GROUP SEQUENTIAL METHODS FOR ROC CURVES

by

Xuan Ye
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
In Partial fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Statistical Science

Committee:

______________________ Dr. Liansheng L. Tang, Dissertation Director

______________________ Dr. Daniel B. Carr, Committee Member

______________________ Dr. Anand N. Vidyashankar, Committee Member

______________________ Dr. Alessandra Luchini, Committee Member

______________________ Dr. William F. Rosenberger, Department Chair

______________________ Dr. Kenneth S. Ball, Dean, Volgenau School of Engineering

Date: ______________________ Fall Semester 2015
George Mason University
Fairfax, VA

Group Sequential Methods for ROC Curves

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Xuan Ye
Master of Science
University of Illinois at Chicago, 2000
Bachelor of Engineering
University of Science and Technology of China, 1994

Director: Dr. Liansheng L. Tang, Associate Professor
Department of Statistics

Fall Semester 2015
George Mason University
Fairfax, VA

Acknowledgments

I would like to thank my advisor, Dr. Liansheng Tang, for his excellent guidance and help on my dissertation. I also would like to thank the members of my PhD committee, Dr. Daniel Carr, Dr. Anand Vidyashankar and Dr. Alessandra Luchini for their time and help throughout the dissertation work. I appreciate the help provided by Dr. William Rosenberger throughout my graduate study. I would like to thank Liz Quigley for her assistance in the department. Finally, I would like to thank my family for their support.

Table of Contents

List of Tables

List of Figures

Abstract

GROUP SEQUENTIAL METHODS FOR ROC CURVES

Xuan Ye, PhD

George Mason University, 2015

Dissertation Director: Dr. Liansheng L. Tang

Comparative diagnostic studies in which each patient has two tests conducted or has several diseased and nondiseased observations for each test will generate correlated or clustered ROC curves. The traditional ROC comparison methods applied on the correlated or clustered data can result in incorrect statistical inference. Furthermore, to design and apply group sequential method in these comparative trials, we need to derive the theoretical variance-covariance structure and the joint distribution of sequential statistics. We first derive the theoretical covariance structure of sequential correlated and clustered ROCs' difference and further verify the findings through simulation studies. Then based on the independent increments covariance structure that we have proved, we conduct group sequential studies for comparing ROC curves on both simulated and real data.

Chapter 1: Introduction

1.1 Diagnostic Tests and ROC, PPV and NPV Curves

Diagnostic tests are important in medical decision makings, such as cancer and glaucoma diagnosis, since they provide reliable information about a patient's health condition and an early diagnosis can possibly save a patient's life. The health care provider can make plans for managing the patient with the diagnosis information (Sox et al. 1989) and possibly better understand the disease mechanism through research (McNeil and Adelstein 1976).

Diagnostic test accuracy is defined as the ability of the test to discriminate the states of health (Zweig and Campbell 1993). Hence the accuracy is measured by comparing the test results to the true disease status. A diagnostic test may have binary, ordinal or continuous results. For a binary test, the accuracy is commonly evaluated using sensitivity and specificity. Sensitivity is the probability of a positive test result when a patient has the disease, and specificity is the probability of a negative test result when a patient does not have the disease. Sensitivity is also known as the true positive rate or TPR, and 1-specificity is also known as the false positive rate or FPR. These classification probabilities are commonly used in diagnostic evaluation study. We denote the disease status by D, with D=1 for a case, and D=0 for a control. Let X denote the binary test result with X=0 for a negative test result, and X=1 for a positive test result. We have that $TPR = P(X = 1|D = 1)$, and $FPR = P(X = 1|D = 0)$.

In addition to the classification probabilities defined above, predictive values reflecting how well the test results predict the disease status are often used to assess the accuracy of a test. The positive predictive value (PPV) and negative predictive value (NPV) are defined as, for a binary test result, $PPV = P(D = 1|X = 1)$ and $NPV = P(D = 0|X = 0)$. The predictive values depend on the prevalence of disease and the performance of the test in two subject groups. Hence, sensitivity and specificity are often used to quantify the inherent accuracy of the test, because they measure how well the test reflects true disease status. Predictive values are used to quantify the clinical value of the test, because the patient and clinician are most interested in how likely the disease is actually present given the test result, that is the measure of how well the test predicts the disease status. These predictive values are also useful for prognostic testing evaluation. A prognostic testing is a prediction about how something such as an illness will develop. A prognostic marker is a marker measured in people with disease used to predict an aspect of their prognosis (Pepe et al. 2004).

A perfect test is one that completely separates the case and control populations and has zero misclassification probabilities with $TPR = 1$ and $FPR = 0$. Consequently, the predict values will also be optimal with $PPV = 1$ and $NPV = 1$. On the other hand, a test with no added value contains no information about true disease status. That is $PPV = P(D = 1)$ or p, and $NPV = P(D = 0)$ or $1 - p$, where p is the prevalence of disease. However, since the test measurements usually follow normal or transformed normal distributions, it is very unlikely to have a perfect test in practice.

For ordinal or continuous test results, the Receiver Operating Characteristic (ROC) curve is commonly used for analysis. An ROC curve is a graphical plot which illustrates the performance of a binary classifier system as we vary the cutoff threshold. It is created by plotting the fraction of true positives out of the total actual positives v.s. the fraction of false positives out of the total actual negatives at various threshold values. Here the sensitivity and specificity depend on how well the test separates the two groups and the threshold we choose. Given a diagnostic test, we let the threshold go from $-\infty$ to ∞, the ROC curve plots all possible pairs of FPR and TPR. Hence, the ROC is a relative operating characteristic curve, because it is a comparison of two operating characteristics, TPR and FPR, as the threshold changes, and the ROC curve is always monotonic.

In the ROC definition, a binary test is defined based on a pre-specified threshold c, and a patient is classified to be positive if $X > c$, or negative if $X \leq c$. Therefore, TPR and FPR are functions of the threshold value c, $TPR(c) = P(X > c|D = 1)$, $FPR(c) = P(X > c|D = 0)$. For c ranging over all possible values, the pairs (FPR(c), TPR(c)) form the ROC curve. By this definition, the ROC curve can be expressed as $R(\cdot) = \{(FPR(c), TPR(c)), c \in \mathbb{R}\}$. Throughout the thesis, we use R to represent the ROC function.

We denote distribution functions on the continuous test result as $F_D(c) = P(X \leq c|D = 1)$ for the case population, and $F_{\bar{D}}(c) = P(X \leq c|D = 0)$ for the control population. Similarly, we denote survival functions on the continuous test result as $S_D(c) = P(X > c|D = 1)$ for the case population, and $S_{\bar{D}}(c) = P(X > c|D = 0)$ for the control population, then the ROC curve can be easily expressed in a function form of FPR as

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

We let D be a Bernoulli random variable with prevalence $p = P(D = 1)$, then $F(x) = pF_D(x) + (1 - p)F_{\bar{D}}(x)$ is the marker distribution function for the entire population.

Under the assumption that the test results follow normal distributions in both the case and the control populations, then this binormal ROC curve has the following property. Assume that the binormal distributions for the test results are, $X|(D = 1) \sim N(\mu_D, \sigma_D^2)$, and $X|(D = 0) \sim N(\mu_{\bar{D}}, \sigma_{\bar{D}}^2)$, for the case and control populations respectively, the ROC curve can be expressed as in Zhou et al. (2002)

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

where $a = \frac{\mu_D - \mu_{\bar{D}}}{\sigma_D}$, $b = \frac{\sigma_{\bar{D}}}{\sigma_D}$.

Furthermore, there exists some monotone transformation of X such that the distributions of the transformed test results are normal. Based on the fact that the ROC curve derived from the monotone transformation on X is identical to the original one, the binormal ROC curve function can be applied to any underlying distributions and is a common function form of ROC curves.

Many statistical analyses for ROC curves are based on the summary statistics which include the area under the curve (AUC), partial area under the curve (pAUC), and the weighted area under the curve (wAUC) (Zhou et al. 2011). The area under ROC curve (AUC) is given by

$$AUC = \int_0^1 R(u)du = P(X_D > X_{\bar{D}}).$$

Wieand et al. (1989) proposed a general method based on the weighted area under the curve. We can apply this method to estimate the area under the curve, partial area under the curve and TPR at a particular FPR using the weighted integration on FPRs. The weighted AUC (wAUC) formula is

$$wAUC = \int_0^1 R(u)dW(u),$$

where $W(u)$ is a probability measure. If we use $W(u) = u$, then the weighted AUC is the same as AUC equation above. Or if we use $W(u)$ equals 0 for $u \in [0, u_0)$ and 1 for $u \in [u_0, 1]$, then the $wAUC$ is the sensitivity at FPR u_0, which equals $R(u_0)$. The partial AUC between FPRs u_0 and u_1 is given by

$$pAUC(u_0, u_1) = \frac{1}{u_1 - u_0} \int_{u_0}^{u_1} R(u)du,$$

which can be achieved by letting $W(u) = (u-u_0)/(u_1-u_0)$ for $u \in (u_0, u_1)$; 0 for $u \in [0, u_0]$; 1 for $u \in [u_1, 1]$; and applying it to the $wAUC$ formula above.

Comparison of the accuracy of two diagnostic tests based on ROC curves is often conducted using fixed sample designs. However, to address the ethics and efficiency concerns of clinical trial studies, there is a need to apply more flexible designs such as a group sequential design. At a series of interim looks during a comparative diagnostic trial, a group sequential test monitors a statistic summarizing the difference in clinical data between the two groups.

For a two-sided test, if the absolute value of this statistic exceeds some specified critical value, the trial is stopped and the null hypothesis of no difference between two groups is rejected. The critical values are the boundaries for the sequence of test statistics. The null hypothesis is accepted if the statistic stays within the test boundaries until the trial's planned termination.

In this thesis, we incorporate group sequential methods into the design of comparative diagnostic study with respect to ROC curves. We estimate ROC curves are estimated empirically without assuming the distributions of the underlying diagnostic test data. We study the difference between sequential empirical ROC curves on the process level. Then we derive the asymptotic distribution theory for the difference between sequential empirical ROC curves and derive the asymptotic covariance structure for comparative ROC statistics. Relating the difference between empirical ROC curves to the Kiefer process, we further show these results can be used to conduct a group sequential design using standard software.

In Figure 1.1, we plot three example ROC curves each evaluating one specific diagnostic test. The ROC curve in red color is generated from the test distribution data shown at the lower-left. Since the diagnostic test separates the case and control populations almost completely with very little overlapping part, the corresponding ROC curve reaches the upper-left corner with AUC close to 1. This indicates that this test is the best one among the three diagnostic tests. The blue one, on the contrary, barely separates the two populations with regard to the test results. Hence it is the least effective diagnostic test with the smallest AUC. The black one lies in between with respect to the separation of the case population and the control population.

Similarly, PPV and NPV for a continuous test results with a given threshold value c are defined as, $PPV(c) = P(D = 1|X > c)$ and $NPV(c) = P(D = 0|X \leq c)$. Furthermore, PPV and NPV curves are defined on PPV(c) and NPV(c) for all $c \in (-\infty, \infty)$. In practice, PPV and NPV curves are usually indexed by a summary of the marker distribution rather than a generic threshold (Pepe 2003; Moskowitz and Pepe 2004; Zheng et al. 2008). Here, we consider the PPV and NPV curves indexed by FPR and by the percentile value in the

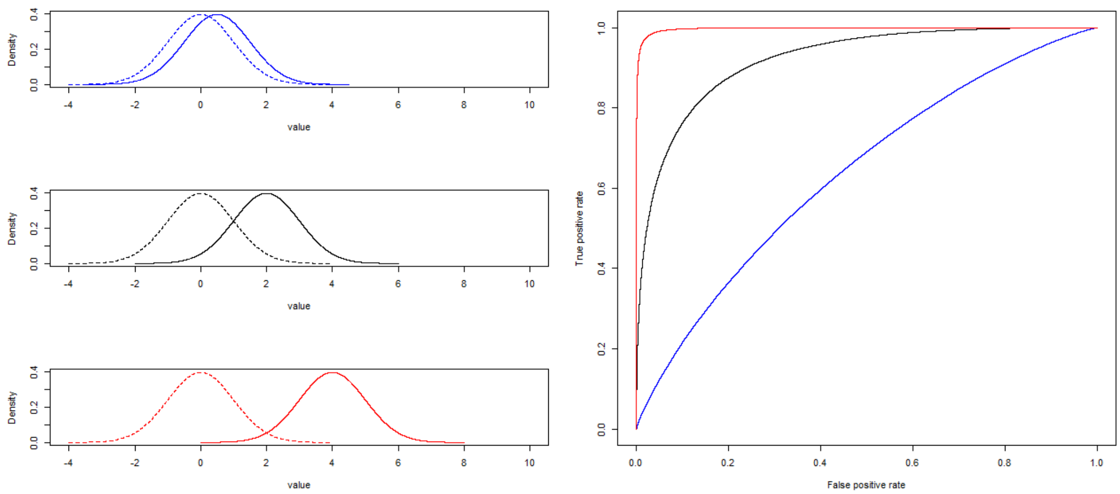

Figure 1.1: Example ROC curves

entire population.

The PPV and NPV curves indexed by FPR are defined as $PPV(t) = P(D = 1|X > S_{\bar{D}}^{-1}(t))$ and $NPV(t) = P(D = 0|X \leq S_{\bar{D}}^{-1}(t))$ for all $t \in (0, 1)$ and can be written as functions of the ROC curve,

$$PPV(t) = \frac{R(t)p}{R(t)p + t(1-p)},$$

and

$$NPV(t) = \frac{(1-t)(1-p)}{(1-R(t))p + (1-t)(1-p)}.$$

The PPV and NPV curves can also be indexed by the percentile value in the entire population. Here we use u to represent the rate of having negative results in the entire population, i.e. $u = P(X \leq c)$ or $F(c)$ with cutoff c, which involves the mixed distribution function for the entire population. In this case, the PPV and NPV curves are defined as $PPV(u) = P(D = 1|X > F^{-1}(u))$ and $NPV(u) = P(D = 0|X \leq F^{-1}(u))$ for all $u \in (0, 1)$.

Under this setting, the PPV curve can be written as

$$PPV(u) = \frac{S_D(F^{-1}(u))p}{1-u},$$

and the NPV curve can be written as

$$NPV(u) = \frac{u-p}{p} + \frac{1-u}{u} PPV(u).$$

These definitions on PPV and NPV are indexed by a variable which involves the biomarker distribution function of the entire population

1.2 Group Sequential Methods for Estimating and Comparison of One ROC, PPV and NPV Curve

The diagnostic accuracy can be evaluated in a fixed sample design or a group sequential design. In a fixed sample design, the ROC statistics are estimated after all subjects are recruited and tests measured. While in a group sequential design, the ROC statistics are estimated at interim analysis points as subjects are being accrued. In fixed sample design approaches, the ROC curves and their comparison based on AUC summaries have been studied (Pepe, Longton, and Janes 2009; Obuchowski 2005; Pepe 2000). The ROC curves study and comparison based on pAUC were also investigated (Obuchowski 2005; Dodd and Pepe 2003). However, since in a fixed sample trial the statistical analysis is conducted at the end when data collection is completed, it has inherent ethical and efficiency issues as patients are involved in these trials. More flexible designs have been proposed, such as adaptive design and group sequential design, to address the ethical and efficiency issues. A group sequential method allows researchers to terminate the study early, if the candidate diagnostic test is clearly superior or non-inferior to the established diagnostic test under comparison (Jennison and Turnbull 2000). A group sequential method also allows early termination for futility based on conditional estimation (Pepe et al. 2009; Jennison and Turnbull 2000; Fleming et al. 1984). The adoption of the group sequential method may substantially save the number of subjects needed, resulting in both time and resource use efficiency and ethical benefits.

Group sequential designs provide a chance to periodically monitor and analyze the accruing data. In many trials in which data accumulate over a period of time, it is an advantage if we can monitor results as they occur and take action accordingly. For trials involving human subjects, there is also an ethical need to monitor results and possibly stop the trial early. In clinical therapeutical trials, this ensures that individuals are not exposed to unsafe, ineffective treatment. In clinical diagnostic trials, this ensures that individuals are not exposed to harmful or intrusive diagnostic procedures such as medical radiological imaging. As the example of the comparative diagnostic trial on CT and PET shows, which will be discussed in detail later, less patients will be exposed to harmful X-ray if we can stop the study earlier at an interim analysis. For administrative reasons, we also need interim analyses to ensure that the experiment is being executed as planned, the study population satisfies inclusion/exclusion criteria and matches the intended use population, and that the study protocol and test procedures are followed. There are also economic benefits conducting group sequential methods as the trials now can be terminated earlier for apparent superiority or futility.

However, if we use the usual critical values for the fixed sample design at each analysis, the type I error rate will be greatly inflated over the nominal α level (Armitage et al. 1969). Hence, Pocock (1977), O'Brien and Fleming (1979), Fleming et al. (1984), Kim and Demets (1992) and Wang and Tsiatis (1987) have developed group sequential methods which adjust the critical values to maintain the overall type I error rate at an acceptable level. Also with further calculations, we can determine the sample size needed for the group sequential test to attain a desired power requirement.

To control the overall type I error rate, we can apply the idea of error spending as demonstrated in the following two-sided testing context. Assuming the maximum number of analyses is J, which is fixed before the study begins. The type I error rate is partitioned into probabilities of $p_1, \cdots, p_J$, with the sum of α, i.e. $\sum_{j=1}^{J} p_j = \alpha$. At each interim analysis point j, critical values c_j for the standardized statistics Z_j, $j = 1, \cdots, J$, are determined such that $P(|Z_1| > c_1) = p_1$, $P(|Z_1| \leq c_1, \cdots, |Z_{j-1}| \leq c_{j-1}, |Z_j| > c_j) = p_j$ for $j = 2, \cdots, J$.

Various methods have been proposed to apply the group sequential methodology in diagnostic test studies (Tang et al. 2008; Tang and Liu 2010; Liu et al. 2008; Pepe et al. 2009; Mazumdar and Liu 2003). The nonparametric sequential methods based on AUC, pAUC and wAUC statistics for ROC curves comparison have been introduced, and the method for sample size recalculation at interim analyses has also been presented. Mazumdar (2004) introduced group sequential design approaches in planning comparative diagnostic accuracy trials. Assuming that the measurements of biomarkers are normally distributed, Mazumdar and Liu (2003) derived the asymptotic distribution of the standardized AUC difference statistic and illustrated the boundary and sample size determination in a group sequential design. To address the bias introduced by allowing early termination for futility in the group sequential study, Koopmeiners et al. (2012) proposed conditional estimators and confidence intervals that correct for the bias if the underlying statistics have an independent increments covariance structure. In the design of a two-stage study to develop and validate a panel of biomarkers where a predictive model is developed in stage 1 and validated in stage 2 using only the samples that were not used for training, Koopmeiners and Vogel (2013) proposed to apply group sequential method with interim analyses in stage 2, resulting in greater savings in the required number of samples. Tang and Liu (2010) developed a nonparametric adaptive method for comparative diagnostic trials which allows early stopping or the sample sizes recalculating based interim data analysis. Hsieh et al. (1996) studied the asymptotic property of the empirical ROC curve and proved it converges to the sum of two independent Brownian bridges. Extending the research, Koopmeiners and Feng (2011) studied the single empirical ROC curve on process level without using summary index. They derived the asymptotic properties of the sequential empirical ROC, PPV and NPV curves, proved the embedded independent increments covariance structure, and applied the theory in group sequential designs.

Related to ROC curves, PPV and NPV curves are extensions of PPV and NPV to continuous markers. Huang et al. (2007) introduced a predictiveness curve that provides a common meaningful scale for comparing markers. Moskowitz and Pepe (2004), Zheng et al. (2008) and Koopmeiners and Feng (2011) studied the PPV and NPV curves for continuous markers.

The asymptotic properties of one sequential empirical ROC curve have been rigorously studied in Koopmeiners and Feng (2011). In their paper, they estimated the ROC, PPV and NPV curves empirically to avoid assumptions about the underlying distributional form of the biomarkers. They derived asymptotic properties of the sequential empirical ROC, PPV and NPV curves under case-control sampling using sequential empirical process theory. They proved that the sequential empirical ROC process converges to the sum of independent Kiefer processes and extended the finding to the empirical PPV and NPV processes. Then they incorporated group sequential methods into the design of diagnostic biomarker studies. They derived the asymptotic property on one sequential empirical ROC curve with J stopping times, and $r_{D,j}, r_{\bar{D},j}$ are the proportions of cases and controls, respectively, available at a given time point j, and proved that $(\widehat{R}_{r_{D,1},r_{\bar{D},1}}(t_1), \widehat{R}_{r_{D,2},r_{\bar{D},2}}(t_2), \cdots, \widehat{R}_{r_{D,J},r_{\bar{D},J}}(t_J))$, is approximately multivariate normal.

Similarly, the asymptotic property on one sequential empirical PPV and NPV curve indexed by FPR with J stopping times, $(\widehat{PPV}_{r_{D,1},r_{\bar{D},1}}(t_1), \widehat{PPV}_{r_{D,2},r_{\bar{D},2}}(t_2), \cdots, \widehat{PPV}_{r_{D,J},r_{\bar{D},J}}(t_J))$ is also approximately multivariate normal. The NPV curve has the same asymptotic property.

Koopmeiners and Feng (2011) further proved the independent increments covariance structure feature and illustrated the implementation of a group sequential study on one ROC, PPV or NPV curve utilizing standard GSD software.

Thorough understanding of the joint asymptotic properties of two sequential empirical

ROC curves, as well as the sequential differences of two empirical ROC curves at any FPR, will help us conduct group sequential designs on the process level instead of the point level. It can be shown that they asymptotically follow special Kiefer processes. This implies that the sequential differences at different FPRs are also asymptotically jointly normal. Furthermore, the existing results on the summary ROC statistics can be obtained from our findings.

1.3 Correlated ROC, PPV and NPV Curves

Correlated ROC data arise when two diagnostic tests are performed on the same set of individuals in the case and control populations. Each patient is examined once using each of the diagnostic tests resulting in correlated ROC curves. Comparison of two ROC curves based on AUCs has been previously studied (Pepe, Longton, and Janes 2009). However, when ROC curves are correlated, the correlated nature of the data must be taken into account in the analysis (DeLong et al. 1988). They presented an approach for the comparison of ROC curves that are correlated. Wieand et al. (1989) studied a broad class of nonparametric statistics for comparing two independent or correlated diagnostic markers. Zhou et al. (2008) discussed the design and application of GSD method to the comparative ROC studies based on the non-parametric Wilcoxon AUC estimators. This approach can be applied to AUC comparisons of two diagnostic tests measured on the same subjects resulting in correlated ROC curves. Liu et al. (2008) developed a nonparametric group sequential method to evaluate and compare the AUCs of clustered ROC curves, which can be either independent or correlated. The procedure relies on the construction of a two-dimensional statistics which are based on Mann-Whitney statistic (Whitehead 1999). Liu et al. (2005) demonstrated that the partial area under the ROC curve (pAUC) is the probability of a constrained stochastic ordering, which can be estimated using a weighted Mann-Whitney statistic. The authors investigated the statistical properties and developed a testing procedure to compare the partial area under two ROC curves, of which the two diagnostic tests are performed on the same group of cases and controls. Tang et al. (2008) derived that the sequential weighted area under the ROC curve (wAUC) statistics has an independent increments covariance structure, and applied it in the GSD for the sequential ROC curves comparison.

In the fields of PPV and NPV, we study two correlated PPV or NPV curves, which can be indexed either by FPR or by the percentile value. For PPV indexed by FPR, with prevalence level p, we define

$$\Delta(t) = PPV_1(t) - PPV_2(t),$$

which is the difference of two markers' PPV at a given FPR t. In general, we add hat to the parameter to denote the estimator of the parameter. Hence, we have $\hat{\Delta}(t) = \widehat{PPV}_1(t) - \widehat{PPV}_2(t)$, which is the estimated difference of two markers' PPV at a given FPR t based on empirical PPV estimation. And $\hat{\Delta}_{r_D, r_{\bar{D}}}(t) = \widehat{PPV}_{1,r_D,r_{\bar{D}}}(t) - \widehat{PPV}_{2,r_D,r_{\bar{D}}}(t)$ represents the estimation at a time point in a sequential trial. We then derive the asymptotic properties of $\hat{\Delta}_{r_D, r_{\bar{D}}}(t)$ and apply it to the group sequential design of PPV curves comparison indexed by FPR.

Similarly, for NPV indexed by FPR, we define

$$\Delta(t) = NPV_1(t) - NPV_2(t).$$

We have $\hat{\Delta}(t) = \widehat{NPV}_1(t) - \widehat{NPV}_2(t)$, and $\hat{\Delta}_{r_D, r_{\bar{D}}}(t) = \widehat{NPV}_{1,r_D,r_{\bar{D}}}(t) - \widehat{NPV}_{2,r_D,r_{\bar{D}}}(t)$. Then following the same steps, we derive the asymptotic properties of $\hat{\Delta}_{r_D, r_{\bar{D}}}(t)$ and apply it to the group sequential comparison study for NPV curves indexed by FPR.

For PPV, NPV indexed by the percentile value, i.e. the rate u of having negative results in the population, we define

$$\Delta(u) = PPV_1(u) - PPV_2(u).$$

We can estimate it using $\hat{\Delta}(u) = \widehat{PPV}_1(u) - \widehat{PPV}_2(u)$, and $\hat{\Delta}_{r_D,r_{\bar{D}}}(u) = \widehat{PPV}_{1,r_D,r_{\bar{D}}}(u) - \widehat{PPV}_{2,r_D,r_{\bar{D}}}(u)$. Then we can derive the asymptotic properties of $\hat{\Delta}_{r_D,r_{\bar{D}}}(u)$ and apply it to the group sequential comparison study for PPV curves indexed by the percentile value.

1.4 Clustered ROC Curves

Clustered ROC data have multiple measurements on both the case and the control units, taken from the same study subject. For example, we might have measurements on both left and right eyes for an ophthalmic diagnostic testing. While in other diagnostic settings, we might get multiple measurements from both normal and diseased tissues of the same subject. In addition, we might need to apply two different diagnostic procedures on the same set of subjects for a comparison study. Thus, there are multiple measurements for each test per subject (Obuchowski 1997). In the paired comparison study design of two ROC curves with clustered data, it is important to take into consideration of two types of correlations. One is the correlation within a cluster, the other is the correlation between the different diagnostic tests from the same cluster (Li and Zhou 2008).

For clustered ROC data, Obuchowski (1997) proposed a nonparametric method using Wilcoxon-Mann-Whitney statistics. Obuchowski (1997) expanded DeLong et al. (1988) nonparametric method and applied the ideas of Rao and Scott (1992) to handle the clustered data. For ℓth biomarker in the ith cluster, we use $X_{\ell ij}$ denotes the jth case result and assume they follow the distribution $F_{\ell,D}$, for $\ell = 1, 2$, $i = 1, \ldots, n$, $j = 1, \ldots, m_{\ell i}$, where n is the total number of clusters and $m_{\ell i}$ represents the number of case results for biomarker ℓ from cluster i. Similarly, $Y_{\ell ik}$ denotes the kth control result of ℓth marker in ith cluster, which has distribution $F_{\ell,\bar{D}}$, for $k = 1, \ldots, n_{\ell i}$ with $n_{\ell i}$ representing the number of control results for biomarker ℓ from cluster i. We also let $S_{\ell,D}$ represents the survival function for $X_{\ell ij}$ and $S_{\ell,\bar{D}}$ the survival function for $Y_{\ell ik}$. The total number of biomarker case results from all clusters is the sum of all $m_{\ell i}$, i.e. $M_\ell = \sum_{i=1}^n m_{\ell i}$, and the total number of biomarker control results from all clusters is the sum of all $n_{\ell i}$, i.e. $N_\ell = \sum_{i=1}^n n_{\ell i}$. The estimated

AUC on the clustered ROC data for $\ell = 1, 2$ is given by

$$\widehat{AUC}_\ell = \frac{1}{M_\ell N_\ell} \sum_{i=1}^{n} \sum_{i'=1}^{n} \sum_{j=1}^{m_{\ell i}} \sum_{k=1}^{n_{\ell i'}} \psi(X_{\ell ij}, Y_{\ell i' k}),$$

where ψ is defined as

$$\psi(X_{\ell ij}, Y_{\ell i'k}) = \begin{cases} 1, & X_{\ell ij} > Y_{\ell i'k} \\ \frac{1}{2}, & X_{\ell ij} = Y_{\ell i'k} \\ 0, & X_{\ell ij} < Y_{\ell i'k} \end{cases}.$$

The case and control biomarker results are transformed into X-components and Y-components, then summed up for the ith cluster. Then the sum of squares of the X-components and Y-components as well as the correlation between the case and control observations within the same cluster are calculated. Based on these, Obuchowski (1997) estimated the variance of $\widehat{AUC}$ and further stated that $(\widehat{AUC} - AUC)/(\widehat{var}(\widehat{AUC}))^{1/2}$ is asymptotically $N(0,1)$.

Obuchowski (1997) further proposed a method to calculate the covariance of two estimated AUCs for comparing cluster-correlated ROC curves. Similarly, based on the sum of the X-components and Y-components for the ith cluster from the ℓth ROC curve, she derived the formula for the covariance between the estimated areas under two ROC curves. The estimator of the variance of the difference between two cluster-correlated ROC curves is given as $\widehat{var}(\widehat{AUC}_1 - \widehat{AUC}_2) = \widehat{var}(\widehat{AUC}_1) + \widehat{var}(\widehat{AUC}_2) - 2\widehat{cov}(\widehat{AUC}_1, \widehat{AUC}_2)$. She further stated that $((\widehat{AUC}_1 - \widehat{AUC}_2) - (AUC_1 - AUC_2))/(\widehat{var}(\widehat{AUC}_1 - \widehat{AUC}_2))^{1/2}$ is asymptotically $N(0,1)$.

Furthermore, Li and Zhou (2008) proposed a unified approach of nonparametric comparison of clustered ROC curves based on empirical ROC curve estimation. The empirical ROC curves are defined by

$$\widehat{R}_\ell(u) = \hat{S}_{\ell,D}(\hat{S}^{-1}_{\ell,\bar{D}}(u)),$$

where $\hat{S}_{\ell,D}(c) = \sum_{i=1}^{n}\sum_{j=1}^{m_{\ell i}} I(X_{\ell ij} > c)/M_\ell$ and $\hat{S}_{\ell,\bar{D}}(c) = \sum_{i=1}^{n}\sum_{k=1}^{n_{\ell i}} I(Y_{\ell ik} > c)/N_\ell$. Assume that as $n \to \infty$, $n^{-1}\sum_{i=1}^{n} n_{\ell i} \to \lambda_\ell$, and $n^{-1}\sum_{i=1}^{n} m_{\ell i} \to \gamma_\ell$ for some positive constants λ_ℓ and γ_ℓ, $\ell = 1, 2$, then

$$\sqrt{n}\begin{pmatrix} \hat{F}_{1,\bar{D}}(c) - F_{1,\bar{D}}(c) \\ \hat{F}_{2,\bar{D}}(c) - F_{2,\bar{D}}(c) \\ \hat{F}_{1,D}(c) - F_{1,D}(c) \\ \hat{F}_{2,D}(c) - F_{2,D}(c) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} W_{F_{1,\bar{D}}}(c) \\ W_{F_{2,\bar{D}}}(c) \\ W_{F_{1,D}}(c) \\ W_{F_{2,D}}(c) \end{pmatrix} \quad as \quad n \to \infty,$$

where $(W_{F_{1,\bar{D}}}(c), W_{F_{2,\bar{D}}}(c), W_{F_{1,D}}(c), W_{F_{2,D}}(c))'$ is a Gaussian processes vector with mean $\mathbf{0}$. Assume that $F_{\ell,\bar{D}}$ and $F_{\ell,D}$ are derivable and have density function $F'_{\ell,\bar{D}}$ and $F'_{\ell,D}$ respectively, then the joint limiting distribution of $(\widehat{R}_1(u), \widehat{R}_2(u))$ is given by,

$$\sqrt{n}\begin{pmatrix} \widehat{R}_1(u) \\ \widehat{R}_2(u) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z_1(1-u) \\ Z_2(1-u) \end{pmatrix} \quad as \quad n \to \infty,$$

where

$$Z_\ell(u) = -\frac{F'_{\ell,D}(F^{-1}_{\ell,\bar{D}}(u))}{F'_{\ell,\bar{D}}(F^{-1}_{\ell,\bar{D}}(u))} W_{F_{\ell,\bar{D}}}(F^{-1}_{\ell,\bar{D}}(u))) + W_{F_{\ell,D}}(F^{-1}_{\ell,\bar{D}}(u))).$$

Let $D(u) = R_1(u) - R_2(u)$, then for comparison of the areas under two ROC curves, as $n \to \infty$,

$$\sqrt{n}(\hat{D}(u) - D(u)) \xrightarrow{d} V(u) = Z_2(1-u) - Z_1(1-u),$$

where $V(u)$ is the limiting process. And the difference between the wAUCs could be estimated by the weighted integration of the two ROC curves' difference,

$$\hat{\Delta} = \int_0^1 \hat{D}(u) dW(u).$$

In summary, although research has been conducted for clustered ROC curves, they are either based on summary statistics or in the field of fixed sample studies. Understanding the sequential properties of ROC curves without relying on summary statistics will give us much flexibility in sequential study designs. Hence, there is necessity to study the sequential statistics theory in clustered ROCs on the process level and further apply the theory in group sequential designs.

1.5 Summary

In this chapter, we give brief introduction on diagnostic tests and ROC, PPV and NPV curves. We introduce the previous research conducted in the field of single sequential empirical ROC, PPV and NPV curve (Koopmeiners and Feng 2011). We also introduce comparison studies conducted in the field of correlated and clustered ROC data on the summary level (Mazumdar and Liu 2003; Zhou et al. 2008; Obuchowski 1997). We talk about correlated and clustered diagnostic data and the importance of group sequential design for comparative diagnostic accuracy studies in the field.

The contributions of the thesis are listed in the following:

- We derive asymptotic property of the sequential difference of two correlated ROC, PPV and NPV curves, and apply the theory in a group sequential method for a comparative diagnostic accuracy trial with correlated data.

- We derive asymptotic property of one sequential clustered ROC curves. We further extend the theory to the sequential difference of two clustered ROC curves, and apply the theory in group sequential designs for clustered ROC curves' comparative studies.

Chapter 2: Group Sequential Method for Comparing Correlated ROC Curves

2.1 Introduction

As introduced in Chapter 1, in a fixed-sample trial, statistical analysis is conducted after all samples' data are collected. However, data usually accumulate steadily over a period of time in a clinical trial, it is natural to analyze the results as they occur and possibly to terminate the trial early for success or futility. With a group sequential design, multiple interim analysis points and rejections boundaries are pre-determined, and we can achieve the specific power requirement with the same type I error rate, but with smaller expected sample size than a fixed-sample method.

Some research has been done in asymptotic sequential property of a single ROC curve (Koopmeiners and Feng 2011). They derived the asymptotic theory for the sequential empirical ROC curve under the case-control sampling. In this chapter, we study the properties of the difference between two correlated empirical ROC curves and present a method to sequentially compare the empirical ROC curves.

In a comparative diagnostic trial, let $X_{i,D}$ and $X_{i,\bar{D}}$ denote the outcome of the ith diagnostic test for the cases and controls, respectively with $i = 1, 2$. Suppose a larger value is more likely to indicate the disease. The cumulative distribution functions of $X_{i,D}$ and $X_{i,\bar{D}}$ are $F_{i,D}$ and $F_{i,\bar{D}}$ for the case and control populations respectively. $S_{i,D}$ and $S_{i,\bar{D}}$ are the survival functions for the case and control populations. The sensitivity and specificity are given by $S_{i,D}(c)$ and $F_{i,\bar{D}}(c)$ for a given cutoff value, c. The ROC curve for the ith diagnostic test is defined by

$$R_i(t) = S_{i,D}(S_{i,\bar{D}}^{-1}(t)), \qquad t \in [0, 1], \tag{2.1}$$

where $S^{-1}(t) = inf\{x : F(x) \geq (1-t)\}$. The ROC curve is a plot of sensitivity against 1-specificity, as the threshold value c varies. Assume that there are a total of n_D case subjects and $n_{\bar{D}}$ control subjects in the study. Suppose that we observe $X_{i,D,j} \sim F_{i,D}$, $j = 1, ..., n_D$, representing the measurements of the ith diagnostic test from n_D subjects, and $X_{i,\bar{D},j} \sim F_{i,\bar{D}}$, $j = 1, ..., n_{\bar{D}}$, the measurements of the ith diagnostic test from $n_{\bar{D}}$ subjects, for $i = 1, 2$. Assume that measurements from different subjects are independent, and measurements of tests 1 and 2 within the same subject are possibly correlated. The survival functions, $S_{i,D}, S_{i,\bar{D}}$, can be empirically estimated to yield the empirical ROC curve

$$\widehat{R}_i(t) = \hat{S}_{i,D}(\hat{S}^{-1}_{i,\bar{D}}(t)), \qquad i = 1, 2, \tag{2.2}$$

where $\hat{S}_{i,D}(t) = \sum_{j=1}^{n_D} I(X_{i,D,j} > t)/n_D$ and $\hat{S}_{i,\bar{D}}(t) = \sum_{j=1}^{n_{\bar{D}}} I(X_{i,\bar{D},j} > t)/n_{\bar{D}}$. Also, $\hat{S}^{-1}_{i,\bar{D}}(t) = inf\{x : \hat{F}_{i,\bar{D}}(x) \geq (1-t)\}$, where $\hat{F}_{i,\bar{D}}(t) = \sum_{j=1}^{n_{\bar{D}}} I(X_{i,\bar{D},j} \leq t)/n_{\bar{D}}$.

2.2 Theoretical Results for Correlated ROC Curves

We give the theoretical results about the difference of two correlated ROC curves in the following, where the theory on single ROC curve can be found in Koopmeiners and Feng (2011) and the theory on correlated ROCs can be found in Ye and Tang (2015).

Suppose we have measurements from two diagnostic tests on n_D case subjects and $n_{\bar{D}}$ control subjects, where all subjects are independent. Let $\Delta(t) = R_1(t) - R_2(t)$, $\hat{\Delta}(t) = \widehat{R_1}(t) - \widehat{R_2}(t)$, and at an interim analysis in a group sequential design when accrued case and control subjects' ratios are $r_D, r_{\bar{D}}$, we define $\hat{\Delta}_{r_D,r_{\bar{D}}}(t) = \widehat{R}_{1,r_D,r_{\bar{D}}}(t) - \widehat{R}_{2,r_D,r_{\bar{D}}}(t)$. For the sequential empirical $\Delta(t)$ at two different analysis points $(r_D, r_{\bar{D}})$ and $(r'_D, r'_{\bar{D}})$, we have vector

$$\begin{pmatrix} n_D^{-1/2}[n_D r_D](\hat{\Delta}_{r_D,r_{\bar{D}}}(t)-\Delta(t)) \\ n_D^{-1/2}[n_D r'_D](\hat{\Delta}_{r'_D,r'_{\bar{D}}}(t)-\Delta(t)) \end{pmatrix}, \tag{2.3}$$

which can be expressed in terms of the empirical $\widehat{R}$ and true ROC curves as

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} n_D^{-1/2}[n_D r_D](\widehat{R}_{1,r_D,r_{\bar{D}}}(t)-R_1(t)) \\ n_D^{-1/2}[n_D r_D](\widehat{R}_{2,r_D,r_{\bar{D}}}(t)-R_2(t)) \\ n_D^{-1/2}[n_D r'_D](\widehat{R}_{1,r'_D,r'_{\bar{D}}}(t)-R_1(t)) \\ n_D^{-1/2}[n_D r'_D](\widehat{R}_{2,r'_D,r'_{\bar{D}}}(t)-R_2(t)) \end{pmatrix}.$$

We have the random vector

$$\begin{pmatrix} n_D^{-1/2}[n_D r_D]q_{1,r_D,r_{\bar{D}}} \\ n_D^{-1/2}[n_D r_D]q_{2,r_D,r_{\bar{D}}} \\ n_D^{-1/2}[n_D r'_D]q_{1,r'_D,r'_{\bar{D}}} \\ n_D^{-1/2}[n_D r'_D]q_{2,r'_D,r'_{\bar{D}}} \end{pmatrix}, \tag{2.4}$$

where $q_{i,r_D,r_{\bar{D}}} = \widehat{R}_{i,r_D,r_{\bar{D}}}(t) - R_i(t)$, for $i=1,2$, are random variables.

As $n_D \to \infty$ and $n_{\bar{D}} \to \infty$, for any diagnostic test i, $i=1,2$, after introducing an additional term, the expressions can be rewritten as (Koopmeiners and Feng 2011)

$$\begin{aligned} & n_D^{-1/2}[n_D r_D](\widehat{R}_{i,r_D,r_{\bar{D}}}(t)-R_i(t)) \\ =& n_D^{-1/2}[n_D r_D](\hat{S}_{i,D,r_D}(\hat{S}^{-1}_{i,\bar{D},r_{\bar{D}}}(t)) - S_{i,D}(\hat{S}^{-1}_{i,\bar{D},r_{\bar{D}}}(t))) \\ &+ n_D^{-1/2}[n_D r_D](S_{i,D}(\hat{S}^{-1}_{i,\bar{D},r_{\bar{D}}}(t)) - S_{i,D}(S^{-1}_{i,\bar{D}}(t))). \end{aligned} \tag{2.5}$$

It was proved by Koopmeiners and Feng (2011) that both terms converge to Kiefer processes. A Kiefer process, K(t,r), is a two-parameter zero-mean Gaussian process in t and r with covariance: $Cov(K(t_1,r_1),K(t_2,r_2)) = (t_1 \wedge t_2 - t_1t_2)(r_1 \wedge r_2)$, where $\wedge$ represents the minimum of two operands. It behaves like a Brownian bridge in t and a Wiener process (Brownian motion) in r. From Koopmeiners and Feng (2011) we have

$$n_D^{-1/2}[n_D r_D](\hat{S}_{D,r_D}(\hat{S}^{-1}_{\bar{D},r_{\bar{D}}}(t)) - S_D(\hat{S}^{-1}_{\bar{D},r_{\bar{D}}}(t))) \xrightarrow{d} K_1(R(t), r_D), \tag{2.6}$$

which is the first term of (2.5). And for the second term of (2.5), we have by Koopmeiners and Feng (2011)

$$n_D^{-1/2}[n_D r_D](S_D(\hat{S}^{-1}_{\bar{D},r_{\bar{D}}}(t)) - S_D(S^{-1}_{\bar{D}}(t))) \xrightarrow{d} \lambda^{1/2}\frac{r_D}{r_{\bar{D}}} \cdot \frac{f_D(S^{-1}_{\bar{D}}(t))}{f_{\bar{D}}(S^{-1}_{\bar{D}}(t))} K_2(t, r_{\bar{D}}). \tag{2.7}$$

Combining the results of both terms of (2.5), from (2.6) and (2.7) it is immediate that

$$\begin{aligned} &n_D^{-1/2}[n_D r_D](\widehat{R}_{i,r_D,r_{\bar{D}}}(t) - R_i(t)) \\ &\xrightarrow{d} K_{i,1}(R_i(t), r_D) + \lambda^{1/2}\frac{r_D}{r_{\bar{D}}}\left(\frac{f_{i,D}(S^{-1}_{i,\bar{D}}(t))}{f_{i,\bar{D}}(S^{-1}_{i,\bar{D}}(t))}\right) K_{i,2}(t, r_{\bar{D}}), \end{aligned} \tag{2.8}$$

where $K_{i,1}$ and $K_{j,2}$ are independent Kiefer processes, for $i,j = \{1,2\}$ representing diagnostic test 1 and test 2 respectively.

