D_i = 1, D_j = 0\big)$ for test l by reader k. Then V_k^{l*} is computed as follows.

$$\begin{aligned}
V_k^{l*} &= Var(\phi(Z_{ik}^l, Z_{jk}^l)|D_i = 1, D_j = 0) = Var(\phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l)|D_i = 1, D_j = 0) \\
&= E[\phi^2(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l)|D_i = 1, D_j = 0] - E^2[\phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l)|D_i = 1, D_j = 0] \\
&= Pr(\epsilon_{ik}^l > \tilde{\epsilon}_{jk}^l|D_i = 1, D_j = 0) + \frac{1}{4}Pr(\epsilon_{ik}^l = \tilde{\epsilon}_{jk}^l|D_i = 1, D_j = 0) \\
&\quad -E^2(\hat{\theta}_k^l|D_i = 1, D_j = 0) \\
&= E(\hat{\theta}_k^l|D_i = 1, D_j = 0) - \frac{1}{4}Pr(\epsilon_{ik}^l = \tilde{\epsilon}_{jk}^l|D_i = 1, D_j = 0) - E^2(\hat{\theta}_k^l|D_i = 1, D_j = 0) \\
&= (a\theta_k^l + b) - \frac{1}{4}Pr(\epsilon_{ik}^l = \tilde{\epsilon}_{jk}^l|D_i = 1, D_j = 0) - (a\theta_k^l + b)^2 \quad (4.10)
\end{aligned}$$

using $E(\hat{\theta}_k^l) = Pr(\epsilon_{ik}^l > \tilde{\epsilon}_{jk}^l|D_i = 1, D_j = 0) + \frac{1}{2}Pr(\epsilon_{ik}^l = \tilde{\epsilon}_{jk}^l|D_i = 1, D_j = 0)$ and $E(\hat{\theta}_k^l) = (a\theta_k^l + b)$ from (4.7).

Finally, by using three types correlations in (4.3), the variance of the empirical ROC curve for test l by reader k is expressed as below.

$$\begin{aligned}
Var(\hat{\theta}_k^l) &= Var\left[\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\phi(Z_{ik}^l, Z_{jk}^l)|D_i = 1, D_j = 0\right] \\
&= Var\left[\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l)|D_i = 1, D_j = 0\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{m^2n^2}\left[\sum_{i=1}^{m}\sum_{j=1}^{n} Var(\phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l)|D_i=1, D_j=0)\right. \\
&\quad + \sum_{i}^{m}\sum_{j\neq j'}^{n} Cov(\phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l), \phi(\epsilon_{ik}^l, \tilde{\epsilon}_{j'k}^l)|D_i=1, D_j=0, D_{j'}=0) \\
&\quad \left. + \sum_{i\neq i'}^{m}\sum_{j}^{n} Cov(\phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l), \phi(\epsilon_{i'k}^l, \tilde{\epsilon}_{jk}^l)|D_i=1, D_j=0, D_{i'}=1)\right] \\
&= \frac{1}{m^2n^2}\left[mnV_k^{l*} + mn(n-1)\rho^*_{1((k,l),(k,l))}V_k^{l*} + mn(m-1)\rho^*_{2((k,l),(k,l))}V_k^{l*}\right] \\
&= \frac{V_k^{l*}}{mn}\left[1 + (n-1)\rho^*_{1((k,l),(k,l))} + (m-1)\rho^*_{2((k,l),(k,l))}\right] \qquad (4.11)
\end{aligned}
$$

Proof of Theorem 4.4. Let $\hat{\boldsymbol{\theta}} = (\hat{\theta}_k^l)_{k=1,\cdots,r;l=1,\cdots,h}$ and $\boldsymbol{\theta} = (\theta_k^l)_{k=1,\cdots,r;l=1,\cdots,h}$. If we denote $U_i = (\epsilon_{ik}^l)_{k=1,\cdots,r;l=1,\cdots,h}$ and $V_j = (\tilde{\epsilon}_{jk}^l)_{k=1,\cdots,r;l=1,\cdots,h}$, $\hat{\boldsymbol{\theta}}$ can be expressed as $\hat{\boldsymbol{\theta}} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} h(U_i, V_j)$. If $\lim_{N\to\infty} m/N = \lambda$ and $\lim_{N\to\infty} n/N = 1-\lambda$ with $0 < \lambda < 1$, and if $E[\phi^2(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l)] < \infty$, by the central limit theorem for U statistics, $\sqrt{N}(\hat{\boldsymbol{\theta}} - (a\boldsymbol{\theta} + b\mathbf{1}))$ converges in distribution to a multivariate normal with zero mean vector and covariance matrix $\boldsymbol{\Sigma} = (\sigma_{((k,l),(k',l'))})$ where