Then we rewrite the random vector components as sums of two terms as of Equation (2.5),

$$\begin{pmatrix} n_D^{-1/2}[n_D r_D](\widehat{R}_{1,r_D,r_{\bar{D}}}(t) - R_1(t)) \\ n_D^{-1/2}[n_D r_D](\widehat{R}_{2,r_D,r_{\bar{D}}}(t) - R_2(t)) \\ n_D^{-1/2}[n_D r'_D](\widehat{R}_{1,r'_D,r'_{\bar{D}}}(t) - R_1(t)) \\ n_D^{-1/2}[n_D r'_D](\widehat{R}_{2,r'_D,r'_{\bar{D}}}(t) - R_2(t)) \end{pmatrix} \tag{2.9}$$

$$= \begin{pmatrix} n_D^{-1/2}[n_D r_D](\hat{S}_{1,D,r_D}(\hat{S}^{-1}_{1,\bar{D},r_{\bar{D}}}(t)) - S_{1,D}(\hat{S}^{-1}_{1,\bar{D},r_{\bar{D}}}(t))) \\ n_D^{-1/2}[n_D r_D](\hat{S}_{2,D,r_D}(\hat{S}^{-1}_{2,\bar{D},r_{\bar{D}}}(t)) - S_{2,D}(\hat{S}^{-1}_{2,\bar{D},r_{\bar{D}}}(t))) \\ n_D^{-1/2}[n_D r'_D](\hat{S}_{1,D,r'_D}(\hat{S}^{-1}_{1,\bar{D},r'_{\bar{D}}}(t)) - S_{1,D}(\hat{S}^{-1}_{1,\bar{D},r'_{\bar{D}}}(t))) \\ n_D^{-1/2}[n_D r'_D](\hat{S}_{2,D,r'_D}(\hat{S}^{-1}_{2,\bar{D},r'_{\bar{D}}}(t)) - S_{2,D}(\hat{S}^{-1}_{2,\bar{D},r'_{\bar{D}}}(t))) \end{pmatrix}$$

$$+ \begin{pmatrix} n_D^{-1/2}[n_D r_D](S_{1,D}(\hat{S}^{-1}_{1,\bar{D},r_{\bar{D}}}(t)) - S_{1,D}(S^{-1}_{1,\bar{D}}(t))) \\ n_D^{-1/2}[n_D r_D](S_{2,D}(\hat{S}^{-1}_{2,\bar{D},r_{\bar{D}}}(t)) - S_{2,D}(S^{-1}_{2,\bar{D}}(t))) \\ n_D^{-1/2}[n_D r'_D](S_{1,D}(\hat{S}^{-1}_{1,\bar{D},r'_{\bar{D}}}(t)) - S_{1,D}(S^{-1}_{1,\bar{D}}(t))) \\ n_D^{-1/2}[n_D r'_D](S_{2,D}(\hat{S}^{-1}_{2,\bar{D},r'_{\bar{D}}}(t)) - S_{2,D}(S^{-1}_{2,\bar{D}}(t))) \end{pmatrix},$$

by (2.6) and (2.7), we know each component converges weakly to a sum of two Kiefer processes.

$$n_D^{-1/2}[n_D r_D](\widehat{R}_{i,r_D,r_{\bar{D}}}(t) - R_i(t))$$

$$\xrightarrow{d} K_{i,1}(R_i(t), r_D) + \lambda^{1/2}\frac{r_D}{r_{\bar{D}}}\left(\frac{f_{i,D}(S^{-1}_{i,\bar{D}}(t))}{f_{i,\bar{D}}(S^{-1}_{i,\bar{D}}(t))}\right) K_{i,2}(t, r_{\bar{D}})$$

However, to prove the convergence of the vector, we will also need to prove the tightness of the left-hand side of (2.9).

Lemma 2.1. *For a multivariate stochastic process of k dimensions, if the marginal univariate stochastic processes are tight, the multivariate stochastic process is also tight.*

Recall that a probability measure P is tight if for each ϵ there exists a compact set X such that $P(X) > 1 - \epsilon$.

Proof: We will prove the lemma with 2-dimensional space. Higher dimensional cases can be proved similarly by induction. Define 2-dimensional random vector $X(t) = (X_1(t), X_2(t))^T$,

Given the condition that the marginal univariate stochastic processes are tight, then at the marginal univariate process level for each component, we have $\forall \epsilon > 0, \exists M_1, M_2$, such that

$$P(\sup_t |X_1(t)| \le M_1) \ge 1 - \epsilon/2, \tag{2.10}$$

and

$$P(\sup_t |X_2(t)| \le M_2) \ge 1 - \epsilon/2. \tag{2.11}$$

Let $M = \max(M_1, M_2)$. We have on the multivariate process level,

$$
\begin{aligned}
&P\left((\sup_t |X_1(t)| \le M) \bigcap (\sup_t |X_2(t)| \le M)\right) \\
=&P(\sup_t |X_1(t)| \le M) + P(\sup_t |X_2(t)| \le M) - P\left((\sup_t |X_1(t)| \le M) \bigcup (\sup_t |X_2(t)| \le M)\right) \\
\ge&(1-\epsilon/2) + (1-\epsilon/2) - 1 \\
=&1-\epsilon,
\end{aligned}
$$

due to the inequalities of (2.10) and (2.11). This proves the multivariate process is tight by definition.

By Lemma 2.1 and Cramér-Wold device (Karr 1993), we can show that the finite dimensional distribution of (2.9) converges in distribution to a multivariate normal distribution, here without loss of generality assuming a vector dimension of four, and that the process on the left-hand side of (2.9) is tight. Furthermore, we know that the vector of (2.9)

$$
\begin{pmatrix}
n_D^{-1/2}[n_D r_D](\widehat{R}_{1,r_D,r_{\bar{D}}}(t) - R_1(t)) \\
n_D^{-1/2}[n_D r_D](\widehat{R}_{2,r_D,r_{\bar{D}}}(t) - R_2(t)) \\
n_D^{-1/2}[n_D r'_D](\widehat{R}_{1,r'_D,r'_{\bar{D}}}(t) - R_1(t)) \\
n_D^{-1/2}[n_D r'_D](\widehat{R}_{2,r'_D,r'_{\bar{D}}}(t) - R_2(t))
\end{pmatrix}
$$

$$\xrightarrow{d} \begin{pmatrix} K_{1,1}(R_1(t), r_D) \\ K_{2,1}(R_2(t), r_D) \\ K_{1,1}(R_1(t), r'_D) \\ K_{2,1}(R_2(t), r'_D) \end{pmatrix} + \begin{pmatrix} \lambda^{1/2} \frac{r_D}{r_{\bar{D}}} \left(\frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))} \right) K_{1,2}(t, r_{\bar{D}}) \\ \lambda^{1/2} \frac{r_D}{r_{\bar{D}}} \left(\frac{f_{2,D}(S^{-1}_{2,\bar{D}}(t))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t))} \right) K_{2,2}(t, r_{\bar{D}}) \\ \lambda^{1/2} \frac{r'_D}{r'_{\bar{D}}} \left(\frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))} \right) K_{1,2}(t, r'_{\bar{D}}) \\ \lambda^{1/2} \frac{r'_D}{r'_{\bar{D}}} \left(\frac{f_{2,D}(S^{-1}_{2,\bar{D}}(t))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t))} \right) K_{2,2}(t, r'_{\bar{D}}) \end{pmatrix},$$

uniformly for $t \in [a, b]$, $r_D \in [c, 1]$, and $r_{\bar{D}} \in [d, 1]$ where $K_{i,1}$ and $K_{j,2}$ are independent Kiefer processes, for $i, j = \{1, 2\}$. Thus the random vector (2.4) is approximately multivariate normal with covariance as derived in the following. The asymptotic covariance matrix of (2.4) is $\Sigma = \{a_{ij}\}_{i=1,\cdots,4;\, j=1,\cdots,4}$, where

$$a_{11} = Var(K_{1,1}(R_1(t), r_D)) + Var\left(\lambda^{1/2} \frac{r_D}{r_{\bar{D}}} \left(\frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))} \right) K_{1,2}(t, r_{\bar{D}}) \right)$$

$$= r_D(R_1(t) - R_1^2(t)) + \lambda \frac{r_D^2}{r_{\bar{D}}} \left(\frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))} \right)^2 (t - t^2),$$

$$\begin{aligned} a_{12} = {} & Cov\Big(n_D^{-1/2}[n_D r_D](\hat{S}_{1,D,r_D}(\hat{S}^{-1}_{1,\bar{D},r_{\bar{D}}}(t)) - S_{1,D}(\hat{S}^{-1}_{1,\bar{D},r_{\bar{D}}}(t))), \\ & \quad n_D^{-1/2}[n_D r_D](\hat{S}_{2,D,r_D}(\hat{S}^{-1}_{2,\bar{D},r_{\bar{D}}}(t)) - S_{2,D}(\hat{S}^{-1}_{2,\bar{D},r_{\bar{D}}}(t)))\Big) \\ & + Cov\Big(n_D^{-1/2}[n_D r_D](S_{1,D}(\hat{S}^{-1}_{1,\bar{D},r_{\bar{D}}}(t)) - S_{1,D}(S^{-1}_{1,\bar{D}}(t))), \\ & \quad n_D^{-1/2}[n_D r_D](S_{2,D}(\hat{S}^{-1}_{2,\bar{D},r_{\bar{D}}}(t)) - S_{2,D}(S^{-1}_{2,\bar{D}}(t)))\Big), \end{aligned}$$

then expanding the empirical survival functions by definitions, we obtain

$$Cov\Big(n_D^{-1/2} \sum_{i=1}^{[n_D r_D]} \Big(I(X_{1,D,i} > \hat{S}^{-1}_{1,\bar{D},r_{\bar{D}}}(t)) - S_{1,D}(\hat{S}^{-1}_{1,\bar{D},r_{\bar{D}}}(t)) \Big),$$

$$n_D^{-1/2}\sum_{i=1}^{[n_D r_D]}\Big(I(X_{2,D,i} > \hat{S}^{-1}_{2,\bar{D},r_{\bar{D}}}(t)) - S_{2,D}(\hat{S}^{-1}_{2,\bar{D},r_{\bar{D}}}(t))\Big)\Big)$$

$$+ Cov\Big(n_D^{-1/2}[n_D r_D](S_{1,D}(\hat{S}^{-1}_{1,\bar{D},r_{\bar{D}}}(t)) - S_{1,D}(S^{-1}_{1,\bar{D}}(t))),$$

$$n_D^{-1/2}[n_D r_D](S_{2,D}(\hat{S}^{-1}_{2,\bar{D},r_{\bar{D}}}(t)) - S_{2,D}(S^{-1}_{2,\bar{D}}(t)))\Big)$$

$$\xrightarrow{d} r_D(S_D(S^{-1}_{1,\bar{D}}(t), S^{-1}_{2,\bar{D}}(t)) - R_1(t)R_2(t)) \tag{2.12}$$

$$+ \lambda\frac{r_D^2}{r_{\bar{D}}}\frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))}\frac{f_{2,D}(S^{-1}_{2,\bar{D}}(t))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t))}(S_{\bar{D}}(S^{-1}_{1,\bar{D}}(t), S^{-1}_{2,\bar{D}}(t)) - t^2).$$

With regard to the previous step, the first term of (2.12) is derived using the sequential empirical process result of section 2.12.1 (van der Vaart and Wellner 1996). For the second term of (2.12), first we derive the following equation by expanding the empirical survival function and then apply the same result of van der Vaart and Wellner (1996)

$$Cov\left(n_{\bar{D}}^{-1/2}[n_{\bar{D}} r_{\bar{D}}](\hat{S}_{1,\bar{D},r_{\bar{D}}}(t_1) - S_{1,\bar{D}}(t_1)), n_{\bar{D}}^{-1/2}[n_{\bar{D}} r_{\bar{D}}](\hat{S}_{2,\bar{D},r_{\bar{D}}}(t_2) - S_{2,\bar{D}}(t_2))\right)$$

$$= Cov\left(n_{\bar{D}}^{-1/2}\sum_{i=1}^{[n_{\bar{D}} r_{\bar{D}}]}\left(I(X_{1,\bar{D},i} > t_1) - S_{1,\bar{D}}(t_1)\right), n_{\bar{D}}^{-1/2}\sum_{i=1}^{[n_{\bar{D}} r_{\bar{D}}]}\left(I(X_{2,\bar{D},i} > t_2) - S_{2,\bar{D}}(t_2)\right)\right)$$

$$\xrightarrow{d} r_{\bar{D}}\left(S_{\bar{D}}(t_1, t_2) - S_{1,\bar{D}}(t_1)S_{2,\bar{D}}(t_2)\right). \tag{2.13}$$

Then by Equation (2.13), Lemma 3.9.20 in van der Vaart and Wellner (1996), and Theorem 3.9.4 in van der Vaart and Wellner (1996), we prove the second term of (2.12) in the following

$$Cov\big(n_D^{-1/2}[n_D r_D](S_{1,D}(\hat{S}^{-1}_{1,\bar{D},r_{\bar{D}}}(t)) - S_{1,D}(S^{-1}_{1,\bar{D}}(t))),$$

$$n_D^{-1/2}[n_D r_D](S_{2,D}(\hat{S}^{-1}_{2,\bar{D},r_{\bar{D}}}(t)) - S_{2,D}(S^{-1}_{2,\bar{D}}(t)))\big)$$

$$= n_{\bar{D}}^{-1}[n_D r_D]^2 n_{\bar{D}}[n_{\bar{D}} r_{\bar{D}}]^{-2} Cov\Big(n_{\bar{D}}^{-1/2}[n_{\bar{D}} r_{\bar{D}}](S_{1,D}(\hat{S}^{-1}_{1,\bar{D},r_{\bar{D}}}(t)) - S_{1,D}(S^{-1}_{1,\bar{D}}(t))),$$

$$n_{\bar{D}}^{-1/2}[n_{\bar{D}} r_{\bar{D}}](S_{2,D}(\hat{S}^{-1}_{2,\bar{D},r_{\bar{D}}}(t)) - S_{2,D}(S^{-1}_{2,\bar{D}}(t)))\Big)$$

$$\xrightarrow{d} \lambda \frac{r_D^2}{r_{\bar{D}}} \frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))} \frac{f_{2,D}(S^{-1}_{2,\bar{D}}(t))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t))} (S_{\bar{D}}(S^{-1}_{1,\bar{D}}(t), S^{-1}_{2,\bar{D}}(t)) - S_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t)) S_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t))$$

$$= \lambda \frac{r_D^2}{r_{\bar{D}}} \frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))} \frac{f_{2,D}(S^{-1}_{2,\bar{D}}(t))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t))} (S_{\bar{D}}(S^{-1}_{1,\bar{D}}(t), S^{-1}_{2,\bar{D}}(t)) - t^2),$$

this concludes the derivation for element a_{12} as shown in Equation(2.12). For the other elements in the asymptotic covariance matrix Σ,

$$a_{13} = Cov\left(K_{1,1}(R_1(t), r_D), K_{1,1}(R_1(t), r'_D)\right)$$

$$+ Cov\left(\lambda^{1/2} \frac{r_D}{r_{\bar{D}}} \left(\frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))}\right) K_{1,2}(t, r_{\bar{D}}), \lambda^{1/2} \frac{r'_D}{r'_{\bar{D}}} \left(\frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))}\right) K_{1,2}(t, r'_{\bar{D}})\right)$$

$$= (r_D \wedge r'_D)(R_1(t) - R_1^2(t)) + (r_{\bar{D}} \wedge r'_{\bar{D}}) \lambda \frac{r_D}{r_{\bar{D}}} \frac{r'_D}{r'_{\bar{D}}} \left(\frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))}\right)^2 (t - t^2),$$

from the covariance structure of Kiefer processes. And

$$a_{14} = Cov\Big(n_D^{-1/2}[n_D r_D](\hat{S}_{1,D,r_D}(\hat{S}^{-1}_{1,\bar{D},r_{\bar{D}}}(t)) - S_{1,D}(\hat{S}^{-1}_{1,\bar{D},r_{\bar{D}}}(t))),$$

$$n_D^{-1/2}[n_D r'_D](\hat{S}_{2,D,r'_D}(\hat{S}^{-1}_{2,\bar{D},r'_{\bar{D}}}(t)) - S_{2,D}(\hat{S}^{-1}_{2,\bar{D},r'_{\bar{D}}}(t)))\Big)$$

$$+ Cov\Big(n_D^{-1/2}[n_D r_D](S_{1,D}(\hat{S}^{-1}_{1,\bar{D},r_{\bar{D}}}(t)) - S_{1,D}(S^{-1}_{1,\bar{D}}(t))),$$

$$n_D^{-1/2}[n_D r'_D](S_{2,D}(\hat{S}^{-1}_{2,\bar{D},r'_{\bar{D}}}(t)) - S_{2,D}(S^{-1}_{2,\bar{D}}(t)))\Big)$$

$$\xrightarrow{d} (r_D \wedge r'_D)(S_D(S^{-1}_{1,\bar{D}}(t), S^{-1}_{2,\bar{D}}(t)) - R_1(t) R_2(t))$$

$$+ (r_{\bar{D}} \wedge r'_{\bar{D}})\lambda \frac{r_D}{r_{\bar{D}}} \frac{r'_D}{r'_{\bar{D}}} \frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))} \frac{f_{2,D}(S^{-1}_{2,\bar{D}}(t))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t))} (S_{\bar{D}}(S^{-1}_{1,\bar{D}}(t), S^{-1}_{2,\bar{D}}(t)) - t^2).$$

The derivation of a_{14} is the same as a_{12} except that when applying the sequential empirical process result of section 2.12.1, (van der Vaart and Wellner 1996), we have to include $r_D \wedge r'_D$ and $r_{\bar{D}} \wedge r'_{\bar{D}}$ terms.

Similarly, we can get the following elements of the covariance matrix.

$$a_{22} = r_D(R_2(t) - R_2^2(t)) + \lambda \tfrac{r_D^2}{r_{\bar{D}}} \left(\tfrac{f_{2,D}(S^{-1}_{2,\bar{D}}(t))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t))} \right)^2 (t - t^2),$$

$$a_{23} = a_{14},$$

$$a_{24} = (r_D \wedge r'_D)(R_2(t) - R_2^2(t)) + (r_{\bar{D}} \wedge r'_{\bar{D}})\lambda \tfrac{r_D}{r_{\bar{D}}} \tfrac{r'_D}{r'_{\bar{D}}} \left(\tfrac{f_{2,D}(S^{-1}_{2,\bar{D}}(t))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t))} \right)^2 (t - t^2),$$

$$a_{33} = r'_D(R_1(t) - R_1^2(t)) + \lambda \tfrac{r'^2_D}{r'_{\bar{D}}} \left(\tfrac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))} \right)^2 (t - t^2),$$

$$a_{34} = r'_D(S_D(S^{-1}_{1,\bar{D}}(t), S^{-1}_{2,\bar{D}}(t)) - R_1(t)R_2(t))$$

$$+ \lambda \frac{r'^2_D}{r'_{\bar{D}}} \frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))} \frac{f_{2,D}(S^{-1}_{2,\bar{D}}(t))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t))} (S_{\bar{D}}(S^{-1}_{1,\bar{D}}(t), S^{-1}_{2,\bar{D}}(t)) - t^2),$$

and

$$a_{44} = r'_D(R_2(t) - R_2^2(t)) + \lambda \frac{r'^2_D}{r'_{\bar{D}}} \left(\frac{f_{2,D}(S^{-1}_{2,\bar{D}}(t))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t))} \right)^2 (t - t^2).$$

Hence the random vector of (2.3) is approximately normal with covariance matrix derived approximately as $A\Sigma A^T$, where Σ is the asymptotic covariance matrix of (2.4) and $A =$

$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$. The approximate covariance matrix of (2.3)

$$Cov \begin{pmatrix} n_D^{-1/2}[n_D r_D](\hat{\Delta}_{r_D, r_{\bar{D}}}(t) - \Delta(t)) \\ n_D^{-1/2}[n_D r'_D](\hat{\Delta}_{r'_D, r'_{\bar{D}}}(t) - \Delta(t)) \end{pmatrix}$$

$$\xrightarrow{d} \begin{pmatrix} a_{11} + a_{22} - 2a_{12} & a_{13} + a_{24} - 2a_{14} \\ a_{13} + a_{24} - 2a_{14} & a_{33} + a_{44} - 2a_{34} \end{pmatrix}.$$

Without the loss of generality, let $r'_D \geq r_D$ and $r'_{\bar{D}} \geq r_{\bar{D}}$, that is, the time point of $(r'_D, r'_{\bar{D}})$ comes after $(r_D, r_{\bar{D}})$. Approximately,

$$Cov(\hat{\Delta}_{r_D, r_{\bar{D}}}(t), \hat{\Delta}_{r'_D, r'_{\bar{D}}}(t)) = Cov(\hat{\Delta}_{r_D, r_{\bar{D}}}(t) - \Delta(t), \hat{\Delta}_{r'_D, r'_{\bar{D}}}(t) - \Delta(t))$$

$$= n_D \frac{1}{n_D r_D} \frac{1}{n_D r'_D} (a_{13} + a_{24} - 2a_{14}).$$

And the variance,

$$Var(\hat{\Delta}_{r'_D, r'_{\bar{D}}}(t)) = Var(\hat{\Delta}_{r'_D, r'_{\bar{D}}}(t) - \Delta(t))$$

$$= n_D \frac{1}{n_D r'_D} \frac{1}{n_D r'_D} (a_{33} + a_{44} - 2a_{34}).$$

It can be shown that

$$Cov(\hat{\Delta}_{r_D, r_{\bar{D}}}(t), \hat{\Delta}_{r'_D, r'_{\bar{D}}}(t)) = Var(\hat{\Delta}_{r'_D, r'_{\bar{D}}}(t)) \tag{2.14}$$

$$\xrightarrow{d} \frac{1}{n_D r'_D}(R_1(t) - R_1^2(t)) + \frac{1}{n_{\bar{D}} r'_{\bar{D}}} \left(\frac{f_{1,D}(S_{1,\bar{D}}^{-1}(t))}{f_{1,\bar{D}}(S_{1,\bar{D}}^{-1}(t))} \right)^2 (t - t^2)$$

$$+ \frac{1}{n_D r'_D}(R_2(t) - R_2^2(t)) + \frac{1}{n_{\bar{D}} r'_{\bar{D}}} \left(\frac{f_{2,D}(S_{2,\bar{D}}^{-1}(t))}{f_{2,\bar{D}}(S_{2,\bar{D}}^{-1}(t))} \right)^2 (t - t^2)$$

$$
- 2\frac{1}{n_D r'_D}(S_D(S^{-1}_{1,\bar{D}}(t), S^{-1}_{2,\bar{D}}(t)) - R_1(t)R_2(t))
$$

$$
- 2\frac{1}{n_{\bar{D}} r'_{\bar{D}}}\frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))}\frac{f_{2,D}(S^{-1}_{2,\bar{D}}(t))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t))}(S_{\bar{D}}(S^{-1}_{1,\bar{D}}(t), S^{-1}_{2,\bar{D}}(t)) - t^2),
$$

for $r'_D \geq r_D$ and $r'_{\bar{D}} \geq r_{\bar{D}}$.

The method above deals only two sequential analysis points and their asymptotic properties. The exact method can be applied to any finite set of sequential analysis points as shown below assuming the number of interim analysis is J.

$$
\begin{pmatrix}
n_D^{-1/2}[n_D r_{D,1}](\hat{\Delta}_{r_{D,1}, r_{\bar{D},1}}(t_1) - \Delta(t_1)) \\
n_D^{-1/2}[n_D r_{D,2}](\hat{\Delta}_{r_{D,2}, r_{\bar{D},2}}(t_2) - \Delta(t_2)) \\
\vdots \\
n_D^{-1/2}[n_D r_{D,J}](\hat{\Delta}_{r_{D,J}, r_{\bar{D},J}}(t_J) - \Delta(t_J))
\end{pmatrix},
$$

which can be expressed in terms of the empirical $\widehat{ROC}$ and true ROC curves as

$$
\begin{pmatrix}
1 & -1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & -1
\end{pmatrix}
\begin{pmatrix}
n_D^{-1/2}[n_D r_{D,1}](\widehat{R}_{1,r_{D,1}, r_{\bar{D},1}}(t_1) - R_1(t_1)) \\
n_D^{-1/2}[n_D r_{D,1}](\widehat{R}_{2,r_{D,1}, r_{\bar{D},1}}(t_1) - R_2(t_1)) \\
\vdots \\
n_D^{-1/2}[n_D r_{D,J}](\widehat{R}_{1,r_{D,J}, r_{\bar{D},J}}(t_J) - R_1(t_J)) \\
n_D^{-1/2}[n_D r_{D,J}](\widehat{R}_{2,r_{D,J}, r_{\bar{D},J}}(t_J) - R_2(t_J))
\end{pmatrix}.
$$

Following the same steps, we will come to the same results with the asymptotic properties of independent increments covariance structure for any finite interim analysis.

The estimated ROC curves has interesting joint asymptotic properties at the process level as indicated above. We then would be able to analyze ROC curves at different FPRs, say $R_1(t_1), R_2(t_2)$. We can do analysis of two ROC curves at multiple points, since they all follow multivariate normal distribution with the variance-covariance stated before.

Furthermore, we can compare multiple points of ROC curves based on weighted average. It can be shown that the sequential weighted average of $\hat{\Delta}(t)$ on several FPRs has similar feature of asymptotic multivariate normality and its asymptotic covariance matrix is given by $Cov(\sum_{i=1}^{K} \omega_i \hat{\Delta}_{r_D, r_{\bar{D}}}(t_i), \sum_{i=1}^{K} \omega_i \hat{\Delta}_{r'_D, r'_{\bar{D}}}(t_i)) = Var(\sum_{i=1}^{K} \omega_i \hat{\Delta}_{r'_D, r'_{\bar{D}}}(t_i))$, where ω_i is the weight on $\hat{\Delta}_{r_D, r_{\bar{D}}}(t_i)$ with $\sum_{i=1}^{K} \omega_i = 1$. This is due to the fact that $Cov(\hat{\Delta}_{r_D, r_{\bar{D}}}(t_i), \hat{\Delta}_{r'_D, r'_{\bar{D}}}(t_j))$= $Cov(\hat{\Delta}_{r'_D, r'_{\bar{D}}}(t_i), \hat{\Delta}_{r'_D, r'_{\bar{D}}}(t_j))$ for $r'_D \geq r_D$ and $r'_{\bar{D}} \geq r_{\bar{D}}$.

To carry out a group sequential test, we analyze the accumulating data in groups rather than after an additional observation as in a fully sequential test or after all data is collected as in a fixed sample test. A group sequential design (GSD) is convenient to conduct and provide an opportunity for stopping the trial earlier than planned. It achieves the goals of lower expected sample sizes and shorter average study lengths. GSD methods utilize different strategies of allocating the overall type I error probability.

From the previous theorem, we know that the sequential empirical difference of two ROC curves is also a Gaussian process. The sequential empirical difference at any finite set of analysis points follow a multivariate normal distribution. And the sequential score statistic has an "independent increment" covariance structure (Jennison and Turnbull 2000), which facilitates the sequential comparison of ROC curves and any standard GSD software can be readily applied.

Suppose we are interested in a two-sided test with the hypothesis of $H_0 : R_1(t_0) - R_2(t_0) = 0$ on a particular FPR t_0, and $H_a : R_1(t_0) - R_2(t_0) \neq 0$. Let $\Delta(t_0) = R_1(t_0) - R_2(t_0)$, and $\hat{\Delta}(t_0) = \widehat{R_1}(t_0) - \widehat{R_2}(t_0)$. Then under H_0, we can do the Z-test with the statistic $Z = \frac{\hat{\Delta}(t_0)}{\sqrt{Var(\hat{\Delta}(t_0))}}$. And for a fixed sample test we reject H_0 if $|Z| > Z_{\alpha/2}$. However, suppose we will do the GSD with a sampling plan of J interim analyses. At the jth analysis, test results are available on the first $n_D r_D^{(j)}$ case subjects and the first $n_{\bar{D}} r_{\bar{D}}^{(j)}$ control subjects, where n_D and $n_{\bar{D}}$ are the maximum case and control sample size respectively, and $r_D^{(j)}$ and $r_{\bar{D}}^{(j)}$ are the ratios of the case and control subjects accrued so far at jth analysis. Given type I error rate α and power $1-\beta$ at $\Delta(t_0) = \pm\delta$, the fixed sample size is calculated based on $\alpha, \beta, \delta, and Var(\hat{\Delta}(t_0))$. The maximum sample size for the GSD are proportional to the fixed sample size, and this ratio $R(J, \alpha, \beta)$ depends only on J, α, β and the particular GSD method used.

Consider a GSD plan involving up to J analyses of sample data. At the time of the jth analysis, let $I_j = 1/\sigma^2_{\hat{\Delta}_j(t)}$, $\tau_j = I_j/I_J = \sigma^2_{\hat{\Delta}_J(t)}/\sigma^2_{\hat{\Delta}_j(t)}$. Define $B(\tau_j) = \sqrt{\tau_j I_j}\hat{\Delta}_j(t)$. For $j < k$, $Cov(B(\tau_j), B(\tau_k)) = \tau_j$. This can be proved using the previous finding of Equation 2.14. Thus $B(\tau_j)$ behaves asymptotically like a Brownian motion process. Then the standard GSD software like R package gsDesign can be readily applied. Similarly, we can apply the transformation on the sequential weighted average of $\hat{\Delta}(t)$ on several FPRs and come up with the same conclusion. The transformation used is $I_j = 1/Var(\sum_{i=1}^K \omega_i \hat{\Delta}_j(t_i))$, $\tau_j = I_j/I_J = Var(\sum_{i=1}^K \omega_i \hat{\Delta}_J(t_i))/Var(\sum_{i=1}^K \omega_i \hat{\Delta}_j(t_i))$. Define $B(\tau_j) = \sqrt{\tau_j I_j}(\sum_{i=1}^K \omega_i \hat{\Delta}_j(t_i))$. Then for $j < k$, again we have $Cov(B(\tau_j), B(\tau_k)) = \tau_j$.

The GSD needs to be specified and the maximum sample sizes need to be determined before conducting the trial. At the first interim analysis, we calculate the Z test statistic based on the empirical estimation of $R_1(t_0)$, $R_2(t_0)$ and $Var(\hat{\Delta}(t_0))$. We compare the Z statistic to the boundaries of Pocock, O'Brien-Flemming, or error spending method that are calculated to control Type I error rate. The boundaries a_j are defined to control the overall type I error rate: $P(|Z_j| > a_j \mid \Delta(t_0) = 0)$ for some $j = 1...J$. If this Z statistic falls in the rejection boundaries, we then reject the null hypothesis, and the clinical trial is stopped with null hypothesis rejection and no more subjects will be accrued. Otherwise, we will continue accruing sufficient subjects to be able to proceed to the next analysis point. At the jth analysis, the first $n_D r_D^{(j)}$ case subjects and the first $n_{\bar{D}} r_{\bar{D}}^{(j)}$ control subjects are used to compute the interim statistic Z_j. We will repeat the process until the last Jth analysis point. At the last analysis, we will either reject the null hypothesis or accept it and stop the clinical trial.

The previous findings and method can also be used to obtain the properties of the sequential wAUC or AUC statistics. Such an extension can be done in comparing summary statistics of two ROC curves through the integration of $\Delta(t)$ from 0 to 1 with regarding to any given weight probability measure function. The AUC and pAUC statistic become special cases, as indicated in Tang et al. (2008). More importantly, because of the results in equation (2.14), we can compare a wide range of ROC summary measures, including curves at different FPRs or their weighted averages of the ROC curves.

2.3 Simulation Studies

2.3.1 Consistency of Covariance Matrix Estimator

We conduct a simulation study to assess the finite sample properties of the results in Theorem 2.14. Diagnostic test data are drawn from bivariate normal distributions. For a case, the bivariate normal model is $(X_1, X_2)^T \sim N\{(10, 6)^T, \Sigma_1\}$, and for a control, the bivariate normal model is $(Y_1, Y_2)^T \sim N\{(0, 4)^T, \Sigma_2\}$, where

$$\Sigma_1 = \begin{pmatrix} 2 & \rho 2\sqrt{2} \\ \rho 2\sqrt{2} & 4 \end{pmatrix} \quad and \quad \Sigma_2 = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \qquad with\ \rho = 0.5\,.$$

We conduct 5000 simulation with $n_D = 200, n_{\bar{D}} = 200$, and for the simulated data, we calculate the variance-covariance of the $\Delta(t)$ at various combinations of $r_D, r_{\bar{D}}$ with t=0.5. Here, the ROC curves are estimated with the empirical functions. Then we compare the simulated covariance matrix to the theoretical covariance matrix derived using the results of Theorem 2.14. Table 2.1 shows that observed variance-covariance values are very close to theoretical values when sample sizes are sufficiently large.

Table 2.1: The values of elements ($\times 10^{-3}$) in observed and theoretical covariance matrix

	Observed covariance matrix				Theoretical covariance matrix			
	$n_D = 200$, $n_{\bar{D}} = 200$							
$\Delta_{0.2,0.3}(0.5)$	3.718	1.850	1.458	0.755	3.720	1.898	1.499	0.782
$\Delta_{0.4,0.5}(0.5)$		1.927	1.490	0.773		1.898	1.499	0.782
$\Delta_{0.5,0.7}(0.5)$			1.529	0.790			1.499	0.782
$\Delta_{1,1}(0.5)$				0.787				0.782

2.3.2 Simulated Type I Error Rate in GSDs

To investigate finite sample performance of the GSD procedure, we conduct a simulation study in a two-group sequential test (J=2), and a five-group sequential test (J=5). The null hypothesis of equal ROC(t) is set to be true and the nominal type I error rate was set to be $\alpha = 0.05$ for two-sided tests. Two set of diagnostic test data are simulated from bivariate normal (Binorm) and bivariate lognormal(Bilognorm) models. The bivariate normal models is $(X_1, X_2)^T \sim N\{(1, 10)^T, \Sigma_1\}$ for the case data. And for the control data, the bivariate normal model is $(Y_1, Y_2)^T \sim N\{(0, 8)^T, \Sigma_2\}$, where

$$\Sigma_1 = \begin{pmatrix} 1 & 2\rho \\ 2\rho & 4 \end{pmatrix} \qquad \Sigma_2 = \begin{pmatrix} 1 & 2\rho \\ 2\rho & 4 \end{pmatrix} \qquad with\ \rho = (0, 0.25, 0.5, 0.75, 0.9)$$

In this case, the ROC curves are identical from the formula of ROC curve under bi-normal models (Zhou et al. 2011): $R(t) = \Phi(a + b\Phi^{-1}(t))$, where $a = (\mu_1 - \mu_0)/\sigma_1$ and $b = \sigma_0/\sigma_1$, (μ_1, σ_1) *and* (μ_0, σ_0) are the normal parameters in the case and control groups. The bivariate lognormal data are generated by taking exponential of the simulated bivariate normal data. Because the ROC curves are invariant to a monotone transformation, the ROC curves under the bivariate lognormal models are also identical. The diagnostic tests distribution comparison and ROC graph are shown in Figure 2.1. Different numbers of case and control subjects, $n_D, n_{\bar{D}} = (50, 250, 500)$, are considered in our simulation study.

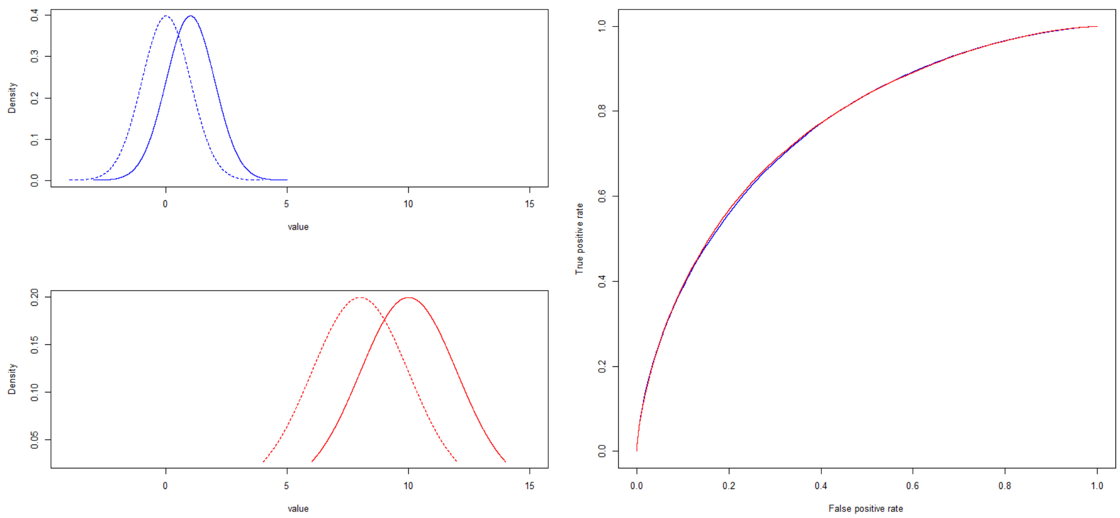

Figure 2.1: Two correlated identical ROC curves

For each simulation setting, 5000 random data sets are generated and the GSD method applied to the simulated data. The Z statistics at each interim analysis point are then calculated based on the empirical ROC difference and estimated variances. The GSD test procedure compares the Z statistics with corresponding test boundaries of design, and the decision of rejection or failing to rejection is obtained for each simulated dataset. We then calculate the overall rejection rates for all simulated datasets. Table 2.2 gives the rejection rates of all different model and sample size combinations with a nominal α level 0.05 under the O'Brien and Fleming's criterion. And Table 2.3 is the results for the Pocock's criterion. As we can see, the simulated Type I error rates are close to the nominal rate and tend to be closer as the overall sample sizes increase. The type I error rates are also plotted in Figure 2.2 and Figure 2.3. In these figures, the type I error rates are plotted as bars showing their deviations from the nominal rate of 0.05 which is the vertical line.