$$
\sigma_{((k,l),(k',l'))} = \left[\frac{1}{\lambda}\xi_{10}^{((k,l),(k',l'))} + \frac{1}{(1-\lambda)}\xi_{01}^{((k,l),(k',l'))}\right]
$$

and

$$
\xi_{10}^{((k,l),(k',l'))} = Cov[\phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l), \phi(\epsilon_{ik'}^{l'}, \tilde{\epsilon}_{j'k'}^{l'})|D_i=1, D_j=0, D_{j'}=0],\ j \neq j'
$$

$$
\xi_{01}^{((k,l),(k',l'))} = Cov[\phi(\epsilon_{ik}^l, \tilde{\epsilon}_{jk}^l), \phi(\epsilon_{i'k'}^{l'}, \tilde{\epsilon}_{jk'}^{l'})|D_i=1, D_j=0, D_{i'}=1],\ i \neq i'.
$$

Let $\hat{\boldsymbol{\theta}}^* = (\hat{\theta}_k^{l*})_{k=1,\cdots,r;l=1,\cdots,h}$ with $\hat{\theta}_k^{l*} = \frac{1}{a}(\hat{\theta}_k^l - b)$. $\hat{\boldsymbol{\theta}}^* = \frac{1}{a}(\hat{\boldsymbol{\theta}} - b\mathbf{1})$, $\mathbf{1} = (1, 1, \cdots, 1)_{rh\times 1}^T$ denotes the bias corrected AUC estimates of $\hat{\boldsymbol{\theta}}$. Then, it follows $\sqrt{N}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta})$ is asymptotically normally distributed with zero mean vector and covariance matrix $\boldsymbol{\Sigma}^*$ where $\boldsymbol{\Sigma}^* = \frac{1}{a^2}\boldsymbol{\Sigma}$.

Proof of Corollary 4.5. $\hat{\boldsymbol{\theta}}^{*} = (\hat{\theta}_k^{l*})_{k=1,\cdots,r;l=1,\cdots,h}$. The bias corrected empirical AUC for diagnostic test l is $\hat{\theta}^{l*} = \frac{1}{r}\sum_{k=1}^{r}\hat{\theta}_k^{l*}$ under the condition that the number of readers r is bounded. Without loss of generality, we assume that $\hat{\theta}^{l*}$ and $\hat{\theta}^{l'*}$ correspond to averages of first and second r elements of $\hat{\boldsymbol{\theta}}^{*}$, respectively. Let g be a linear function of $\hat{\boldsymbol{\theta}}^{*}$ that has bounded second derivatives in a neighborhood of $\boldsymbol{\theta}$. If the same conditions of Theorem 4.4 hold, $\sqrt{N}(g(\hat{\boldsymbol{\theta}}^{*}) - g(\boldsymbol{\theta}))$ is asymptotically normally distributed with mean zero and variance $\sigma^{*}_{l,l'}$. In this case, $g(\hat{\boldsymbol{\theta}}^{*}) = \boldsymbol{L}'\hat{\boldsymbol{\theta}}^{*} = \hat{\theta}^{l*} - \hat{\theta}^{l'*}$ and $\boldsymbol{L} = (\frac{1}{r}\mathbf{1}_{r\times 1}, -\frac{1}{r}\mathbf{1}_{r\times 1}, \mathbf{0}_{(h-2)r\times 1})^{T}$. Note that $E[g(\hat{\boldsymbol{\theta}}^{*})] = g(\boldsymbol{\theta}) = \boldsymbol{L}'\boldsymbol{\theta} = \frac{1}{r}\sum_{k=1}^{r}\theta_k^{l} - \frac{1}{r}\sum_{k=1}^{r}\theta_k^{l'} =$ let $(\theta^{l} - \theta^{l'})$ and