We take the same two identical ROC curves as mentioned above and the null hypothesis of H_0: $\sum_{t=\{0.2,0.5,0.8\}} \Delta(t)/3 = 0$ as an example for the sequential weighted average test. For the simulation with $n_D = 250, n_{\bar{D}} = 250$ and $J = 5$, we get the type I error rates

Table 2.2: Type I error rates ($\times 10^{-2}$) using the O'Brien-Fleming GSD with $\alpha = 0.05$

			$\rho = 0$		$\rho = 0.25$		$\rho = 0.5$		$\rho = 0.75$		$\rho = 0.9$	
n_D	$n_{\bar{D}}$	t	Binorm	Bilog	Binorm	Bilog	Binorm	Bilog	Binorm	Bilog	Binorm	Bilog
			Two-group sequential design (J=2)									
50	50	0.2	6.94	10.96	7.02	10.30	6.20	9.74	5.98	8.60	4.26	5.80
		0.4	5.38	8.98	4.90	8.34	4.36	7.44	4.08	6.80	3.38	5.26
		0.5	3.86	9.14	4.14	8.78	3.86	8.74	3.60	6.90	2.60	4.64
		0.6	3.84	9.00	3.74	7.94	3.18	7.22	2.52	5.72	1.64	3.56
		0.8	1.58	3.62	1.56	3.46	1.24	2.58	0.82	1.80	0.42	0.62
250	250	0.2	6.10	7.38	5.72	7.00	5.28	6.42	5.00	5.86	5.54	5.84
		0.4	4.36	5.56	4.24	5.60	4.42	5.58	4.72	5.92	4.58	5.62
		0.5	3.98	6.96	4.70	6.80	4.62	7.14	4.74	7.18	4.44	6.18
		0.6	4.32	7.88	4.08	7.14	4.74	7.78	4.06	6.60	4.02	6.62
		0.8	3.52	6.20	3.60	5.90	3.86	5.90	3.02	5.06	2.68	3.90
250	500	0.2	5.30	6.04	5.36	5.82	5.24	5.70	5.52	5.42	5.34	5.46
		0.4	4.72	5.42	5.36	6.04	5.54	5.92	5.18	5.74	4.74	5.28
		0.5	5.10	6.64	4.84	6.04	5.48	6.96	4.88	6.38	4.82	5.88
		0.6	4.86	6.44	5.32	7.20	4.88	6.58	5.08	6.70	4.78	6.14
		0.8	3.80	5.80	3.88	5.32	3.72	5.28	4.10	5.42	3.60	4.24
500	500	0.2	5.44	7.00	5.68	7.04	5.32	6.60	5.18	5.80	5.42	5.62
		0.4	5.14	6.10	5.00	5.62	4.56	5.28	4.66	5.08	4.48	4.58
		0.5	4.52	6.08	4.76	5.94	5.32	6.36	4.38	5.78	4.70	5.72
		0.6	4.30	7.14	4.84	7.36	4.36	6.68	4.40	6.08	4.60	6.46
		0.8	4.14	6.84	4.18	6.42	4.00	6.16	3.86	5.50	3.84	5.02
			Five-group sequential design (J=5)									
50	50	0.2	8.06	13.94	8.06	12.98	7.16	11.22	6.42	10.26	4.84	6.96
		0.4	5.90	9.86	5.00	9.38	4.80	8.42	4.44	7.36	3.38	5.50
		0.5	4.26	9.82	4.28	9.42	4.20	9.56	3.60	7.58	2.46	4.44
		0.6	3.68	9.80	3.66	8.84	3.20	7.82	2.52	5.68	1.54	3.24
		0.8	1.38	3.20	1.34	2.80	0.86	2.08	0.68	1.30	0.30	0.46
250	250	0.2	6.64	8.18	6.38	7.90	5.84	7.62	5.70	6.58	5.74	6.42
		0.4	4.54	6.12	4.76	6.24	4.54	6.16	5.08	6.36	4.84	6.08
		0.5	4.38	7.70	4.84	7.60	5.10	7.90	4.90	7.32	4.42	6.52
		0.6	4.68	8.66	4.18	7.66	5.12	8.46	4.34	7.52	4.36	6.88
		0.8	3.32	6.34	3.46	6.04	3.82	6.18	3.04	4.84	2.26	3.60
250	500	0.2	5.58	6.48	5.60	6.38	5.76	6.20	5.72	5.72	5.42	5.68
		0.4	4.86	5.72	5.80	6.38	5.64	6.36	5.48	6.00	4.82	5.50
		0.5	5.36	7.12	5.18	6.78	5.66	7.40	5.32	6.74	5.00	6.34
		0.6	4.76	7.08	5.18	7.40	5.14	7.18	5.32	7.24	4.98	6.32
		0.8	3.90	5.86	3.68	5.76	3.92	5.58	4.00	5.86	3.06	4.56
500	500	0.2	5.60	7.70	5.80	7.58	5.56	6.98	5.48	6.44	5.84	6.16
		0.4	5.32	6.16	5.12	5.88	4.84	5.52	4.74	5.36	4.50	4.88
		0.5	4.70	6.38	4.76	6.30	5.46	7.04	4.62	6.20	4.82	6.20
		0.6	4.36	7.36	5.00	7.78	4.62	7.42	4.56	6.70	4.62	6.74
		0.8	4.04	7.24	4.28	6.76	4.16	6.70	4.00	6.18	3.86	5.52

Table 2.3: Type I error rates ($\times 10^{-2}$) using the Pocock GSD with $\alpha = 0.05$

			$\rho = 0$		$\rho = 0.25$		$\rho = 0.5$		$\rho = 0.75$		$\rho = 0.9$	
n_D	$n_{\bar{D}}$	t	Binorm	Bilog	Binorm	Bilog	Binorm	Bilog	Binorm	Bilog	Binorm	Bilog
			Two-group sequential design (J=2)									
50	50	0.2	7.40	12.50	7.40	12.28	6.90	11.74	6.26	9.84	4.12	6.70
		0.4	5.36	9.90	5.06	9.66	4.88	8.86	3.78	7.18	2.74	4.56
		0.5	4.16	9.54	4.00	9.22	3.82	8.66	2.80	6.66	1.88	3.58
		0.6	3.54	8.84	3.12	7.78	2.64	6.56	1.86	4.56	0.88	2.42
		0.8	1.14	2.34	0.92	2.04	0.72	1.70	0.38	0.98	0.16	0.32
250	250	0.2	6.36	8.12	6.06	7.30	5.88	7.00	5.28	6.56	5.20	5.70
		0.4	4.44	6.40	4.84	6.38	4.94	6.68	4.72	6.30	4.46	5.86
		0.5	4.34	7.90	4.74	7.70	4.80	7.94	4.16	7.20	3.88	6.54
		0.6	4.32	8.96	4.18	7.84	4.20	8.06	3.82	6.82	3.54	6.48
		0.8	3.00	6.30	3.06	5.84	3.14	5.90	2.34	4.22	1.60	2.64
250	500	0.2	5.74	6.48	5.72	6.68	6.12	6.46	5.90	6.04	5.78	5.32
		0.4	4.92	5.92	5.48	6.70	5.42	6.28	4.98	5.82	4.72	5.44
		0.5	5.30	7.28	5.40	7.24	5.30	7.18	4.64	6.58	5.04	6.48
		0.6	4.94	7.68	4.98	7.62	4.72	7.08	4.84	6.76	4.70	6.28
		0.8	3.72	5.54	3.72	5.34	3.58	5.26	3.20	4.80	2.28	3.14
500	500	0.2	5.72	7.24	5.98	7.42	5.26	6.60	5.40	6.14	5.86	6.14
		0.4	5.38	6.58	5.00	5.96	4.82	5.78	4.62	5.64	4.38	4.86
		0.5	4.58	6.30	4.68	6.50	5.08	7.14	4.48	6.16	4.66	6.28
		0.6	4.66	7.38	4.76	7.68	4.36	7.06	4.04	6.66	4.52	6.86
		0.8	4.00	6.90	4.14	6.88	4.00	6.50	3.56	5.84	3.24	5.06
			Five-group sequential design (J=5)									
50	50	0.2	11.02	19.42	10.16	18.52	9.24	16.34	6.68	12.38	4.44	7.14
		0.4	6.60	11.92	5.80	11.06	5.08	9.36	3.46	6.64	1.78	3.42
		0.5	4.66	9.74	4.08	9.14	3.78	8.10	2.54	5.62	1.36	2.60
		0.6	3.06	7.68	2.96	7.14	2.56	5.94	1.68	3.96	0.50	1.56
		0.8	0.64	1.62	0.54	1.30	0.34	0.70	0.24	0.38	0.08	0.14
250	250	0.2	7.44	10.28	7.58	10.82	6.74	9.42	6.16	8.86	5.38	7.06
		0.4	4.70	7.80	5.34	8.32	4.94	7.84	4.62	7.22	3.76	6.06
		0.5	4.76	9.14	5.14	9.10	5.06	9.44	4.36	8.42	3.30	6.40
		0.6	4.22	10.32	3.96	8.78	4.16	9.06	3.60	7.30	2.96	6.22
		0.8	2.10	5.06	2.48	4.78	1.98	4.62	1.66	3.32	0.86	1.84
250	500	0.2	6.88	7.82	6.46	7.74	6.54	7.58	6.28	7.00	5.26	5.58
		0.4	5.06	6.56	5.98	7.30	5.20	6.52	5.34	6.84	4.32	6.10
		0.5	5.16	8.04	5.04	7.98	4.68	7.68	4.68	7.24	3.96	5.96
		0.6	5.04	8.20	4.76	7.82	4.28	7.56	4.60	6.94	3.54	5.78
		0.8	2.88	5.08	2.76	4.54	2.48	4.22	2.18	3.62	1.22	1.80
500	500	0.2	6.10	8.40	6.26	8.66	5.68	7.96	5.98	7.42	5.36	6.52
		0.4	5.22	6.88	4.86	7.22	4.96	6.62	4.88	6.10	4.62	5.92
		0.5	4.72	7.80	4.88	8.02	4.70	8.54	5.12	7.38	4.34	7.02
		0.6	4.74	9.08	4.90	9.08	4.30	8.52	3.76	7.16	3.98	6.98
		0.8	3.28	6.52	3.16	6.40	3.44	6.30	2.72	5.00	2.02	3.72

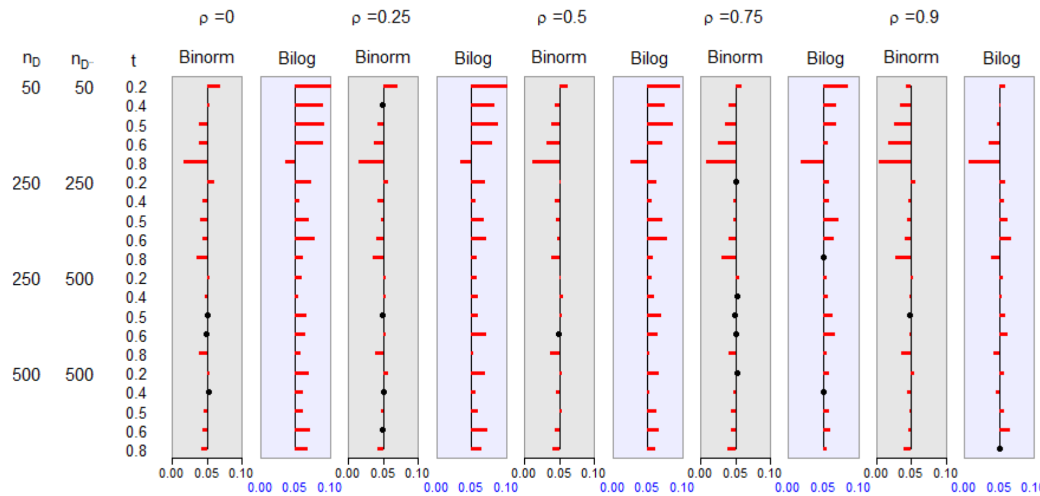

Figure 2.2: Type I error rates plot using the O'Brien-Fleming GSD with $\alpha = 0.05$, $J = 2$

as following. When $\rho = 0$, error=0.0526 for bi-normal distribution, error=0.0768 for bi-lognormal distribution. When $\rho = 0.25$, error=0.053 and 0.0694 for bi-normal and bi-lognormal distributions respectively. When $\rho = 0.5$, error=0.0514 and 0.07 for bi-normal and bi-lognormal distributions respectively. When $\rho = 0.75$, error=0.0546 and 0.0654; when $\rho = 0.9$, error=0.062 and 0.0668 for bi-normal and bi-lognormal distributions respectively. More results are shown in Table 2.4.

2.3.3 Expected Sample Size in GSDs

Furthermore, we conduct simulation studies on two correlated ROC curves that are not equal at certain FPR under investigation. While maintaining the α level and specific power $(1 - \beta)$ requirement, we show that the expected sample size with GSD is substantially less than the one with fixed sample size design. We use both the formula of (2.14) and bootstrap method to estimate the variance, and both results from the two methods are presented. This would be an additional verification of our variance covariance formula.

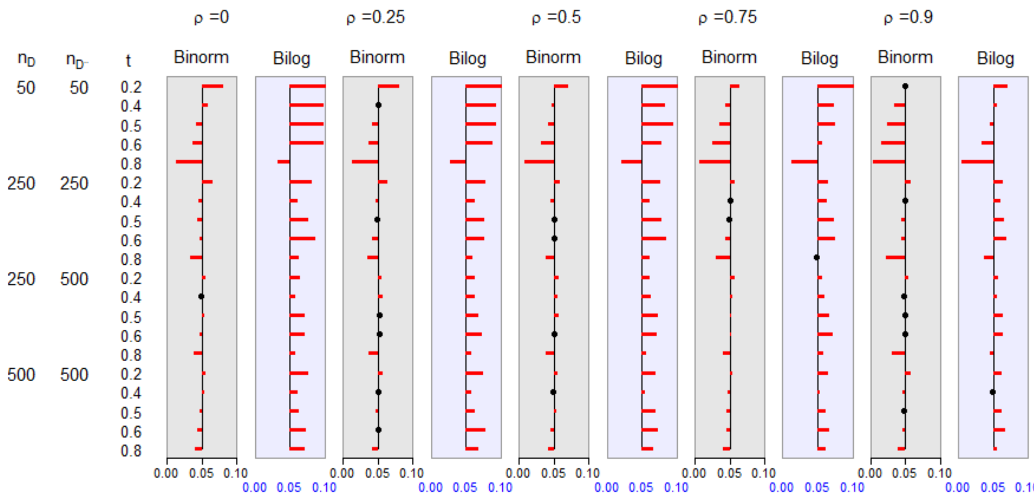

Figure 2.3: Type I error rates plot using the O'Brien-Fleming GSD with $\alpha = 0.05$, $J = 5$

Since the data are not independent and identically distributed, we conduct re-sampling in such a way that it preserves the underlying correlation. Hence, we perform re-sampling on subjects in bootstrap method.

Given two correlated ROC curves, with pre-specified α and specific power requirement, using the following formula we can determine the fixed sample size for a two-sided hypothesis testing study:

$$n \geq (\Phi^{-1}(1-\alpha/2) + \Phi^{-1}(1-\beta))^2 \frac{\sigma^2}{\delta^2},$$

where δ is the difference of two ROC curves at investigational FPR t_0. Let $\alpha = 0.05$, power $(1-\beta) = 90\%$. We simulate the correlated ROC data from a bivariate normal model, for the case data $(X_1, X_2)^T \sim N\{(6, 5.5)^T, \Sigma_1\}$, and for the control data the bivariate normal

Table 2.4: Test based on average: type I error rates ($\times 10^{-2}$) using the O'Brien-Fleming GSD with $\alpha = 0.05$

			$\rho = 0$		$\rho = 0.25$		$\rho = 0.5$		$\rho = 0.75$		$\rho = 0.9$	
n_D	$n_{\bar{D}}$	t	Binorm	Bilog	Binorm	Bilog	Binorm	Bilog	Binorm	Bilog	Binorm	Bilog
Two-group sequential design (J=2)												
50	50	0.2,0.5,0.8	5.52	9.74	5.54	9.34	5.52	9.04	5.92	8.92	6.70	8.64
250	250	0.2,0.5,0.8	5.12	7.04	5.14	6.58	4.88	6.28	5.02	5.92	5.02	5.34
250	500	0.2,0.5,0.8	5.28	6.38	5.22	6.26	5.40	6.22	5.56	5.92	5.82	5.88
500	500	0.2,0.5,0.8	4.96	6.80	5.48	6.70	5.66	6.66	5.42	6.52	6.02	6.94
Five-group sequential design (J=5)												
50	50	0.2,0.5,0.8	6.02	11.30	5.84	11.06	6.40	11.00	7.42	12.02	10.30	12.46
250	250	0.2,0.5,0.8	5.26	7.68	5.30	6.94	5.14	7.00	5.46	6.54	6.20	6.68
250	500	0.2,0.5,0.8	5.54	6.88	5.60	6.48	5.92	6.74	6.10	6.66	6.96	6.84
500	500	0.2,0.5,0.8	5.28	7.10	5.70	7.14	5.62	7.22	5.68	6.74	6.40	8.02

model is $(Y_1, Y_2)^T \sim N\{(3,3)^T, \Sigma_2\}$, where

$$\Sigma_1 = \begin{pmatrix} 4 & 4\rho \\ 4\rho & 4 \end{pmatrix}, \qquad \Sigma_2 = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \qquad \textit{with } \rho = \{0, 0.25, 0.5, 0.75, 0.9\}$$

The corresponding distributions and ROC curves are shown in Figure 2.4.

For the scenario with $\rho = 0.5$, $\delta = 0.0387$ at $t_0 = 0.5$, with the α level and 90% power requirement at this δ , we determine that sample size need to be 923 of both the case and control subjects for a fixed sample study. Then with the ratios provided in Jennison and Turnbull (2000), where with O'Brien-Fleming method, for J=2, the ratio is 1.007; for J=5, the ratio is 1.026. With Pocock method, for J=2, the ratio is 1.1; for J=5, the ratio is 1.207. Multiply the fixed sample size with the corresponding ratio, we know that to maintain the α and power level, for a group sequential study assuming equal group sizes, the total sample sizes are: with O'Brien-Fleming method, for J=2, the sample size is 929; for J=5, the sample size is 947. With Pocock method, for J=2, the sample size is 1015; for J=5, the sample size is 1114. The following simulation results, either using formula or using Bootstrap method (Table 2.7), show that the expected sample sizes of GSDs are less than the fixed sample size (923), while still meet the $\alpha(0.05)$ and power (90%) requirements.

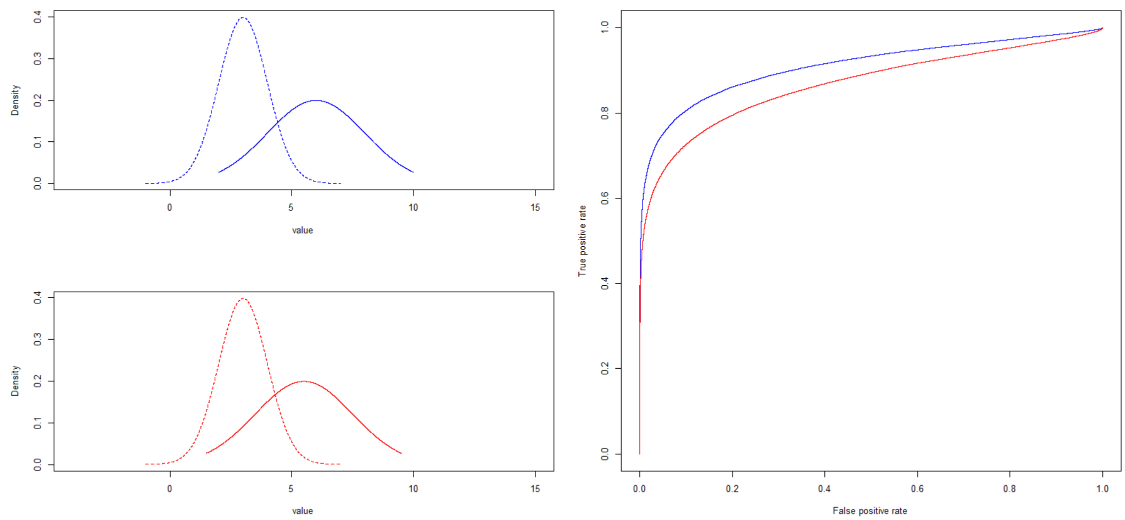

Figure 2.4: An example correlated ROC curves for GSD

With the same setting except the power requirement set to 80%, we determine that sample size need to be 689 for a fixed sample study. Then with the ratios provided in Jennison and Turnbull (2000), where with O'Brien-Fleming method, for J=2, the ratio is 1.008; for J=5, the ratio is 1.028. With Pocock method, for J=2, the ratio is 1.11; for J=5, the ratio is 1.229. Similarly, we calculated the sample sizes needed for group sequential studies assuming equal interim group sizes and obtain the following. With O'Brien-Fleming method, for J=2, the sample size is 695; for J=5, the sample size is 709. With Pocock method, for J=2, the sample size is 765; for J=5, the sample size is 847. The following simulation results, both using formula and using Bootstrap (Table 2.7), show that the expected sample sizes of GSDs are less than the fixed sample size (689), while still meet the $\alpha(0.05)$ and power (80%) requirements.

In both scenarios, the expected sample sizes in GSD are smaller than fixed sample design size, such that the trials utilizing GSD method are expected to end earlier with less subjects than using the fixed sample design. This has advantages both economically and ethically.

Table 2.5: Power(%) using the O'Brien-Fleming GSD with $\alpha = 0.05$

		Power=80%				**Power=90%**			
		Analytical Method		Bootstrap		Analytical Method		Bootstrap	
ρ	t	Normal	Lognormal	Normal	Lognormal	Normal	Lognormal	Normal	Lognormal
				Two-group sequential design (J=2)					
	0.2	79.5	80.2	77.9	80.0	89.5	89.6	89.1	88.7
	0.4	80.7	81.1	81.1	80.4	90.8	91.3	90.8	90.8
0	0.5	79.2	81.3	79.7	79.7	90.2	90.9	90.3	89.1
	0.6	81.1	81.1	80.0	79.4	90.3	90.9	90.3	90.0
	0.8	79.9	81.1	80.5	81.2	91.2	90.4	89.6	90.2
	0.2	80.6	80.3	79.7	80.7	90.0	90.8	91.3	89.2
	0.4	80.7	80.2	78.6	78.9	90.1	89.7	89.3	89.7
0.25	0.5	79.0	80.7	76.3	78.3	88.5	90.4	88.2	88.5
	0.6	79.3	80.8	79.2	78.6	89.8	90.4	89.6	89.8
	0.8	79.4	81.8	80.6	81.0	90.9	91.7	90.9	91.0
	0.2	79.3	82.1	79.8	79.1	90.1	90.8	89.4	89.9
	0.4	80.6	80.8	78.4	79.2	89.9	90.5	90.2	89.5
0.5	0.5	79.3	81.5	78.9	79.0	89.7	91.4	90.0	89.2
	0.6	78.9	80.4	78.7	77.8	90.7	89.3	88.7	88.9
	0.8	80.2	82.0	80.3	80.7	90.2	91.4	90.4	90.0
	0.2	81.1	81.5	78.0	77.8	89.6	90.8	89.6	89.0
	0.4	81.3	82.3	79.1	78.9	91.0	91.4	90.3	89.1
0.75	0.5	80.8	82.4	78.8	80.3	90.2	90.6	89.8	89.7
	0.6	81.0	82.0	80.3	80.3	91.1	92.1	90.6	91.2
	0.8	79.5	80.9	79.7	79.5	90.9	91.4	90.6	89.4
	0.2	80.7	82.0	76.3	76.7	91.2	92.5	87.4	88.1
	0.4	82.4	84.4	79.5	78.8	91.6	92.8	90.2	89.9
0.9	0.5	82.3	84.9	80.5	80.4	92.0	93.0	90.3	89.9
	0.6	82.3	85.4	81.1	81.3	92.8	92.7	91.2	91.3
	0.8	83.8	84.3	81.2	81.1	92.1	93.3	91.3	91.9
				Five-group sequential design (J=5)					
	0.2	78.3	80.8	78.4	78.3	89.5	89.3	88.9	90.2
	0.4	80.5	82.5	80.8	80.0	91.1	90.9	90.2	90.3
0	0.5	80.4	80.9	79.2	80.6	90.4	90.4	89.1	89.9
	0.6	82.1	82.1	80.4	80.9	90.1	91.9	89.5	89.9
	0.8	80.3	81.5	80.2	80.2	90.4	90.0	90.3	90.1
	0.2	80.3	80.6	79.4	78.8	90.3	90.9	89.7	90.2
	0.4	79.7	80.6	78.8	79.1	89.6	89.7	88.8	90.3
0.25	0.5	78.5	80.6	77.0	77.9	88.6	89.5	89.1	90.1
	0.6	80.3	80.8	79.4	79.2	90.0	89.9	89.1	89.1
	0.8	81.5	81.7	80.9	80.7	90.7	91.1	90.7	90.5
	0.2	80.6	81.9	79.8	79.1	90.7	90.8	89.5	89.3
	0.4	79.7	80.6	79.5	79.6	89.8	90.4	90.2	89.4
0.5	0.5	79.3	80.7	78.9	79.3	89.6	90.5	89.6	88.4
	0.6	78.3	80.9	77.9	77.4	90.2	89.9	89.0	89.8
	0.8	81.5	81.8	80.5	81.4	90.6	91.3	90.2	89.5
	0.2	80.3	82.8	78.1	78.7	90.7	90.9	88.5	88.5
	0.4	81.1	81.4	79.4	79.6	90.4	91.2	90.1	90.7
0.75	0.5	81.0	82.6	79.8	79.6	90.7	91.8	89.9	90.1
	0.6	82.6	83.2	80.3	80.0	91.3	92.2	90.9	91.3
	0.8	80.4	82.6	78.9	80.3	90.8	90.5	90.1	89.5
	0.2	81.3	83.0	76.9	77.0	91.4	91.2	88.6	88.1
	0.4	82.2	84.4	79.2	79.8	92.2	93.1	90.7	90.3
0.9	0.5	82.8	84.7	80.5	81.0	92.4	93.4	90.8	90.9
	0.6	83.3	85.9	81.7	80.7	92.0	93.3	91.3	91.4
	0.8	83.5	84.2	81.1	80.6	92.2	93.0	91.3	91.1

Table 2.6: Power(%) using the Pocock GSD with $\alpha = 0.05$

		Power=80%				**Power=90%**			
		Analytical Method		Bootstrap		Analytical Method		Bootstrap	
ρ	t	Normal	Lognormal	Normal	Lognormal	Normal	Lognormal	Normal	Lognormal
		Two-group sequential design (J=2)							
	0.2	80.2	79.5	79.1	78.5	89.5	89.9	88.5	89.9
	0.4	80.8	82.6	79.8	80.6	90.6	91.4	90.5	91.6
0	0.5	80.0	80.9	78.8	79.5	90.3	90.8	89.7	90.5
	0.6	81.1	81.3	80.4	80.3	91.0	91.3	90.3	89.9
	0.8	80.3	81.3	79.4	80.3	91.2	90.8	89.9	90.4
	0.2	81.1	81.7	79.7	79.6	91.3	90.6	90.3	89.1
	0.4	79.2	79.8	79.9	79.6	90.2	91.2	90.1	89.6
0.25	0.5	78.2	79.3	78.8	77.1	89.3	90.1	89.2	88.3
	0.6	78.3	80.2	78.9	78.9	89.4	90.6	89.4	89.2
	0.8	80.8	82.5	80.8	79.8	90.0	91.8	90.8	91.0
	0.2	80.1	81.8	77.9	79.4	90.3	90.9	89.6	88.8
	0.4	80.0	81.9	79.7	79.4	89.9	90.8	89.1	89.4
0.5	0.5	79.2	80.7	78.7	79.4	90.4	90.3	89.6	88.9
	0.6	79.3	80.3	78.2	79.0	89.1	89.4	88.8	88.8
	0.8	80.8	81.3	80.2	80.6	90.5	91.0	90.3	90.2
	0.2	80.2	81.6	76.7	77.6	90.1	90.7	88.6	90.2
	0.4	79.4	81.7	79.1	79.5	90.5	91.5	90.4	89.2
0.75	0.5	80.6	82.9	79.6	78.4	90.2	91.1	89.6	90.3
	0.6	80.7	83.0	80.8	78.8	91.7	92.5	90.6	90.5
	0.8	80.5	80.7	79.2	79.3	91.3	91.0	90.4	90.0
	0.2	81.3	83.1	77.0	77.8	90.8	91.0	87.9	87.9
	0.4	83.2	84.2	79.9	78.7	92.1	92.6	90.3	90.0
0.9	0.5	83.2	84.5	80.4	80.6	92.7	93.2	91.4	90.5
	0.6	82.3	84.5	80.6	80.0	92.2	93.4	90.9	91.0
	0.8	82.6	84.5	80.3	80.8	92.0	93.5	91.2	91.0
		Five-group sequential design (J=5)							
	0.2	80.2	80.8	78.8	79.8	90.6	90.3	90.4	88.5
	0.4	80.3	82.1	80.4	80.4	91.7	91.8	90.1	89.7
0	0.5	80.4	82.1	80.3	80.1	90.6	91.1	90.4	89.6
	0.6	80.3	82.1	80.1	81.0	90.1	91.2	91.3	90.7
	0.8	79.9	80.6	79.9	79.5	90.7	91.1	90.0	90.0
	0.2	80.6	81.7	80.8	79.7	90.2	91.5	90.4	90.4
	0.4	79.6	81.3	78.8	80.9	89.6	90.5	89.8	89.8
0.25	0.5	79.7	80.3	78.1	78.4	88.8	90.3	88.1	89.2
	0.6	79.2	81.0	78.4	79.7	89.9	90.8	88.8	90.0
	0.8	81.8	82.8	80.2	81.2	90.9	91.9	90.1	90.4
	0.2	81.4	82.0	79.7	80.1	90.6	90.9	89.9	89.8
	0.4	79.5	82.1	78.2	78.5	90.2	90.5	89.6	89.6
0.5	0.5	79.7	81.3	77.9	78.2	89.9	90.7	88.9	90.5
	0.6	78.6	80.4	78.1	77.4	89.9	90.0	89.7	89.8
	0.8	80.3	82.3	80.1	80.9	90.2	91.4	90.0	90.6
	0.2	81.5	82.7	77.8	78.4	90.7	90.4	88.5	88.6
	0.4	81.7	83.1	78.8	79.9	90.8	91.3	89.9	90.8
0.75	0.5	81.3	83.0	80.4	79.3	90.9	91.6	90.0	89.8
	0.6	81.8	83.7	80.6	80.4	91.8	92.4	91.1	90.6
	0.8	80.5	81.4	80.2	79.2	90.3	91.7	91.1	90.5
	0.2	81.4	83.5	77.1	77.2	90.6	92.0	87.8	89.2
	0.4	81.8	85.6	78.9	78.7	91.6	93.3	90.8	90.6
0.9	0.5	83.2	84.2	80.4	80.4	92.9	93.6	91.0	92.0
	0.6	83.2	84.1	80.1	80.4	92.2	93.6	91.2	91.4
	0.8	83.3	85.5	80.5	80.9	91.8	92.6	91.2	90.6

Table 2.7: Expected sample sizes using GSD with $\alpha = 0.05$

		Power=80%				**Power=90%**			
		Analytical Method		Bootstrap		Analytical Method		Bootstrap	
ρ	t	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock
				Two-group sequential design (J=2)					
	0.2	613	582	617	585	775	709	781	716
	0.4	754	708	757	716	950	859	954	870
0	0.5	839	781	838	791	1042	954	1050	951
	0.6	967	900	974	915	1218	1104	1221	1099
	0.8	1414	1337	1424	1331	1789	1605	1793	1624
	0.2	549	517	555	523	695	623	698	635
	0.4	660	620	659	624	830	748	837	759
0.25	0.5	729	690	733	695	923	838	931	844
	0.6	859	816	857	811	1084	984	1083	989
	0.8	1337	1252	1331	1254	1678	1508	1684	1519
	0.2	450	424	453	430	569	515	575	522
	0.4	551	523	555	523	697	629	699	635
0.5	0.5	628	592	630	596	794	718	801	725
	0.6	725	688	732	689	912	826	918	842
	0.8	1153	1079	1161	1085	1462	1312	1462	1323
	0.2	319	304	323	311	404	368	411	373
	0.4	407	387	414	390	516	471	524	473
0.75	0.5	473	442	481	452	598	540	606	546
	0.6	564	533	574	537	715	638	727	647
	0.8	876	827	884	830	1105	991	1106	993
	0.2	214	199	219	208	269	241	280	251
	0.4	283	259	286	268	354	316	364	323
0.9	0.5	335	303	340	315	420	366	428	382
	0.6	398	366	403	377	498	436	507	452
	0.8	634	580	644	600	797	700	815	721
				Five-group sequential design (J=5)					
	0.2	554	541	560	551	683	624	685	634
	0.4	683	675	685	668	833	748	840	782
0	0.5	752	733	754	742	924	834	929	848
	0.6	868	857	875	861	1069	976	1085	988
	0.8	1277	1255	1282	1256	1559	1425	1571	1444
	0.2	498	482	502	485	605	555	611	562
	0.4	595	584	603	596	738	673	737	671
0.25	0.5	664	646	663	661	816	744	818	756
	0.6	774	763	779	774	949	879	963	889
	0.8	1198	1169	1203	1187	1465	1346	1469	1360
	0.2	406	393	410	405	498	452	507	463
	0.4	501	493	503	503	615	555	614	570
0.5	0.5	567	556	572	570	698	642	700	656
	0.6	658	650	662	658	806	738	815	744
	0.8	1041	1018	1046	1033	1274	1178	1285	1180
	0.2	290	280	296	297	354	324	366	341
	0.4	369	358	374	372	454	413	457	425
0.75	0.5	428	411	435	430	525	478	531	492
	0.6	508	497	519	508	624	564	633	584
	0.8	789	768	804	791	966	884	978	897
	0.2	192	189	200	204	236	218	245	234
	0.4	255	251	260	267	310	286	319	297
0.9	0.5	298	295	306	310	364	327	375	351
	0.6	356	347	362	370	434	396	445	416
	0.8	570	559	581	587	692	640	710	656

Table 2.8: GSD design sample sizes (maximum) with $\alpha = 0.05$

		Power=80%		**Power=90%**	
$\boldsymbol{\rho}$	t	O'Brien-Fleming	Pocock	O'Brien-Fleming	Pocock
		Two-group sequential design (J=2)			
	0.2	683	752	913	997
	0.4	843	928	1127	1231
0	0.5	925	1018	1236	1350
	0.6	1076	1185	1439	1572
	0.8	1571	1730	2100	2294
	0.2	614	676	821	897
	0.4	731	805	977	1067
0.25	0.5	804	885	1075	1174
	0.6	948	1044	1267	1384
	0.8	1477	1627	1976	2158
	0.2	501	552	670	732
	0.4	612	674	818	894
0.5	0.5	695	765	929	1015
	0.6	799	880	1068	1167
	0.8	1278	1407	1709	1866
	0.2	355	391	474	518
	0.4	452	498	604	660
0.75	0.5	525	578	701	766
	0.6	627	691	839	916
	0.8	961	1058	1284	1403
	0.2	233	257	312	340
	0.4	308	339	412	450
0.9	0.5	364	400	486	531
	0.6	434	478	580	633
	0.8	690	760	923	1008
		Five-group sequential design (J=5)			
	0.2	696	832	930	1094
	0.4	860	1028	1148	1351
0	0.5	943	1127	1260	1482
	0.6	1098	1312	1466	1725
	0.8	1602	1915	2140	2517
	0.2	626	748	836	984
	0.4	745	891	996	1171
0.25	0.5	820	980	1095	1288
	0.6	967	1155	1291	1519
	0.8	1507	1801	2013	2368
	0.2	511	611	683	803
	0.4	624	746	834	981
0.5	0.5	709	847	947	1114
	0.6	815	974	1088	1280
	0.8	1303	1558	1741	2048
	0.2	362	432	483	568
	0.4	461	551	616	724
0.75	0.5	535	639	715	840
	0.6	640	765	855	1006
	0.8	980	1171	1309	1539
	0.2	238	284	318	373
	0.4	314	376	420	494
0.9	0.5	371	443	495	583
	0.6	442	529	591	695
	0.8	704	842	940	1106

Table 2.9: Fixed sample design sample sizes with $\alpha = 0.05$

$\boldsymbol{\rho}$	t	**Power=80%**	**Power=90%**
	0.2	677	906
	0.4	836	1119
0	0.5	917	1228
	0.6	1068	1429
	0.8	1558	2086
	0.2	609	815
	0.4	725	970
0.25	0.5	798	1067
	0.6	940	1259
	0.8	1466	1962
	0.2	497	666
	0.4	607	813
0.5	0.5	689	923
	0.6	792	1061
	0.8	1267	1697
	0.2	352	471
	0.4	448	600
0.75	0.5	520	696
	0.6	623	833
	0.8	953	1275
	0.2	231	309
	0.4	306	409
0.9	0.5	361	483
	0.6	430	576
	0.8	685	917

2.4 A Hypothetical Sequential Diagnostic Trial

In this section, we illustrate the GSD in a hypothetical lung cancer diagnostic trial. Both CT and PET can be used for diagnosing the staging of non-small cell lung cancer. The AUC for staging non-small cell lung cancer is between 52% and 85% for CT and between 81% and 96% for PET (Lardinois et al. 2003; Silvestri et al. 2003). In our example, we choose the AUCs to be 75% for CT and 90% for PET from the reasonable range. Consider testing the null hypothesis of $\Delta(t) = 0$ for t={0.2,0.4,0.5,0.6,0.8} and correlation between two diagnostic tests' data as $\rho = 0.5$ and are bi-normally distributed. Our example is a possible case under the alternative hypothesis condition, with $\Delta(t) = \{0.289, 0.182, 0.135, 0.094, 0.032\}$ for t={0.2,0.4,0.5,0.6,0.8} respectively. In Table 2.10, we show the interim looks of one simulation data with statistics and corresponding boundaries (O'Brien-Fleming) displayed at the bottom.

Suppose $n_D = 250$, $n_{\bar{D}} = 250$, FPR =0.5, and the number of looks is 5. At the first endpoint, with $n_D = 50$, $n_{\bar{D}} = 50$ subjects recruited and tested, the Z-statistic is 2.202, which is within the rejection boundaries for the null hypothesis. Thus we fail to reject the null hypothesis, and continue to recruit 50 additional cases and 50 additional controls. The difference between the ROC curves at FPR=0.5 and its variance can be estimated using the derived formula on the accruing data from the 100 cases and controls. The statistic of 1.247 is calculated and is smaller than the boundary 3.23. Again, we fail to reject the null hypothesis, and continue to recruit another 50 cases and controls. At the third interim analysis with overall 150 cases and controls, we calculate the Z-statistic to be 2.637, which is greater than the boundary 2.63. Therefor, we reject the null hypothesis of $\Delta(0.5) = 0$ at this step, and conclude that the two imaging tests are significantly different in their accuracy at the false positive rate of 0.5.

We also experiment with an example of comparing the average of three ROC points at different FPRs. Suppose FPR=(0.2,0.5,0.8) are of interest, and $n_D = 250$, $n_{\bar{D}} = 250$. All other settings remain the same as the previous example. The AUCs are set to be 75% for

Table 2.10: Interim test statistics of the diagnostic trial example

	Interim Z-Statistic				
FPR	1	2	3	4	5
0.2	1.562	2.174	3.544		
0.4	1.632	2.364	3.386		
0.5	2.202	1.247	2.637		
0.6	1.424	2.019	2.557	2.791	
0.8	1.472	1.692	1.885	2.269	2.218
Boundaries	±4.56	±3.23	±2.63	±2.28	±2.04

CT and 90% for PET with $\Delta(t) = \{0.289, 0.135, 0.032\}$ for t={0.2,0.5,0.8}, respectively. The average of the $\Delta(t)$ at the three FPRs is 0.152. We also reject the null hypothesis, $H_0 : \sum_{t \in \{0.2,0.5,0.8\}} (R_1(t)/3 - R_2(t)/3) = 0$, with the expected sample size to be 111 for either cases or controls.

2.5 Discussion

In this chapter, we have derived asymptotic properties of the sequential differences of two empirical ROC curves at the process level. We then used these results to develop distribution theory for the sequential difference of two empirical ROC curves at a FPR. We also extended the work to the asymptotic properties of the sequential difference of weighted ROC averages at several FPRs. Our approach not only enables us to investigate the difference of two correlated ROC curves, but also enables us to investigate the joint behavior of multiple points of two correlated ROC curves' differences and their weighted averages. Based on this, standard GSD software can be readily applied to design group sequential comparative diagnostic tests studies.

Based on the theorems developed, we conducted a simulation study to assess the finite sample properties of the results in Theorem 2.14. The simulation study verified the asymptotic variance-covariance matrix by comparing the theoretical covariance matrix to the observed covariance matrix from the simulated data. We verified that they match each other closely when sample size n is sufficiently large. We also conducted simulation studies, both for one point and for average of multiple points on ROC curves. With α level set to 0.05, the test Type I error rate is approximately 0.05 and tend to be closer to the number as we increase the sample sizes.

Furthermore, we demonstrate that the expected sample size of group sequential design can be substantially smaller than that of a fixed sample size design while maintaining the pre-specified α level and power requirement. We also conduct the simulation studies using both the formula method and the bootstrap method, which serves as an additional verification of our derived variance formula.

We further applied the GSD to a lung cancer diagnosis example, and our results clearly illustrate the advantage of sequentially monitoring the comparative diagnostic trial based on our theorem. The example shows that we are able to reject the null hypothesis under the alternative hypothesis with a substantially smaller expected sample size.

In our study, we used empirical cumulative distribution functions and Kernel density estimation to generate an estimate of $\sigma_{\hat{\Delta}(t)}$. Due to the limitation of Kernel density estimation, it will be desirable if we can develop a new non-parametric estimation method for variance without involving density estimation. Currently, we mainly deal with two correlated ROC curves and provide the variance covariance formula. We will extend the research to more general cases like clustered ROC curves and their differences. We can also apply a similar approach to compare multiple ROC curves.

Chapter 3: Group Sequential Method for Comparing Correlated PPV, NPV Curves

3.1 Introduction

The diagnostic test's accuracy can also be quantified by how well the test result predicts true disease status, which leads to the predictive values definition of PPV and NPV. Most of the time, we are more concerned in how likely the disease is present given the test result. Hence PPV and NPV quantify the clinical value of the diagnostic test. On the other hand, the classification probabilities, TPR and FPR, quantify the inherent accuracy of the test or how well the diagnostic test reflects true disease status. In many studies, the predictive values are reported in addition to the disease-specific classification probabilities. It is worth noting that the predictive values depend on not only the performance of the diagnostic test in diseased and non-diseased subjects, but also the prevalence of disease in population. Pepe (2003) points out that there is a direct relationship between predictive values and the classification probabilities as long as the prevalence is known. In fact, the complete joint distribution of (D, X) requires three parameters, which could be either (TPR, FPR, p) or (PPV, NPV, u), where p represents the prevalence and u represents the proportion of the population that are classified as negative. The relationship between two parameterizations can be derived by application of Bayes' theorem (Pepe 2003). PPV can be expressed as a function of TPR, FPR, and the prevalence p,

$$PPV = p \cdot TPR/(p \cdot TPR + (1-p)FPR).$$

Similarly for NPV as a function of the three parameters,

$$NPV = (1-p)(1-FPR)/((1-p)(1-FPR) + p(1-TPR)).$$

So is the proportion of the population that are classified as negative,

$$1 - u = p \cdot TPR + (1-p)FPR.$$

In this chapter, we will derived the asymptotic properties of correlated PPV and NPV curves both indexed by the FPR and indexed by the percentile value u. Then we will use simulation studies to show the consistency of covariance matrix estimator. We will also apply the results in a group sequential study and present the type I error rates through simulation.

3.2 Theoretical Results of Correlated PPV, NPV Curves

3.2.1 PPV and NPV indexed by the FPR

For PPV indexed by the FPR, we define the following difference of two correlated PPV at any given FPR of t,

$$\Delta(t) = PPV_1(t) - PPV_2(t),$$

and the estimated difference of two correlated PPV based on proportions of the accrued case and control subjects,

$$\hat{\Delta}_{r_D, r_{\bar{D}}}(t) = \widehat{PPV}_{1,r_D,r_{\bar{D}}}(t) - \widehat{PPV}_{2,r_D,r_{\bar{D}}}(t),$$

where $r_D, r_{\bar{D}}$ represents the proportions of the case and control subjects that has been accrued with test results available, respectively.

Let p be the disease status prevalence for the entire population, since PPV(t) is a function of ROC curve, we can write

$$PPV(t) = \frac{R(t)p}{R(t)p + t(1-p)}.$$

We put the derivation of the asymptotic distribution theory on one $PPV(t)$ curve in the following, which can be found in Koopmeiners and Feng (2011).

$$\begin{aligned}
& n_D^{-1/2}[n_D r_D](\widehat{PPV}_{r_D, r_{\bar{D}}}(t) - PPV(t)) \\
=& n_D^{-1/2}[n_D r_D]\left(\frac{\hat{R}_{r_D, r_{\bar{D}}}(t)p}{\hat{R}_{r_D, r_{\bar{D}}}(t)p + t(1-p)} - \frac{R(t)p}{R(t)p + t(1-p)}\right) \\
=& \frac{\left(\frac{\hat{R}_{r_D, r_{\bar{D}}}(t)p}{\hat{R}_{r_D, r_{\bar{D}}}(t)p + t(1-p)} - \frac{R(t)p}{R(t)p + t(1-p)}\right)}{\hat{R}_{r_D, r_{\bar{D}}}(t) - R(t)} n_D^{-1/2}[n_D r_D](\hat{R}_{r_D, r_{\bar{D}}}(t) - R(t))
\end{aligned}$$

Next, we will need to show that $\hat{R}_{r_D, r_{\bar{D}}}(t) \xrightarrow{a.s.} R(t)$ uniformly for $t \in [a, b]$, $r_D \in [c, 1]$ and $r_{\bar{D}} \in [d, 1]$,

$$\begin{aligned}
& \sup_{c \le r_D \le 1} \sup_{d \le r_{\bar{D}} \le 1} \sup_{a \le t \le b} |\hat{R}_{r_D, r_{\bar{D}}}(t) - R(t)| \\
= & \sup_{c \le r_D \le 1} \sup_{d \le r_{\bar{D}} \le 1} \sup_{a \le t \le b} |\hat{S}_{D, r_D}(\hat{S}^{-1}_{\bar{D}, r_{\bar{D}}}(t)) - S_D(S^{-1}_{\bar{D}}(t)))| \\
\le & \sup_{c \le r_D \le 1} \sup_{d \le r_{\bar{D}} \le 1} \sup_{a \le t \le b} |\hat{S}_{D, r_D}(\hat{S}^{-1}_{\bar{D}, r_{\bar{D}}}(t)) - S_D(\hat{S}^{-1}_{\bar{D}, r_{\bar{D}}}(t)))| \\
& + \sup_{c \le r_D \le 1} \sup_{d \le r_{\bar{D}} \le 1} \sup_{a \le t \le b} |S_D(\hat{S}^{-1}_{\bar{D}, r_{\bar{D}}}(t))) - S_D(S^{-1}_{\bar{D}}(t))|
\end{aligned}$$

$$= \frac{n_D}{[n_D c]} \sup_{c \le r_D \le 1} \sup_{d \le r_{\bar{D}} \le 1} \sup_{a \le t \le b} \frac{[n_D c]}{n_D} |\hat{S}_{D, r_D}(\hat{S}^{-1}_{\bar{D}, r_{\bar{D}}}(t)) - S_D(\hat{S}^{-1}_{\bar{D}, r_{\bar{D}}}(t)))|$$

$$+\frac{n_{\bar{D}}}{n_{\bar{D}}d}\sup_{c\le r_D\le 1}\sup_{d\le r_{\bar{D}}\le 1}\sup_{a\le t\le b}\frac{n_{\bar{D}}d}{n_{\bar{D}}}|S_D(S_{\bar{D}}^{-1}(S_{\bar{D}}(\hat{S}_{\bar{D},r_{\bar{D}}}^{-1}(t))))-S_D(S_{\bar{D}}^{-1}(t))|$$

$$\le\frac{n_D}{[n_D c]}\sup_{c\le r_D\le 1}\sup_{d\le r_{\bar{D}}\le 1}\sup_{a\le t\le b}\frac{[n_D r_D]}{n_D}|\hat{S}_{D,r_D}(\hat{S}_{\bar{D},r_{\bar{D}}}^{-1}(t))-S_D(\hat{S}_{\bar{D},r_{\bar{D}}}^{-1}(t)))|$$

$$+\frac{n_{\bar{D}}}{n_{\bar{D}}d}\sup_{c\le r_D\le 1}\sup_{d\le r_{\bar{D}}\le 1}\sup_{a\le t\le b}\frac{n_{\bar{D}}r_{\bar{D}}}{n_{\bar{D}}}|S_D(S_{\bar{D}}^{-1}(S_{\bar{D}}(\hat{S}_{\bar{D},r_{\bar{D}}}^{-1}(t))))-S_D(S_{\bar{D}}^{-1}(t))|$$

Using the Glivenko-Cantelli theorem Theorem 1.51, 1.52 of Csörgő and Szyszkowicz (1998), and as $n_D\to\infty$ and $n_{\bar{D}}\to\infty$, $\frac{n_D}{[n_D c]}\to\frac{1}{c}$ and $\frac{n_{\bar{D}}}{[n_{\bar{D}}d]}\to\frac{1}{d}$ respectively, we have

$$\frac{n_D}{[n_D c]}\sup_{c\le r_D\le 1}\sup_{d\le r_{\bar{D}}\le 1}\sup_{a\le t\le b}\frac{[n_D r_D]}{n_D}|\hat{S}_{D,r_D}(\hat{S}_{\bar{D},r_{\bar{D}}}^{-1}(t))-S_D(\hat{S}_{\bar{D},r_{\bar{D}}}^{-1}(t)))|\xrightarrow{a.s.}0,$$

and

$$\frac{n_{\bar{D}}}{n_{\bar{D}}d}\sup_{c\le r_D\le 1}\sup_{d\le r_{\bar{D}}\le 1}\sup_{a\le t\le b}\frac{n_{\bar{D}}r_{\bar{D}}}{n_{\bar{D}}}|S_D(S_{\bar{D}}^{-1}(S_{\bar{D}}(\hat{S}_{\bar{D},r_{\bar{D}}}^{-1}(t))))-S_D(S_{\bar{D}}^{-1}(t))|\xrightarrow{a.s.}0,$$

where the uniform continuity feature of $S_D(S_{\bar{D}}^{-1}(t))$ is applied to get the second statement. Combining two results gives that,

$$\sup_{c\le r_D\le 1}\sup_{d\le r_{\bar{D}}\le 1}\sup_{a\le t\le b}|\hat{R}_{r_D,r_{\bar{D}}}(t)-R(t)|\xrightarrow{a.s.}0. \tag{3.1}$$

By Mean Value Theorem, we know there is a value $\widetilde{R}(t)$ between $\hat{R}_{r_D,r_{\bar{D}}}(t)$ and $R(t)$ such that

$$\frac{\left(\frac{\hat{R}_{r_D,r_{\bar{D}}}(t)p}{\hat{R}_{r_D,r_{\bar{D}}}(t)p+t(1-p)}-\frac{R(t)p}{R(t)p+t(1-p)}\right)}{\hat{R}_{r_D,r_{\bar{D}}}(t)-R(t)}=\frac{t(1-p)p}{\left(\widetilde{R}(t)p+t(1-p)\right)^2}$$

And by Euation(3.1), and the definition of $\widetilde{R}(t)$ above, we know that $\widetilde{R}(t)\xrightarrow{a.s.}R(t)$. This feature and the uniform continuity of $\frac{t(1-p)p}{(R(t)p+t(1-p))^2}$, gives us that

$$\sup_{c\le r_D\le 1}\ \sup_{d\le r_{\bar{D}}\le 1}\ \sup_{a\le t\le b}\left|\frac{t(1-p)p}{\left(\widetilde{R}(t)p+t(1-p)\right)^2}-\frac{t(1-p)p}{(R(t)p+t(1-p))^2}\right|\xrightarrow{a.s.}0,$$

which will give the following equation,

$$\frac{\left(\frac{\hat{R}_{r_D,r_{\bar{D}}}(t)p}{\hat{R}_{r_D,r_{\bar{D}}}(t)p+t(1-p)}-\frac{R(t)p}{R(t)p+t(1-p)}\right)}{\hat{R}_{r_D,r_{\bar{D}}}(t)-R(t)}\xrightarrow{a.s.}\frac{t(1-p)p}{(R(t)p+t(1-p))^2}, \tag{3.2}$$

uniformly for $t\in[a,b]$, $r_D\in[c,1]$ and $r_{\bar{D}}\in[d,1]$. This and the Equation(2.8) gives us the result of

$$\begin{aligned}
&n_D^{-1/2}[n_D r_D](\widehat{PPV}_{\ell,r_D,r_{\bar{D}}}(t)-PPV_\ell(t)) \qquad (3.3)\\
&\xrightarrow{d}\frac{t(1-p)p}{(R_\ell(t)p+t(1-p))^2}\left(K_{\ell,1}(R_\ell(t),r_D)+\lambda^{1/2}\frac{r_D}{r_{\bar{D}}}\left(\frac{f_{\ell,D}(S^{-1}_{\ell,\bar{D}}(t))}{f_{\ell,\bar{D}}(S^{-1}_{\ell,\bar{D}}(t))}\right)K_{\ell,2}(t,r_{\bar{D}})\right)
\end{aligned}$$

Alternatively we can use the delta method to prove the asymptotic property in the following. Let map $\phi:\ D[0,1]\mapsto D[0,1]$, where D[0,1] is the set of all functions $z:[0,1]\mapsto\mathbb{R}$ that are right continuous and whose limits from the left exist everywhere in [0,1]. In which, the functions in D[0,1] are called *càdlàg*. Here, ϕ is a map from a ROC function to a PPV function.

$$\begin{aligned}
PPV &= \phi(R)\\
&= \frac{R\cdot p}{R\cdot p+t(1-p)}
\end{aligned}$$

This functional ϕ is Hadamard differentiable as shown in the the following using the definition in section 3.9.1 of van der Vaart and Wellner (1996), again we let R represents ROC function,

$$
\begin{aligned}
&\frac{\phi(R+t_n h_n)-\phi(R)}{t_n} \\
&= \frac{1}{t_n}\left(\frac{(R+t_n h_n)p}{(R+t_n h_n)p+t(1-p)} - \frac{R\cdot p}{R\cdot p+t(1-p)}\right) \\
&= \frac{p\cdot t(1-p)h_n}{((R+t_n h_n)p+t(1-p))(R\cdot p+t(1-p))} \\
&\to \frac{t(1-p)p}{(R\cdot p+t(1-p))^2}\cdot h, \qquad n\to\infty,
\end{aligned}
$$

for all converging sequences $t_n \to 0$ and $h_n \to h$. And the ϕ'_R is continuous linear map with

$$
\phi'_R(h) = \frac{t(1-p)p}{(R\cdot p+t(1-p))^2}\cdot h.
$$

Since ϕ is Hadamard differentiable, by Theorem 3.9.4 (van der Vaart and Wellner 1996), and based on the results on correlated ROC curves in Equation(2.8), we obtain

$$
\begin{aligned}
&n_D^{-1/2}[n_D r_D](\widehat{PPV}_{\ell,r_D,r_{\bar{D}}}(t) - PPV_\ell(t)) \qquad (3.4)\\
=&n_D^{-1/2}[n_D r_D]\left(\phi(\widehat{R}_{\ell,r_D,r_{\bar{D}}}(t)) - \phi(R_\ell(t))\right) \\
\xrightarrow{d}&\frac{t(1-p)p}{(R_\ell(t)p+t(1-p))^2}\left(K_{\ell,1}(R_\ell(t),r_D) + \lambda^{1/2}\frac{r_D}{r_{\bar{D}}}\left(\frac{f_{\ell,D}(S^{-1}_{\ell,\bar{D}}(t))}{f_{\ell,\bar{D}}(S^{-1}_{\ell,\bar{D}}(t))}\right)K_{\ell,2}(t,r_{\bar{D}})\right)
\end{aligned}
$$

For a PPVs' comparison study, by (3.4) and Cramér-Wold device, applying to the following vector,

$$\mathbf{V} = \begin{pmatrix} n_D^{-1/2}[n_D r_D](\widehat{PPV}_{1,r_D,r_{\bar{D}}}(t) - PPV_1(t)) \\ n_D^{-1/2}[n_D r_D](\widehat{PPV}_{2,r_D,r_{\bar{D}}}(t) - PPV_2(t)) \\ n_D^{-1/2}[n_D r'_D](\widehat{PPV}_{1,r'_D,r'_{\bar{D}}}(t') - PPV_1(t')) \\ n_D^{-1/2}[n_D r'_D](\widehat{PPV}_{2,r'_D,r'_{\bar{D}}}(t') - PPV_2(t')) \end{pmatrix}$$

$$\xrightarrow{d} \begin{pmatrix} \frac{t(1-p)p}{(R_1(t)p+t(1-p))^2}\left(K_{1,1}(R_1(t), r_D) + \lambda^{1/2}\frac{r_D}{r_{\bar{D}}}\left(\frac{f_{1,D}(S_{1,\bar{D}}^{-1}(t))}{f_{1,\bar{D}}(S_{1,\bar{D}}^{-1}(t))}\right)K_{1,2}(t, r_{\bar{D}})\right) \\ \frac{t(1-p)p}{(R_2(t)p+t(1-p))^2}\left(K_{2,1}(R_2(t), r_D) + \lambda^{1/2}\frac{r_D}{r_{\bar{D}}}\left(\frac{f_{2,D}(S_{2,\bar{D}}^{-1}(t))}{f_{2,\bar{D}}(S_{2,\bar{D}}^{-1}(t))}\right)K_{2,2}(t, r_{\bar{D}})\right) \\ \frac{t'(1-p)p}{(R_1(t')p+t'(1-p))^2}\left(K_{1,1}(R_1(t'), r'_D) + \lambda^{1/2}\frac{r'_D}{r'_{\bar{D}}}\left(\frac{f_{1,D}(S_{1,\bar{D}}^{-1}(t'))}{f_{1,\bar{D}}(S_{1,\bar{D}}^{-1}(t'))}\right)K_{1,2}(t', r'_{\bar{D}})\right) \\ \frac{t'(1-p)p}{(R_2(t')p+t'(1-p))^2}\left(K_{2,1}(R_2(t'), r'_D) + \lambda^{1/2}\frac{r'_D}{r'_{\bar{D}}}\left(\frac{f_{2,D}(S_{2,\bar{D}}^{-1}(t'))}{f_{2,\bar{D}}(S_{2,\bar{D}}^{-1}(t'))}\right)K_{2,2}(t', r'_{\bar{D}})\right) \end{pmatrix},$$

And

$$\mathbf{Y} = \begin{pmatrix} n_D^{-1/2}[n_D r_D](\hat{\Delta}_{r_D,r_{\bar{D}}}(t) - \Delta(t)) \\ n_D^{-1/2}[n_D r'_D](\hat{\Delta}_{r'_D,r'_{\bar{D}}}(t') - \Delta(t')) \end{pmatrix},$$

which can be expressed in terms of the empirical $\widehat{PPV}$ and true PPV curves as

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} n_D^{-1/2}[n_D r_D](\widehat{PPV}_{1,r_D,r_{\bar{D}}}(t) - PPV_1(t)) \\ n_D^{-1/2}[n_D r_D](\widehat{PPV}_{2,r_D,r_{\bar{D}}}(t) - PPV_2(t)) \\ n_D^{-1/2}[n_D r'_D](\widehat{PPV}_{1,r'_D,r'_{\bar{D}}}(t') - PPV_1(t')) \\ n_D^{-1/2}[n_D r'_D](\widehat{PPV}_{2,r'_D,r'_{\bar{D}}}(t') - PPV_2(t')) \end{pmatrix}.$$

Thus the random vector $\mathbf{V}$ is approximately multivariate normal with covariance as derived in the following. We use Σ to represent the asymptotic covariance matrix $Cov(\mathbf{V})$,

$\Sigma = \{a_{ij}\}_{i=1,\cdots,4;\,j=1,\cdots,4}$. Hence the random vector $\mathbf{Y}$ is approximately normal with covariance matrix derived approximately in the following.

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \Sigma \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} + a_{22} - 2a_{12} & a_{13} + a_{24} - a_{14} - a_{23} \\ a_{13} + a_{24} - a_{14} - a_{23} & a_{33} + a_{44} - 2a_{34} \end{pmatrix}.$$

It can be shown that

$$Cov(\hat{\Delta}_{r_D, r_{\bar{D}}}(t), \hat{\Delta}_{r'_D, r'_{\bar{D}}}(t')) = Cov(\hat{\Delta}_{r'_D, r'_{\bar{D}}}(t), \hat{\Delta}_{r'_D, r'_{\bar{D}}}(t')), \tag{3.5}$$

and as a special case when $t' = t$,

$$Cov(\hat{\Delta}_{r_D, r_{\bar{D}}}(t), \hat{\Delta}_{r'_D, r'_{\bar{D}}}(t)) = Var(\hat{\Delta}_{r'_D, r'_{\bar{D}}}(t)), \tag{3.6}$$

for $r'_D \geq r_D$ and $r'_{\bar{D}} \geq r_{\bar{D}}$.

The proof of the asymptotic property is given in the following. For simplicity, we define

$$C_\ell(t) \triangleq \frac{t(1-p)p}{(R_\ell(t)p + t(1-p))^2}.$$

We then derive the elements in Σ as:

$$a_{11} = C_1^2(t) \left(r_D(R_1(t) - R_1^2(t)) + \lambda \frac{r_D^2}{r_{\bar{D}}} \left(\frac{f_{1,D}(S_{1,\bar{D}}^{-1}(t))}{f_{1,\bar{D}}(S_{1,\bar{D}}^{-1}(t))} \right)^2 (t - t^2) \right),$$

$$a_{12} = C_1(t)C_2(t)\{r_D(S_D(S^{-1}_{1,\bar{D}}(t), S^{-1}_{2,\bar{D}}(t)) - R_1(t)R_2(t))$$

$$+ \lambda\frac{r_D^2}{r_{\bar{D}}}\frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))}\frac{f_{2,D}(S^{-1}_{2,\bar{D}}(t))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t))}(S_{\bar{D}}(S^{-1}_{1,\bar{D}}(t), S^{-1}_{2,\bar{D}}(t)) - t^2)\},$$

$$a_{13} =C_1(t)C_1(t')\big((r_D \wedge r'_D)(R_1(t) \wedge R_1(t') - R_1(t)R_1(t'))$$

$$+ (r_{\bar{D}} \wedge r'_{\bar{D}})\lambda\frac{r_D}{r_{\bar{D}}}\frac{r'_D}{r'_{\bar{D}}}\frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))}\frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t'))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t'))}(t \wedge t' - tt')\big),$$

$$a_{14} = C_1(t)C_2(t')\{(r_D \wedge r'_D)(S_D(S^{-1}_{1,\bar{D}}(t), S^{-1}_{2,\bar{D}}(t')) - R_1(t)R_2(t'))$$

$$+ (r_{\bar{D}} \wedge r'_{\bar{D}})\lambda\frac{r_D}{r_{\bar{D}}}\frac{r'_D}{r'_{\bar{D}}}\frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))}\frac{f_{2,D}(S^{-1}_{2,\bar{D}}(t'))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t'))}(S_{\bar{D}}(S^{-1}_{1,\bar{D}}(t), S^{-1}_{2,\bar{D}}(t')) - tt')\}.$$

Similarly, we can obtain the following elements of the covariance matrix.

$$a_{22} = C_2^2(t)\left(r_D(R_2(t) - R_2^2(t)) + \lambda\frac{r_D^2}{r_{\bar{D}}}\left(\frac{f_{2,D}(S^{-1}_{2,\bar{D}}(t))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t))}\right)^2(t - t^2)\right),$$

$$a_{23} = C_1(t')C_2(t)\{(r_D \wedge r'_D)(S_D(S^{-1}_{1,\bar{D}}(t'), S^{-1}_{2,\bar{D}}(t)) - R_1(t')R_2(t))$$

$$+ (r_{\bar{D}} \wedge r'_{\bar{D}})\lambda\frac{r_D}{r_{\bar{D}}}\frac{r'_D}{r'_{\bar{D}}}\frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t'))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t'))}\frac{f_{2,D}(S^{-1}_{2,\bar{D}}(t))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t))}(S_{\bar{D}}(S^{-1}_{1,\bar{D}}(t'), S^{-1}_{2,\bar{D}}(t)) - tt')\},$$

$$a_{24} = C_2(t)C_2(t')\{(r_D \wedge r'_D)(R_2(t) \wedge R_2(t') - R_2(t)R_2(t'))$$

$$+ (r_{\bar{D}} \wedge r'_{\bar{D}})\lambda \frac{r_D}{r_{\bar{D}}} \frac{r'_D}{r'_{\bar{D}}} \frac{f_{2,D}(S^{-1}_{2,\bar{D}}(t))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t))} \frac{f_{2,D}(S^{-1}_{2,\bar{D}}(t'))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t'))}(t \wedge t' - tt')\},$$

$$a_{33} = C_1^2(t')\left(r'_D(R_1(t') - R_1^2(t')) + \lambda \frac{r'^2_D}{r'_{\bar{D}}}\left(\frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t'))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t'))}\right)^2 (t' - t'^2)\right),$$

$$a_{34} = C_1(t')C_2(t')\{r'_D(S_D(S^{-1}_{1,\bar{D}}(t'), S^{-1}_{2,\bar{D}}(t')) - R_1(t')R_2(t'))$$

$$+ \lambda \frac{r'^2_D}{r'_{\bar{D}}} \frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t'))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t'))} \frac{f_{2,D}(S^{-1}_{2,\bar{D}}(t'))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t'))}(S_{\bar{D}}(S^{-1}_{1,\bar{D}}(t'), S^{-1}_{2,\bar{D}}(t')) - t'^2)\},$$

$$a_{44} = C_2^2(t')\left(r'_D(R_2(t') - R_2^2(t')) + \lambda \frac{r'^2_D}{r'_{\bar{D}}}\left(\frac{f_{2,D}(S^{-1}_{2,\bar{D}}(t'))}{f_{2,\bar{D}}(S^{-1}_{2,\bar{D}}(t'))}\right)^2 (t' - t'^2)\right).$$

With regard to Equation(3.5), it can be shown that the LHS converges to RHS. First, both sides can be expressed in the following formula:

$$LHS = \frac{1}{r_D n_D r'_D}(a_{13} + a_{24} - a_{14} - a_{23})$$

$$RHS = \frac{1}{r'_D n_D r'_D}(a^*_{13} + a^*_{24} - a^*_{14} - a^*_{23})$$

where a^*_{ij} is a_{ij} with $r_D, r_{\bar{D}}$ substituted by $r'_D, r'_{\bar{D}}$ respectively. Then substitute the covariance elements with the formula we derived above, we obtain

$$LHS = \tag{3.7}$$

$$C_1(t)C_1(t')\Bigg(\frac{1}{n_D r'_D}(R_1(t) \wedge R_1(t') - R_1(t)R_1(t'))$$

$$+ \frac{1}{n_{\bar{D}} r'_{\bar{D}}} \frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t))} \frac{f_{1,D}(S^{-1}_{1,\bar{D}}(t'))}{f_{1,\bar{D}}(S^{-1}_{1,\bar{D}}(t'))}(t \wedge t' - tt')\Bigg)$$

$$+ C_2(t)C_2(t')\Bigg(\frac{1}{n_D r'_D}(R_2(t) \wedge R_2(t') - R_2(t)R_2(t'))$$

$$+\frac{1}{n_{\bar{D}}r'_{\bar{D}}}\frac{f_{2,D}(S_{2,\bar{D}}^{-1}(t))}{f_{2,\bar{D}}(S_{2,\bar{D}}^{-1}(t))}\frac{f_{2,D}(S_{2,\bar{D}}^{-1}(t'))}{f_{2,\bar{D}}(S_{2,\bar{D}}^{-1}(t'))}(t\wedge t'-tt')\Bigg)$$

$$-C_1(t)C_2(t')\Bigg(\frac{1}{n_Dr'_D}(S_D(S_{1,\bar{D}}^{-1}(t),S_{2,\bar{D}}^{-1}(t'))-R_1(t)R_2(t'))$$

$$+\frac{1}{n_{\bar{D}}r'_{\bar{D}}}\frac{f_{1,D}(S_{1,\bar{D}}^{-1}(t))}{f_{1,\bar{D}}(S_{1,\bar{D}}^{-1}(t))}\frac{f_{2,D}(S_{2,\bar{D}}^{-1}(t'))}{f_{2,\bar{D}}(S_{2,\bar{D}}^{-1}(t'))}(S_{\bar{D}}(S_{1,\bar{D}}^{-1}(t),S_{2,\bar{D}}^{-1}(t'))-tt')\Bigg)$$

$$-C_1(t')C_2(t)\Bigg(\frac{1}{n_Dr'_D}(S_D(S_{1,\bar{D}}^{-1}(t'),S_{2,\bar{D}}^{-1}(t))-R_1(t')R_2(t))$$

$$+\frac{1}{n_{\bar{D}}r'_{\bar{D}}}\frac{f_{1,D}(S_{1,\bar{D}}^{-1}(t'))}{f_{1,\bar{D}}(S_{1,\bar{D}}^{-1}(t'))}\frac{f_{2,D}(S_{2,\bar{D}}^{-1}(t))}{f_{2,\bar{D}}(S_{2,\bar{D}}^{-1}(t))}(S_{\bar{D}}(S_{1,\bar{D}}^{-1}(t'),S_{2,\bar{D}}^{-1}(t))-tt')\Bigg)$$

$$=RHS$$

And as a special case where $t'=t$, we obtain the following:

$$Cov(\hat{\Delta}_{r_D,r_{\bar{D}}}(t),\hat{\Delta}_{r'_D,r'_{\bar{D}}}(t))=Var(\hat{\Delta}_{r'_D,r'_{\bar{D}}}(t)) \tag{3.8}$$

$$=C_1^2(t)\left(\frac{1}{n_Dr'_D}(R_1(t)-R_1^2(t))+\frac{1}{n_{\bar{D}}r'_{\bar{D}}}\frac{f_{1,D}(S_{1,\bar{D}}^{-1}(t))}{f_{1,\bar{D}}(S_{1,\bar{D}}^{-1}(t))}\frac{f_{1,D}(S_{1,\bar{D}}^{-1}(t))}{f_{1,\bar{D}}(S_{1,\bar{D}}^{-1}(t))}(t-t^2)\right)$$

$$+C_2^2(t)\left(\frac{1}{n_Dr'_D}(R_2(t)-R_2^2(t))+\frac{1}{n_{\bar{D}}r'_{\bar{D}}}\frac{f_{2,D}(S_{2,\bar{D}}^{-1}(t))}{f_{2,\bar{D}}(S_{2,\bar{D}}^{-1}(t))}\frac{f_{2,D}(S_{2,\bar{D}}^{-1}(t))}{f_{2,\bar{D}}(S_{2,\bar{D}}^{-1}(t))}(t-t^2)\right)$$

$$-2C_1(t)C_2(t)\Bigg(\frac{1}{n_Dr'_D}(S_D(S_{1,\bar{D}}^{-1}(t),S_{2,\bar{D}}^{-1}(t))-R_1(t)R_2(t))$$

$$+\frac{1}{n_{\bar{D}}r'_{\bar{D}}}\frac{f_{1,D}(S_{1,\bar{D}}^{-1}(t))}{f_{1,\bar{D}}(S_{1,\bar{D}}^{-1}(t))}\frac{f_{2,D}(S_{2,\bar{D}}^{-1}(t))}{f_{2,\bar{D}}(S_{2,\bar{D}}^{-1}(t))}(S_{\bar{D}}(S_{1,\bar{D}}^{-1}(t),S_{2,\bar{D}}^{-1}(t))-t^2)\Bigg)$$

for $r'_D\geq r_D$ and $r'_{\bar{D}}\geq r_{\bar{D}}$. This completes the proof of Equation (3.5) and (3.6).

The method above deals only two sequential analysis points and their asymptotic properties. In fact, the exact method can be applied to any finite set of sequential analysis points as shown below assuming the number of interim analysis is J.

$$\mathbf{Y} = \begin{pmatrix} n_D^{-1/2}[n_D r_{D,1}](\hat{\Delta}_{r_{D,1},r_{\bar{D},1}}(t_1) - \Delta(t_1)) \\ n_D^{-1/2}[n_D r_{D,2}](\hat{\Delta}_{r_{D,2},r_{\bar{D},2}}(t_2) - \Delta(t_2)) \\ \vdots \\ n_D^{-1/2}[n_D r_{D,J}](\hat{\Delta}_{r_{D,J},r_{\bar{D},J}}(t_J) - \Delta(t_J)) \end{pmatrix},$$

which can be expressed in terms of the empirical $\widehat{PPV}$ and true PPV curves as

$$\begin{pmatrix} 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{pmatrix} \begin{pmatrix} n_D^{-1/2}[n_D r_{D,1}](\widehat{PPV}_{1,r_{D,1},r_{\bar{D},1}}(t_1) - PPV_1(t_1)) \\ n_D^{-1/2}[n_D r_{D,1}](\widehat{PPV}_{2,r_{D,1},r_{\bar{D},1}}(t_1) - PPV_2(t_1)) \\ \vdots \\ n_D^{-1/2}[n_D r_{D,J}](\widehat{PPV}_{1,r_{D,J},r_{\bar{D},J}}(t_J) - PPV_1(t_J)) \\ n_D^{-1/2}[n_D r_{D,J}](\widehat{PPV}_{2,r_{D,J},r_{\bar{D},J}}(t_J) - PPV_2(t_J)) \end{pmatrix}.$$

Following the same steps, we will come to the same results of the asymptotic properties with independent increments covariance structure for any finite interim analysis.

Similarly, for NPV indexed by the FPR we define the following difference of two correlated NPV at any given FPR of t,

$$\Delta(t) = NPV_1(t) - NPV_2(t).$$

We know that

$$NPV(t) = \frac{(1-t)(1-p)}{(1-R(t))p + (1-t)(1-p)},$$

and its estimator

$$\widehat{NPV}(t) = \frac{(1-t)(1-p)}{(1-\widehat{R}(t))p + (1-t)(1-p)}.$$

Hence we define map $\phi : D[0,1] \mapsto D[0,1]$, where D[0,1] is the set of all *càdlàg* functions $z : [0,1] \mapsto \mathbb{R}$. Here, ϕ is a map from a ROC function to a NPV function.

$$NPV = \phi(R) = \frac{(1-t)(1-p)}{(1-R)p + (1-t)(1-p)}.$$

This functional ϕ is also Hadamard differentiable using the definition in section 3.9.1 of van der Vaart and Wellner (1996). Then by Theorem 3.9.4 (van der Vaart and Wellner 1996), and the results on correlated ROC curves in Equation(2.8), we can derive the asymptotic property of sequential empirical process of NPV. We can further prove that Equation (3.5) and (3.6) also hold true for correlated NPV curves indexed by FPR.

Due to the independent increments covariance structure for any finite interim analysis points, as shown in Equation (3.5) and (3.6) for correlated PPV and NPV curves indexed by FPR, we can readily apply these in group sequential designs using standard method to calculate the rejection boundaries at each interim analysis point. This will be demonstrated in the simulation study section with a covariance matrix estimator study and a type I error rate simulation study.

3.2.2 PPV and NPV indexed by the Percentile Value

We now consider PPV and NPV curves indexed by the proportion of the population that are classified as negative, u.

We know from Bayes' theorem that

$$PPV(u) = \frac{S_D(F^{-1}(u))p}{1-u}.$$

The PPV estimator at an interim point is given as

$$\widehat{PPV}_{r_D, r_{\bar{D}}}(u) = \frac{\hat{S}_{D,r_D}(\hat{F}^{-1}_{r_D, r_{\bar{D}}}(u))p}{1-u},$$

where $r_D, r_{\bar{D}}$ represents the proportions of the case and control subjects that has been accrued with test result available at the interim point respectively. We have

$$\widehat{PPV}_{r_D,r_{\bar{D}}}(u) - PPV(u) = \frac{p}{1-u}\left(\hat{S}_{D,r_D}(\hat{F}^{-1}_{r_D,r_{\bar{D}}}(u) - S_D(F^{-1}(u))\right),$$

Now define the difference of two correlated PPV at any given proportion of u,

$$\Delta(u) = PPV_1(u) - PPV_2(u),$$

and at the interim point noted by $r_D, r_{\bar{D}}$,

$$\hat{\Delta}_{r_D,r_{\bar{D}}}(u) = \widehat{PPV}_{1,r_D,r_{\bar{D}}}(u) - \widehat{PPV}_{2,r_D,r_{\bar{D}}}(u).$$

To derive the asymptotic properties of the sequential differences $\hat{\Delta}_{r_D,r_{\bar{D}}}(u)$, first we have for the following random vector,

$$\mathbf{V} = \begin{pmatrix} n_D^{-1/2}[n_D r_D](\widehat{PPV}_{1,r_D,r_{\bar{D}}}(u) - PPV_1(u)) \\ n_D^{-1/2}[n_D r_D](\widehat{PPV}_{2,r_D,r_{\bar{D}}}(u) - PPV_2(u)) \\ n_D^{-1/2}[n_D r'_D](\widehat{PPV}_{1,r'_D,r'_{\bar{D}}}(u') - PPV_1(u')) \\ n_D^{-1/2}[n_D r'_D](\widehat{PPV}_{2,r'_D,r'_{\bar{D}}}(u') - PPV_2(u')) \end{pmatrix}$$

$$\xrightarrow{d} \begin{pmatrix} r_1(u)K_{1,1}(F_{1,D}(F_1^{-1}(u)), r_D) + q_1(u)\frac{r_D}{r_{\bar{D}}}K_{1,2}(F_{1,\bar{D}}(F_1^{-1}(u)), r_{\bar{D}}) \\ r_2(u)K_{2,1}(F_{2,D}(F_2^{-1}(u)), r_D) + q_2(u)\frac{r_D}{r_{\bar{D}}}K_{2,2}(F_{2,\bar{D}}(F_2^{-1}(u)), r_{\bar{D}}) \\ r_1(u')K_{1,1}(F_{1,D}(F_1^{-1}(u')), r'_D) + q_1(u')\frac{r'_D}{r'_{\bar{D}}}K_{1,2}(F_{1,\bar{D}}(F_1^{-1}(u')), r'_{\bar{D}}) \\ r_2(u')K_{2,1}(F_{2,D}(F_2^{-1}(u')), r'_D) + q_2(u')\frac{r'_D}{r'_{\bar{D}}}K_{2,2}(F_{2,\bar{D}}(F_2^{-1}(u')), r'_{\bar{D}}) \end{pmatrix},$$

where for simplicity we define

$$r_v(u) = -\frac{p(1-p)}{1-u}\frac{f_{v,\bar{D}}(F_v^{-1}(u))}{f_v(F_v^{-1}(u))},$$

and

$$q_v(u) = \frac{p(1-p)}{1-u}\frac{f_{v,D}(F_v^{-1}(u))}{f_v(F_v^{-1}(u))}\sqrt{\lambda}.$$

We also define the following for simplicity, which will be used later,

$$h_{v,D}(u) = F_{v,D}(F_v^{-1}(u)),$$

and

$$h_{v,\bar{D}}(u) = F_{v,\bar{D}}(F_v^{-1}(u)).$$

To derive the asymptotic variance-covariance matrix, please note that $K_{v,1}(F_{v,D}(F_v^{-1}(u)), r_D)$ can be replaced by

$$n_D^{-1/2}[n_D r_D](\hat{F}_{v,D,r_D}(F_v^{-1}(u)) - F_{v,D}(F_v^{-1}(u))),$$

and that $K_{v,2}(F_{v,\bar{D}}(F_v^{-1}(u)), r_{\bar{D}})$ can be replace by

$$n_{\bar{D}}^{-1/2}[n_{\bar{D}} r_{\bar{D}}](\hat{F}_{v,\bar{D},r_{\bar{D}}}(F_v^{-1}(u)) - F_{v,\bar{D}}(F_v^{-1}(u))).$$

Then the random vector

$$\mathbf{Y} = \begin{pmatrix} n_D^{-1/2}[n_D r_D](\hat{\Delta}_{r_D, r_{\bar{D}}}(u) - \Delta(u)) \\ n_D^{-1/2}[n_D r'_D](\hat{\Delta}_{r'_D, r'_{\bar{D}}}(u') - \Delta(u')) \end{pmatrix}$$

can be expressed in terms of the empirical $\widehat{PPV}$ and true PPV curves as

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} n_D^{-1/2}[n_D r_D](\widehat{PPV}_{1,r_D,r_{\bar{D}}}(u) - PPV_1(u)) \\ n_D^{-1/2}[n_D r_D](\widehat{PPV}_{2,r_D,r_{\bar{D}}}(u) - PPV_2(u)) \\ n_D^{-1/2}[n_D r'_D](\widehat{PPV}_{1,r'_D,r'_{\bar{D}}}(u') - PPV_1(u')) \\ n_D^{-1/2}[n_D r'_D](\widehat{PPV}_{2,r'_D,r'_{\bar{D}}}(u') - PPV_2(u')) \end{pmatrix}.$$

The random vector $\mathbf{V}$ is approximately multivariate normal with covariance as derived in the following. We write the asymptotic covariance $Cov(\mathbf{V})$ as Σ, and $\Sigma = \{a_{ij}\}_{i=1,\cdots,4;\, j=1,\cdots,4}$.

Hence the random vector $\mathbf{Y}$ is approximately normal with covariance matrix derived with the following formula.

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \Sigma \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} + a_{22} - 2a_{12} & a_{13} + a_{24} - a_{14} - a_{23} \\ a_{13} + a_{24} - a_{14} - a_{23} & a_{33} + a_{44} - 2a_{34} \end{pmatrix}.$$

Furthermore, it can be shown that

$$Cov(\hat{\Delta}_{r_D,r_{\bar{D}}}(u), \hat{\Delta}_{r'_D,r'_{\bar{D}}}(u')) = Cov(\hat{\Delta}_{r'_D,r'_{\bar{D}}}(u), \hat{\Delta}_{r'_D,r'_{\bar{D}}}(u')), \tag{3.9}$$

and as a special case when $u' = u$,

$$Cov(\hat{\Delta}_{r_D,r_{\bar{D}}}(u), \hat{\Delta}_{r'_D,r'_{\bar{D}}}(u)) = Var(\hat{\Delta}_{r'_D,r'_{\bar{D}}}(u)), \tag{3.10}$$

for $r'_D \geq r_D$ and $r'_{\bar{D}} \geq r_{\bar{D}}$.

We provide the details of the proof in the following. First, we derive each element in the covariance matrix Σ,

$$a_{11} = r_1^2(u) r_D (h_{1,D}(u) - h_{1,D}^2(u)) + q_1^2(u) \frac{r_D^2}{r_{\bar{D}}} (h_{1,\bar{D}}(u) - h_{1,\bar{D}}^2(u)),$$

$$\begin{aligned}
a_{12} =& r_1(u) r_2(u) Cov\big(n_D^{-1/2}[n_D r_D](\hat{F}_{1,D,r_D}(F_1^{-1}(u)) - F_{1,D}(F_1^{-1}(u))),\\
& n_D^{-1/2}[n_D r_D](\hat{F}_{2,D,r_D}(F_2^{-1}(u)) - F_{2,D}(F_2^{-1}(u)))\big)\\
& + q_1(u) q_2(u) \frac{r_D^2}{r_{\bar{D}}^2} Cov\big(n_{\bar{D}}^{-1/2}[n_{\bar{D}} r_{\bar{D}}](\hat{F}_{1,\bar{D},r_{\bar{D}}}(F_1^{-1}(u)) - F_{1,\bar{D}}(F_1^{-1}(u))),\\
& n_{\bar{D}}^{-1/2}[n_{\bar{D}} r_{\bar{D}}](\hat{F}_{2,\bar{D},r_{\bar{D}}}(F_2^{-1}(u)) - F_{2,\bar{D}}(F_2^{-1}(u)))\big)\\
=& r_1(u) r_2(u) r_D (F_D(F_1^{-1}(u), F_2^{-1}(u)) - h_{1,D}(u) h_{2,D}(u))\\
& + q_1(u) q_2(u) \frac{r_D^2}{r_{\bar{D}}} (F_{\bar{D}}(F_1^{-1}(u), F_2^{-1}(u)) - h_{1,\bar{D}}(u) h_{2,\bar{D}}(u)),
\end{aligned}$$

$$\begin{aligned}
a_{13} =& r_1(u) r_1(u') (r_D \wedge r'_D)(h_{1,D}(u) \wedge h_{1,D}(u') - h_{1,D}(u) h_{1,D}(u'))\\
& + q_1(u) q_1(u') \frac{r_D}{r_{\bar{D}}} \frac{r'_D}{r'_{\bar{D}}} (r_{\bar{D}} \wedge r'_{\bar{D}})(h_{1,\bar{D}}(u) \wedge h_{1,\bar{D}}(u') - h_{1,\bar{D}}(u) h_{1,\bar{D}}(u')),
\end{aligned}$$

$$\begin{aligned}
a_{14} =& r_1(u) r_2(u') (r_D \wedge r'_D)(F_D(F_1^{-1}(u), F_2^{-1}(u')) - h_{1,D}(u) h_{2,D}(u'))\\
& + q_1(u) q_2(u') \frac{r_D}{r_{\bar{D}}} \frac{r'_D}{r'_{\bar{D}}} (r_{\bar{D}} \wedge r'_{\bar{D}})(F_{\bar{D}}(F_1^{-1}(u), F_2^{-1}(u')) - h_{1,\bar{D}}(u) h_{2,\bar{D}}(u')),
\end{aligned}$$

$$a_{22} = r_2^2(u) r_D (h_{2,D}(u) - h_{2,D}^2(u)) + q_2^2(u) \frac{r_D^2}{r_{\bar{D}}} (h_{2,\bar{D}}(u) - h_{2,\bar{D}}^2(u)),$$

$$a_{23} = r_1(u')r_2(u)(r_D \wedge r'_D)(F_D(F_1^{-1}(u'), F_2^{-1}(u)) - h_{1,D}(u')h_{2,D}(u))$$

$$+ q_1(u')q_2(u)\frac{r_D}{r_{\bar{D}}}\frac{r'_D}{r'_{\bar{D}}}(r_{\bar{D}} \wedge r'_{\bar{D}})(F_{\bar{D}}(F_1^{-1}(u'), F_2^{-1}(u)) - h_{1,\bar{D}}(u')h_{2,\bar{D}}(u)),$$

$$a_{24} = r_2(u)r_2(u')(r_D \wedge r'_D)(h_{2,D}(u) \wedge h_{2,D}(u') - h_{2,D}(u)h_{2,D}(u'))$$

$$+ q_2(u)q_2(u')\frac{r_D}{r_{\bar{D}}}\frac{r'_D}{r'_{\bar{D}}}(r_{\bar{D}} \wedge r'_{\bar{D}})(h_{2,\bar{D}}(u) \wedge h_{2,\bar{D}}(u') - h_{2,\bar{D}}(u)h_{2,\bar{D}}(u')),$$

$$a_{33} = r_1^2(u')r'_D(h_{1,D}(u') - h_{1,D}^2(u')) + q_1^2(u')\frac{r'^2_D}{r'_{\bar{D}}}(h_{1,\bar{D}}(u') - h_{1,\bar{D}}^2(u')),$$

$$a_{34} = r_1(u')r_2(u')r'_D(F_D(F_1^{-1}(u'), F_2^{-1}(u')) - h_{1,D}(u')h_{2,D}(u'))$$

$$+ q_1(u')q_2(u')\frac{r'^2_D}{r'_{\bar{D}}}(F_{\bar{D}}(F_1^{-1}(u'), F_2^{-1}(u')) - h_{1,\bar{D}}(u')h_{2,\bar{D}}(u')),$$

$$a_{44} = r_2^2(u')r'_D(h_{2,D}(u') - h_{2,D}^2(u')) + q_2^2(u')\frac{r'^2_D}{r'_{\bar{D}}}(h_{2,\bar{D}}(u') - h_{2,\bar{D}}^2(u')).$$

With regard to Equation(3.9), we have that

$$Cov(\hat{\Delta}_{r_D,r_{\bar{D}}}(u), \hat{\Delta}_{r'_D,r'_{\bar{D}}}(u')) = \frac{1}{r_D n_D r'_D}(a_{13} + a_{24} - a_{14} - a_{23}),$$

and

$$Cov(\hat{\Delta}_{r'_D,r'_{\bar{D}}}(u), \hat{\Delta}_{r'_D,r'_{\bar{D}}}(u')) = \frac{1}{r'_D n_D r'_D}(a^*_{13} + a^*_{24} - a^*_{14} - a^*_{23}).$$

where a^*_{ij} is a_{ij} with $r_D, r_{\bar{D}}$ substituted by $r'_D, r'_{\bar{D}}$ respectively. Then,

$$
\begin{aligned}
LHS =& \frac{1}{n_D}\{r_1(u)r_1(u')\frac{1}{r'_D}(h_{1,D}(u)\wedge h_{1,D}(u') - h_{1,D}(u)h_{1,D}(u')) \\
& + q_1(u)q_1(u')\frac{1}{r'_{\bar{D}}}(h_{1,\bar{D}}(u)\wedge h_{1,\bar{D}}(u') - h_{1,\bar{D}}(u)h_{1,\bar{D}}(u')) \\
& + r_2(u)r_2(u')\frac{1}{r'_D}(h_{2,D}(u)\wedge h_{2,D}(u') - h_{2,D}(u)h_{2,D}(u')) \\
& + q_2(u)q_2(u')\frac{1}{r'_{\bar{D}}}(h_{2,\bar{D}}(u)\wedge h_{2,\bar{D}}(u') - h_{2,\bar{D}}(u)h_{2,\bar{D}}(u')) \\
& - r_1(u)r_2(u')\frac{1}{r'_D}(F_D(F_1^{-1}(u), F_2^{-1}(u')) - h_{1,D}(u)h_{2,D}(u')) \\
& - q_1(u)q_2(u')\frac{1}{r'_{\bar{D}}}(F_{\bar{D}}(F_1^{-1}(u), F_2^{-1}(u')) - h_{1,\bar{D}}(u)h_{2,\bar{D}}(u')) \\
& - r_1(u')r_2(u)\frac{1}{r'_D}(F_D(F_1^{-1}(u'), F_2^{-1}(u)) - h_{1,D}(u')h_{2,D}(u)) \\
& - q_1(u')q_2(u)\frac{1}{r'_{\bar{D}}}(F_{\bar{D}}(F_1^{-1}(u'), F_2^{-1}(u)) - h_{1,\bar{D}}(u')h_{2,\bar{D}}(u))\} \\
=& RHS.
\end{aligned}
$$

This completes the proof of Equation (3.9). And as a special case when $u' = u$, for Equation (3.10) we have

$$
\begin{aligned}
& Cov(\hat{\Delta}_{r_D, r_{\bar{D}}}(u), \hat{\Delta}_{r'_D, r'_{\bar{D}}}(u)) = Var(\hat{\Delta}_{r'_D, r'_{\bar{D}}}(u)) \\
=& \frac{1}{n_D}\{r_1^2(u)\frac{1}{r'_D}(h_{1,D}(u) - h_{1,D}^2(u)) + q_1^2(u)\frac{1}{r'_{\bar{D}}}(h_{1,\bar{D}}(u) - h_{1,\bar{D}}^2(u)) \\
& + r_2^2(u)\frac{1}{r'_D}(h_{2,D}(u) - h_{2,D}^2(u)) + q_2^2(u)\frac{1}{r'_{\bar{D}}}(h_{2,\bar{D}}(u) - h_{2,\bar{D}}^2(u))
\end{aligned}
$$

$$
- 2r_1(u)r_2(u)\frac{1}{r'_D}(F_D(F_1^{-1}(u), F_2^{-1}(u)) - h_{1,D}(u)h_{2,D}(u))
$$

$$
- 2q_1(u)q_2(u)\frac{1}{r'_{\bar{D}}}(F_{\bar{D}}(F_1^{-1}(u), F_2^{-1}(u)) - h_{1,\bar{D}}(u)h_{2,\bar{D}}(u))\}
$$

for $r'_D \geq r_D$ and $r'_{\bar{D}} \geq r_{\bar{D}}$.