$$
\begin{aligned}
\sigma^{*}_{l,l'} &= \frac{1}{a^2r^2}\sum_{t=l,l'}\left[\sum_{k=1}^{r}\left(\frac{1}{\lambda}\xi_{10}^{((k,t),(k,t))} + \frac{1}{(1-\lambda)}\xi_{01}^{((k,t),(k,t))}\right) + \sum_{k\neq k'}\left(\frac{1}{\lambda}\xi_{10}^{((k,t),(k',t))}\right.\right. \\
&\quad \left.\left. + \frac{1}{(1-\lambda)}\xi_{01}^{((k,t),(k',t))}\right)\right] - \frac{2}{a^2r^2}\left[\sum_{k=1}^{r}\left(\frac{1}{\lambda}\xi_{10}^{((k,l),(k,l'))} + \frac{1}{(1-\lambda)}\xi_{01}^{((k,l),(k,l'))}\right)\right. \\
&\quad \left. + \sum_{k\neq k'}^{r}\left(\frac{1}{\lambda}\xi_{10}^{((k,l),(k',l'))} + \frac{1}{(1-\lambda)}\xi_{01}^{((k,l),(k',l'))}\right)\right].
\end{aligned}
$$

Proof of Corollary 4.6. First $Var(\hat{\theta}_k^{l}) = \frac{V_k^{l*}}{mn}\left[1+(n-1)\rho^{*}_{1((k,l),(k,l))}+(m-1)\rho^{*}_{2((k,l),(k,l))}\right]$ from (4.11) and similarly, $Cov(\hat{\theta}_k^{l*}, \hat{\theta}_{k'}^{l'*}) = \frac{V_k^{l*}}{mn}\left[(n-1)\rho^{*}_{1((k,l),(k',l'))}+(m-1)\rho^{*}_{2((k,l),(k',l'))}+\rho^{*}_{3((k,l),(k',l'))}\right]$ $(k \neq k'$ or $l \neq l')$.

Using $Var(\hat{\theta}^{l*} - \hat{\theta}^{l'*}) = Var(\hat{\theta}^{l*}) + Var(\hat{\theta}^{l'*}) - 2Cov(\hat{\theta}^{l*}, \hat{\theta}^{l'*})$ and the above two equations,

$$
\begin{aligned}
Var(\hat{\theta}^{l*}) &= Var\left[\frac{1}{r}\sum_{k=1}^{r}\left(\frac{\hat{\theta}_k^{l} - b}{a}\right)\right] = \frac{1}{a^2r^2}\left[\sum_{k=1}^{r}Var(\hat{\theta}_k^{l}) + \sum_{k\neq k'}Cov(\hat{\theta}_k^{l}, \hat{\theta}_{k'}^{l})\right] \\
&= \frac{1}{mna^2r^2}\left[\sum_{k=1}^{r}V_k^{l*}(1 + (n-1)\rho^{*}_{1((k,l),(k,l))} + (m-1)\rho^{*}_{2((k,l),(k,l))})\right. \\
&\quad \left. + \sum_{k\neq k'}\sqrt{V_k^{l*}}\sqrt{V_{k'}^{l*}}((n-1)\rho^{*}_{1((k,l),(k',l))} + (m-1)\rho^{*}_{2((k,l),(k',l))} + \rho^{*}_{3((k,l),(k',l))})\right].
\end{aligned}
$$

Similarly, for diagnostic test l',

$$\begin{aligned} Var(\hat{\theta}^{l'*}) &= \frac{1}{mna^2r^2}\left[\sum_{k=1}^{r} V_k^{l'*}(1+(n-1)\rho^*_{1((k,l'),(k,l'))}+(m-1)\rho^*_{2((k,l'),(k,l'))})\right. \\ &\quad \left.+\sum_{k\neq k'}\sqrt{V_k^{l'*}}\sqrt{V_{k'}^{l'*}}((n-1)\rho^*_{1((k,l'),(k',l'))}+(m-1)\rho^*_{2((k,l'),(k',l'))}+\rho^*_{3((k,l'),(k',l'))})\right]. \end{aligned}$$