The method above handles two sequential analysis points and their asymptotic properties. In fact, the method can be applied to any finite set of sequential analysis points as shown below assuming the number of interim analysis is J.

$$
\mathbf{Y} = \begin{pmatrix} n_D^{-1/2}[n_D r_{D,1}](\hat{\Delta}_{r_{D,1},r_{\bar{D},1}}(u_1) - \Delta(u_1)) \\ n_D^{-1/2}[n_D r_{D,2}](\hat{\Delta}_{r_{D,2},r_{\bar{D},2}}(u_2) - \Delta(u_2)) \\ \vdots \\ n_D^{-1/2}[n_D r_{D,J}](\hat{\Delta}_{r_{D,J},r_{\bar{D},J}}(u_J) - \Delta(u_J)) \end{pmatrix},
$$

which can be expressed in terms of the empirical $\widehat{PPV}$ and true PPV curves as

$$
\begin{pmatrix} 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{pmatrix} \begin{pmatrix} n_D^{-1/2}[n_D r_{D,1}](\widehat{PPV}_{1,r_{D,1},r_{\bar{D},1}}(u_1) - PPV_1(u_1)) \\ n_D^{-1/2}[n_D r_{D,1}](\widehat{PPV}_{2,r_{D,1},r_{\bar{D},1}}(u_1) - PPV_2(u_1)) \\ \vdots \\ n_D^{-1/2}[n_D r_{D,J}](\widehat{PPV}_{1,r_{D,J},r_{\bar{D},J}}(u_J) - PPV_1(u_J)) \\ n_D^{-1/2}[n_D r_{D,J}](\widehat{PPV}_{2,r_{D,J},r_{\bar{D},J}}(u_J) - PPV_2(u_J)) \end{pmatrix}.
$$

Following the same steps, we will come to the same results of the asymptotic properties.

For NPV indexed by the percentile value, we can either follow the steps in the previous subsection to derive the asymptotic properties of NPV indexed by the percentile value, or apply functional delta method to the previous subsection's results since NPV curve can be expressed as a function of PPV curve as

$$NPV(u) = \frac{u-p}{u} + \frac{1-u}{u} PPV(u).$$

3.3 Simulation Studies

3.3.1 Consistency of Covariance Matrix Estimator

We conduct a simulation study to assess the finite sample properties of the results in Theorem 3.6. Diagnostic test data are drawn from bivariate normal distributions. For a case, the bivariate normal model is $(X_1, X_2)^T \sim N\{(10,6)^T, \Sigma_1\}$, and for a control, the bivariate normal model is $(Y_1, Y_2)^T \sim N\{(0,4)^T, \Sigma_2\}$, where

$$\Sigma_1 = \begin{pmatrix} 2 & \rho 2\sqrt{2} \\ \rho 2\sqrt{2} & 4 \end{pmatrix} \quad and \quad \Sigma_2 = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \qquad with\ \rho = 0.5\,.$$

We conduct 5000 simulation with $n_D = 200, n_{\bar{D}} = 200$, and for the simulated data, we calculate the variance-covariance of the $\Delta(t) = PPV_1(t) - PPV_2(t)$ at various combinations of $r_D, r_{\bar{D}}$ with FPR t=0.5. Here, the PPV functions are estimated with the empirical functions. Then we compare the simulated covariance matrix to the theoretical covariance matrix derived using the results of Theorem 3.6. The results are presented in Table 3.1, for prevalence $p \in \{0.1, 0.2, 0.3\}$.

3.3.2 Simulated Type I Error Rate in GSDs

To investigate finite sample performance of the GSD procedure, we conduct a simulation study in a two-group sequential test (J=2), and a five-group sequential test (J=5). The null hypothesis of equal PPV(t) is set to be true and the nominal type I error rate was set to be $\alpha = 0.05$ for two-sided tests. The diagnostic test data are simulated from bivariate

Table 3.1: The values of elements ($\times 10^{-5}$) in observed and theoretical Δ_{PPV} covariance matrix

	Observed covariance matrix				Theoretical covariance matrix			
	$p = 0.1,\ n_D = 200, n_{\bar{D}} = 200$							
$\Delta_{0.2,0.3}(0.5)$	9.493	4.711	3.711	1.898	8.302	4.151	3.321	1.660
$\Delta_{0.4,0.5}(0.5)$		4.814	3.712	1.906		4.151	3.321	1.660
$\Delta_{0.5,0.7}(0.5)$			3.772	1.929			3.321	1.660
$\Delta_{1,1}(0.5)$				1.932				1.660
	$p = 0.2,\ n_D = 200, n_{\bar{D}} = 200$							
$\Delta_{0.2,0.3}(0.5)$	23.634	11.701	9.216	4.706	20.480	10.240	8.192	4.096
$\Delta_{0.4,0.5}(0.5)$		11.937	9.195	4.716		10.240	8.192	4.096
$\Delta_{0.5,0.7}(0.5)$			9.340	4.769			8.192	4.096
$\Delta_{1,1}(0.5)$				4.771				4.096
	$p = 0.3,\ n_D = 200, n_{\bar{D}} = 200$							
$\Delta_{0.2,0.3}(0.5)$	32.550	16.077	12.657	6.454	27.938	13.969	11.175	5.588
$\Delta_{0.4,0.5}(0.5)$		16.370	12.600	6.453		13.969	11.175	5.588
$\Delta_{0.5,0.7}(0.5)$			12.789	6.522			11.175	5.588
$\Delta_{1,1}(0.5)$				6.516				5.588

normal models. The bivariate normal models is $(X_1, X_2)^T \sim N\{(1,10)^T, \Sigma_1\}$ for the case data. And for the control data, the bivariate normal model is $(Y_1, Y_2)^T \sim N\{(0,8)^T, \Sigma_2\}$, where

$$\Sigma_1 = \begin{pmatrix} 1 & 2\rho \\ 2\rho & 4 \end{pmatrix} \qquad \Sigma_2 = \begin{pmatrix} 1 & 2\rho \\ 2\rho & 4 \end{pmatrix} \qquad with\ \rho = (0, 0.25, 0.5, 0.75, 0.9)$$

With the above setting, we also simulate two cases with prevalence level p set to be 0.1 and 0.2 respectively. In all cases, the ROC curves are identical from the formula of ROC curve under bi-normal models (Zhou et al. 2011). Hence the PPV curves are identical according to formula $PPV(t) = \frac{R(t)p}{R(t)p+t(1-p)}$. Different numbers of case and control subjects, $n_D, n_{\bar{D}} = (50, 250, 500)$, are considered in our simulation study.

For each simulation setting, 5000 random data sets are generated and the GSD method applied to the simulated data. The Z statistics at each interim analysis point are then calculated based on the empirical ROC difference and estimated variances. The GSD test procedure compares the Z statistics with corresponding test boundaries of design, and the decision of rejection or failing to rejection is obtained for each simulated dataset. We then calculate the overall rejection rates for all simulated datasets. Table 3.2 gives the rejection rates of all different model and sample size combinations with a nominal α level 0.05 under the O'Brien and Fleming's criterion. And Table 3.3 is the results for the Pocock's criterion. As we can see, the simulated type I error rates are close to the nominal rate and tend to be closer as the overall sample sizes increase. The type I error rates are also plotted in Figure 3.1 and Figure 3.2. In these figures, the type I error rates are plotted as bars showing their deviations from the nominal rate of 0.05 which is the vertical line.

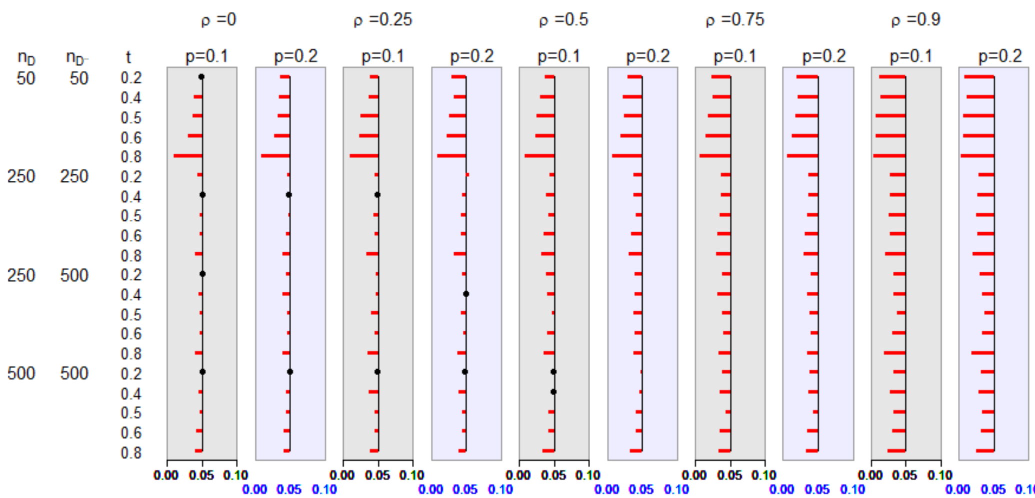

Figure 3.1: PPV indexed by FPR, type I error rates plot using the O'Brien-Fleming GSD with $\alpha = 0.05$, $J = 2$

Table 3.2: PPV indexed by FPR, type I error rates ($\times 10^{-2}$) using the O'Brien-Fleming GSD with $\alpha = 0.05$

			$\rho = 0$		$\rho = 0.25$		$\rho = 0.5$		$\rho = 0.75$		$\rho = 0.9$	
n_D	$n_{\bar{D}}$	t	p=0.1	0.2	p=0.1	0.2	p=0.1	0.2	p=0.1	0.2	p=0.1	0.2
			Two-group sequential design (J=2)									
50	50	0.2	4.82	3.66	3.82	2.90	3.78	2.88	2.36	2.02	1.16	0.84
		0.4	3.84	3.46	3.72	3.36	3.02	2.32	2.46	2.08	1.40	1.16
		0.5	3.68	3.26	2.56	2.66	2.56	2.40	1.88	1.76	0.74	0.60
		0.6	3.00	2.72	2.42	2.20	2.36	1.98	1.48	1.20	0.72	0.58
		0.8	0.98	0.88	0.96	0.86	0.86	0.76	0.68	0.52	0.32	0.28
250	250	0.2	4.40	4.56	4.46	5.46	4.44	3.82	3.80	3.70	2.80	2.74
		0.4	5.00	4.86	4.96	4.44	3.84	3.82	3.72	3.60	2.80	2.70
		0.5	4.72	4.72	4.38	4.22	4.28	4.16	3.56	3.48	2.54	2.48
		0.6	4.64	4.50	4.48	4.36	3.56	3.42	3.16	3.14	2.64	2.60
		0.8	3.98	3.94	3.42	3.34	3.22	3.18	2.96	2.94	2.06	2.04
250	500	0.2	5.10	4.50	4.66	4.46	3.92	3.96	3.96	3.90	3.22	3.06
		0.4	4.56	3.94	4.70	4.92	4.06	3.96	3.24	3.54	3.30	3.26
		0.5	4.74	4.64	4.04	4.26	4.66	3.84	3.96	3.86	3.72	3.66
		0.6	4.74	4.70	4.62	4.60	4.04	3.94	3.98	3.96	3.12	3.26
		0.8	3.96	3.86	3.48	3.80	3.52	3.72	3.46	3.50	1.92	1.86
500	500	0.2	4.98	4.90	4.90	4.84	4.94	4.78	3.92	3.80	3.28	3.22
		0.4	4.60	4.44	3.76	3.96	4.86	4.60	3.50	3.58	2.76	2.80
		0.5	4.76	4.38	4.46	4.42	4.16	4.10	4.32	4.26	3.26	3.34
		0.6	4.22	4.18	4.50	4.48	4.22	4.06	3.56	3.54	3.02	3.08
		0.8	4.10	4.10	3.94	3.92	3.30	3.28	3.42	3.38	2.42	2.40
			Five-group sequential design (J=5)									
50	50	0.2	5.02	3.90	4.22	3.78	3.06	2.74	2.12	1.84	1.18	0.70
		0.4	3.92	3.86	3.80	2.94	3.06	2.88	2.00	1.88	0.88	0.62
		0.5	3.94	3.02	3.12	2.70	2.46	2.66	1.98	1.44	0.76	0.56
		0.6	3.00	2.70	2.70	2.36	2.16	2.14	1.78	1.30	0.60	0.54
		0.8	0.96	0.96	0.74	0.78	0.50	0.42	0.26	0.30	0.08	0.08
250	250	0.2	5.10	4.10	4.34	3.86	4.80	4.04	4.04	4.10	3.32	3.14
		0.4	4.32	4.08	4.72	4.40	4.12	4.08	3.42	3.34	2.84	2.56
		0.5	4.32	4.50	4.22	4.32	4.16	4.16	3.60	3.56	2.30	2.12
		0.6	4.54	4.46	3.62	3.82	4.12	3.64	3.50	3.06	2.38	2.34
		0.8	3.88	3.20	3.80	3.40	2.96	2.96	2.96	2.44	1.78	1.76
250	500	0.2	5.32	4.66	4.80	4.66	4.58	4.20	4.16	3.88	3.02	2.88
		0.4	4.58	4.96	4.80	4.56	4.10	4.64	3.66	3.66	3.32	3.52
		0.5	4.52	4.52	4.68	4.54	4.34	4.48	4.02	4.00	3.16	3.24
		0.6	4.72	4.82	4.36	4.16	4.08	4.30	3.86	4.04	2.86	2.84
		0.8	4.06	3.92	3.56	3.56	3.78	3.84	2.88	2.86	2.36	1.98
500	500	0.2	5.02	4.96	5.18	5.00	4.68	4.52	4.44	4.40	3.66	3.46
		0.4	4.46	4.32	4.92	4.26	4.70	4.44	4.16	4.10	3.46	3.16
		0.5	4.64	4.58	4.66	4.76	4.66	4.48	4.22	4.22	3.30	3.42
		0.6	4.54	4.76	4.90	4.78	4.16	4.00	4.42	4.36	3.14	3.16
		0.8	4.58	4.56	4.24	4.22	3.76	3.56	3.32	3.28	2.58	2.56

Table 3.3: PPV indexed by FPR, type I error rates ($\times 10^{-2}$) using the Pocock GSD with $\alpha = 0.05$

			$\rho = 0$		$\rho = 0.25$		$\rho = 0.5$		$\rho = 0.75$		$\rho = 0.9$	
n_D	$n_{\bar{D}}$	t	p=0.1	0.2	p=0.1	0.2	p=0.1	0.2	p=0.1	0.2	p=0.1	0.2
Two-group sequential design (J=2)												
50	50	0.2	5.06	3.62	3.94	2.74	3.08	2.16	1.84	1.40	0.72	0.64
		0.4	3.50	3.02	3.30	2.72	2.74	2.04	1.66	1.38	0.78	0.66
		0.5	3.72	2.72	2.78	2.18	2.08	1.94	1.14	1.08	0.24	0.26
		0.6	2.40	2.28	2.02	1.78	1.58	1.46	0.82	0.68	0.28	0.22
		0.8	0.52	0.46	0.58	0.64	0.44	0.36	0.14	0.14	0.06	0.06
250	250	0.2	5.00	4.52	4.30	4.02	4.10	3.50	3.64	2.68	2.16	2.06
		0.4	4.96	4.52	4.28	3.86	3.78	3.62	3.50	3.38	2.00	1.96
		0.5	3.90	3.78	4.52	4.44	3.78	3.56	2.98	2.96	2.08	2.22
		0.6	4.50	4.36	3.96	3.50	3.46	3.40	2.88	2.88	1.80	1.64
		0.8	3.32	2.84	3.02	2.84	2.42	2.42	1.96	1.96	0.98	0.96
250	500	0.2	5.08	4.92	4.48	3.98	3.82	3.72	3.30	3.32	3.18	3.00
		0.4	5.14	5.32	4.16	4.10	3.80	3.78	3.34	2.96	3.12	2.74
		0.5	4.46	4.48	4.16	4.22	3.94	3.86	3.62	3.50	2.68	2.72
		0.6	4.54	4.44	4.00	3.78	4.22	3.52	3.66	3.60	2.40	2.40
		0.8	3.08	3.14	3.34	3.08	3.28	3.20	2.56	2.50	1.52	1.36
500	500	0.2	4.72	4.82	4.52	4.44	4.82	4.04	3.68	3.80	3.34	3.18
		0.4	4.54	4.44	4.38	4.42	4.10	4.12	4.34	4.02	2.90	2.46
		0.5	4.56	4.40	4.76	4.06	3.64	3.08	3.94	3.92	2.94	2.64
		0.6	4.54	4.18	4.62	4.46	3.64	3.56	3.14	3.08	3.04	3.28
		0.8	3.54	3.52	3.46	3.44	2.98	2.92	2.86	2.78	2.06	2.04
Five-group sequential design (J=5)												
50	50	0.2	4.26	3.22	3.84	2.40	3.00	1.28	1.20	0.96	0.36	0.30
		0.4	3.86	2.44	2.92	2.12	2.14	1.46	1.08	0.72	0.24	0.16
		0.5	3.02	1.80	2.36	1.68	1.62	0.96	0.50	0.48	0.20	0.18
		0.6	1.68	1.58	1.58	0.90	1.10	0.70	0.42	0.38	0.06	0.06
		0.8	0.38	0.32	0.24	0.22	0.16	0.08	0.06	0.02	0.02	0.02
250	250	0.2	4.30	4.16	4.92	4.06	3.94	3.52	3.36	2.88	1.90	1.60
		0.4	4.72	3.90	4.28	4.20	3.66	3.56	2.54	2.10	1.78	1.50
		0.5	3.28	3.40	3.74	3.36	3.42	3.16	2.78	2.42	1.34	1.24
		0.6	4.04	3.68	4.20	3.04	3.04	2.92	2.32	2.34	1.18	1.22
		0.8	2.30	1.98	1.88	1.80	1.72	1.42	1.00	1.16	0.60	0.60
250	500	0.2	5.04	4.06	4.24	3.82	4.50	3.48	3.34	3.36	2.34	1.86
		0.4	4.58	3.78	0.045	4.14	4.28	3.14	3.42	3.34	2.44	2.10
		0.5	4.36	3.82	3.96	4.06	3.22	3.22	3.50	2.90	2.26	2.10
		0.6	4.04	4.36	4.16	3.54	3.28	3.04	2.68	2.82	1.88	1.46
		0.8	2.58	2.50	2.38	2.18	2.06	2.20	1.14	1.18	0.72	0.64
500	500	0.2	5.50	4.90	4.62	3.88	5.02	4.26	3.80	3.20	2.32	2.12
		0.4	4.28	4.86	4.50	4.88	3.78	3.52	3.72	3.24	2.42	2.32
		0.5	4.96	4.32	4.60	3.84	3.58	3.76	3.32	2.70	2.44	2.10
		0.6	4.34	4.44	4.24	3.74	3.52	3.52	3.16	2.90	2.26	2.26
		0.8	3.10	3.08	3.24	3.18	2.98	2.40	1.90	1.90	1.28	1.00

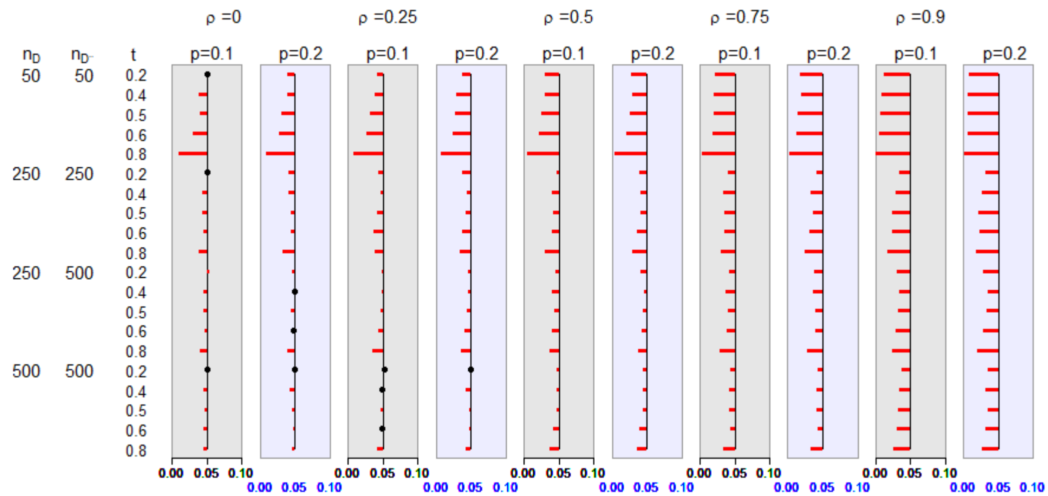

Figure 3.2: PPV indexed by FPR, type I error rates plot using the O'Brien-Fleming GSD with $\alpha = 0.05$, $J = 5$

3.4 Discussion

In this chapter, we have derived asymptotic properties of the sequential differences of two empirical PPV or NPV curves at the process level. We have studied both cases of indexed by FPR or indexed by percentile value. We then used these results to develop distribution theory for the sequential difference of two empirical PPV or NPV curves at a FPR or percentile value. Our approach not only enables us to investigate the difference of two correlated PPV/NPV curves, but also enables us to investigate the joint behavior of multiple points of two correlated ROC curves' differences. Based on this, standard GSD software can be readily applied to design group sequential comparative diagnostic tests studies for correlated PPV and NPV.

Based on the theorems developed, we conducted a simulation study to assess the finite sample properties of the results in Theorem 3.6. The simulation study verified the asymptotic variance-covariance matrix by comparing the theoretical covariance matrix to the observed covariance matrix from the simulated data. We verified that they match each other closely when sample size n is sufficiently large. We also conducted simulation studies on correlated PPV curves. With α level set to 0.05, the test Type I error rate is approximately 0.05 and tend to be closer to the number as we increase the sample sizes.

Chapter 4: Group Sequential Method for Comparing Clustered ROC Curves

4.1 Introduction

We define the clustered ROC sequential empirical process in the following. First, the empirical distribution function defined in the clustered case based on proportion of subjects as

$$\hat{F}_{[nt]}(x) = \frac{1}{M_{[nt]}} \sum_{i=1}^{[nt]} \sum_{j=1}^{m_i} I(X_{ij} \leq x),$$

where t is the percentage of subjects accrued so far at this interim analysis point, and $M_{[nt]} = \sum\limits_{i=1}^{[nt]} m_i$. For simplicity, $M_{[nt]}$ can be written as M_t, and $\hat{F}_{[nt]}(x)$ written as $\hat{F}_t(x)$.

The sequential empirical process is defined as

$$M_n^{-1/2} M_{[nt]}(\hat{F}_{[nt]}(x) - F(x))$$

$$= \sqrt{\frac{M_{[nt]}}{M_n}} \sqrt{M_{[nt]}} (\hat{F}_{[nt]}(x) - F(x)).$$

With the assumption that as $n \to \infty$, $n^{-1} \sum_{i=1}^{n} m_i \to \lambda$ for some positive constant λ, we have that as $n \to \infty$,

$$\frac{M_{[nt]} \wedge M_{[ns]}}{M_n} \to (t \wedge s).$$

4.2 Theoretical Results for Clustered ROC Vector

4.2.1 One Clustered ROC Result

In a diagnostic study with clustered data, suppose we have a total of n subjects in the study. Within each subject i, we observe X_{ij}, $j = 1, \cdots, m_i$, which are the measurements from m_i healthy units within subject i. We also observe Y_{ij}, $j = 1, \cdots, n_i$, which are the measurements from n_i diseased units within subject i, for $i = 1, \cdots, n$. We further assume that the observations X_{ij}, $j = 1, \cdots, m_i$ follow the survival function $S_{\bar{D}}$, and the observations Y_{ij}, $j = 1, \cdots, n_i$ follow the survival function S_D.

It is reasonable to assume that measurements from different subjects are independent and measurements within the same subject are possibly correlated. There are correlations within the same disease status group as well as between two groups within the same subject. This kind of study will generate clustered ROC data, hence any statistical inference will need to account for the within-subject correlations.

In a group sequential study scenario, we define $M_r = \sum_{i=1}^{[nr]} m_i$, $M = \sum_{i=1}^{n} m_i$, and $N_r = \sum_{i=1}^{[nr]} n_i$, $N = \sum_{i=1}^{n} n_i$, where r represents the percentage of subjects accrued so far at this analysis point. And assume that as $n \to \infty$, $n^{-1} \sum_{i=1}^{n} m_i \to \lambda$, and $n^{-1} \sum_{i=1}^{n} n_i \to \gamma$ for some positive constants λ and γ. The following theory is needed to establish the limiting distribution of $\widehat{R}(t)$ at any finite number of interim analysis points.

The proof of the univariate process convergence is presented in the following. First we verify that the Theorems 1.51, 1.52 of Csörgő and Szyszkowicz (1998) are also valid for clustered case. We need to prove that Dvoretzky-Kiefer-Wolfowitz inequality, which had been proved for i.i.d case, is also valid for clustered data. We have the following lemma.

Lemma 4.1. *For a clustered dataset, in which multiple samples can be collected from the same subject, let m_i be the number of samples collected from subject i and the total number $M = \sum_{i=1}^{n} m_i$. Within each subject i, we observe X_{ij}, $j = 1, \cdots, m_i$, which are the m_i observations within subject i and assume that they all follow the same distribution function*

$F(x)$, then we have

$$P\{\sup_{x\in\mathbb{R}} \frac{1}{M}\Big|\sum_{i=1}^{n}\sum_{j=1}^{m_i}(I(X_{ij} \le x) - F(x))\Big| > \epsilon\} \le C\exp(-2n\epsilon^2),$$

i.e.

$$P\{\sup_{x\in\mathbb{R}} \big|\hat{F}_n(x) - F(x)\big| > \epsilon\} \le C\exp(-2n\epsilon^2),$$

for all $\epsilon > 0$ and $n \ge 1$.

Proof: By Dvoretzky-Kiefer-Wolfowitz Inequality, given any natural number n, let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random variables with distribution function F. Let $\tilde{F}_n(x)$ be the associated empirical distribution function defined by $\tilde{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x),\ for\ x \in \mathbb{R}$. The inequality bounds the probability that the random function $\tilde{F}_n$ differs from F by more than a given constant $\epsilon > 0$ anywhere on the real line. Specifically, by Dvoretzky et al. (1956), we know there is a constant C such that

$$P\{\sup_{x\in\mathbb{R}} \sqrt{n}|\tilde{F}_n(x) - F(x)| > \epsilon\} \le C\exp(-2\epsilon^2) \tag{4.1}$$

for all $\epsilon > 0$. By Massart (1990), the optimal choice of C is obtained by C=2, which is

$$P\{\sup_{x\in\mathbb{R}} \sqrt{n}|\tilde{F}_n(x) - F(x)| > \epsilon\} \le 2\exp(-2\epsilon^2), \tag{4.2}$$

as well as

$$P\{\sup_{x\in\mathbb{R}} |\tilde{F}_n(x) - F(x)| > \epsilon\} \le 2\exp(-2n\epsilon^2). \tag{4.3}$$

We now prove that the Dvoretzky-Kiefer-Wolfowitz inequality also holds for the clustered empirical process. i.e. (4.1), (4.2), (4.3) is also true for clustered estimator $\hat{F}_n(x)$. We first consider a special case, $m_i \equiv m$, for $i = 1, \cdots, n$, then $\hat{F}_n(x) = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} I(X_{ij} \leq x)$.

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}(\hat{F}_n(x) - F(x)) \\
=&\sup_{x\in\mathbb{R}} \frac{1}{m} \sum_{j=1}^{m} (\frac{1}{n} \sum_{i=1}^{n} I(X_{ij} \leq x) - F(x)) \\
\leq&\frac{1}{m} \sum_{j=1}^{m} \sup_{x\in\mathbb{R}} (\frac{1}{n} \sum_{i=1}^{n} I(X_{ij} \leq x) - F(x))
\end{aligned} \tag{4.4}$$

Next, we consider the positive and negative parts separately, define $x^+ = max(x, 0)$ and $x^- = -min(x, 0)$. Both are non-negative and have that $|x| = x^+ + x^-$. For the positive part, since $\sup_{x\in\mathbb{R}}(\hat{F}_n(x) - F(x))^+ = \sup_{x\in\mathbb{R}}(\hat{F}_n(x) - F(x))$, hence

$$\begin{aligned}
&P\{\sup_{x\in\mathbb{R}}(\hat{F}_n(x) - F(x))^+ > \epsilon\} \\
=&P\{\sup_{x\in\mathbb{R}}(\hat{F}_n(x) - F(x)) > \epsilon\} \\
\leq&P\{\frac{1}{m} \sum_{j=1}^{m} \sup_{x\in\mathbb{R}} (\frac{1}{n} \sum_{i=1}^{n} I(X_{ij} \leq x) - F(x)) > \epsilon\},
\end{aligned}$$

by applying (4.4). Since the average of m values greater than ϵ implies that at lease one of the m values should be greater than ϵ, which can be easily proved by contradiction. Hence

$$\begin{aligned}
&P\{\frac{1}{m} \sum_{j=1}^{m} \sup_{x\in\mathbb{R}} (\frac{1}{n} \sum_{i=1}^{n} I(X_{ij} \leq x) - F(x)) > \epsilon\} \\
\leq&P\{\bigcup_{j=1}^{m} \sup_{x\in\mathbb{R}} (\frac{1}{n} \sum_{i=1}^{n} I(X_{ij} \leq x) - F(x)) > \epsilon\} \\
\leq&\sum_{j=1}^{m} P\{\sup_{x\in\mathbb{R}} (\frac{1}{n} \sum_{i=1}^{n} I(X_{ij} \leq x) - F(x)) > \epsilon\}
\end{aligned}$$

$$\leq \sum_{j=1}^{m} P\{\sup_{x\in\mathbb{R}} |\frac{1}{n}\sum_{i=1}^{n} I(X_{ij} \leq x) - F(x)| > \epsilon\}.$$

Note that each of the m elements consists of *i.i.d.* samples from n subjects, applying Dvoretzky-Kiefer-Wolfowitz Inequality (4.3),

$$\sum_{j=1}^{m} P\{\sup_{x\in\mathbb{R}} |\frac{1}{n}\sum_{i=1}^{n} I(X_{ij} \leq x) - F(x)| > \epsilon\}$$

$$\leq m \cdot c\exp(-2n\epsilon^2)$$

$$= c\exp(-2n\epsilon^2).$$

Note that $\sup_{x\in\mathbb{R}}(\frac{1}{n}\sum_{i=1}^{n} I(X_{ij} \leq x) - F(x))$, for $j = 1, \cdots, m$, are identically distributed. Similarly, we have

$$P\{\sup_{x\in\mathbb{R}}(\hat{F}_n(x) - F(x))^- > \epsilon\}$$

$$= P\{\sup_{x\in\mathbb{R}}(F(x) - \hat{F}_n(x)) > \epsilon\}$$

$$\leq c\exp(-2n\epsilon^2).$$

Hence combining the two results,

$$P\left\{\sup_{x\in\mathbb{R}}\left|\hat{F}_n(x) - F(x)\right| > \epsilon\right\}$$

$$\leq P\left\{\sup_{x\in\mathbb{R}}(\hat{F}_n(x) - F(x))^+ > \epsilon\right\} + P\left\{\sup_{x\in\mathbb{R}}(\hat{F}_n(x) - F(x))^- > \epsilon\right\}$$

$$\leq C\exp(-2n\epsilon^2).$$

For general cases that m_i are not constant, let $m = \max(m_i)$. For $i = 1, \cdots, n$, $j =$

$1, \cdots, m$, define $O_{ij} = 1$ if jth value is observed for subject i, otherwise $O_{ij} = 0$. Here we assume the missing indicator variable is independent of X_i for $i = 1, \cdots, n$, that is we assume missing completely at random (MCAR). Then $\hat{F}_n(x) = \frac{1}{M} \sum_{i=1}^{n} \sum_{j=1}^{m} I(X_{ij} \leq x, O_{ij} = 1)$, where $M = \sum_{i=1}^{n} \sum_{j=1}^{m} I(O_{ij} = 1)$. Following the previous steps, we can come to the same conclusion.

Hence (4.2), (4.3) are also true for clustered empirical process. We summarize the conclusion in Lemma 4.1.

In a sequential test setting, we have the next lemma.

Lemma 4.2. *With the same cluster setting as in Lemma 4.1, but in sequential test scenario with t represents the percentage of subjects accrued so far at current analysis point, and $M_{[nt]} = \sum_{i=1}^{[nt]} m_i$, $M = \sum_{i=1}^{n} m_i$,*

$$P\{\sup_{0 \leq t \leq 1} \sup_{x \in \mathbb{R}} M^{-1} M_{[nt]} |\hat{F}_{[nt]}(x) - F(x)| > \epsilon\} \leq Cn \exp(-2n\epsilon^2),$$

for all $\epsilon > 0$ and $n \geq 1$.

Proof:

$$P\{\sup_{0 \leq t \leq 1} \sup_{x \in \mathbb{R}} M^{-1} M_{[nt]} |\hat{F}_{[nt]}(x) - F(x)| > \epsilon\}$$

$$= P\{\sup_{0 \leq t \leq 1} \sup_{x \in \mathbb{R}} M^{-1} \Big| \sum_{i=1}^{[nt]} \sum_{j=1}^{m_i} (I(X_{ij} \leq x) - F(x)) \Big| > \epsilon\},$$

where t changes the supremum only at certain values, hence we have

$$P\{\sup_{0 \leq t \leq 1} \sup_{x \in \mathbb{R}} M^{-1} \Big| \sum_{i=1}^{[nt]} \sum_{j=1}^{m_i} (I(X_{ij} \leq x) - F(x)) \Big| > \epsilon\}$$

$$\leq P\{\sup_{x\in\mathbb{R}} M^{-1}\big|\sum_{i=1}^{nt_1}\sum_{j=1}^{m_i}(I(X_{ij}\leq x)-F(x))\big|>\epsilon\}$$

$$+P\{\sup_{x\in\mathbb{R}} M^{-1}\big|\sum_{i=1}^{nt_2}\sum_{j=1}^{m_i}(I(X_{ij}\leq x)-F(x))\big|>\epsilon\}$$

$$\cdots$$

$$+P\{\sup_{x\in\mathbb{R}} M^{-1}\big|\sum_{i=1}^{n}\sum_{j=1}^{m_i}(I(X_{ij}\leq x)-F(x))\big|>\epsilon\},$$

where $nt_1 = 1, \cdots, nt_k = k, \cdots, nt_n = n$, for $k = 1, \cdots, n$. In other words, $t_k = k/n$, for $k = 1, \cdots, n$. Of which, each of the n item has the following property derived by applying Lemma 4.1,

$$P\{\sup_{x\in\mathbb{R}} M^{-1}\big|\sum_{i=1}^{nt_k}\sum_{j=1}^{m_i}(I(X_{ij}\leq x)-F(x))\big|>\epsilon\}$$

$$=P\{\sup_{x\in\mathbb{R}} M_{[nt_k]}^{-1}\big|\sum_{i=1}^{nt_k}\sum_{j=1}^{m_i}(I(X_{ij}\leq x)-F(x))\big|>\epsilon\frac{M}{M_{[nt_k]}}\}$$

$$\leq C\exp(-2nt_k\epsilon^2\frac{M^2}{M_{[nt_k]}^2})$$

$$\leq C\exp(-2n\epsilon^2).$$

Hence,

$$LHS \leq Cn\exp(-2n\epsilon^2),$$

which proved the lemma. $\square$

Applying Lemma 4.2 and summing up all items with n from 1 to ∞,

$$\sum_{n=1}^{\infty} P\{\sup_{0\leq t\leq 1}\sup_{x\in\mathbb{R}} M^{-1}M_{[nt]}|\hat{F}_{[nt]}(x)-F(x)|>\epsilon\} \leq \sum_{n=1}^{\infty} Cn\exp(-2n\epsilon^2)<\infty, \tag{4.5}$$

then apply Borel-Cantelli lemma, we have

Lemma 4.3. *In clustered data setting and in sequential scenario with t represents the percentage of subjects accrued so far at current analysis point, and $M_{[nt]} = \sum_{i=1}^{[nt]} m_i$, , $M = \sum_{i=1}^{n} m_i$,*

$$\sup_{0\leq t\leq 1}\sup_{x\in\mathbb{R}} M^{-1}M_{[nt]}|\hat{F}_{[nt]}(x)-F(x)| \xrightarrow{a.s.} 0,$$

which is the clustered version of Theorem(1.52) of Csörgő and Szyszkowicz (1998). And

Lemma 4.4. *If F is continuous,*

$$\sup_{0\leq t\leq 1}\sup_{0\leq y\leq 1} M^{-1}M_{[nt]}|F(\hat{F}_{[nt]}^{-1}(y))-y| \xrightarrow{a.s.} 0.$$

We need two additional lemmas which are presented in the following with proofs. We start with the expression

$$\begin{aligned}
&\sup_{c\leq r\leq 1}\sup_{a\leq t\leq b}|F(\hat{F}_r^{-1}(t))-t| \\
&= \frac{M}{M_r}\sup_{c\leq r\leq 1}\sup_{a\leq t\leq b}\frac{M_r}{M}|F(\hat{F}_r^{-1}(t))-t| \\
&\leq \frac{M}{M_c}\sup_{c\leq r\leq 1}\sup_{a\leq t\leq b}\frac{M_r}{M}|F(\hat{F}_r^{-1}(t))-t|.
\end{aligned}$$

By Lemma 4.4, we know that

$$\sup_{c\leq r\leq 1}\sup_{a\leq t\leq b}\frac{M_t}{M}|F(\hat{F}_r^{-1}(t))-t| \rightarrow_{a.s.} 0,$$

and $\frac{M}{M_c} \to \frac{1}{c}$, therefore,

$$\sup_{c \le r \le 1} \sup_{a \le t \le b} |F(\hat{F}_r^{-1}(t)) - t| \to_{a.s.} 0. \tag{4.6}$$

Furthermore, $F^{-1}(t)$ is continuous by the assumptions that distribution function $F(x)$ is continuous and strictly increasing, and hence is uniformly continuous on [a,b] by Heine-Cantor theorem. Hence,

$$\sup_{c \le r \le 1} \sup_{a \le t \le b} |\hat{F}_r^{-1}(t) - F^{-1}(t)| \to_{a.s.} 0\ . \tag{4.7}$$

Due to continuity of $F(x)$, $S^{-1} = F^{-1}(1-t)$, so (4.6),(4.7) also apply to $S^{-1}(t)$. Hence we have

Lemma 4.5.

$$\sup_{c \le r \le 1} \sup_{a \le t \le b} |S(\hat{S}_r^{-1}(t)) - t| \to_{a.s.} 0\ ,$$

and

Lemma 4.6.

$$\sup_{c \le r \le 1} \sup_{a \le t \le b} |\hat{S}_r^{-1}(t) - S^{-1}(t)| \to_{a.s.} 0\ .$$

Lemma 4.7. *With the same clustered setting as in Lemma 4.2. Let $\hat{F}_n(t)$ be the empirical distribution function based on the cluster correlated samples from subject $1, \cdots, n$; and let $B_{1j}(t), B_{2j}(t), \cdots$ be a sequence of independent Brownian bridges, for $j = 1, \cdots, m$. There is a version of the sequence $B_{nj}(t)$ such that*

$$P\left(\sup_{0 \le r \le 1} \sup_{t \in \mathbb{R}} \left| M_{[nr]}(\hat{F}_{[nr]}(t) - F(t)) - \sum_{i=1}^{[nr]} \sum_{j=1}^{m_i} B_{ij}(t) \right| > (C_1 \log n + x) \log n \right) < C_2 \exp(-C_3 x).$$

Proof: First, from Theorem 4 of Komlós et al. (1975), we have the following. Let

$X_1, X_2, \cdots$ be a sequence of i.i.d. random variables with the same distribution function $F(t)$. Let $\tilde{F}_n(t)$ be the empirical distribution function based on the sample $X_1, X_2, \cdots, X_n$; and let $B_1(t), B_2(t), \cdots$ be a sequence of independent Brownian bridges. There is a version of the sequence $B_n(t)$ such that

$$P\left(\sup_{0\leq r\leq 1}\sup_{t\in\mathbb{R}}\left|[nr](\tilde{F}_{[nr]}(t)-F(t))-\sum_{i=1}^{[nr]}B_i(t)\right|>(C_1\log n+x)\log n\right)<C_2\exp(-C_3x), \tag{4.8}$$

for all $x > 0$, where C_1, C_2 and C_3 are positive constants.

Hence if $m_i \equiv m$, then for each jth measurement of all subjects, we have by applying (4.8) for $j = 1, \cdots, m$,

$$P\left(\sup_{0\leq r\leq 1}\sup_{t\in\mathbb{R}}\left|[nr](\sum_{i=1}^{[nr]}I(X_{ij}\leq t)/[nr]-F(t))-\sum_{i=1}^{[nr]}B_{ij}(t)\right|>(C_1\log n+x)\log n\right)$$
$$<C_2\exp(-C_3x) \tag{4.9}$$

where $B_{1j}(t), B_{2j}(t), \cdots$ be a sequence of independent Brownian bridges, for $j = 1, \cdots, m$. For the supremum item, we have

$$\begin{aligned}
&\sup_{0\leq r\leq 1}\sup_{t\in\mathbb{R}}\big(M_{[nr]}(\hat{F}_{[nr]}(t)-F(t))-\sum_{i=1}^{[nr]}\sum_{j=1}^{m_i}B_{ij}(t)\big)\\
&=\sup_{0\leq r\leq 1}\sup_{t\in\mathbb{R}}\big(m[nr](\frac{1}{m[nr]}\sum_{i=1}^{[nr]}\sum_{j=1}^{m}I(X_{ij}\leq t)-F(t))-\sum_{i=1}^{[nr]}\sum_{j=1}^{m}B_{ij}(t)\big)\\
&=\sup_{0\leq r\leq 1}\sup_{t\in\mathbb{R}}\sum_{j=1}^{m}\big([nr](\sum_{i=1}^{[nr]}I(X_{ij}\leq t)/[nr]-F(t))-\sum_{i=1}^{[nr]}B_{ij}(t)\big)\\
&\leq\sum_{j=1}^{m}\sup_{0\leq r\leq 1}\sup_{t\in\mathbb{R}}\big([nr](\sum_{i=1}^{[nr]}I(X_{ij}\leq t)/[nr]-F(t))-\sum_{i=1}^{[nr]}B_{ij}(t)\big).
\end{aligned} \tag{4.10}$$

For the positive part, we have that

$$P\{\sup_{0\le r\le 1}\sup_{t\in\mathbb{R}}\big(M_{[nr]}(\hat{F}_{[nr]}(t)-F(t))-\sum_{i=1}^{[nr]}\sum_{j=1}^{m_i}B_{ij}(t)\big)^+ > (C_1\log n+x)\log n\}$$

$$=P\{\sup_{0\le r\le 1}\sup_{t\in\mathbb{R}}\big(M_{[nr]}(\hat{F}_{[nr]}(t)-F(t))-\sum_{i=1}^{[nr]}\sum_{j=1}^{m_i}B_{ij}(t)\big) > (C_1\log n+x)\log n\},$$

then by (4.10) we know that

$$P\{\sup_{0\le r\le 1}\sup_{t\in\mathbb{R}}\big(M_{[nr]}(\hat{F}_{[nr]}(t)-F(t))-\sum_{i=1}^{[nr]}\sum_{j=1}^{m_i}B_{ij}(t)\big) > (C_1\log n+x)\log n\}$$

$$\le P\{\sum_{j=1}^{m}\sup_{0\le r\le 1}\sup_{t\in\mathbb{R}}\big([nr](\sum_{i=1}^{[nr]}I(X_{ij}\le t)/[nr]-F(t))-\sum_{i=1}^{[nr]}B_{ij}(t)\big) > (C_1\log n+x)\log n\}$$

$$=P\{\frac{1}{m}\sum_{j=1}^{m}\sup_{0\le r\le 1}\sup_{t\in\mathbb{R}}\big([nr](\sum_{i=1}^{[nr]}I(X_{ij}\le t)/[nr]-F(t))-\sum_{i=1}^{[nr]}B_{ij}(t)\big) > (\frac{C_1}{m}\log n+\frac{x}{m})\log n\}.$$

Since if the average of m values is greater than a constant, it implies that at least one of the m values should be greater than the constant. Hence, the probability

$$P\{\frac{1}{m}\sum_{j=1}^{m}\sup_{0\le r\le 1}\sup_{t\in\mathbb{R}}\big([nr](\sum_{i=1}^{[nr]}I(X_{ij}\le t)/[nr]-F(t))-\sum_{i=1}^{[nr]}B_{ij}(t)\big) > (\frac{C_1}{m}\log n+\frac{x}{m})\log n\}$$

$$\le P\{\bigcup_{j=1}^{m}\sup_{0\le r\le 1}\sup_{t\in\mathbb{R}}\big([nr](\sum_{i=1}^{[nr]}I(X_{ij}\le t)/[nr]-F(t))-\sum_{i=1}^{[nr]}B_{ij}(t)\big) > (\frac{C_1}{m}\log n+\frac{x}{m})\log n\}$$

$$\le \sum_{j=1}^{m}P\{\sup_{0\le r\le 1}\sup_{t\in\mathbb{R}}\big([nr](\sum_{i=1}^{[nr]}I(X_{ij}\le t)/[nr]-F(t))-\sum_{i=1}^{[nr]}B_{ij}(t)\big) > (\frac{C_1}{m}\log n+\frac{x}{m})\log n\}$$

$$\le \sum_{j=1}^{m}P\{\sup_{0\le r\le 1}\sup_{t\in\mathbb{R}}\big|[nr](\sum_{i=1}^{[nr]}I(X_{ij}\le t)/[nr]-F(t))-\sum_{i=1}^{[nr]}B_{ij}(t)\big| > (\frac{C_1}{m}\log n+\frac{x}{m})\log n\}.$$

Applying (4.9), we know the sum of the probabilities

$$\sum_{j=1}^{m} P\{\sup_{0\le r\le 1}\sup_{t\in\mathbb{R}}\big|[nr](\sum_{i=1}^{[nr]} I(X_{ij}\le t)/[nr]-F(t))-\sum_{i=1}^{[nr]} B_{ij}(t)\big| > (\frac{C_1}{m}\log n+\frac{x}{m})\log n\}$$

$$\le m\cdot C_2\exp(-C_3\frac{x}{m})$$

$$=C_2\exp(-C_3 x).$$

Similarly, for the negative part we have

$$P\{\sup_{0\le r\le 1}\sup_{t\in\mathbb{R}}\big(M_{[nr]}(\hat{F}_{[nr]}(t)-F(t))-\sum_{i=1}^{[nr]}\sum_{j=1}^{m_i} B_{ij}(t)\big)^- > (C_1\log n+x)\log n\}$$

$$=P\{\sup_{0\le r\le 1}\sup_{t\in\mathbb{R}}\big(\sum_{i=1}^{[nr]}\sum_{j=1}^{m_i} B_{ij}(t)-M_{[nr]}(\hat{F}_{[nr]}(t)-F(t))\big) > (C_1\log n+x)\log n\}$$

$$\le C_2\exp(-C_3 x).$$

Hence combining both positive and negative parts,

$$P\{\sup_{0\le r\le 1}\sup_{t\in\mathbb{R}}\big|M_{[nr]}(\hat{F}_{[nr]}(t)-F(t))-\sum_{i=1}^{[nr]}\sum_{j=1}^{m_i} B_{ij}(t)\big| > (C_1\log n+x)\log n\}$$

$$=P\{\sup_{0\le r\le 1}\sup_{t\in\mathbb{R}}\big(M_{[nr]}(\hat{F}_{[nr]}(t)-F(t))-\sum_{i=1}^{[nr]}\sum_{j=1}^{m_i} B_{ij}(t)\big)^+ > (C_1\log n+x)\log n\}$$

$$+P\{\sup_{0\le r\le 1}\sup_{t\in\mathbb{R}}\big(M_{[nr]}(\hat{F}_{[nr]}(t)-F(t))-\sum_{i=1}^{[nr]}\sum_{j=1}^{m_i} B_{ij}(t)\big)^- > (C_1\log n+x)\log n\}$$

$$\le C_2\exp(-C_3 x).$$

This proved the lemma. □

By Lemma 4.7 and Borel-Cantelli lemma gives that with any $\epsilon > 0$,

$$\limsup_{n\to\infty} \frac{n^{1/2}}{(\log n)^2} \sup_{0\le r\le 1} \sup_{t\in\mathbb{R}} \left| M^{-1/2} M_{[nr]}(\hat{F}_{[nr]}(t) - F(t)) - M^{-1/2} \sum_{i=1}^{[nr]} \sum_{j=1}^{m} B_{ij}(t) \right| \le c + \epsilon \qquad a.s. \tag{4.11}$$

Theorem 4.1. *With the same clustered setting as in Lemma 4.2, as $n \to \infty$, we have*

$$M^{-1/2} M_{[nr]}(\hat{F}_{[nr]}(t) - F(t)) \to_d \tilde{K}(t,r),$$

where $\tilde{K}$ process $\{\tilde{K}(t,r); t \in \mathbb{R}, 0 \le r \le 1\}$ is a separable 2-time parameter real-valued Gaussian process with $\tilde{K}(t,0) = 0$, $E\tilde{K}(t,r) = 0$. And for all $(t_i, r_i) \in \mathbb{R} \times [0,1]$, $i = 1,2$,

$$E\tilde{K}(t_1,r_1)\tilde{K}(t_2,r_2)$$

$$=(r_1 \wedge r_2) \cdot \frac{1}{n} \sum_{i=1}^{n} Cov\left(\sqrt{\frac{n}{M}} \sum_{j=1}^{m_i} \{I(X_{ij} \le t_1) - F(t_1)\}, \sqrt{\frac{n}{M}} \sum_{j=1}^{m_i} \{I(X_{ij} \le t_2) - F(t_2)\} \right),$$

$n \to \infty$.

Proof: Because of the equation (4.11), also because that $B_{ij}(t)$ in (4.11) are Brownian Bridges, and $B_{ij}(t)$ and $B_{i'j'}(t)$ are independent for any $i \ne i'$, we have $M^{-1/2} \sum_{i=1}^{[nr]} \sum_{j=1}^{m} B_{ij}(t)$ converges in distribution to a Gaussian process $\tilde{K}$ indexed by (t,r), which behaves like a Brownian motion in r. □

From Theorem 4.1 we can derived the following properties of some special cases. If we assume each subject has the same number of observations, i.e. $m_i \equiv m$, and observations of every subject follow the same joint distribution with identical correlation coefficient ρ between observations, then the covariance formula stated in the theorem will be simplified as

$$(r_1 \wedge r_2)\cdot\Big(F(t_1)\wedge F(t_2) - F(t_1)F(t_2)$$

$$+ (m-1)\rho\sqrt{F(t_1)(1-F(t_1))}\sqrt{F(t_2)(1-F(t_2))}\Big).$$

Furthermore, if $m = 1$ or $\rho = 0$ it has Kiefer process variance-covariance structure

$$(r_1 \wedge r_2)\cdot\left(F(t_1)\wedge F(t_2) - F(t_1)F(t_2)\right).$$

With previously proved lemmas and theorem, now we look at one clustered ROC sequential empirical process as $n \to \infty$. We have the following by adding and subtracting an intermediate term,

$$\begin{aligned} & N^{-1/2}N_r(\widehat{R}_r(t) - R(t)) && (4.12)\\ =& N^{-1/2}N_r(\hat{S}_{D,r}(\hat{S}^{-1}_{\bar{D},r}(t) - S_D(S^{-1}_{\bar{D}}(t)))) \\ =& N^{-1/2}N_r(\hat{S}_{D,r}(\hat{S}^{-1}_{\bar{D},r}(t)) - S_D(\hat{S}^{-1}_{\bar{D},r}(t))) \\ &+ N^{-1/2}N_r(S_D(\hat{S}^{-1}_{\bar{D},r}(t)) - S_D(S^{-1}_{\bar{D}}(t))). \end{aligned}$$

For the one ROC sequential empirical expression in (4.12), first we know from Theorem 4.1 that

$$N^{-1/2}N_r(\hat{S}_{D,r}(x) - S_D(x)) \to_d W_{S_D}(x, r).$$

where W_{S_D} is a 2-time parameter real-valued Gaussian process indexed by (x, r) as $\tilde{K}$ process defined in Theorem 4.1.

By letting $x = S^{-1}_{\bar{D}}(t)$, we have:

$$N^{-1/2}N_r(\hat{S}_{D,r}(S^{-1}_{\bar{D}}(t)) - S_D(S^{-1}_{\bar{D}}(t))) \to_d W_{S_D}(S^{-1}_{\bar{D}}(t), r).$$

This equation along with Lemma 4.6 and the uniform continuity of the Gaussian process, we have:

$$N^{-1/2}N_r(\hat{S}_{D,r}(\hat{S}^{-1}_{\bar{D},r}(t)) - T(\hat{S}^{-1}_{\bar{D},r}(t))) \to_d W_{S_D}(S^{-1}_{\bar{D}}(t), r), \tag{4.13}$$

which is the first term of (4.12). The second term of (4.12) can be transformed in the following

$$\begin{aligned}
&N^{-1/2}N_r(S_D(\hat{S}^{-1}_{\bar{D},r}(t)) - S_D(S^{-1}_{\bar{D}}(t)))\\
=&N^{-1/2}N_r(S_D(S^{-1}_{\bar{D}}(S_{\bar{D}}(\hat{S}^{-1}_{\bar{D},r}(t)))) - S_D(S^{-1}_{\bar{D}}(t)))\\
=&\frac{N^{-1/2}N_r}{M^{-1/2}M_r}\frac{(S_D(S^{-1}_{\bar{D}}(S_{\bar{D}}(\hat{S}^{-1}_{\bar{D},r}(t)))) - S_D(S^{-1}_{\bar{D}}(t)))}{S_{\bar{D}}(\hat{S}^{-1}_{\bar{D},r}(t)) - t}M^{-1/2}M_r(S_{\bar{D}}(\hat{S}^{-1}_{\bar{D},r}(t)) - t)\\
=&\frac{N^{-1/2}N_r}{M^{-1/2}M_r}\frac{(S_D(S^{-1}_{\bar{D}}(S_{\bar{D}}(\hat{S}^{-1}_{\bar{D},r}(t)))) - S_D(S^{-1}_{\bar{D}}(t)))}{S_{\bar{D}}(\hat{S}^{-1}_{\bar{D},r}(t)) - t}M^{-1/2}M_r(S_{\bar{D}}(\hat{S}^{-1}_{\bar{D},r}(t)) - \hat{S}_{\bar{D},r}(\hat{S}^{-1}_{\bar{D},r}(t)))\\
&+\frac{N^{-1/2}N_r}{M^{-1/2}M_r}\frac{(S_D(S^{-1}_{\bar{D}}(S_{\bar{D}}(\hat{S}^{-1}_{\bar{D},r}(t)))) - S_D(S^{-1}_{\bar{D}}(t)))}{S_{\bar{D}}(\hat{S}^{-1}_{\bar{D},r}(t)) - t}M^{-1/2}M_r(\hat{S}_{\bar{D},r}(\hat{S}^{-1}_{\bar{D},r}(t)) - t).
\end{aligned}$$

Applying Mean Value Theorem and the fact that

$$\frac{d(S_D(S^{-1}_{\bar{D}}(x)))}{dx} = \frac{S'_D(S^{-1}_{\bar{D}}(x))}{S'_{\bar{D}}(S^{-1}_{\bar{D}}(x))},$$

we know there exists a value $S_{\bar{D}}(\tilde{S}^{-1}_{\bar{D},r}(t))$ between $S_{\bar{D}}(\hat{S}^{-1}_{\bar{D},r}(t))$ and t that meets the following condition. Note that $S_{\bar{D}}(\tilde{S}^{-1}_{\bar{D},r}(t))$ can be deemed as the c in the Mean Value Theorem stated in Chapter 2.

$$\frac{(S_D(S_{\bar{D}}^{-1}(S_{\bar{D}}(\hat{S}_{\bar{D},r}^{-1}(t)))) - S_D(S_{\bar{D}}^{-1}(t)))}{S_{\bar{D}}(\hat{S}_{\bar{D},r}^{-1}(t)) - t} = \frac{S'_D(S_{\bar{D}}^{-1}(S_{\bar{D}}(\tilde{S}_{\bar{D},r}^{-1}(t))))}{S'_{\bar{D}}(S_{\bar{D}}^{-1}(S_{\bar{D}}(\tilde{S}_{\bar{D},r}^{-1}(t))))}. \tag{4.14}$$

By Lemma 4.5, we have that $S_{\bar{D}}(\hat{S}_{\bar{D},r}^{-1}(t)) \to_{a.s.} t$, uniformly for $t \in [a,b]$, $and\, r \in [c,1]$. Hence, $S_{\bar{D}}(\tilde{S}_{\bar{D},r}^{-1}(t)) \to_{a.s.} t$, uniformly for $t \in [a,b]$, $r_D \in [c,1]$. Then using the uniform continuity of $\frac{S'_D(S_{\bar{D}}^{-1}(t))}{S'_{\bar{D}}(S_{\bar{D}}^{-1}(t))}$, we have

$$\sup_{c\le r\le 1}\sup_{a\le t\le b}\left|\frac{S'_D(S_{\bar{D}}^{-1}(S_{\bar{D}}(\tilde{S}_{\bar{D},r}^{-1}(t))))}{S'_{\bar{D}}(S_{\bar{D}}^{-1}(S_{\bar{D}}(\tilde{S}_{\bar{D},r}^{-1}(t))))} - \frac{S'_D(S_{\bar{D}}^{-1}(t))}{S'_{\bar{D}}(S_{\bar{D}}^{-1}(t))}\right| \to_{a.s.} 0,$$

by (4.14) it implies,

$$\sup_{c\le r\le 1}\sup_{a\le t\le b}\left|\frac{(S_D(S_{\bar{D}}^{-1}(S_{\bar{D}}(\hat{S}_{\bar{D},r}^{-1}(t)))) - S_D(S_{\bar{D}}^{-1}(t)))}{S_{\bar{D}}(\hat{S}_{\bar{D},r}^{-1}(t)) - t} - \frac{S'_D(S_{\bar{D}}^{-1}(t))}{S'_{\bar{D}}(S_{\bar{D}}^{-1}(t))}\right| \to_{a.s.} 0. \tag{4.15}$$

By definition of $\hat{S}_{\bar{D},r}$, $\hat{S}_{\bar{D},r}^{-1}$, we have for all $r \in [c,1]$,

$$\sup_{a\le t\le b}|\hat{S}_{\bar{D},r}(\hat{S}_{\bar{D},r}^{-1}(t)) - t| \le_{a.s.} \frac{1}{M_r},$$

Therefore,

$$\sup_{c\le r\le 1}\sup_{a\le t\le b} M^{-1/2}M_r|\hat{S}_{\bar{D},r}(\hat{S}_{\bar{D},r}^{-1}(t)) - t| \le_{a.s.} \frac{1}{M^{1/2}},$$

Hence

$$\sup_{c\le r\le 1}\sup_{a\le t\le b} M^{-1/2}M_r|\hat{S}_{\bar{D},r}(\hat{S}_{\bar{D},r}^{-1}(t)) - t| \to_{a.s.} 0. \tag{4.16}$$

And from Lemma 4.6 and the uniform continuity of Gaussian process, we have

$$M^{-1/2}M_r(S_{\bar{D}}(\hat{S}^{-1}_{\bar{D},r}(t)) - \hat{S}_{\bar{D},r}(\hat{S}^{-1}_{\bar{D},r}(t))) \to_d W_{S_{\bar{D}}}(S^{-1}_{\bar{D}}(t)), r). \tag{4.17}$$

By (4.15),(4.16),(4.17), it is easy to see that

$$N^{-1/2}N_r(S_D(\hat{S}^{-1}_{\bar{D},r}(t)) - S_D(S^{-1}_{\bar{D}}(t))) \to_d (\frac{\gamma}{\lambda})^{1/2} \cdot \frac{S'_D(S^{-1}_{\bar{D}}(t))}{S'_{\bar{D}}(S^{-1}_{\bar{D}}(t))} W_{S_{\bar{D}}}(S^{-1}_{\bar{D}}(t), r). \tag{4.18}$$

Applying (4.13) , (4.18) to (4.12) gives the result.

$$N^{-1/2}N_r(\hat{S}_{D,r}(\hat{S}^{-1}_{\bar{D},r}(t) - S_D(S^{-1}_{\bar{D}}(t))))$$

$$\xrightarrow{d} W_{S_D}(S^{-1}_{\bar{D}}(t), r) + (\frac{\gamma}{\lambda})^{1/2} \cdot \frac{S'_D(S^{-1}_{\bar{D}}(t))}{S'_{\bar{D}}(S^{-1}_{\bar{D}}(t))} W_{S_{\bar{D}}}(S^{-1}_{\bar{D}}(t), r), \tag{4.19}$$

where W_{S_D} and $W_{S_{\bar{D}}}$ are Gaussian processes.

From (4.19) we have the following theorem.

Theorem 4.2. *If $S_{\bar{D}}$ and S_D are absolutely continuous survival function (with respect to Lebesgue measure) with a strictly negative derivative functions $S'_{\bar{D}}$ and S'_D on the real line. For $t_1, t_2, \cdots, t_J \in (0,1)$, $r_1, r_2, \cdots, r_J \in (0,1]$, and a vector of arbitrary points on the sequential empirical clustered ROC curve, $(\hat{R}_{r_1}(t_1), \hat{R}_{r_2}(t_2), \cdots, \hat{R}_{r_J}(t_J))^T$ is approximately multivariate normal*

$$\hat{R}_{r_j}(t_j) \sim N(R(t_j), Var(\hat{R}_{r_j}(t_j)), \;\; j = 1, \cdots, J,$$

which has the variance-covariance structure as shown in (4.20), and has the property of $Cov(\hat{R}_{r_i}(t_i), \hat{R}_{r_j}(t_j)) = Cov(\hat{R}_{r_j}(t_i), \hat{R}_{r_j}(t_j))$ for $r_i \leq r_j$.

For simplicity, we define the following notations

$$Cov(X, t_1, X, t_2)$$

$$\triangleq \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} Cov\left(\sqrt{\frac{n}{M}} \sum_{j=1}^{m_i} \{I(X_{ij} > t_1) - S_{\bar{D}}(t_1)\}, \sqrt{\frac{n}{M}} \sum_{j=1}^{m_i} \{I(X_{ij} > t_2) - S_{\bar{D}}(t_2)\}\right),$$

which is the limit of within subject covariance between healthy unit measurements.

$$Cov(Y, t_1, Y, t_2)$$

$$\triangleq \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} Cov\left(\sqrt{\frac{n}{N}} \sum_{j=1}^{n_i} \{I(Y_{ij} > t_1) - S_D(t_1)\}, \sqrt{\frac{n}{N}} \sum_{j=1}^{n_i} \{I(Y_{ij} > t_2) - S_D(t_2)\}\right),$$

which is the limit of within subject covariance between diseased unit measurements.

$$Cov(X, t_1, Y, t_2)$$

$$\triangleq \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} Cov\left(\sqrt{\frac{n}{M}} \sum_{j=1}^{m_i} \{I(X_{ij} > t_1) - S_{\bar{D}}(t_1)\}, \sqrt{\frac{n}{N}} \sum_{j=1}^{n_i} \{I(Y_{ij} > t_2) - S_D(t_2)\}\right),$$

which is the limit of within subject covariance between diseased unit measurements and healthy unit measurements.

Based on the assumption that study subjects are independent, we get the covariance equation,

$$Cov(\hat{R}_{r_i}(t_i), \hat{R}_{r_j}(t_j)) \tag{4.20}$$

$$=\frac{1}{n\gamma r_i r_j}(r_i \wedge r_j)\Big(Cov(Y, S_{\bar{D}}^{-1}(t_i), Y, S_{\bar{D}}^{-1}(t_j))$$

$$+\frac{\gamma}{\lambda}\cdot\left(\frac{S_D'(S_{\bar{D}}^{-1}(t_i))}{S_{\bar{D}}'(S_{\bar{D}}^{-1}(t_i))}\right)\left(\frac{S_D'(S_{\bar{D}}^{-1}(t_j))}{S_{\bar{D}}'(S_{\bar{D}}^{-1}(t_j))}\right)Cov(X,S_{\bar{D}}^{-1}(t_i),X,S_{\bar{D}}^{-1}(t_j))$$

$$+(\frac{\gamma}{\lambda})^{1/2}\cdot\left(\frac{S_D'(S_{\bar{D}}^{-1}(t_j))}{S_{\bar{D}}'(S_{\bar{D}}^{-1}(t_j))}\right)Cov(X,S_{\bar{D}}^{-1}(t_j),Y,S_{\bar{D}}^{-1}(t_i))$$

$$+(\frac{\gamma}{\lambda})^{1/2}\cdot\left(\frac{S_D'(S_{\bar{D}}^{-1}(t_i))}{S_{\bar{D}}'(S_{\bar{D}}^{-1}(t_i))}\right)Cov(X,S_{\bar{D}}^{-1}(t_i),Y,S_{\bar{D}}^{-1}(t_j))\Big).$$

For a special case with $t_i = t_j = t$ as we are often interested in a particular point t on the sequential empirical ROC curvers, we have the following corollary.

Corollary 4.1. *For $t \in (0,1]$, and a vector of points on the sequential empirical clustered ROC curve, $(\hat{R}_{r_1}(t), \hat{R}_{r_2}(t), \cdots, \hat{R}_{r_J}(t))^T$ is approximately multivariate normal*

$$\hat{R}_{r_j}(t) \sim N(R(t), Var(\hat{R}_{r_j}(t)), \;\; j = 1, \cdots, J,$$

and has the variance-covariance structure as shown in (4.21).

$$Cov(\hat{R}_{r_i}(t), \hat{R}_{r_j}(t)) = Var(\hat{R}_{r_j}(t)) \tag{4.21}$$

$$=\frac{1}{n\gamma r_j}\Big(Cov(Y,S_{\bar{D}}^{-1}(t),Y,S_{\bar{D}}^{-1}(t))$$

$$+\frac{\gamma}{\lambda}\cdot\left(\frac{S_D'(S_{\bar{D}}^{-1}(t))}{S_{\bar{D}}'(S_{\bar{D}}^{-1}(t))}\right)^2 Cov(X,S_{\bar{D}}^{-1}(t),X,S_{\bar{D}}^{-1}(t))$$

$$+2\cdot(\frac{\gamma}{\lambda})^{1/2}\cdot\left(\frac{S_D'(S_{\bar{D}}^{-1}(t))}{S_{\bar{D}}'(S_{\bar{D}}^{-1}(t))}\right)Cov(X,S_{\bar{D}}^{-1}(t),Y,S_{\bar{D}}^{-1}(t))\Big),$$

for $r_i \leq r_j$.

Proof: Immediate from Theorem 4.2. □

Corollary 4.2. *For a special clustered dataset where $m_i \equiv n_i \equiv 1$, then for $t_1, t_2, \cdots, t_J \in$*

$(0,1)$, $r_1, r_2, \cdots, r_J \in (0,1]$, *and a vector of arbitrary points on the sequential empirical clustered ROC curve,* $(\hat{R}_{r_1}(t_1), \hat{R}_{r_2}(t_2), \cdots, \hat{R}_{r_J}(t_J))^T$ *is approximately multivariate normal*

$$\hat{R}_{r_j}(t_j) \sim N(R(t_j), Var(\hat{R}_{r_j}(t_j)), \;\; j = 1, \cdots, J,$$

which has the variance-covariance structure as shown in (4.22)*, and also has the property of* $Cov(\hat{R}_{r_i}(t_i), \hat{R}_{r_j}(t_j)) = Cov(\hat{R}_{r_j}(t_i), \hat{R}_{r_j}(t_j))$ *for* $r_i \leq r_j$.

$$\begin{aligned}
&Cov(\hat{R}_{r_i}(t_i), \hat{R}_{r_j}(t_j)) \qquad (4.22)\\
=&\frac{1}{nr_ir_j}(r_i \wedge r_j)\Bigg((R(t_i) \wedge R(t_j) - R(t_i)R(t_j))\\
&+\left(\frac{S'_D(S_{\bar{D}}^{-1}(t_i))}{S'_{\bar{D}}(S_{\bar{D}}^{-1}(t_i))}\right)\left(\frac{S'_D(S_{\bar{D}}^{-1}(t_j))}{S'_{\bar{D}}(S_{\bar{D}}^{-1}(t_j))}\right)(t_i \wedge t_j - t_it_j)\\
&+\left(\frac{S'_D(S_{\bar{D}}^{-1}(t_j))}{S'_{\bar{D}}(S_{\bar{D}}^{-1}(t_j))}\right)(S_{\bar{D},D}(S_{\bar{D}}^{-1}(t_j), S_{\bar{D}}^{-1}(t_i)) - t_jR(t_i))\\
&+\left(\frac{S'_D(S_{\bar{D}}^{-1}(t_i))}{S'_{\bar{D}}(S_{\bar{D}}^{-1}(t_i))}\right)(S_{\bar{D},D}(S_{\bar{D}}^{-1}(t_i), S_{\bar{D}}^{-1}(t_j)) - t_iR(t_j))\Bigg).
\end{aligned}$$

where $S_{\bar{D},D}$ is the joint survival function on healthy and diseased unit measurements.

Proof: Immediate from Theorem 4.2. □

4.2.2 Comparison of Clustered ROCs

In a comparison study of clustered ROCs, we have a total of n subjects in the study. Within each subject i, we observe $X_{ij}^{(v)}$, $j = 1, \cdots, m_i^{(v)}$, which are the measurements of vth marker from $m_i^{(v)}$ healthy units within subject i. And we observe $Y_{ij}^{(v)}$, $j = 1, \cdots, n_i^{(v)}$, which are the measurements of vth marker from $n_i^{(v)}$ diseased units within subject i, $i = 1, \cdots, n$

and $v = 1, 2$ representing different biomarkers. We further assume that the observations $X_{ij}^{(v)}$, $j = 1, \cdots, m_i^{(v)}$ follow the survival function $S_{\bar{D}}^{(v)}$, and the observations $Y_{ij}^{(v)}$, $j = 1, \cdots, n_i^{(v)}$ follow the survival function $S_D^{(v)}$.

And we assume that measurements from different subjects are independent and measures within the same subject are possibly correlated. These will generate clustered ROC data. In this setting, we allows for both between-biomarker and within-biomarker within-subject correlations. Different markers might have different numbers of measurements per disease/non-disease group per subject.

For simplicity, we define $M_r^{(v)} = \sum_{i=1}^{[nr]} m_i^{(v)}$, $M^{(v)} = \sum_{i=1}^{n} m_i^{(v)}$, $N_r^{(v)} = \sum_{i=1}^{[nr]} n_i^{(v)}$, $N^{(v)} = \sum_{i=1}^{n} n_i^{(v)}$, where r represents the percentage of subjects accrued so far at this analysis point. And assume that as $n \to \infty$, $n^{-1}\sum_{i=1}^{n} m_i^{(v)} \to \lambda^{(v)}$, and $n^{-1}\sum_{i=1}^{n} n_i^{(v)} \to \gamma^{(v)}$ for some positive constants $\lambda^{(v)}$ and $\gamma^{(v)}$, for $v = 1, 2$. The following theory is needed to establish the limiting distribution of $(\widehat{R^{(1)}}(t), \widehat{R^{(2)}}(t))$ at finite number of interim analysis points.

$$\begin{pmatrix} M^{(1)^{-1/2}} M_{r_1}^{(1)} (\widehat{S}_{\bar{D},r_1}^{(1)}(t) - S_{\bar{D}}^{(1)}(t)) \\ M^{(2)^{-1/2}} M_{r_2}^{(2)} (\widehat{S}_{\bar{D},r_2}^{(2)}(t) - S_{\bar{D}}^{(2)}(t)) \\ N^{(1)^{-1/2}} N_{r_3}^{(1)} (\widehat{S}_{D,r_3}^{(1)}(t) - S_{D}^{(1)}(t)) \\ N^{(2)^{-1/2}} N_{r_4}^{(2)} (\widehat{S}_{D,r_4}^{(2)}(t) - S_{D}^{(2)}(t)) \end{pmatrix} \tag{4.23}$$

$$= \begin{pmatrix} \sqrt{\frac{M_{r_1}^{(1)}}{M^{(1)}}} \cdot [nr_1]^{-1/2} \sum_{i=1}^{[nr_1]} \sqrt{\frac{[nr_1]}{M_{r_1}^{(1)}}} \sum_{j=1}^{m_i^{(1)}} \{I(X_{ij}^{(1)} > t) - S_{\bar{D}}^{(1)}(t)\} \\ \sqrt{\frac{M_{r_2}^{(2)}}{M^{(2)}}} \cdot [nr_2]^{-1/2} \sum_{i=1}^{[nr_2]} \sqrt{\frac{[nr_2]}{M_{r_2}^{(2)}}} \sum_{j=1}^{m_i^{(2)}} \{I(X_{ij}^{(2)} > t) - S_{\bar{D}}^{(2)}(t)\} \\ \sqrt{\frac{N_{r_3}^{(1)}}{N^{(1)}}} \cdot [nr_3]^{-1/2} \sum_{i=1}^{[nr_3]} \sqrt{\frac{[nr_3]}{N_{r_3}^{(1)}}} \sum_{j=1}^{n_i^{(1)}} \{I(Y_{ij}^{(1)} > t) - S_{D}^{(1)}(t)\} \\ \sqrt{\frac{N_{r_4}^{(2)}}{N^{(2)}}} \cdot [nr_4]^{-1/2} \sum_{i=1}^{[nr_4]} \sqrt{\frac{[nr_4]}{N_{r_4}^{(2)}}} \sum_{j=1}^{n_i^{(2)}} \{I(Y_{ij}^{(2)} > t) - S_{D}^{(2)}(t)\} \end{pmatrix}$$

Let

$$V_i(t) = \begin{pmatrix} \sqrt{\frac{n}{M^{(1)}}} \sum_{j=1}^{m_i^{(1)}} \{I(X_{ij}^{(1)} > t) - S_{\bar{D}}^{(1)}(t)\} \\ \sqrt{\frac{n}{M^{(2)}}} \sum_{j=1}^{m_i^{(2)}} \{I(X_{ij}^{(2)} > t) - S_{\bar{D}}^{(2)}(t)\} \\ \sqrt{\frac{n}{N^{(1)}}} \sum_{j=1}^{n_i^{(1)}} \{I(Y_{ij}^{(1)} > t) - S_{D}^{(1)}(t)\} \\ \sqrt{\frac{n}{N^{(2)}}} \sum_{j=1}^{n_i^{(2)}} \{I(Y_{ij}^{(2)} > t) - S_{D}^{(2)}(t)\} \end{pmatrix}, \qquad i = 1, \cdots, n,$$

which are independent random vectors for $i = 1, \cdots, n$. Applying the Cramer-Wold device and the Lyapunov central limit theorem, and the result of Csörgő and Szyszkowicz (1998) for sequential empirical distribution processes, it can be show that (4.23) $\xrightarrow{d}$ $\mathbf{W}(t, r)$ in $D(\mathbb{R} \times [0, 1])^4$, where

$$\mathbf{W}(t, r) = \begin{pmatrix} W_{S_{\bar{D}}^{(1)}}(t, r_1) \\ W_{S_{\bar{D}}^{(2)}}(t, r_2) \\ W_{S_{D}^{(1)}}(t, r_3) \\ W_{S_{D}^{(2)}}(t, r_4) \end{pmatrix} \tag{4.24}$$

is a mean-zero Gaussian process in $D(\mathbb{R} \times [0, 1])^4$, whose variance-covariance function is as following. The scalar part is

$$\begin{pmatrix} r_1 & r_1 \wedge r_2 & r_1 \wedge r_3 & r_1 \wedge r_4 \\ r_2 \wedge r_1 & r_2 & r_2 \wedge r_3 & r_2 \wedge r_4 \\ r_3 \wedge r_1 & r_3 \wedge r_2 & r_3 & r_3 \wedge r_4 \\ r_4 \wedge r_1 & r_4 \wedge r_2 & r_4 \wedge r_3 & r_4 \end{pmatrix},$$

and the covariance part, a 4×4 matrix, is the limit of

$$\frac{1}{n}\sum_{i=1}^{n} Cov(V_i(t), V_i(t)) \quad \text{as } n \to \infty.$$

Each component of the vector is a marginal mean-zero Gaussian process. Take $W_{S_{\bar{D}}^{(1)}}(t, r)$ as an example, at any two index (t_1, r_1) and (t_2, r_2) of this process, its covariance is the limit of

$$(r_1 \wedge r_2) \cdot \frac{1}{n}\sum_{i=1}^{n} Cov\left(\sqrt{\frac{n}{M^{(1)}}}\sum_{j=1}^{m_i^{(1)}}\{I(X_{ij}^{(1)} > t_1) - S_{\bar{D}}^{(1)}(t_1)\}, \sqrt{\frac{n}{M^{(1)}}}\sum_{j=1}^{m_i^{(1)}}\{I(X_{ij}^{(1)} > t_2) - S_{\bar{D}}^{(1)}(t_2)\}\right)$$

as $n \to \infty$.

If assume each subject has the same number of observations, i.e. $m_i^{(1)} \equiv m$, and observations of every subject follow the same joint distribution with identical correlation coefficient ρ between observations, then the aforementioned formula will be simplified as

$$(r_1 \wedge r_2) \cdot \Big(S_{\bar{D}}^{(1)}(t_1) \wedge S_{\bar{D}}^{(1)}(t_2) - S_{\bar{D}}^{(1)}(t_1) S_{\bar{D}}^{(1)}(t_2)$$
$$+ (m-1)\rho \sqrt{S_{\bar{D}}^{(1)}(t_1)(1 - S_{\bar{D}}^{(1)}(t_1))} \sqrt{S_{\bar{D}}^{(1)}(t_2)(1 - S_{\bar{D}}^{(1)}(t_2))} \Big).$$

When $m = 1$ or $\rho = 0$ it has Kiefer process variance-covariance structure

$$(r_1 \wedge r_2) \cdot \left(S_{\bar{D}}^{(1)}(t_1) \wedge S_{\bar{D}}^{(1)}(t_2) - S_{\bar{D}}^{(1)}(t_1) S_{\bar{D}}^{(1)}(t_2) \right).$$

We further assume that for $v = 1, 2$, $S_{\bar{D}}^{(v)}$ and $S_D^{(v)}$ have derivatives $S_{\bar{D}}^{(v)\prime}$ and $S_D^{(v)\prime}$ respectively which are negative and continuous on $[S_{\bar{D}}^{(v)^{-1}}(b) - \epsilon, S_{\bar{D}}^{(v)^{-1}}(a) + \epsilon]$, for some $0 < a < b < 1$ and $\epsilon > 0$. Then as $n \to \infty$, by (4.24), the compact differentiability of the inverse function and the functional delta method Theorem 3.9.4 of van der Vaart and Wellner (1996),

$$\begin{pmatrix} M^{(1)^{-1/2}} M_r^{(1)} (\widehat{S}_{\bar{D},r}^{(1)^{-1}}(t) - S_{\bar{D}}^{(1)^{-1}}(t)) \\ M^{(2)^{-1/2}} M_r^{(2)} (\widehat{S}_{\bar{D},r}^{(2)^{-1}}(t) - S_{\bar{D}}^{(2)^{-1}}(t)) \\ N^{(1)^{-1/2}} N_r^{(1)} (\widehat{S}_{D,r}^{(1)}(t) - S_D^{(1)}(t)) \\ N^{(2)^{-1/2}} N_r^{(2)} (\widehat{S}_{D,r}^{(2)}(t) - S_D^{(2)}(t)) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \frac{W_{S_{\bar{D}}^{(1)}}(S_{\bar{D}}^{(1)^{-1}}(t),r)}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))} \\ \frac{W_{S_{\bar{D}}^{(2)}}(S_{\bar{D}}^{(2)^{-1}}(t),r)}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))} \\ W_{S_D^{(1)}}(t,r) \\ W_{S_D^{(2)}}(t,r) \end{pmatrix}$$

in $D([a,b]\times[0,1])\times D([a,b]\times[0,1])\times D([S_{\bar{D}}^{(1)^{-1}}(b), S_{\bar{D}}^{(1)^{-1}}(a)]\times[0,1])\times D([S_{\bar{D}}^{(2)^{-1}}(b), S_{\bar{D}}^{(2)^{-1}}(a)]\times [0,1])$ as $n \to \infty$. Furthermore,

$$\begin{pmatrix} N^{(1)^{-1/2}} N_r^{(1)} (\widehat{S}_{\bar{D},r}^{(1)^{-1}}(t) - S_{\bar{D}}^{(1)^{-1}}(t)) \\ N^{(2)^{-1/2}} N_r^{(2)} (\widehat{S}_{\bar{D},r}^{(2)^{-1}}(t) - S_{\bar{D}}^{(2)^{-1}}(t)) \\ N^{(1)^{-1/2}} N_r^{(1)} (\widehat{S}_{D,r}^{(1)}(t) - S_D^{(1)}(t)) \\ N^{(2)^{-1/2}} N_r^{(2)} (\widehat{S}_{D,r}^{(2)}(t) - S_D^{(2)}(t)) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} (\frac{\gamma^{(1)}}{\lambda^{(1)}})^{1/2} \cdot \frac{W_{S_{\bar{D}}^{(1)}}(S_{\bar{D}}^{(1)^{-1}}(t),r)}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))} \\ (\frac{\gamma^{(2)}}{\lambda^{(2)}})^{1/2} \cdot \frac{W_{S_{\bar{D}}^{(2)}}(S_{\bar{D}}^{(2)^{-1}}(t),r)}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))} \\ W_{S_D^{(1)}}(t,r) \\ W_{S_D^{(2)}}(t,r) \end{pmatrix}$$

Combining this result and Lemma 3.9.27 of van der Vaart and Wellner (1996) and the functional delta method implies that,

$$\begin{pmatrix} N^{(1)^{-1/2}} N_r^{(1)} (\widehat{S}_{D,r}^{(1)}(\widehat{S}_{\bar{D},r}^{(1)^{-1}}(t)) - S_D^{(1)}(S_{\bar{D}}^{(1)^{-1}}(t))) \\ N^{(2)^{-1/2}} N_r^{(2)} (\widehat{S}_{D,r}^{(2)}(\widehat{S}_{\bar{D},r}^{(2)^{-1}}(t)) - S_D^{(2)}(S_{\bar{D}}^{(2)^{-1}}(t))) \end{pmatrix}$$

$$\xrightarrow{d} \begin{pmatrix} W_{S_D^{(1)}}(S_{\bar{D}}^{(1)^{-1}}(t), r) + (\frac{\gamma^{(1)}}{\lambda^{(1)}})^{1/2} \cdot \frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))} W_{S_{\bar{D}}^{(1)}}(S_{\bar{D}}^{(1)^{-1}}(t), r) \\ \\ W_{S_D^{(2)}}(S_{\bar{D}}^{(2)^{-1}}(t), r) + (\frac{\gamma^{(2)}}{\lambda^{(2)}})^{1/2} \cdot \frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))} W_{S_{\bar{D}}^{(2)}}(S_{\bar{D}}^{(2)^{-1}}(t), r) \\ \end{pmatrix}$$

in $D([a, b] \times [0, 1])^2$. By expanding the vector to include two analysis points, r and r', we have

$$\begin{pmatrix} N^{(1)^{-1/2}} N_r^{(1)}(\widehat{S}_{D,r}^{(1)}(\widehat{S}_{\bar{D},r}^{(1)^{-1}}(t)) - S_D^{(1)}(S_{\bar{D}}^{(1)^{-1}}(t))) \\ N^{(2)^{-1/2}} N_r^{(2)}(\widehat{S}_{D,r}^{(2)}(\widehat{S}_{\bar{D},r}^{(2)^{-1}}(t)) - S_D^{(2)}(S_{\bar{D}}^{(2)^{-1}}(t))) \\ N^{(1)^{-1/2}} N_{r'}^{(1)}(\widehat{S}_{D,r'}^{(1)}(\widehat{S}_{\bar{D},r'}^{(1)^{-1}}(t)) - S_D^{(1)}(S_{\bar{D}}^{(1)^{-1}}(t))) \\ N^{(2)^{-1/2}} N_{r'}^{(2)}(\widehat{S}_{D,r'}^{(2)}(\widehat{S}_{\bar{D},r'}^{(2)^{-1}}(t)) - S_D^{(2)}(S_{\bar{D}}^{(2)^{-1}}(t))) \end{pmatrix} \tag{4.25}$$

$$\xrightarrow{d} \begin{pmatrix} W_{S_D^{(1)}}(S_{\bar{D}}^{(1)^{-1}}(t), r) + (\frac{\gamma^{(1)}}{\lambda^{(1)}})^{1/2} \cdot \frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))} W_{S_{\bar{D}}^{(1)}}(S_{\bar{D}}^{(1)^{-1}}(t), r) \\ W_{S_D^{(2)}}(S_{\bar{D}}^{(2)^{-1}}(t), r) + (\frac{\gamma^{(2)}}{\lambda^{(2)}})^{1/2} \cdot \frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))} W_{S_{\bar{D}}^{(2)}}(S_{\bar{D}}^{(2)^{-1}}(t), r) \\ W_{S_D^{(1)}}(S_{\bar{D}}^{(1)^{-1}}(t), r') + (\frac{\gamma^{(1)}}{\lambda^{(1)}})^{1/2} \cdot \frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))} W_{S_{\bar{D}}^{(1)}}(S_{\bar{D}}^{(1)^{-1}}(t), r') \\ W_{S_D^{(2)}}(S_{\bar{D}}^{(2)^{-1}}(t), r') + (\frac{\gamma^{(2)}}{\lambda^{(2)}})^{1/2} \cdot \frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))} W_{S_{\bar{D}}^{(2)}}(S_{\bar{D}}^{(2)^{-1}}(t), r') \end{pmatrix}$$

The proof of the marginal univariate process convergence is presented in Section 4.2.1. To prove the convergence of the random vector, we will also need to prove the tightness of the left-hand side of (4.25). By the Lemma 2.1 in Chapter 2, we can prove that the multivariate stochastic process is tight from the fact that each marginal univariate stochastic process is tight.