$$\begin{aligned} & Cov(\hat{\theta}^{l*},\hat{\theta}^{l'*}) \\ = \; & Cov\left[\frac{1}{r}\sum_{k=1}^{r}\frac{\hat{\theta}_k^l-b}{a},\frac{1}{r}\sum_{k=1}^{r}\frac{\hat{\theta}_k^{l'}-b}{a}\right] \\ = \; & \frac{1}{a^2r^2}\left[\sum_{k=1}^{r}Cov(\hat{\theta}_k^l,\hat{\theta}_k^{l'})+\sum_{k\neq k'}Cov(\hat{\theta}_k^l,\hat{\theta}_{k'}^{l'})\right] \\ = \; & \frac{1}{mna^2r^2}\left[\sum_{k=1}^{r}\sqrt{V_k^{l*}}\sqrt{V_k^{l'*}}((n-1)\rho^*_{1((k,l),(k,l'))}+(m-1)\rho^*_{2((k,l),(k,l'))}+\rho^*_{3((k,l),(k,l'))})\right. \\ & \left.+\sum_{k\neq k'}^{r}\sqrt{V_k^{l*}}\sqrt{V_{k'}^{l'*}}((n-1)\rho^*_{1((k,l),(k',l'))}+(m-1)\rho^*_{2((k,l),(k',l'))}+\rho^*_{3((k,l),(k',l'))})\right]. \end{aligned}$$

Therefore,

$$\begin{aligned} & Var(\hat{\theta}^{l*}-\hat{\theta}^{l'*}) \\ = \; & \frac{1}{mna^2r^2}\sum_{t=l,l'}\left[\sum_{k=1}^{r}V_k^{t*}((1+(n-1)\rho^*_{1((k,l),(k,l))}+(m-1)\rho^*_{2((k,l),(k,l))})\right. \\ & \left.+\sum_{k\neq k'}\sqrt{V_k^{t*}}\sqrt{V_{k'}^{t*}}((n-1)\rho^*_{1((k,l),(k',l))}+(m-1)\rho^*_{2((k,l),(k',l))}+\rho^*_{3((k,l),(k',l))})\right] \\ & -\frac{2}{mna^2r^2}\left[\sum_{k=1}^{r}\sqrt{V_k^{l*}}\sqrt{V_k^{l'*}}((n-1)\rho^*_{1((k,l),(k,l'))}+(m-1)\rho^*_{2((k,l),(k,l'))}+\rho^*_{3((k,l),(k,l'))})\right. \\ & \left.+\sum_{k\neq k'}^{r}\sqrt{V_k^{l*}}\sqrt{V_{k'}^{l'*}}((n-1)\rho^*_{1((k,l),(k',l'))}+(m-1)\rho^*_{2((k,l),(k',l'))}+\rho^*_{3((k,l),(k',l'))})\right]. \end{aligned}$$

Proof of Equation (4.5). The asymptotic power for a hypothesis of the difference

in accuracy.

$$\mathrm{H}_0: \ \delta_0 = \theta^l - \theta^{l'} = 0 \quad \mathrm{H}_1: \ \delta_1 = \theta^l - \theta^{l'} \neq 0$$

Based on the test statistic $((\hat{\theta}^{l*} - \hat{\theta}^{l'*}) - (\theta^l - \theta^{l'}))/\sqrt{Var(\hat{\theta}^{l*} - \hat{\theta}^{l'*})}$, the power is computed as below.

$$Rejection\ region = \left\{ \left| \frac{\hat{\theta}^{l*} - \hat{\theta}^{l'*} - 0}{\sqrt{Var_0(\hat{\theta}^{l*} - \hat{\theta}^{l'*})}} \right| > Z_{1-\alpha/2} \right\}$$

which is same as

$$\{(\hat{\theta}^{l*} - \hat{\theta}^{l'*}) > z_{1-\alpha/2}\sqrt{Var_0(\hat{\theta}^{l*} - \hat{\theta}^{l'*})}\} \cup \{(\hat{\theta}^{l*} - \hat{\theta}^{l'*}) < -z_{1-\alpha/2}\sqrt{Var_0(\hat{\theta}^{l*} - \hat{\theta}^{l'*})}\}.$$

$$\begin{aligned}
Power &= 1 - \beta = Pr\Big((\hat{\theta}^{l*} - \hat{\theta}^{l'*}) > z_{1-\alpha/2}\sqrt{Var_0(\hat{\theta}^{l*} - \hat{\theta}^{l'*})}\Big|H_1\Big) \\
&\quad + Pr\Big((\hat{\theta}^{l*} - \hat{\theta}^{l'*}) < -z_{1-\alpha/2}\sqrt{Var_0(\hat{\theta}^{l*} - \hat{\theta}^{l'*})}\Big|H_1\Big) \\
&= \Phi\left(\frac{-z_{1-\alpha/2}\sqrt{Var_0(\hat{\theta}^{l*} - \hat{\theta}^{l'*})} + \delta_1}{\sqrt{Var_1(\hat{\theta}^{l*} - \hat{\theta}^{l'*})}}\right) + \Phi\left(\frac{-z_{1-\alpha/2}\sqrt{Var_0(\hat{\theta}^{l*} - \hat{\theta}^{l'*})} - \delta_1}{\sqrt{Var_1(\hat{\theta}^{l*} - \hat{\theta}^{l'*})}}\right) \\
&\approx \Phi\left(\frac{-z_{1-\alpha/2}\sqrt{Var_0(\hat{\theta}^{l*} - \hat{\theta}^{l'*})} + \delta_1}{\sqrt{Var_0(\hat{\theta}^{l*} - \hat{\theta}^{l'*})}}\right) + \Phi\left(\frac{-z_{1-\alpha/2}\sqrt{Var_0(\hat{\theta}^{l*} - \hat{\theta}^{l'*})} - \delta_1}{\sqrt{Var_0(\hat{\theta}^{l*} - \hat{\theta}^{l'*})}}\right) \\
&= \Phi\left(\frac{-z_{1-\alpha/2}\sqrt{2(m+n)V^*((1-\rho_a) + (r-1)(\rho_b - \rho_c))/mna^2r} + \delta_1}{\sqrt{2(m+n)V^*((1-\rho_a) + (r-1)(\rho_b - \rho_c))/mna^2r}}\right) \\
&\quad + \Phi\left(\frac{-z_{1-\alpha/2}\sqrt{2(m+n)V^*((1-\rho_a) + (r-1)(\rho_b - \rho_c))/mna^2r} - \delta_1}{\sqrt{2(m+n)V^*((1-\rho_a) + (r-1)(\rho_b - \rho_c))/mna^2r}}\right),
\end{aligned}$$

with $V^* = (a\bar{\theta} + b) - (a\bar{\theta} + b)^2$.

Chapter 5

Conclusions and Future Research

In this dissertation, we have proposed semiparametric and nonparametric methods for evaluating biomarkers and diagnostic tests for a specific disease or conditions. We utilized the ROC analysis as a framework for developing the methods.

In Chapter 2, we proposed a semiparametric transformation model to combine multiple biomarkers in order to optimize diagnostic accuracy. The simulation studies and applications to the real data suggest that the proposed method performs well in small-sample settings, and the obtained optimal AUC is comparable to those using the nonparametric and logistic regression methods when biomarkers have a weak association, and is superior to the existing approaches for highly correlated biomarkers. As mentioned in the discussion section of Chapter 2, the proposed transformation model has several advantages over the methods that use the linear combinations of original biomarkers in the previous work: First, we allow a completely unknown transformation, so the diagnostic rule is less sensitive to the extreme values of biomarkers. Second, since our method is applied to the biomarkers characterized by left or right censoring and accounts for the correlated structure of biomarkers by introducing random effects, it handles practical data more properly and uses more data information for inference. Third, in our method, finding the optimal linear combination of biomarkers is straightforward and the final result does not depend on any monotone transformation of the

biomarkers.

In Chapter 3, we focused on assessing the accuracy of biomarkers by adjusting for covariates that could influence the performance of biomarkers. We developed an accelerated ROC model by generalizing the usual accelerated failure time model in the survival context to the ROC analysis. Our method models the covariate effects on the ROC curves, so that the interpretation of model parameters pertains directly to the rescaling of the ROC curves. Comparatively, the traditional model models such effects as the location-shift of the ROC curves. Thus, our model provides a useful alternative to the traditional method.