Through some modification, we have

$$\begin{pmatrix} N^{(1)^{-1/2}} N_r^{(1)} (\widehat{S}_{D,r}^{(1)} (\widehat{S}_{\bar{D},r}^{(1)^{-1}}(t)) - S_D^{(1)}(S_{\bar{D}}^{(1)^{-1}}(t))) \\ N^{(1)^{-1/2}} N_r^{(1)} (\widehat{S}_{D,r}^{(2)} (\widehat{S}_{\bar{D},r}^{(2)^{-1}}(t)) - S_D^{(2)}(S_{\bar{D}}^{(2)^{-1}}(t))) \\ N^{(1)^{-1/2}} N_{r'}^{(1)} (\widehat{S}_{D,r'}^{(1)} (\widehat{S}_{\bar{D},r'}^{(1)^{-1}}(t)) - S_D^{(1)}(S_{\bar{D}}^{(1)^{-1}}(t))) \\ N^{(1)^{-1/2}} N_{r'}^{(1)} (\widehat{S}_{D,r'}^{(2)} (\widehat{S}_{\bar{D},r'}^{(2)^{-1}}(t)) - S_D^{(2)}(S_{\bar{D}}^{(2)^{-1}}(t))) \end{pmatrix} \triangleq \mathbf{V}$$

$$\xrightarrow{d} \begin{pmatrix} W_{S_D^{(1)}}(S_{\bar{D}}^{(1)^{-1}}(t), r) + (\frac{\gamma^{(1)}}{\lambda^{(1)}})^{1/2} \cdot \frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))} W_{S_{\bar{D}}^{(1)}}(S_{\bar{D}}^{(1)^{-1}}(t), r) \\ (\frac{\gamma^{(1)}}{\gamma^{(2)}})^{1/2} W_{S_D^{(2)}}(S_{\bar{D}}^{(2)^{-1}}(t), r) + (\frac{\gamma^{(1)}}{\lambda^{(2)}})^{1/2} \cdot \frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))} W_{S_{\bar{D}}^{(2)}}(S_{\bar{D}}^{(2)^{-1}}(t), r) \\ W_{S_D^{(1)}}(S_{\bar{D}}^{(1)^{-1}}(t), r') + (\frac{\gamma^{(1)}}{\lambda^{(1)}})^{1/2} \cdot \frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))} W_{S_{\bar{D}}^{(1)}}(S_{\bar{D}}^{(1)^{-1}}(t), r') \\ (\frac{\gamma^{(1)}}{\gamma^{(2)}})^{1/2} W_{S_D^{(2)}}(S_{\bar{D}}^{(2)^{-1}}(t), r') + (\frac{\gamma^{(1)}}{\lambda^{(2)}})^{1/2} \cdot \frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))} W_{S_{\bar{D}}^{(2)}}(S_{\bar{D}}^{(2)^{-1}}(t), r') \end{pmatrix}$$

Now we want to prove that $Cov(\hat{\Delta}_r(t), \hat{\Delta}_{r'}(t)) = Var(\hat{\Delta}_{r'}(t))$, for $r \leq r'$. We define

$$\mathbf{Y} \triangleq \begin{pmatrix} N^{(1)^{-1/2}} N_r^{(1)} (\hat{\Delta}_r(t) - \hat{\Delta}_r(t)) \\ N^{(1)^{-1/2}} N_{r'}^{(1)} (\hat{\Delta}_{r'}(t) - \hat{\Delta}_{r'}(t)) \end{pmatrix}$$

$$= \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \mathbf{V} \, .$$

The random vector $\mathbf{V}$ is asymptotically multivariate normal with covariance $Cov(\mathbf{V})$, or $\Sigma = \{a_{ij}\}_{i=1,\cdots,4;\, j=1,\cdots,4.}$ Hence the random vector $\mathbf{Y}$ is asymptotically normal with covariance matrix derived approximately in the following.

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \Sigma \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} + a_{22} - 2a_{12} & a_{13} + a_{24} - a_{14} - a_{23} \\ a_{13} + a_{24} - a_{14} - a_{23} & a_{33} + a_{44} - 2a_{34} \end{pmatrix}.$$

Then we have,

$$Cov(\hat{\Delta}_r(t), \hat{\Delta}_{r'}(t)) = N^{(1)} \frac{1}{N_r^{(1)}} \frac{1}{N_{r'}^{(1)}} (a_{13} + a_{24} - a_{14} - a_{23}), \tag{4.26}$$

and

$$Var(\hat{\Delta}_{r'}(t)) = N^{(1)} \frac{1}{N_{r'}^{(1)}} \frac{1}{N_{r'}^{(1)}} (a_{33} + a_{44} - 2a_{34}). \tag{4.27}$$

For simplicity, we define the following notations for $v_1, v_2 = 1, 2$. For the limit of within subject covariance between healthy unit measurements, whether of the same marker or not, we define

$$Cov(X^{(v_1)}, t_1, X^{(v_2)}, t_2)$$

$$\triangleq \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} Cov\Big(\sqrt{\frac{n}{M^{(v_1)}}} \sum_{j=1}^{m_i^{(v_1)}} \{I(X_{ij}^{(v_1)} > t_1) - S_{\bar{D}}^{(v_1)}(t_1)\},$$

$$\sqrt{\frac{n}{M^{(v_2)}}} \sum_{j=1}^{m_i^{(v_2)}} \{I(X_{ij}^{(v_2)} > t_2) - S_{\bar{D}}^{(v_2)}(t_2)\}\Big).$$

For the limit of within subject covariance between diseased unit measurements, whether of the same marker or not, we define

$$Cov(Y^{(v_1)}, t_1, Y^{(v_2)}, t_2)$$

$$\triangleq \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} Cov\Big(\sqrt{\frac{n}{N^{(v_1)}}} \sum_{j=1}^{n_i^{(v_1)}} \{I(Y_{ij}^{(v_1)} > t_1) - S_{D}^{(v_1)}(t_1)\},$$

$$\sqrt{\frac{n}{N^{(v_2)}}}\sum_{j=1}^{n_i^{(v_2)}}\{I(Y_{ij}^{(v_2)} > t_2) - S_D^{(v_2)}(t_2)\}\Big).$$

For the limit of within subject covariance between diseased unit measurements and healthy unit measurements, whether of the same marker or not, we define

$$\begin{aligned}
&Cov(X^{(v_1)}, t_1, Y^{(v_2)}, t_2) \\
&\triangleq \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} Cov\Big(\sqrt{\frac{n}{M^{(v_1)}}}\sum_{j=1}^{m_i^{(v_1)}}\{I(X_{ij}^{(v_1)} > t_1) - S_{\bar{D}}^{(v_1)}(t_1)\}, \\
&\qquad \sqrt{\frac{n}{N^{(v_2)}}}\sum_{j=1}^{n_i^{(v_2)}}\{I(Y_{ij}^{(v_2)} > t_2) - S_D^{(v_2)}(t_2)\}\Big).
\end{aligned}$$

Expanding the (4.26), for each item in the equation we can derive the following based on the assumption that study subjects are independent.

$$\begin{aligned}
&\frac{1}{r}a_{13} \\
=&\frac{1}{r}(r\wedge r')\Bigg(Cov(Y^{(1)}, S_{\bar{D}}^{(1)^{-1}}(t), Y^{(1)}, S_{\bar{D}}^{(1)^{-1}}(t)) \\
&+\frac{\gamma^{(1)}}{\lambda^{(1)}}\cdot\left(\frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}\right)^2 Cov(X^{(1)}, S_{\bar{D}}^{(1)^{-1}}(t), X^{(1)}, S_{\bar{D}}^{(1)^{-1}}(t)) \\
&+2\cdot(\frac{\gamma^{(1)}}{\lambda^{(1)}})^{1/2}\cdot\left(\frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}\right) Cov(X^{(1)}, S_{\bar{D}}^{(1)^{-1}}(t), Y^{(1)}, S_{\bar{D}}^{(1)^{-1}}(t))\Bigg),
\end{aligned}$$

$$\frac{1}{r}a_{24}$$

$$=\frac{1}{r}(r\wedge r')\Bigg(\frac{\gamma^{(1)}}{\gamma^{(2)}}Cov(Y^{(2)},S_{\bar{D}}^{(2)^{-1}}(t),Y^{(2)},S_{\bar{D}}^{(2)^{-1}}(t))$$

$$+\frac{\gamma^{(1)}}{\lambda^{(2)}}\cdot\left(\frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}\right)^2 Cov(X^{(2)},S_{\bar{D}}^{(2)^{-1}}(t),X^{(2)},S_{\bar{D}}^{(2)^{-1}}(t))$$

$$+2\cdot\gamma^{(1)}(\frac{1}{\gamma^{(2)}\lambda^{(2)}})^{1/2}\cdot\left(\frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}\right)Cov(X^{(2)},S_{\bar{D}}^{(2)^{-1}}(t),Y^{(2)},S_{\bar{D}}^{(2)^{-1}}(t))\Bigg),$$

$$\frac{1}{r}a_{14}$$

$$=\frac{1}{r}(r\wedge r')\Bigg((\frac{\gamma^{(1)}}{\lambda^{(2)}})^{1/2}Cov(Y^{(1)},S_{\bar{D}}^{(1)^{-1}}(t),Y^{(2)},S_{\bar{D}}^{(2)^{-1}}(t))$$

$$+\gamma^{(1)}(\frac{1}{\lambda^{(1)}\lambda^{(2)}})^{1/2}\cdot\left(\frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}\right)\left(\frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}\right)Cov(X^{(1)},S_{\bar{D}}^{(1)^{-1}}(t),X^{(2)},S_{\bar{D}}^{(2)^{-1}}(t))$$

$$+(\frac{\gamma^{(1)}}{\lambda^{(2)}})^{1/2}\cdot\left(\frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}\right)Cov(X^{(2)},S_{\bar{D}}^{(2)^{-1}}(t),Y^{(1)},S_{\bar{D}}^{(1)^{-1}}(t))$$

$$+\gamma^{(1)}(\frac{1}{\lambda^{(1)}\gamma^{(2)}})^{1/2}\cdot\left(\frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}\right)Cov(X^{(1)},S_{\bar{D}}^{(1)^{-1}}(t),Y^{(2)},S_{\bar{D}}^{(2)^{-1}}(t))\Bigg),$$

and

$$\frac{1}{r}a_{23}$$

$$=\frac{1}{r}(r\wedge r')\Bigg((\frac{\gamma^{(1)}}{\lambda^{(2)}})^{1/2}Cov(Y^{(1)},S_{\bar{D}}^{(1)^{-1}}(t),Y^{(2)},S_{\bar{D}}^{(2)^{-1}}(t))$$

$$+ \gamma^{(1)}(\frac{1}{\lambda^{(1)}\lambda^{(2)}})^{1/2} \cdot \left(\frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}\right) \left(\frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}\right) Cov(X^{(1)}, S_{\bar{D}}^{(1)^{-1}}(t), X^{(2)}, S_{\bar{D}}^{(2)^{-1}}(t))$$

$$+ (\frac{\gamma^{(1)}}{\lambda^{(2)}})^{1/2} \cdot \left(\frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}\right) Cov(X^{(2)}, S_{\bar{D}}^{(2)^{-1}}(t), Y^{(1)}, S_{\bar{D}}^{(1)^{-1}}(t))$$

$$+ \gamma^{(1)}(\frac{1}{\lambda^{(1)}\gamma^{(2)}})^{1/2} \cdot \left(\frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}\right) Cov(X^{(1)}, S_{\bar{D}}^{(1)^{-1}}(t), Y^{(2)}, S_{\bar{D}}^{(2)^{-1}}(t))\Bigg).$$

And similarly for (4.27), we expand for each item in the equation in the following:

$$\frac{1}{r'}a_{33}$$

$$=\frac{1}{r'}r'\Bigg(Cov(Y^{(1)}, S_{\bar{D}}^{(1)^{-1}}(t), Y^{(1)}, S_{\bar{D}}^{(1)^{-1}}(t))$$

$$+ \frac{\gamma^{(1)}}{\lambda^{(1)}} \cdot \left(\frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}\right)^2 Cov(X^{(1)}, S_{\bar{D}}^{(1)^{-1}}(t), X^{(1)}, S_{\bar{D}}^{(1)^{-1}}(t))$$

$$+ 2 \cdot (\frac{\gamma^{(1)}}{\lambda^{(1)}})^{1/2} \cdot \left(\frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}\right) Cov(X^{(1)}, S_{\bar{D}}^{(1)^{-1}}(t), Y^{(1)}, S_{\bar{D}}^{(1)^{-1}}(t))\Bigg),$$

$$\frac{1}{r'}a_{44}$$

$$=\frac{1}{r'}r'\Bigg(\frac{\gamma^{(1)}}{\gamma^{(2)}}Cov(Y^{(2)}, S_{\bar{D}}^{(2)^{-1}}(t), Y^{(2)}, S_{\bar{D}}^{(2)^{-1}}(t))$$

$$+ \frac{\gamma^{(1)}}{\lambda^{(2)}} \cdot \left(\frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}\right)^2 Cov(X^{(2)}, S_{\bar{D}}^{(2)^{-1}}(t), X^{(2)}, S_{\bar{D}}^{(2)^{-1}}(t))$$

$$+ 2 \cdot \gamma^{(1)}(\frac{1}{\gamma^{(2)}\lambda^{(2)}})^{1/2} \cdot \left(\frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}\right) Cov(X^{(2)}, S_{\bar{D}}^{(2)^{-1}}(t), Y^{(2)}, S_{\bar{D}}^{(2)^{-1}}(t))\Bigg),$$

and

$$
\begin{aligned}
&\frac{1}{r}a_{34}\\
=&\frac{1}{r'}r'\Bigg((\frac{\gamma^{(1)}}{\lambda^{(2)}})^{1/2}Cov(Y^{(1)},S_{\bar{D}}^{(1)^{-1}}(t),Y^{(2)},S_{\bar{D}}^{(2)^{-1}}(t))\\
&+\gamma^{(1)}(\frac{1}{\lambda^{(1)}\lambda^{(2)}})^{1/2}\cdot\left(\frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}\right)\left(\frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}\right)Cov(X^{(1)},S_{\bar{D}}^{(1)^{-1}}(t),X^{(2)},S_{\bar{D}}^{(2)^{-1}}(t))\\
&+(\frac{\gamma^{(1)}}{\lambda^{(2)}})^{1/2}\cdot\left(\frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}\right)Cov(X^{(2)},S_{\bar{D}}^{(2)^{-1}}(t),Y^{(1)},S_{\bar{D}}^{(1)^{-1}}(t))\\
&+\gamma^{(1)}(\frac{1}{\lambda^{(1)}\gamma^{(2)}})^{1/2}\cdot\left(\frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}\right)Cov(X^{(1)},S_{\bar{D}}^{(1)^{-1}}(t),Y^{(2)},S_{\bar{D}}^{(2)^{-1}}(t))\Bigg).
\end{aligned}
$$

Summing up the above expanded items, we then get (4.26) and prove that it equals (4.27) as following.

$$
\begin{aligned}
&Cov(\hat{\Delta}_r(t),\hat{\Delta}_{r'}(t)) \qquad (4.28)\\
=&N^{(1)}\frac{1}{N_r^{(1)}}\frac{1}{N_{r'}^{(1)}}(a_{13}+a_{24}-a_{14}-a_{23})\\
=&\frac{1}{nr'\gamma^{(1)}}\Bigg\{\Bigg(Cov(Y^{(1)},S_{\bar{D}}^{(1)^{-1}}(t),Y^{(1)},S_{\bar{D}}^{(1)^{-1}}(t))\\
&+\frac{\gamma^{(1)}}{\lambda^{(1)}}\cdot\left(\frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}\right)^2Cov(X^{(1)},S_{\bar{D}}^{(1)^{-1}}(t),X^{(1)},S_{\bar{D}}^{(1)^{-1}}(t))\\
&+2\cdot(\frac{\gamma^{(1)}}{\lambda^{(1)}})^{1/2}\cdot\left(\frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}\right)Cov(X^{(1)},S_{\bar{D}}^{(1)^{-1}}(t),Y^{(1)},S_{\bar{D}}^{(1)^{-1}}(t))\Bigg)\\
&+\Bigg(\frac{\gamma^{(1)}}{\gamma^{(2)}}Cov(Y^{(2)},S_{\bar{D}}^{(2)^{-1}}(t),Y^{(2)},S_{\bar{D}}^{(2)^{-1}}(t))
\end{aligned}
$$

$$+\frac{\gamma^{(1)}}{\lambda^{(2)}}\cdot\left(\frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}\right)^2 Cov(X^{(2)},S_{\bar{D}}^{(2)^{-1}}(t),X^{(2)},S_{\bar{D}}^{(2)^{-1}}(t))$$

$$+2\cdot\gamma^{(1)}(\frac{1}{\gamma^{(2)}\lambda^{(2)}})^{1/2}\cdot\left(\frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}\right) Cov(X^{(2)},S_{\bar{D}}^{(2)^{-1}}(t),Y^{(2)},S_{\bar{D}}^{(2)^{-1}}(t))\Bigg)$$

$$-2\Bigg((\frac{\gamma^{(1)}}{\lambda^{(2)}})^{1/2}Cov(Y^{(1)},S_{\bar{D}}^{(1)^{-1}}(t),Y^{(2)},S_{\bar{D}}^{(2)^{-1}}(t))$$

$$+\gamma^{(1)}(\frac{1}{\lambda^{(1)}\lambda^{(2)}})^{1/2}\cdot\left(\frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}\right)\left(\frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}\right) Cov(X^{(1)},S_{\bar{D}}^{(1)^{-1}}(t),X^{(2)},S_{\bar{D}}^{(2)^{-1}}(t))$$

$$+(\frac{\gamma^{(1)}}{\lambda^{(2)}})^{1/2}\cdot\left(\frac{S_D^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}{S_{\bar{D}}^{(2)'}(S_{\bar{D}}^{(2)^{-1}}(t))}\right) Cov(X^{(2)},S_{\bar{D}}^{(2)^{-1}}(t),Y^{(1)},S_{\bar{D}}^{(1)^{-1}}(t))$$

$$+\gamma^{(1)}(\frac{1}{\lambda^{(1)}\gamma^{(2)}})^{1/2}\cdot\left(\frac{S_D^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}{S_{\bar{D}}^{(1)'}(S_{\bar{D}}^{(1)^{-1}}(t))}\right) Cov(X^{(1)},S_{\bar{D}}^{(1)^{-1}}(t),Y^{(2)},S_{\bar{D}}^{(2)^{-1}}(t))\Bigg)\Bigg\}$$

$$=N^{(1)}\frac{1}{N_{r'}^{(1)}}\frac{1}{N_{r'}^{(1)}}(a_{33}+a_{44}-2a_{34})$$

$$=Var(\hat{\Delta}_{r'}(t)),\quad for\ \ r\leq r'.$$

Hence, $Cov(\hat{\Delta}_r(t),\hat{\Delta}_{r'}(t))=Var(\hat{\Delta}_{r'}(t))$, for $r\leq r'$. The variance / covariance formula consists of ten components with each represents the correlation within diseased or non-diseased group within the same marker, the correlation within diseased or non-diseased group between markers, the correlation between diseased and non-diseased group within the same marker, and the correlation between diseased and non-diseased group between markers. All the above correlations are within the same subject, and data between subjects are independent according the assumption.

4.3 Simulation Studies

4.3.1 Consistency of Covariance Matrix Estimator

We conduct a simulation study to assess the finite sample properties of the results in Theorem 4.28. We generated the clustered measurements using a setting similar to Emir et al. (2000). First, we generate $\mathbf{X}^{(1)} = \mathbf{Y}_1\sqrt{\lambda}+\mathbf{Y}_2\sqrt{1-\lambda}$ and $\mathbf{X}^{(2)} = \mathbf{Y}_1\sqrt{\lambda}+\mathbf{Y}_3\sqrt{1-\lambda}$, where $\mathbf{Y}_i = (Y_{i1}, \cdots, Y_{i,2m})^T$, $i = 1, 2, 3$, are i.i.d. multivariate normal random vectors with mean $\mathbf{0}$ and $cov(Y_{ij}, Y_{ik}) = \rho^{|j-k|}$, for $j, k = 1, \cdots, 2m$. Here we assume $m_i^{(v)} \equiv n_i^{(v)} \equiv m$, for $i = 1, \cdots, n$ and $v = 1, 2$. For the covariance matrix simulation study, we let $m = 4$, and randomly assign m values to be from diseased tissues, and the other m values to be from nondiseased tissues. The values for subject i form marker v at location j is $X_{ij}^{(v)}$ if the location is "nondiseased", and is $X_{ij}^{(v)} + 1$ if it is "diseased". Here, the λ and ρ measure the between-marker and within-marker correlations.

We conduct 5000 simulation with $n = 400$, and for the simulated data, we calculate the variance-covariance of the $\Delta(t)$ at various proportions of r with t=0.5. Here, the ROC curves are estimated with the empirical functions. Then we compare the simulated covariance matrix to the theoretical covariance matrix derived using the results of Theorem (4.28). The results are presented in Table 4.1, which illustrates that the observed variance-covariance values are very close to the theoretical values when sample size is sufficiently large.

4.3.2 Simulated Type I Error Rate in GSDs

To investigate finite sample performance of the GSD procedure, we conduct a simulation study in a two-group sequential test (J=2), and a five-group sequential test (J=5). The data generating procedure is similar to the setting for the covariance simulation in Section 4.3. The null hypothesis of equal ROC(t) is set to be true and the nominal type I error rate was set to be $\alpha = 0.05$ for two-sided tests. Two set of diagnostic test data

Table 4.1: The values of elements ($\times 10^{-4}$) in observed and theoretical clustered covariance matrix

	Observed covariance matrix				Theoretical covariance matrix			
	$n = 400$							
$\Delta_{0.2}(0.5)$	9.475	4.628	3.668	1.821	9.302	4.651	3.721	1.860
$\Delta_{0.4}(0.5)$		4.894	3.794	1.839		4.651	3.721	1.860
$\Delta_{0.5}(0.5)$			3.874	1.831			3.721	1.860
$\Delta_{1}(0.5)$				1.897				1.860

are simulated from the aforementioned model, and the ROC curves are identical. Various combination of ρ, λ and subject number n are considered in our simulation study, where $n = (50, 100, 250, 500)$, $\rho = (0, 0.25, 0.5, 0.75)$, $\lambda = (0, 0.5, 0.75)$. The FPR points investigated are $t = (0.2, 0.4, 0.5, 0.6, 0.8)$.

For each simulation setting, 5000 random data sets are generated and the GSD method applied to the simulated data. The Z statistics at each interim analysis point are then calculated based on the empirical ROC difference and estimated variances. The GSD test procedure compares the Z statistics with corresponding test boundaries of design, and the decision of rejection or failing to rejection is obtained for each simulated dataset. Then we can calculate the overall rejection rates for all simulated datasets. Table 4.2 gives the rejection rates for all parameter and sample size combinations with a nominal α level 0.05 under the O'Brien and Fleming's criterion. And Table 4.3 is the results for the Pocock's criterion. Furthermore, simulation results for lognormal data are presented in Table 4.4 and 4.5. Lognormal data has similar results as normal data due to invariance to monotone transformation. All these tables show that the simulated Type I error rates are close to the nominal rate and tend to be closer as the overall sample sizes increase. Note that this is true for all ρ and λ combinations and for all FPR points we analyze. The type I error rates are also plotted in Figure 4.1 and Figure 4.2. In these figures, the type I error rates are plotted as bars showing their deviations from the nominal rate of 0.05 which is the vertical line.

Table 4.2: Type I error rates ($\times 10^{-2}$) using the O'Brien-Fleming GSD with $\alpha = 0.05$, normal data

		$\rho = 0$			$\rho = 0.25$			$\rho = 0.5$			$\rho = 0.75$		
n	t	λ=0	0.5	0.75	λ=0	0.5	0.75	λ=0	0.5	0.75	λ=0	0.5	0.75
Two-group sequential design (J=2)													
50	0.2	5.32	4.98	3.80	5.36	4.46	4.06	5.28	4.50	3.30	4.54	4.16	3.60
	0.4	4.58	4.68	3.66	4.90	4.58	3.92	4.94	4.46	3.34	5.06	4.04	3.96
	0.5	4.90	4.62	4.00	5.18	3.80	3.54	4.54	4.54	2.96	4.62	3.72	2.96
	0.6	5.04	4.14	3.36	4.72	4.16	3.36	4.34	3.40	3.04	4.08	3.64	3.24
	0.8	3.72	2.84	2.26	3.94	2.76	2.54	4.00	3.26	2.36	3.92	3.76	2.64
100	0.2	4.60	4.90	4.34	4.74	4.36	4.50	5.14	4.34	4.36	5.26	4.42	3.82
	0.4	5.36	4.08	4.18	5.16	4.84	3.70	4.68	4.40	3.72	4.96	4.12	3.88
	0.5	5.02	4.24	4.12	4.68	4.70	4.32	4.62	4.94	3.62	4.52	4.80	3.32
	0.6	4.38	4.36	3.44	4.70	4.16	4.34	4.08	4.42	3.72	4.10	4.08	3.90
	0.8	3.90	3.72	3.72	4.28	3.98	3.44	4.06	3.64	3.02	4.40	4.14	3.36
200	0.2	4.86	4.96	4.58	5.10	4.50	4.46	4.42	4.90	4.62	5.00	4.62	3.78
	0.4	5.14	4.64	4.26	5.44	4.44	4.70	4.90	4.52	4.10	4.50	4.62	3.94
	0.5	5.06	4.62	3.80	4.82	4.30	4.54	4.48	4.24	4.50	4.44	4.38	3.96
	0.6	4.20	4.24	3.58	5.18	4.30	3.96	4.32	4.20	4.10	4.78	4.40	3.76
	0.8	4.34	4.32	3.56	4.18	4.24	3.66	4.34	3.78	3.72	4.68	4.12	3.96
Five-group sequential design (J=5)													
50	0.2	6.42	5.26	4.60	6.10	5.56	4.08	5.94	5.02	3.50	4.94	4.62	3.34
	0.4	5.54	4.92	3.70	5.18	4.62	4.08	4.96	4.60	3.18	4.46	4.14	3.24
	0.5	4.78	4.72	3.86	5.34	4.20	3.62	4.94	4.98	3.62	5.16	4.18	3.68
	0.6	5.60	4.44	3.28	4.72	4.06	3.24	4.64	4.22	2.98	4.62	4.48	3.12
	0.8	3.46	3.56	2.12	3.68	2.88	2.26	3.58	3.10	2.54	3.44	3.94	2.22
100	0.2	5.28	5.24	4.54	5.94	4.48	4.12	5.12	5.10	4.30	4.36	4.64	4.00
	0.4	4.64	4.72	5.10	5.02	4.80	3.98	4.72	4.68	4.10	4.64	4.52	3.50
	0.5	5.04	5.58	3.84	5.68	4.82	3.90	4.96	4.58	3.80	4.52	3.84	3.62
	0.6	5.02	4.44	4.12	4.68	4.58	3.80	5.36	3.92	4.12	4.42	4.36	3.36
	0.8	3.98	3.58	3.02	3.98	3.56	2.96	3.82	3.70	3.26	4.56	4.26	2.94
200	0.2	4.86	5.38	4.38	4.86	4.64	4.24	4.88	4.74	3.94	5.20	5.06	3.96
	0.4	4.98	4.60	4.60	4.70	4.84	4.80	5.64	5.20	4.52	4.54	4.22	3.94
	0.5	4.68	5.02	4.18	5.32	4.40	4.34	4.88	5.00	3.74	4.46	4.84	4.74
	0.6	5.18	4.24	4.08	4.54	4.48	4.20	4.90	5.26	4.20	5.08	4.86	4.00
	0.8	4.50	4.50	4.14	4.18	4.34	4.42	4.52	4.58	3.42	4.96	4.00	3.76

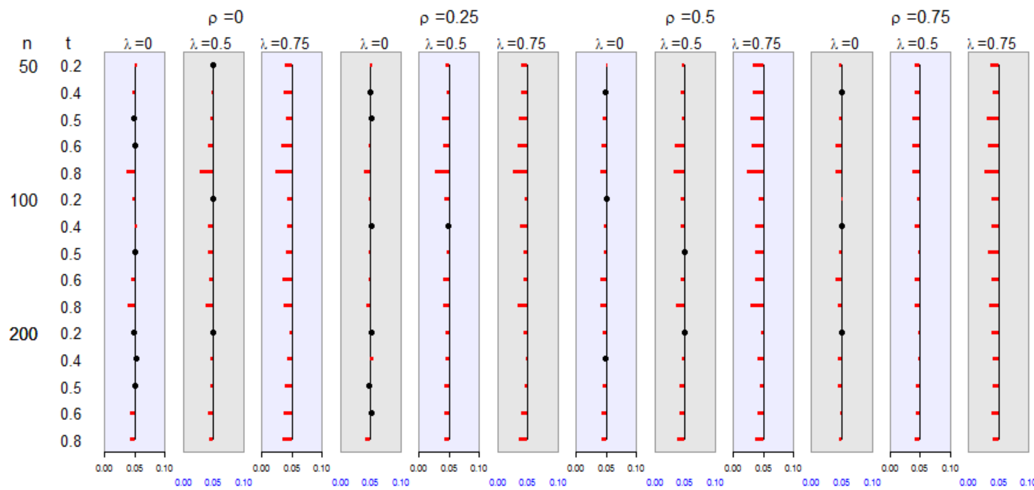

Figure 4.1: Type I error rates plot using the O'Brien-Fleming GSD with $\alpha = 0.05$,$J = 2$, normal data

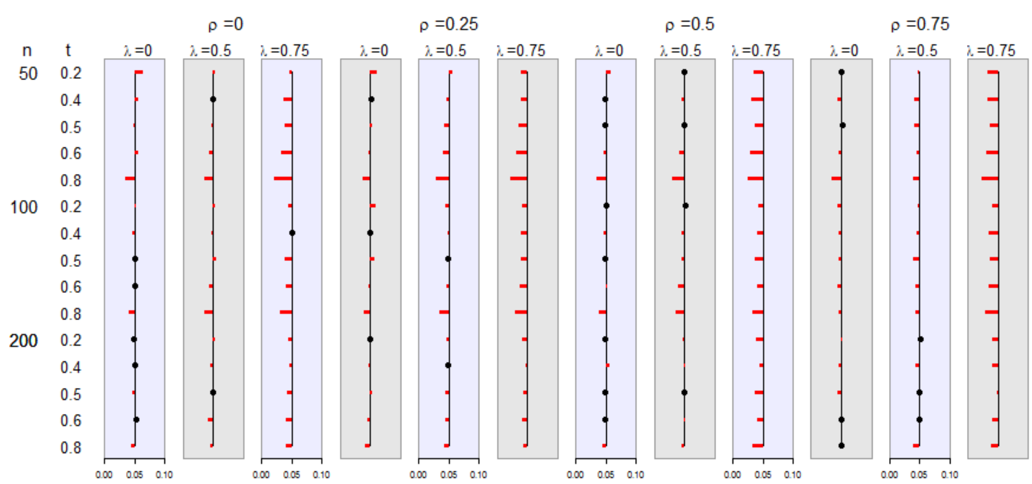

Figure 4.2: Type I error rates plot using the O'Brien-Fleming GSD with $\alpha = 0.05$,$J = 5$, normal data

Table 4.3: Type I error rates ($\times 10^{-2}$) using the Pocock GSD with $\alpha = 0.05$, normal data

		$\rho = 0$			$\rho = 0.25$			$\rho = 0.5$			$\rho = 0.75$		
n	t	λ=0	0.5	0.75	λ=0	0.5	0.75	λ=0	0.5	0.75	λ=0	0.5	0.75
Two-group sequential design (J=2)													
50	0.2	5.46	5.14	3.78	5.18	4.50	3.64	5.38	4.86	3.46	4.56	4.14	2.92
	0.4	5.38	4.68	4.02	5.50	4.66	3.38	4.82	4.00	3.22	4.64	3.62	3.10
	0.5	5.30	4.52	3.54	4.80	4.76	3.44	4.68	3.74	3.14	4.82	3.92	3.06
	0.6	4.82	4.70	3.32	5.20	4.24	3.10	4.48	4.28	3.00	4.08	4.50	3.08
	0.8	2.82	2.42	1.68	3.22	2.32	1.58	2.96	2.84	1.76	3.42	2.60	1.98
100	0.2	5.42	4.58	3.64	5.40	4.24	4.14	5.20	4.52	4.22	4.34	4.00	3.18
	0.4	5.08	4.20	4.36	5.02	4.72	4.02	4.70	3.92	3.52	4.14	4.14	3.70
	0.5	5.08	4.18	3.90	4.80	3.94	3.26	5.14	4.22	3.32	4.56	4.24	3.56
	0.6	4.76	4.42	3.74	4.56	3.96	3.20	4.06	4.60	3.62	5.20	4.42	3.64
	0.8	3.26	3.28	2.50	3.96	3.54	2.54	3.70	3.72	2.84	3.94	3.38	2.92
200	0.2	5.22	4.70	3.88	5.36	4.80	3.96	5.02	4.44	3.98	4.26	4.92	3.76
	0.4	5.16	4.84	3.92	5.20	5.20	3.88	4.64	4.52	3.94	4.90	4.02	4.22
	0.5	5.28	4.96	3.84	5.04	4.84	3.82	4.64	4.22	3.64	5.10	4.40	3.88
	0.6	4.80	4.74	3.96	5.10	4.32	4.04	4.62	4.56	4.06	4.70	3.70	3.90
	0.8	4.36	3.82	3.72	4.16	3.72	3.50	4.34	3.18	3.40	4.10	4.88	3.40
Five-group sequential design (J=5)													
50	0.2	7.90	6.02	4.58	7.40	6.14	3.96	6.94	5.26	3.92	5.64	4.38	2.90
	0.4	7.00	5.60	3.90	6.62	4.90	3.40	5.58	4.56	2.98	5.64	3.96	2.98
	0.5	5.56	4.68	3.28	6.38	4.92	3.08	6.00	4.32	2.42	5.22	4.72	2.72
	0.6	5.34	4.30	3.14	4.92	3.84	2.90	5.20	3.58	2.44	5.64	3.86	2.50
	0.8	2.26	1.26	1.14	2.38	1.70	0.76	2.64	2.10	1.26	3.10	2.58	1.32
100	0.2	6.44	5.68	4.18	5.90	4.66	3.86	5.22	4.58	3.62	4.92	4.00	2.84
	0.4	5.70	5.58	4.06	5.98	5.56	4.06	5.66	4.82	3.62	4.98	3.60	2.98
	0.5	5.40	4.28	4.04	5.20	4.42	3.68	4.48	4.44	3.02	4.76	4.00	2.94
	0.6	4.80	4.32	3.08	5.20	4.10	3.58	4.88	3.84	2.66	4.78	4.26	3.54
	0.8	3.16	2.56	1.64	3.04	2.64	1.74	3.18	2.60	1.78	3.76	2.54	2.10
200	0.2	5.88	5.56	5.00	5.88	4.86	4.20	5.92	4.78	4.06	4.78	4.82	3.64
	0.4	5.40	4.98	4.52	5.30	5.12	3.76	5.08	4.36	3.74	4.78	4.18	3.50
	0.5	5.50	4.50	4.08	4.50	4.44	3.90	5.10	4.62	3.80	4.26	4.46	3.34
	0.6	5.34	4.38	3.78	4.78	4.24	3.52	5.10	4.24	3.88	4.20	4.18	3.26
	0.8	3.60	3.08	2.80	4.14	3.20	3.02	4.22	3.52	2.36	3.76	3.26	2.50

Table 4.4: Type I error rates ($\times 10^{-2}$) using the O'Brien-Fleming GSD with $\alpha = 0.05$, lognormal data

		$\rho = 0$			$\rho = 0.25$			$\rho = 0.5$			$\rho = 0.75$		
n	t	λ=0	0.5	0.75	λ=0	0.5	0.75	λ=0	0.5	0.75	λ=0	0.5	0.75
Two-group sequential design (J=2)													
50	0.2	5.76	5.10	3.66	5.62	4.88	3.84	5.36	4.68	3.92	4.72	3.78	3.34
	0.4	5.02	4.46	3.62	5.72	4.38	3.60	5.18	4.34	3.50	4.10	4.20	3.46
	0.5	5.68	4.12	3.54	4.66	4.04	3.60	4.70	3.88	3.38	4.74	4.60	3.48
	0.6	5.06	4.08	3.12	4.74	4.22	2.88	4.68	3.76	2.70	4.28	4.14	3.00
	0.8	3.88	3.48	2.42	3.70	3.04	2.12	3.82	2.68	2.62	3.76	3.52	2.38
100	0.2	5.74	4.84	4.30	5.00	4.82	3.86	4.32	4.76	4.02	4.82	4.64	4.08
	0.4	4.56	4.42	4.04	4.90	4.34	3.98	4.90	4.56	3.86	4.30	4.26	3.82
	0.5	4.82	4.66	3.84	4.54	4.74	4.30	5.16	4.70	3.92	4.42	3.82	3.58
	0.6	4.74	4.08	4.24	4.48	4.10	4.24	4.30	4.42	3.90	4.40	4.48	3.46
	0.8	4.10	3.36	3.50	4.00	3.12	3.66	4.04	3.36	3.02	3.84	3.74	3.46
200	0.2	5.20	5.04	4.34	4.78	5.00	3.76	4.84	4.50	3.88	4.36	4.40	3.84
	0.4	4.96	5.24	4.22	4.90	4.90	4.22	4.64	4.94	3.78	4.46	4.26	4.26
	0.5	5.36	4.68	4.88	4.88	4.78	4.26	5.10	4.16	3.86	4.40	3.92	4.12
	0.6	4.92	4.68	4.38	4.44	4.08	4.18	4.84	4.06	4.26	5.00	4.48	4.36
	0.8	4.06	3.78	4.18	4.00	4.10	4.10	4.38	3.84	3.06	4.68	4.34	3.78
Five-group sequential design (J=5)													
50	0.2	5.66	5.16	3.76	5.32	5.06	3.70	5.24	4.66	3.78	4.86	4.28	3.76
	0.4	5.80	4.64	4.10	5.68	4.10	3.48	4.96	4.86	3.74	4.86	4.62	3.18
	0.5	5.20	5.22	3.72	5.14	4.68	3.70	5.42	4.36	3.56	4.82	4.14	3.62
	0.6	5.08	4.82	4.00	4.50	4.50	3.46	4.22	4.16	3.32	4.66	3.98	3.08
	0.8	3.76	3.24	2.18	4.00	2.72	2.28	3.22	2.72	2.70	3.86	3.42	1.94
100	0.2	5.26	4.92	3.96	5.40	4.92	3.90	5.32	4.68	3.96	5.34	4.12	3.82
	0.4	5.80	4.88	4.14	5.08	4.46	3.80	4.74	4.20	3.94	4.40	4.82	2.92
	0.5	4.82	4.72	4.36	5.38	5.12	3.72	4.52	4.44	3.72	4.88	3.98	3.60
	0.6	4.52	4.88	3.88	5.06	4.58	4.28	4.46	4.48	3.72	4.62	4.14	3.36
	0.8	3.98	3.46	3.14	3.68	3.72	3.34	3.80	3.38	3.48	4.72	3.50	3.60
200	0.2	5.32	5.64	4.58	4.70	4.60	3.94	4.66	4.34	4.06	4.70	4.88	3.86
	0.4	5.34	4.02	4.88	5.14	4.46	4.74	4.40	4.66	4.40	4.38	4.56	4.24
	0.5	5.48	4.28	4.14	5.26	5.02	4.70	4.86	4.16	4.02	4.80	4.54	4.00
	0.6	4.74	5.00	4.66	5.26	4.22	4.44	4.62	4.30	3.84	5.12	4.28	4.02
	0.8	4.82	4.44	3.54	3.90	3.70	3.42	4.70	4.26	3.32	4.28	4.74	3.82

Table 4.5: Type I error rates ($\times 10^{-2}$) using the Pocock GSD with $\alpha = 0.05$, lognormal data

		$\rho = 0$			$\rho = 0.25$			$\rho = 0.5$			$\rho = 0.75$		
n	t	λ=0	0.5	0.75	λ=0	0.5	0.75	λ=0	0.5	0.75	λ=0	0.5	0.75
		Two-group sequential design (J=2)											
50	0.2	5.94	4.90	3.48	6.94	4.64	4.00	4.98	4.32	3.56	5.22	3.78	3.00
	0.4	6.14	4.76	3.48	5.68	4.66	3.38	4.74	4.08	3.32	4.64	3.92	3.54
	0.5	5.08	4.68	3.74	4.72	4.18	3.76	4.94	4.18	2.86	4.86	3.56	2.90
	0.6	4.68	3.92	2.94	4.42	4.30	2.78	4.88	3.86	3.00	4.40	3.52	2.54
	0.8	2.84	2.22	1.38	3.04	2.64	1.68	2.66	2.42	1.44	3.50	2.54	1.48
100	0.2	5.42	4.58	3.56	5.88	4.92	4.22	4.84	4.58	3.56	4.56	3.74	3.34
	0.4	5.10	4.72	3.78	4.98	4.14	4.08	5.04	4.34	3.80	4.86	4.62	3.42
	0.5	4.92	3.94	3.76	4.42	4.40	3.90	4.20	4.40	3.68	4.38	4.48	3.18
	0.6	4.80	4.00	3.78	4.32	4.38	3.22	4.50	3.68	3.30	4.46	4.20	2.98
	0.8	4.30	3.08	2.26	4.06	3.26	2.78	3.72	3.30	2.86	4.26	3.16	2.58
200	0.2	4.74	5.18	4.20	5.00	5.00	3.86	4.30	5.08	4.34	4.40	4.40	3.74
	0.4	4.78	4.68	3.88	4.64	4.16	4.26	5.56	4.58	4.32	4.48	4.56	3.64
	0.5	4.84	4.36	4.86	4.72	4.68	4.22	4.84	4.30	3.68	4.18	4.78	3.36
	0.6	5.32	4.06	3.50	5.02	4.98	3.64	4.62	4.40	3.76	4.70	4.32	3.74
	0.8	4.08	3.96	3.18	3.78	3.88	3.26	3.84	4.08	2.78	4.52	3.88	3.54
		Five-group sequential design (J=5)											
50	0.2	8.38	6.22	4.50	7.90	5.84	3.76	6.74	5.04	3.56	5.48	4.44	3.00
	0.4	7.06	5.68	3.92	6.86	4.98	3.04	6.02	4.42	2.86	5.76	4.08	2.82
	0.5	6.08	4.78	3.44	5.46	4.44	2.84	5.18	4.06	2.82	5.02	4.16	2.44
	0.6	5.24	4.20	2.84	5.98	3.80	2.58	4.78	3.66	2.24	4.96	3.48	2.56
	0.8	2.42	2.18	0.94	2.16	1.64	1.02	2.46	2.12	0.86	3.02	2.08	1.50
100	0.2	6.88	5.76	4.46	5.60	5.34	3.96	5.48	4.32	3.26	5.20	4.84	3.28
	0.4	6.20	5.36	3.78	5.68	4.80	4.00	5.40	4.46	3.66	4.60	4.78	2.96
	0.5	5.68	4.78	3.94	5.74	4.80	3.14	5.04	3.82	3.06	4.60	4.36	3.02
	0.6	4.76	4.50	3.48	4.38	4.38	2.94	4.76	3.82	3.36	4.56	3.66	2.94
	0.8	3.52	2.50	2.20	3.14	2.26	2.22	3.50	2.92	1.70	3.28	2.92	2.28
200	0.2	5.60	5.16	4.32	5.16	4.36	4.56	5.08	4.30	4.00	4.56	4.24	3.44
	0.4	5.94	5.00	4.62	5.10	4.76	4.24	5.00	4.72	3.52	5.36	4.42	3.66
	0.5	5.36	4.72	4.40	5.04	4.06	4.12	5.16	4.08	4.18	4.82	4.32	3.60
	0.6	5.18	3.90	4.02	5.28	4.00	3.46	4.94	4.28	3.30	4.98	4.22	3.00
	0.8	3.94	3.08	2.64	3.76	3.34	2.94	4.02	3.48	2.98	4.26	3.40	2.62

4.3.3 Expected Sample Size in GSDs

Furthermore, we conduct simulation studies on two clustered ROC curves that are not equal at certain FPR under investigation. While maintain the α level and specific power requirement, we show that the expected sample size with GSD is less than the one with fixed sample size design. Given two clustered ROC curves, with pre-specified α and specific power requirement, using the following formula, we can determine the sample size for a fixed sample study, for a two-sided test:

$$n \geq (\Phi^{-1}(1-\alpha/2) + \Phi^{-1}(1-\beta))^2 \frac{\sigma^2}{\delta^2},$$

where δ is the difference of two ROC curves at FPR t_0 and σ^2 can be estimated using the simulation method. Let $\alpha = 0.05$, power $(1-\beta) = 90\%$ *or* 80%, using similar data generating setting as previous, we try three different scenarios where the value increase for diseased units varies. In the following result tables, we use abbreviation "OF" for O'Brien-Fleming method and "LogN" for lognormal datasets.