Finally, in Chapter 4, we developed a latent model to estimate and compare correlated AUCs in a paired-patient, paired-reader design. We assumed diagnostic test results come from some unknown and monotone functions of continuous latent variables, and further assumed reader variability is characterized by random effects due to a specific reader from a given diagnostic test. We also presented a method for correcting an imperfect gold standard bias and sample size formula in this design.

High-throughput technologies such as microarrays allow researchers to gather tens of thousands of genes simultaneously. Thus, there is a need for developing statistical methods to select biomarkers from thousands of genes and construct a classification rule of disease. Currently, few methods exist for evaluating high dimensional biomarkers using ROC techniques. We will extend our transformation models in Chapter 2 and will address biomarker selection as well as classification of disease.

Another application of our approach is to combine multiple biomarkers for diagnosing disease outcomes which may have more than two levels or even take ordinal values. Moreover, we will develop methods to evaluate longitudinal or repeated measures of biomarkers for disease detection. For one example, we are considering methods to combine multiple biomarkers in which different types of biomarkers are accumulated

at different time points. Finally, the model proposed in Chapter 3 can be extended to evaluate two or more biomarkers, so that one can evaluate the covariate effects on multiple biomarkers simultaneously. Furthermore, since the parameter estimates of covariates based on the log-rank estimating equation may not be efficient, we will explore other methods to attain the semiparametric efficiency.

The paired-patient and paired-reader design is the most popular and efficient to compare diagnostic tests. In Section 4.4, we developed a sample size formula based on the asymptotic normality of the AUC and by simplifying various correlations due to same readers or tests. However, we found that the theoretical powers are very conservative compared to empirical powers, especially when the imperfect gold standard bias is present as discussed in Section 4.7. Moreover, currently available methods do not accommodate this design properly for sample size calculations. The very next extension to this research will be to develop a more accurate sample size formula in this design. In broad perspective, we will develop statistical methods to assess and compare diagnostic techniques accommodating a particular situation or a study design.

Bibliography

Alonzo, T. A., and Pepe, M. S. (2000). Distribution-free ROC analysis using binary regression techniques. *Biostatistics* 3, 421-433.

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating graph. *Journal of Mathematical Psychology* 12, 387.

Beam, C.A. (1995). Random-effects models in the receiver operating characteristic curve-based assessment of the effectiveness of diagnostic imaging technology: concepts, approaches, and issues. *Academic Radiology*, 2: supple: 4-13.

Beam, C.A. (1998). Analysis of clustered data in receiver operating characteristic studies. *Statistical Methods in Medical Research*, 7:324-336.

Beiden, S.V., Wagner, R.F. and Campbell, G. (2000). Components-of-variance models and multiple-bootstrap experiments. *Academic Radiology*, 7:341-349.

Bickel, P.J., Klaassen, C.A.J., Ritov, Y., and Wellner, J.A. (1993). *Efficient and Adaptive Estimation Semiparametric Models*. Johns Hopkins University Press, Baltimore.

Bullitt E., Gerig G., Pizer S.M., Lin, W., Aylward S.R. (2003). Measuring Tortuosity of the Intracerebral Vasculature from MRA Images. *IEEE-TMI*, 22:1163-1171.

Cai, T. and Moskowitz, C. (2004). Semi-parametric estimation of the binormal ROC curve for a continuous diagnostic test. *Biostatistics*, 5(4), 573-86.

Cai, T. and Pepe, M.S. (2002). Semi-parametric ROC analysis to evaluate biomarkers for disease. *Journal of the American Statistical Association*, 97, 1099-1107.

Cole, E.B., Pisano, E.D., Zeng, D., Muller, K., Aylward, S.R., Park, S., Kuzmiak, C., Koomen M., Pavic, D., Walsh, R., Baker, J., Gimenez, E.I., Freimanis, R. (2005). The effects of gray scale image processing on digital mammography interpretation performance. *Academic Radiology*, 12(5):585-595.

Coleman, T. F. and Li, Y. (1994). On the Convergence of Reflective Newton Methods for Large-scale Nonlinear Minimization Subject to Bounds. *Mathematical Programming*, 67, 189-224.

Coleman, T. F. and Li, Y. (1996). An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds. *SIAM Journal on Optimization*, 6, 418-445.

DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44:837.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society: Series B*, 39, 1-38.

Dorfman, D.D., Alf, E. (1969). Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals rating method data. *Journal of Mathematical Psychology*, 6:487.

Dorfman, D.D., Berbaum, K.S., Metz, C.E. (1992). ROC rating analysis: generalization to the population of readers and cases with the jackknife method. *Investigative Radiology*, 27:723.

Egan, J.P. (1975). *Signal detection theory and ROC analysis*. New York: Academic Press, 1975.

Etzioni, R., Pepe. M. S., Longton, G., Hu, C., Goodman, G. (1999). Incorporating the time dimension in receiver operating characteristic curves: a case study of prostate cancer. *Medical Decision Making*, 19, 242-251.

Etzioni, R., Kooperberg, C., Pepe, M.S., Smith R., Gann, P.H. (2003). Combining biomarkers to detect disease with application to prostate cancer. *Biostatistics*, 4:523-538.

Gatsonis, C.A. (1995). Random effects models for diagnostic test accuracy. *Academic Radiology*. 2:S14-S21.

Goddard, M.J., Hinberg, I. (1990). Receiver operator characteristic (ROC) curves and non-normal data: An empirical study. *Statistics in Medicine*. 9, 325-337.

Goodman, G., Omenn, G. S., Thornquist, M., Lund, B., Metch, B., and Gylys-Colwell, I. (1993). The Carotene and Retinol Efficacy Trial (CARET) to prevent lung cancer in high-risk populations: pilot study with cigarette smokers. *Cancer Epidemiology Biomarkers and Prevention*, 2, 389-396.

Green, D. M., Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley and Sons.

Grey, D.R., Morgan, B.J.T. (1972). Some aspects of ROC curve-fitting: normal and logistic models. *Journal of Mathematical Psychology*, 9:128.

Hanley, J.A. and McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143:29.

Hanley, J.A. and McNeil, B.J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148:839.

Hanley, J.A. (1988). The robustness of the binormal assumptions used in fitting ROC curves. *Medical Decision Making* 8, 197-203.

Hanley, J. A. (1989). Receiver Operating Characteristic (ROC) methodology: the state of the art. *Clinical Reviews in Diagnostic Imaging*, 29, 307-335.

Heagerty, P.J., and Pepe, M.S. (1999). Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children. *Journal of the Royal Statistical Society*, (C) 48, 533-551.

Hellmich, M., Abrams, K.R., Jones, D.R., Lambert, P.C. (1998). A Bayesian approach to a General Regression Model for ROC curves. *Medical Decision Making*, 18:436-443.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19, 293-325.

Hoeffding, W. (1951). 'Optimum nonparametric tests. *Proceedings of 2nd Berkeley Symposium on Mathematical Statistics and Probability*, 83-92 (Berkeley, University of California Press).

Hsieh, F.S. and Turnbull, B.W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Annals of Statistics*, 24, 25-40.

Ishwaran, H. and Gatsonis, C. (2000). A general class of hierarchical ordinal regression models with applications to correlated ROC analysis. *Canadian Journal of Statistics*, 28, 731-750.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 44, 226-233.

McClish D.K. (1989). Analyzing a portion of the ROC curve. *Medical Decision Making*, 9:190.

Metz, C.E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8:283.

Metz, C.E. (1998). Herman BA, Shen J-H. Maximum-likelihood estimation of ROC curves from continuously-distributed data. *Statistics in Medicine*, 17:1033.

Obuchowski, N.A. (1995a). Multi-reader Multi-modality ROC Studies: Hypothesis Testing and Sample Size Estimation Using an ANOVA Approach with Dependent Observations. *Academic Radiology*, 2:522-529.

Obuchowski, N.A. (1995b). Multireader receiver operating characteristic studies: a comparison of study designs. *Academic Radiology*, 2:709-716.

Obuchowski, N.A. and Rockette, H.E. (1995). Hypothesis testing of diagnostic accuracy for multiple readers and multiple tests: An ANOVA approach with dependent observations. *Communications in Statistics - Simulation and Computation*, 24(2), 285-308.

Obuchowski, N.A. (1997). Nonparametric Analysis of Clustered ROC Curve Data. *Biometrics*, 53:170-180.

Peng, F., Hall, W.J. (1996). Bayesian analysis of ROC curves using Markov-Chain Monte Carlo methods. *Medical Decision Making*, 16:404-411.

Pepe, M.S. (1997). A regression modelling framework for receiver operating characteristic curves in medical diagnostic testing. *Biometrika*, 84(3):595-608.

Pepe, M.S. (1998). Three Approaches to Regression Analysis of Receiver Operating Characteristic Curves for Continuous Test Results. *Biometrics*, 54, 124-135.

Pepe, M. S. (2000a). An interpretation for the ROC curve and inference using GLM procedures. *Biometrics*, 56, 352-359.

Pepe, M. S. (2000b). Receiver Operating Characteristic Methodology. *Journal of the American Statistical Association*, 95, 308-311.

Pepe, M.S. and Thompson, M.L. (2000). Combining diagnostic test results to increase accuracy. *Biostatistics*, 1:123-140.

Pepe, M.S. (2003). *The Statistical Evaluation of Medical Tetsts for Classification and Prediction*. Oxford: Oxford University Press.

Rao, J.N.K., and Scott, A.J. (1992). A simple method for the analysis of clustered binary data. *Biometrics*, 15, 385-397.

Sen, P. K. (1960). On some convergence properties of U-statistics. *Calcutta Statistical Association Bulletin*, 10, 1-18.

Su, J. Q. and Liu, J. S. (1993). Linear combinations of multiple diagnostic markers. *Journal of the American Statistical Association*, 88:1350-1355.

Swets,J.A. and Pickett, R.M. (1982). *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. Academic Press, New York.

Thompson, M.L., Zucchini, W. (1989). On the statistical analysis of ROC curves. *Statistics in Medicine*, 8:1277.

Toledano, A. and Gatsonis, C.A. (1995). Regression analysis of correlated ROC data. *Academic Radiology*. 2:S30-S36.

Tosteson, A., Begg, C. (1998). A general regression methodology for ROC curve estimations. *Medical Decision Making*, 8:204.

Wieand, S., Gail, M.H., James, B.R., James, K.L. (1989). A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika*, 76:585.

Zeng, D. (2004). Estimating Marginal Survival Function by Adjusting for Dependent Censoring Using Many Covariates. *Annals of Statistics*, 32, 1533-1555.

Zeng, D., and Lin, D.Y. (2008). Efficient resampling methods for non-smooth estimating functions. *Biostatistics*, 9, 355-363.

Zhang, D.D., Zhou, X.-H., Freman, D.H. Jr., and Freeman, J. (2002). A non-parametric method for the comparison of partial areas under ROC curves and its application to large health care data sets. *Statistics in Medicine*, 21, 701-715.

Zhou, X.-H., Obuchowski, N.A., and McClish, D.K. (2002). *Statistical Methods in Diagnostic Medicine*. New York:Wiley.

Zou, K.H. and Hall, W.J. (2000). Two transformation models for estimating an ROC curve derived from continuous data. *Journal of Applied Statistics*, 27, 621-631.

Zweig, M.H., Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39:561.

Swets, J.A. (1986). Indices of discrimination or diagnostic accuracy: their ROCs and implied models. *Psychological Bulletin*, 99, 100-117.

van der Vaart, A. W., and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

Zeng, D., and Lin, D. Y. (2007). Maximum Likelihood Estimation in Semiparametric Models with Censored Data (with discussion). *Journal of the Royal Statistical Society B*, 69, 507-564.

Zeng, D., and Lin, D. Y. (2009). General Asymptotic Theory for Maximum Likelihood Estimation in Semiparametric Regression Models With Censored Data. *Statistica Sinica*, in press.

Zhou, X.-H., Obuchowski, N.A., and McClish, D.K. (2002). *Statistical Methods in Diagnostic Medicine*. New York:Wiley.

Zou, K. and Hall, W. J. (2000). Two transformation models for estimating an ROC curve derived from continuous data. *Journal of Applied Statistics*, 27, 621-633.