We apply the method on three cases in comparing two clustered ROC curves. As show in Figure 4.3-4.5, case I has the biggest difference between two investigational ROC curves, while case III has the smallest difference.

The simulation results in Tables 4.6 - 4.20 illustrate several points. The simulated powers are close to the expected values, 80% or 90%, where the sample sizes are determined to achieve the pre-specified power requirement. In each case we find that the power goals are closely met for both O'Brien-Fleming and Pocock methods with different number of interim looks and also for different ρ, λ combinations and at different FPR. Lognormal and normal data yield similar result as we know that ROC is invariant to monotone transformation. Inspection of Tables 4.18 4.20 reveals that the fixed sample design sample sizes for different FPR vary substantially due to the difference in variance and δ at different FPR. Consequently, the GSD design sizes and GSD expected sample sizes vary substantially at different FPR. This is the case for both OBrien-Fleming and Pocock GSD methods. Furthermore, Pocock method tend to have larger GSD design size and smaller expected sample size compared to OBrien-Fleming method.

We give detailed steps for a GSD study using an example in the following, which explains GSD design maximum sample size determination with specific power requirement, as well as the calculations of the expected GSD sample size and actual achieved power through simulation. Case I given the setting with $\rho = 0.5$, $\lambda = 0.5$, $FPR = 0.5$ and 90% power requirement for the predefined value increase for diseased units. In this case where $\delta = 0.0614$, we determine that sample size need to be 227.84 for a fixed sample study. All fixed sample design sample size requirements for this case are shown in Table 4.18. Then with the ratios provided in Jennison and Turnbull (2000), where with O'Brien-Fleming method, for J=2 the ratio is 1.007; for J=5 the ratio is 1.026. With Pocock method, for J=2 the ratio is 1.1; for J=5 the ratio is 1.207. Multiply the fixed sample size with the corresponding ratio, we know to maintain the α and power level, for a group sequential study assuming equal group sizes, the maximum sample sizes needed are: with O'Brien-Fleming method, for J=2 the sample size is 230; for J=5 the sample size is 234. With Pocock method, for J=2 the sample size is 251; for J=5 the sample size is 276. The following simulation results (Table 4.6, 4.7, 4.12, 4.15), shows that the expected sample sizes of GSDs are less than the fixed sample size (228), while still meet the $\alpha(0.05)$ and power requirements.

With the same setting except the power requirement set to 80%, we determine that sample size need to be 170.19 for a fixed sample study (Table 4.18). Then with the ratios provided in Jennison and Turnbull (2000), where with O'Brien-Fleming method, for J=2 the ratio is 1.008; for J=5 the ratio is 1.028. With Pocock method, for J=2 the ratio is 1.11; for J=5 the ratio is 1.229. Similarly, we calculated the sample sizes needed for group sequential studies assuming equal interim group sizes. With O'Brien-Fleming method, for J=2 the sample size is 172; for J=5 the sample size is 175. With Pocock method, for J=2 the sample size is 189; for J=5 the sample size is 210. The simulation results (Table 4.6, 4.7, 4.12, 4.15) shows that the expected sample sizes of GSDs are less than the fixed sample size (171), while still meet the $\alpha(0.05)$ and power (80%) requirements.

Similarly, the sample size determination and simulation results for Case II can be found in Tables 4.8, 4.9, 4.13, 4.16 and 4.19. And Case III in Tables 4.10, 4.11, 4.14, 4.17 and 4.20.

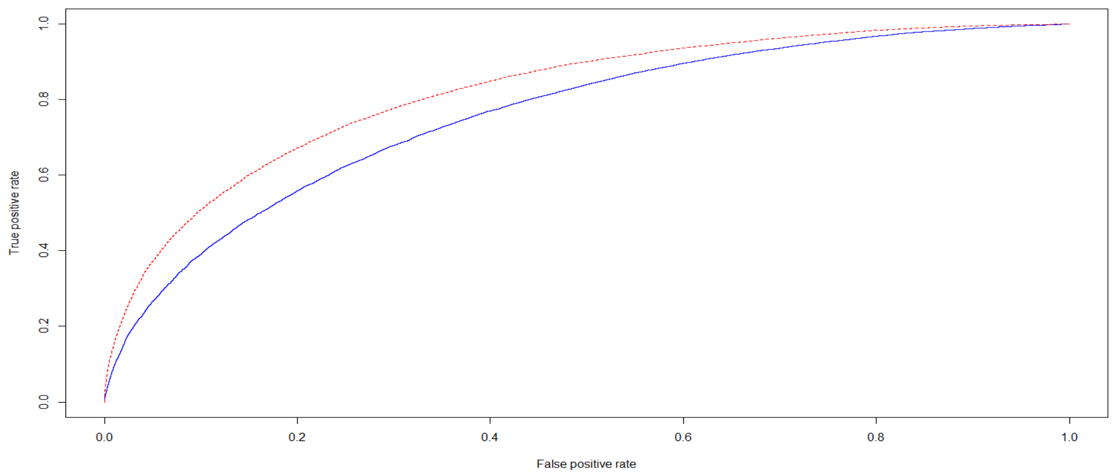

Figure 4.3: Empirical ROC curves of clustered data, case I

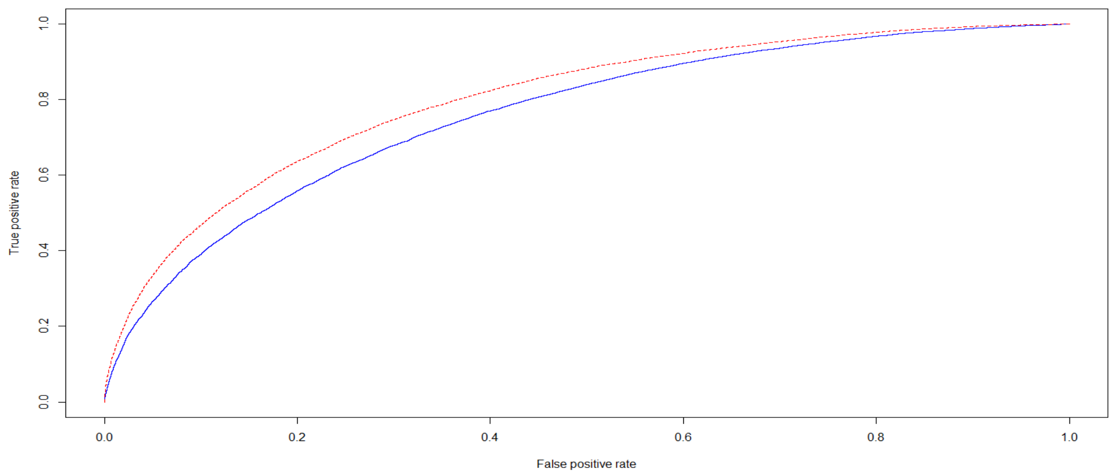

Figure 4.4: Empirical ROC curves of clustered data, case II

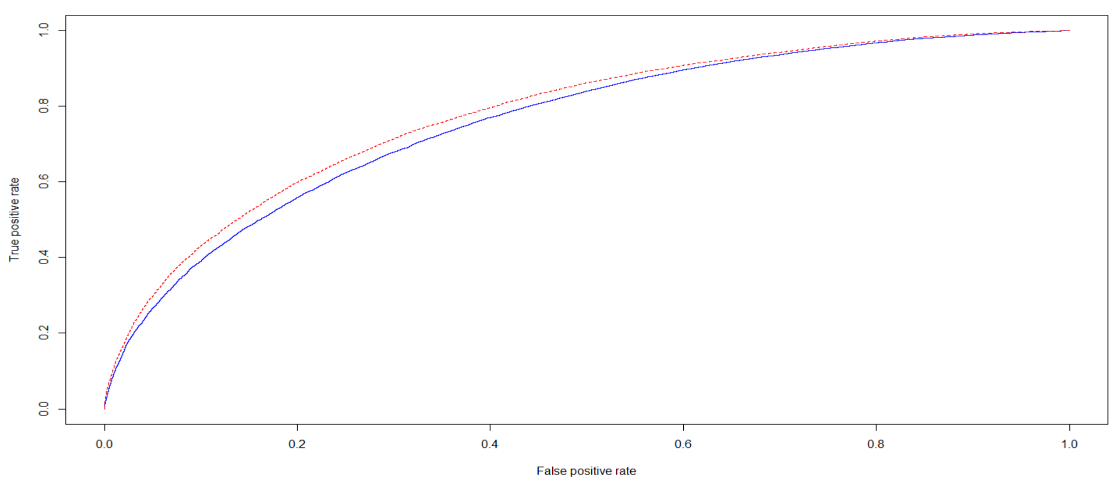

Figure 4.5: Empirical ROC curves of clustered data, case III

Table 4.6: Power(%) using the O'Brien-Fleming GSD with $\alpha = 0.05$, case I

		Power=80%						Power=90%					
		$\lambda = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\lambda = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
ρ	t	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN
Two-group sequential design (J=2)													
	0.2	76.7	77.6	77.2	77.3	77.6	78.0	88.4	88.3	87.9	87.2	88.3	89.0
	0.4	78.1	78.7	77.8	79.0	75.4	78.0	89.3	89.1	88.8	88.3	88.6	88.1
0	0.5	78.2	77.8	79.0	78.4	78.3	78.6	89.5	89.6	89.2	90.2	88.2	89.6
	0.6	77.7	77.2	79.2	79.6	80.5	80.0	88.6	89.0	90.0	90.2	90.8	90.5
	0.8	78.8	79.2	81.2	80.8	79.7	79.6	89.2	89.3	90.6	91.3	91.3	90.7
	0.2	79.2	78.1	77.6	78.8	75.7	76.6	88.3	87.9	88.1	88.2	88.1	88.8
	0.4	77.9	78.4	77.6	78.6	77.6	77.4	89.6	88.8	88.0	88.3	89.6	88.8
0.25	0.5	78.3	79.3	79.3	78.9	77.3	77.1	89.2	88.9	89.5	89.6	89.3	88.3
	0.6	79.7	78.7	80.2	80.5	80.2	79.4	89.4	89.7	89.6	91.0	90.7	90.9
	0.8	78.5	78.9	79.4	78.9	80.7	80.7	89.5	88.5	90.1	90.3	90.8	90.8
	0.2	78.0	77.7	76.5	76.8	75.7	74.9	88.7	89.0	86.7	87.0	86.9	86.9
	0.4	77.3	79.2	78.5	77.4	78.5	77.9	88.7	88.2	89.7	89.4	88.8	89.0
0.5	0.5	78.9	78.5	78.9	78.6	78.2	78.2	89.1	89.2	89.5	89.2	88.6	89.2
	0.6	78.5	80.0	78.9	78.3	79.8	78.9	88.8	89.4	90.2	89.9	89.2	89.3
	0.8	80.2	80.8	79.1	77.7	79.1	78.7	90.1	90.9	89.4	89.9	89.6	90.0
	0.2	77.3	78.0	76.1	76.6	76.2	77.2	88.4	88.6	87.1	87.8	87.0	87.8
	0.4	77.2	77.9	77.5	77.7	77.1	77.2	88.3	88.5	88.0	87.5	88.1	87.4
0.75	0.5	78.9	79.2	77.3	77.6	77.2	78.3	89.4	88.0	89.5	88.6	88.7	88.4
	0.6	78.2	77.6	78.9	77.5	78.5	79.0	88.7	87.9	89.7	90.0	89.5	88.9
	0.8	78.1	79.1	80.7	80.2	77.3	78.5	89.7	89.3	91.1	90.1	88.9	88.7
Five-group sequential design (J=5)													
	0.2	78.6	77.9	77.1	78.2	78.7	78.1	87.9	87.9	88.3	87.9	89.5	89.6
	0.4	77.6	78.3	77.8	78.7	78.9	78.9	88.2	88.5	89.2	89.0	87.8	88.6
0	0.5	78.3	78.0	78.9	78.3	79.5	79.7	88.6	88.8	89.9	88.8	90.1	89.7
	0.6	78.5	77.9	79.9	80.5	81.0	80.0	87.7	88.6	89.9	90.2	91.2	91.0
	0.8	79.0	77.9	81.6	80.4	80.6	80.2	88.9	89.4	91.0	91.1	90.6	91.1
	0.2	77.6	77.8	77.6	77.2	77.1	77.3	87.9	88.0	88.7	88.0	88.0	89.2
	0.4	77.5	78.1	77.0	76.7	78.6	78.2	88.3	88.8	89.1	88.5	88.6	89.1
0.25	0.5	78.1	79.9	79.2	78.6	78.7	77.9	88.5	89.1	89.7	90.0	89.4	88.5
	0.6	79.4	79.1	80.6	80.8	80.0	80.3	89.5	89.9	91.1	90.4	90.5	91.3
	0.8	78.2	78.9	79.4	79.8	80.4	80.5	89.3	89.5	90.5	89.7	91.6	91.6
	0.2	78.0	78.0	77.4	77.2	76.7	77.0	89.5	90.4	87.0	88.1	88.2	87.9
	0.4	78.0	77.5	79.7	78.3	78.3	78.7	88.8	88.3	88.7	89.6	88.5	89.5
0.5	0.5	79.4	79.9	79.2	79.8	78.3	78.8	90.3	89.3	89.4	89.4	88.6	89.7
	0.6	80.3	80.7	79.0	79.7	78.8	79.4	89.4	89.3	89.5	89.8	90.3	90.0
	0.8	80.2	80.3	79.4	78.7	78.6	79.3	90.9	89.7	89.4	90.2	89.9	89.7
	0.2	78.5	78.5	76.6	77.3	77.1	77.6	88.8	88.9	87.5	87.4	88.4	88.1
	0.4	79.1	77.7	77.8	77.6	76.5	76.3	88.9	88.7	89.5	89.0	87.9	88.4
0.75	0.5	78.6	76.3	76.5	76.4	78.7	77.8	88.3	88.8	87.8	88.2	89.4	90.0
	0.6	78.5	78.2	78.7	79.3	78.9	77.7	89.4	89.7	88.7	89.5	89.1	89.5
	0.8	78.9	79.1	80.5	81.1	78.8	78.2	89.5	90.3	90.4	90.4	89.3	88.3

Table 4.7: Power(%) using the Pocock GSD with $\alpha = 0.05$, case I

		Power=80%						**Power=90%**					
		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	t	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN
Two-group sequential design (J=2)													
	0.2	77.4	78.5	77.3	77.4	78.3	77.5	87.5	87.8	89.0	87.8	89.0	89.0
	0.4	79.1	78.1	78.5	78.3	77.4	76.9	88.7	88.4	89.4	89.3	88.1	89.0
0	0.5	77.0	79.0	77.8	80.0	78.7	79.5	89.1	89.0	88.9	89.3	90.2	90.4
	0.6	78.4	78.8	79.1	80.1	79.4	80.7	88.5	88.1	90.1	89.8	91.2	90.8
	0.8	78.6	78.9	80.1	79.5	79.8	80.5	89.6	89.0	91.5	90.6	91.2	91.1
	0.2	77.2	78.4	77.0	78.0	76.3	76.8	88.4	87.5	88.6	88.0	87.8	88.0
	0.4	78.4	78.5	77.7	78.2	77.8	77.4	88.4	88.4	88.6	89.1	88.8	88.7
0.25	0.5	77.5	78.8	79.1	77.3	77.0	76.5	88.7	88.5	89.5	89.4	89.9	88.9
	0.6	79.5	78.6	79.6	80.3	80.5	81.2	89.4	89.2	90.0	90.7	91.2	91.4
	0.8	77.5	78.9	79.7	79.8	80.9	80.2	89.1	89.3	89.8	89.7	91.5	90.9
	0.2	78.6	78.2	75.6	75.9	75.9	76.9	89.4	89.1	87.1	87.6	87.4	88.1
	0.4	77.7	77.7	78.6	79.2	79.2	76.7	88.4	88.3	89.2	88.4	88.8	88.7
0.5	0.5	78.4	78.8	79.2	78.3	76.5	76.6	89.9	89.6	89.9	90.4	89.4	88.8
	0.6	79.5	78.1	78.2	79.1	79.2	78.9	90.2	89.1	90.0	89.6	89.6	89.2
	0.8	81.4	80.5	78.2	79.6	78.2	78.8	90.7	89.8	89.8	89.4	89.9	89.6
	0.2	79.2	79.2	75.6	77.2	77.1	76.1	88.3	88.9	87.6	87.1	87.3	88.4
	0.4	78.2	77.6	77.3	77.1	76.4	77.0	88.4	89.3	89.2	88.6	87.4	88.3
0.75	0.5	78.0	78.9	77.5	77.8	78.4	78.1	88.5	87.6	88.5	88.7	89.5	89.1
	0.6	79.5	78.1	78.8	78.4	78.7	77.4	88.7	88.6	89.2	88.9	89.4	89.7
	0.8	79.5	78.8	80.7	80.4	77.9	78.2	89.4	89.8	91.0	90.7	89.3	89.9
Five-group sequential design (J=5)													
	0.2	77.8	78.7	78.8	77.7	79.1	78.8	88.5	88.8	89.3	88.8	90.0	89.6
	0.4	78.2	78.8	78.0	78.4	79.4	78.2	88.5	89.3	88.6	90.0	89.2	88.1
0	0.5	78.6	77.7	79.3	78.7	79.3	78.9	88.9	89.0	89.0	88.8	89.5	89.7
	0.6	78.0	78.8	79.1	79.7	81.7	80.0	88.6	88.4	91.1	90.2	90.6	90.7
	0.8	78.9	79.6	79.7	80.0	80.5	79.5	89.3	89.3	91.2	91.2	90.8	91.1
	0.2	77.1	78.5	78.5	77.7	78.1	78.0	89.1	89.5	88.4	88.1	88.4	88.6
	0.4	78.8	78.6	79.9	77.2	78.8	77.7	88.8	89.5	89.1	89.2	89.9	90.0
0.25	0.5	79.6	78.4	79.1	79.1	79.2	77.3	89.8	89.0	89.5	89.8	89.5	89.1
	0.6	79.1	79.1	80.4	80.0	80.9	79.7	89.7	89.8	91.0	89.9	91.6	90.8
	0.8	78.9	77.4	79.2	79.9	82.0	80.7	89.3	88.9	90.2	89.7	91.2	91.5
	0.2	79.5	80.1	76.1	77.3	77.4	77.3	88.6	88.9	88.0	87.8	88.3	88.2
	0.4	79.4	78.5	78.0	79.6	78.9	78.9	88.8	88.9	89.3	89.1	88.9	90.1
0.5	0.5	78.9	79.5	80.4	79.0	78.1	79.0	89.5	89.7	89.3	89.9	89.4	90.1
	0.6	79.4	79.5	80.1	79.7	78.8	79.0	89.5	88.4	90.1	89.8	89.1	89.2
	0.8	80.2	80.8	79.1	77.7	79.1	79.6	90.7	90.6	89.8	89.5	89.9	90.1
	0.2	78.4	79.1	76.6	76.3	77.9	78.4	89.3	89.2	88.1	87.9	89.0	88.9
	0.4	78.8	78.5	76.6	78.2	77.7	78.7	89.0	89.4	89.1	88.0	88.2	88.9
0.75	0.5	77.7	78.9	76.9	78.8	78.7	78.8	89.4	88.7	88.9	88.2	89.4	89.4
	0.6	77.5	79.1	79.4	79.4	78.8	78.7	89.3	89.9	89.1	90.0	89.5	89.1
	0.8	79.9	79.7	81.6	82.1	78.4	78.1	89.8	89.2	90.8	90.7	89.0	89.0

Table 4.8: Power(%) using the O'Brien-Fleming GSD with $\alpha = 0.05$, case II

		Power=80%						**Power=90%**					
		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	t	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN
						Two-group sequential design (J=2)							
	0.2	77.3	77.3	77.6	78.9	77.8	78.2	89.1	87.6	89.4	88.9	89.9	87.9
	0.4	79.2	79.2	77.6	79.2	77.7	76.3	89.2	89.1	89.2	89.4	88.4	88.1
0	0.5	78.5	77.6	78.3	78.7	78.3	79.0	88.2	88.4	88.9	89.5	88.7	90.1
	0.6	77.5	76.6	78.4	79.2	79.8	79.7	89.4	88.5	89.9	90.0	90.8	91.0
	0.8	77.8	78.2	79.7	79.2	79.7	79.3	89.5	88.5	89.8	89.9	88.7	90.1
	0.2	78.7	78.6	78.1	77.9	77.0	76.7	89.0	88.8	88.6	88.4	88.6	87.8
	0.4	78.8	77.8	77.7	79.3	78.4	78.4	89.0	89.2	88.6	88.2	88.6	89.7
0.25	0.5	79.6	79.5	78.6	78.9	78.1	77.5	89.5	88.9	90.7	89.3	88.7	88.5
	0.6	79.0	79.1	80.8	80.3	80.8	80.7	89.9	89.2	91.4	90.4	91.6	90.5
	0.8	77.8	78.5	79.6	79.9	79.6	80.9	88.8	88.4	90.8	90.5	90.1	90.7
	0.2	80.2	79.6	76.6	77.6	76.2	77.0	89.6	89.4	87.8	87.6	87.1	87.1
	0.4	77.7	78.7	78.2	79.0	77.4	77.8	88.8	89.2	88.4	88.8	88.9	89.7
0.5	0.5	79.8	78.8	78.6	79.0	78.4	77.2	89.9	89.4	90.6	89.7	89.2	88.5
	0.6	79.5	79.3	78.0	77.9	78.9	79.0	90.0	89.6	88.9	88.9	90.0	90.0
	0.8	81.2	79.8	79.3	79.0	80.7	79.8	90.0	90.7	89.6	89.4	90.5	90.9
	0.2	78.8	78.1	77.6	77.4	76.3	76.3	90.1	89.5	87.7	87.9	87.8	87.6
	0.4	77.8	77.8	78.5	77.9	75.8	76.5	88.4	88.2	89.1	88.0	86.8	87.4
0.75	0.5	77.8	77.8	76.6	78.2	77.8	78.2	88.9	88.5	88.5	88.1	88.4	88.3
	0.6	78.0	78.8	78.3	78.3	77.6	77.4	89.0	89.1	88.9	89.3	88.3	88.4
	0.8	79.2	79.6	79.6	80.6	79.9	78.9	89.6	90.0	90.8	90.2	89.5	88.7
						Five-group sequential design (J=5)							
	0.2	77.2	77.2	79.1	79.1	78.0	78.1	88.1	87.6	89.2	89.2	88.4	88.6
	0.4	77.7	78.2	78.1	78.3	76.9	76.9	89.0	88.9	88.7	89.6	88.2	89.0
0	0.5	79.0	77.8	78.6	79.1	78.7	79.3	89.1	89.5	89.1	89.4	89.6	89.1
	0.6	78.5	78.4	80.4	79.8	79.5	80.7	88.9	88.8	90.5	89.7	90.3	90.1
	0.8	78.2	78.1	79.3	79.6	80.0	79.1	88.2	88.3	89.9	89.3	89.3	90.0
	0.2	78.3	80.0	78.5	77.4	78.4	77.5	88.3	89.4	88.5	89.0	88.5	89.5
	0.4	79.1	79.4	78.7	79.8	76.7	77.9	88.8	88.8	89.5	89.2	88.9	89.8
0.25	0.5	78.2	79.1	79.3	80.1	77.2	76.8	88.7	88.1	89.7	89.5	88.7	89.2
	0.6	79.5	79.5	80.7	80.8	81.7	81.1	89.2	90.0	91.1	89.8	90.8	91.3
	0.8	78.1	78.9	79.8	79.6	79.3	80.3	89.0	89.5	90.8	90.4	91.1	89.9
	0.2	79.4	79.5	76.8	76.6	76.6	76.8	90.3	90.3	87.8	88.8	87.5	88.2
	0.4	79.3	78.4	79.0	78.8	79.2	78.1	89.5	89.0	89.5	89.3	90.0	88.7
0.5	0.5	79.9	80.1	79.7	79.6	77.4	78.3	90.3	90.2	89.9	90.4	88.4	88.6
	0.6	79.3	79.9	79.5	78.3	80.0	79.6	89.9	89.9	89.8	88.9	90.7	90.2
	0.8	81.6	81.4	78.9	80.2	81.0	79.8	90.2	90.5	89.5	89.6	90.7	90.2
	0.2	79.6	79.0	78.3	78.6	76.6	77.5	90.1	90.1	89.0	87.9	88.8	88.3
	0.4	80.0	79.4	79.0	77.5	77.7	77.0	89.4	89.1	89.0	88.7	87.3	88.7
0.75	0.5	78.6	79.4	78.2	78.7	78.5	77.8	88.4	88.7	87.5	88.3	89.2	89.3
	0.6	78.2	78.7	78.3	78.7	78.0	78.8	89.3	89.6	89.1	89.2	88.5	87.9
	0.8	79.2	79.4	80.1	81.7	79.4	78.8	89.4	89.8	91.2	90.6	89.5	89.8

Table 4.9: Power(%) using the Pocock GSD with $\alpha = 0.05$, case II

		Power=80%						**Power=90%**					
		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	t	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN
		Two-group sequential design (J=2)											
	0.2	76.6	77.3	78.2	79.1	77.1	78.7	88.4	87.6	89.9	89.8	88.7	88.7
	0.4	78.6	79.0	77.3	78.5	78.0	77.2	88.2	88.9	89.7	89.0	88.1	88.4
0	0.5	78.7	77.8	78.5	78.7	78.9	78.4	88.7	89.3	89.1	88.4	88.9	89.2
	0.6	78.0	77.0	79.2	78.8	79.9	80.8	88.4	88.8	89.3	90.1	90.6	90.9
	0.8	77.5	77.9	78.8	79.3	79.8	78.2	88.8	89.1	90.6	89.3	89.3	90.0
	0.2	79.0	77.4	78.3	78.4	77.6	78.7	88.9	90.0	88.2	89.1	88.6	88.6
	0.4	78.6	77.7	78.8	78.6	77.1	77.9	89.6	89.3	89.3	88.5	89.1	88.6
0.25	0.5	79.3	79.2	79.4	79.3	77.9	77.2	90.2	88.4	89.6	89.6	88.8	88.5
	0.6	78.4	80.3	80.0	79.8	81.0	81.6	90.0	89.1	90.7	90.9	90.6	90.4
	0.8	78.0	78.2	79.3	79.6	80.5	80.9	89.4	89.1	90.8	91.2	90.1	90.1
	0.2	78.7	79.8	76.9	76.1	77.2	76.4	89.2	89.8	88.8	87.4	87.3	87.4
	0.4	79.0	78.7	78.3	79.3	78.1	78.6	89.1	89.9	89.4	89.2	89.5	88.7
0.5	0.5	81.2	79.4	79.9	79.4	78.1	77.8	89.7	89.5	89.4	89.8	89.2	88.5
	0.6	78.8	78.6	78.7	78.7	79.0	79.7	89.7	89.6	88.9	89.8	89.7	89.3
	0.8	80.1	80.2	79.1	77.7	79.8	80.0	90.5	90.6	88.6	90.3	90.4	90.4
	0.2	80.1	78.0	77.8	78.2	77.0	77.1	89.8	89.4	87.6	88.7	88.2	88.7
	0.4	78.5	78.0	77.7	78.5	78.0	76.6	89.2	88.7	89.4	89.5	88.2	88.2
0.75	0.5	77.1	77.7	77.6	77.3	78.3	77.6	89.0	89.0	89.1	88.5	89.8	89.5
	0.6	78.7	77.7	77.5	78.5	77.2	77.8	89.4	88.6	88.9	89.5	88.9	87.8
	0.8	79.2	78.8	80.9	80.9	79.3	78.5	89.8	89.6	90.6	90.6	89.0	89.3
		Five-group sequential design (J=5)											
	0.2	77.7	78.3	78.7	79.4	78.7	78.4	88.2	88.7	89.3	89.3	89.9	88.9
	0.4	78.2	78.0	79.0	79.0	77.5	77.6	88.9	89.0	89.6	88.9	88.7	88.7
0	0.5	78.6	79.1	79.6	79.8	79.0	79.1	89.1	88.3	88.6	89.0	89.5	90.2
	0.6	77.9	78.7	80.2	79.6	81.0	81.0	88.7	88.6	90.0	89.8	90.6	91.0
	0.8	77.8	78.7	79.0	79.4	79.2	79.8	88.9	88.2	90.3	89.8	90.8	89.8
	0.2	78.8	79.1	79.4	78.0	78.7	78.4	89.1	88.6	89.2	88.4	88.7	89.5
	0.4	77.7	79.8	78.4	79.0	77.8	78.7	88.9	88.7	89.4	90.0	88.8	88.6
0.25	0.5	79.5	79.7	80.0	79.0	79.0	79.1	90.1	89.3	89.6	89.2	88.8	88.5
	0.6	79.6	79.7	80.4	80.6	80.0	80.4	89.9	89.9	91.1	90.5	90.4	91.2
	0.8	77.6	77.7	80.0	81.6	80.6	80.4	89.2	89.0	90.2	90.2	90.5	90.6
	0.2	79.1	79.8	76.4	77.3	76.4	77.0	90.4	89.9	88.1	88.7	87.1	87.6
	0.4	78.6	79.4	78.6	77.7	77.7	78.8	88.4	88.8	88.4	89.3	89.8	90.0
0.5	0.5	79.3	80.6	79.7	80.2	78.0	77.6	90.2	89.9	89.6	89.0	89.4	89.7
	0.6	78.6	80.2	79.3	78.7	79.8	79.3	89.6	89.4	89.2	89.1	89.7	89.7
	0.8	81.2	80.3	79.7	78.3	79.8	80.5	89.8	90.5	90.3	88.8	91.1	90.4
	0.2	79.4	80.3	76.8	78.0	77.7	76.6	89.6	90.1	88.9	88.5	88.2	88.1
	0.4	78.5	78.3	78.9	79.4	78.0	77.4	89.0	88.9	89.7	88.9	87.4	87.4
0.75	0.5	78.4	78.5	78.6	77.9	79.0	79.5	88.6	88.5	89.1	89.0	89.5	89.5
	0.6	79.2	78.5	79.1	78.7	77.7	77.8	89.7	89.2	89.6	89.7	89.0	89.1
	0.8	80.1	78.1	80.1	80.7	80.8	78.7	89.2	89.8	90.9	90.5	88.8	88.8

Table 4.10: Power(%) using the O'Brien-Fleming GSD with $\alpha = 0.05$, case III

		Power=80%						**Power=90%**					
		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	t	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN
Two-group sequential design (J=2)													
	0.2	77.2	77.4	79.7	78.6	78.0	78.7	87.5	87.7	89.5	90.4	89.0	89.0
	0.4	77.9	78.4	79.1	77.7	78.0	78.7	88.8	87.8	89.4	89.2	89.0	89.0
0	0.5	77.8	78.5	77.7	78.5	79.0	78.6	87.9	88.7	89.3	88.8	89.8	90.1
	0.6	77.6	77.9	79.2	77.9	80.4	79.9	88.4	88.5	88.8	89.1	90.0	89.3
	0.8	75.6	76.4	79.3	78.2	79.7	79.3	86.9	87.4	89.9	89.8	89.6	89.5
	0.2	80.4	78.8	79.2	79.3	80.0	79.9	89.9	90.1	89.0	89.9	89.6	89.7
	0.4	80.1	80.5	79.5	79.1	80.1	79.2	90.4	91.1	89.2	89.0	89.2	89.0
0.25	0.5	80.8	79.1	79.2	78.8	79.2	78.4	89.7	88.9	88.6	89.9	89.4	89.5
	0.6	80.5	81.6	81.9	81.6	82.0	81.7	90.7	90.8	91.9	91.0	91.3	91.2
	0.8	79.9	80.0	81.9	80.4	81.4	81.8	90.1	90.0	90.5	91.0	90.7	90.8
	0.2	81.3	80.1	79.1	78.6	77.1	77.0	90.1	89.9	89.2	88.9	88.9	88.2
	0.4	78.9	79.8	79.5	80.1	79.7	79.0	89.8	89.4	90.5	90.4	90.2	89.7
0.5	0.5	80.3	79.9	80.1	79.7	79.0	78.8	90.3	89.9	91.0	90.7	89.8	89.4
	0.6	81.5	81.1	79.3	80.0	80.8	81.2	89.7	90.1	90.4	89.9	90.2	90.7
	0.8	81.4	80.7	79.1	79.9	81.1	81.1	90.9	91.7	90.3	89.2	90.7	91.2
	0.2	80.0	79.6	78.0	77.8	79.0	77.9	90.7	90.8	89.1	89.6	89.4	89.3
	0.4	78.5	76.2	78.3	79.5	77.2	77.2	89.3	87.9	89.2	89.5	87.3	87.5
0.75	0.5	78.4	78.4	77.4	76.8	78.0	78.9	89.9	89.4	88.8	88.7	89.2	89.1
	0.6	79.9	79.0	80.3	79.3	78.0	78.8	90.2	90.0	89.3	90.0	89.1	88.7
	0.8	78.6	78.7	80.4	81.3	79.0	79.0	89.6	90.5	90.5	90.5	89.1	89.4
Five-group sequential design (J=5)													
	0.2	77.5	77.2	80.1	79.6	78.3	79.5	87.9	87.8	90.4	89.6	90.1	88.7
	0.4	77.4	77.0	78.2	79.4	78.9	78.2	88.7	88.3	89.2	89.9	88.6	89.2
0	0.5	76.8	77.3	78.7	78.4	78.9	79.2	88.2	87.5	88.1	88.4	89.5	89.9
	0.6	77.3	77.4	78.1	77.5	80.8	80.0	88.1	88.2	89.5	89.6	90.2	90.4
	0.8	74.8	74.1	79.1	78.9	80.0	79.2	86.1	87.7	89.5	89.3	89.2	89.8
	0.2	79.3	79.0	79.4	79.9	79.9	79.2	89.7	90.3	89.8	90.1	89.7	89.5
	0.4	79.9	80.2	79.9	79.1	79.5	80.2	89.1	90.1	89.5	90.7	89.5	90.7
0.25	0.5	80.0	79.6	80.0	80.5	79.5	78.4	89.8	90.6	89.8	88.9	89.4	88.9
	0.6	79.8	81.2	81.4	81.7	81.3	80.7	90.9	90.8	91.1	91.9	90.9	91.5
	0.8	79.4	79.3	81.2	81.3	81.0	81.6	90.2	90.3	90.4	90.6	91.4	91.2
	0.2	81.6	80.2	78.4	78.4	78.4	78.1	89.7	90.4	89.6	88.4	88.8	88.5
	0.4	80.3	80.3	79.9	80.0	79.8	79.5	89.6	89.4	89.9	89.7	90.0	90.0
0.5	0.5	80.3	80.0	81.4	82.0	78.8	79.3	90.2	90.4	89.9	91.1	89.2	89.5
	0.6	81.1	81.4	80.7	80.7	80.9	81.0	90.8	90.3	90.9	90.1	91.0	90.3
	0.8	80.9	82.1	80.3	81.1	81.2	82.1	91.4	91.0	90.4	90.1	90.6	91.1
	0.2	80.7	79.2	79.6	78.5	78.3	78.8	90.0	90.3	89.7	89.0	88.7	89.4
	0.4	78.6	77.9	79.0	79.1	77.8	77.9	89.1	88.5	89.7	89.8	88.5	87.8
0.75	0.5	78.7	78.0	78.1	78.1	79.1	79.7	88.6	89.0	89.2	88.4	89.1	89.1
	0.6	78.9	78.8	78.5	79.8	79.2	78.2	89.9	89.6	89.7	89.4	89.7	89.9
	0.8	80.0	79.3	80.1	80.8	80.0	79.0	89.9	89.0	89.8	90.4	89.8	89.1

Table 4.11: Power(%) using the Pocock GSD with $\alpha = 0.05$, case III

		Power=80%						**Power=90%**					
		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	t	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN
						Two-group sequential design (J=2)							
	0.2	77.7	76.6	79.1	79.1	78.9	78.4	87.6	87.8	90.6	89.2	89.1	89.6
	0.4	76.3	78.0	78.8	78.4	78.8	78.7	87.8	88.0	89.6	89.7	87.9	88.9
0	0.5	77.5	78.3	77.4	77.4	79.2	79.6	87.8	87.4	88.2	88.8	89.0	89.6
	0.6	78.6	76.5	78.9	79.1	80.0	79.8	87.6	87.9	88.7	89.3	90.3	89.8
	0.8	74.7	76.3	78.8	79.4	79.0	79.5	87.6	86.0	89.3	89.2	89.6	89.5
	0.2	79.2	80.2	79.4	79.3	79.6	80.0	89.8	90.3	89.5	89.5	90.2	89.8
	0.4	80.2	80.4	79.0	79.0	79.6	78.6	89.4	90.1	89.8	89.1	89.9	90.5
0.25	0.5	79.2	79.5	80.4	79.4	78.5	78.7	90.2	90.5	90.2	90.4	89.9	88.7
	0.6	81.3	80.2	81.7	82.4	80.9	80.8	90.2	90.7	92.0	91.5	90.8	91.1
	0.8	79.2	79.7	81.3	80.7	80.5	81.6	89.8	90.1	90.6	90.1	90.8	90.9
	0.2	80.8	81.0	77.9	78.9	77.1	77.2	90.0	90.7	88.6	89.7	87.5	88.2
	0.4	79.0	78.8	79.0	80.0	79.1	80.1	89.8	90.2	89.4	91.4	90.5	90.4
0.5	0.5	79.9	80.0	80.5	80.3	79.2	78.8	90.4	90.7	90.4	90.7	90.4	89.9
	0.6	81.8	80.8	80.2	81.2	80.1	81.9	90.0	91.5	89.6	90.1	91.6	90.9
	0.8	82.0	81.1	79.6	80.0	80.3	80.2	90.9	90.8	89.8	90.0	90.6	90.7
	0.2	80.4	79.6	78.8	77.1	79.0	78.7	89.1	90.2	88.8	89.0	89.1	89.5
	0.4	78.0	78.1	78.9	79.4	76.2	77.0	88.1	87.4	89.9	89.5	87.9	88.5
0.75	0.5	78.2	78.2	77.4	77.9	79.0	79.1	89.2	89.2	88.4	88.0	89.6	89.7
	0.6	79.6	80.3	79.2	78.4	78.2	78.4	89.4	89.5	89.6	89.5	89.6	88.7
	0.8	80.5	79.0	80.3	81.0	77.7	79.0	89.2	90.3	90.8	90.8	88.8	88.9
						Five-group sequential design (J=5)							
	0.2	77.2	77.9	80.0	80.1	79.1	79.9	88.3	88.1	90.7	89.3	89.1	89.6
	0.4	78.6	77.1	79.9	78.4	77.8	78.8	89.3	89.1	88.5	88.9	88.9	89.5
0	0.5	78.9	77.4	79.2	78.1	79.7	78.3	88.1	88.4	88.8	88.7	91.0	89.8
	0.6	78.3	77.6	79.9	80.0	81.1	79.8	88.2	87.7	89.1	89.1	90.6	89.7
	0.8	76.1	75.7	78.3	79.0	79.0	79.2	86.0	87.1	89.2	89.0	90.3	89.5
	0.2	80.5	79.6	80.2	81.4	79.7	80.2	90.1	89.8	90.3	90.1	89.3	89.7
	0.4	80.4	80.5	80.5	79.8	80.1	79.8	90.3	90.0	90.2	89.2	90.0	90.1
0.25	0.5	79.1	79.8	79.5	80.4	78.7	79.8	89.3	90.5	89.4	90.2	89.5	89.5
	0.6	80.3	81.4	82.2	82.1	81.1	81.6	90.7	90.0	91.8	91.7	90.6	90.6
	0.8	79.4	78.7	80.5	80.6	81.0	81.4	89.9	90.0	91.3	90.9	91.4	90.9
	0.2	80.4	80.3	78.3	78.7	77.3	77.2	90.2	90.3	88.7	88.3	87.9	88.7
	0.4	80.3	79.3	79.6	81.3	78.7	79.6	90.1	89.7	90.9	90.1	90.2	89.9
0.5	0.5	80.7	80.2	80.0	81.4	79.5	80.0	90.7	90.5	90.6	90.7	89.7	89.4
	0.6	80.6	80.4	80.0	79.9	81.4	81.2	91.1	90.6	90.8	89.7	90.7	91.0
	0.8	81.4	82.0	78.9	80.1	81.7	81.4	90.7	90.7	91.2	89.8	91.3	91.4
	0.2	80.5	80.9	77.9	78.4	78.1	78.9	90.0	90.6	89.4	88.5	89.0	89.2
	0.4	78.6	79.5	79.3	78.9	77.9	77.3	88.6	89.1	90.0	89.3	88.5	88.5
0.75	0.5	78.7	78.5	77.9	78.7	79.3	79.8	89.2	89.4	88.9	87.9	88.9	89.5
	0.6	80.5	78.7	80.1	79.3	77.8	78.5	90.1	89.2	89.3	90.1	88.6	89.1
	0.8	80.4	80.4	80.7	79.8	78.9	78.5	90.1	89.6	90.2	90.3	89.1	90.2

Table 4.12: Expected sample sizes using GSD with $\alpha = 0.05$, case I

		Power=80%						**Power=90%**					
		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	t	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock
Two-group sequential design (J=2)													
	0.2	187	178	132	127	98	93	238	217	168	154	124	114
	0.4	195	184	140	132	102	97	248	227	175	162	128	118
0	0.5	219	209	161	155	121	115	277	255	205	186	154	139
	0.6	260	248	205	192	160	149	328	301	258	237	199	180
	0.8	523	496	453	425	357	335	662	604	573	514	451	401
	0.2	178	172	129	123	94	91	228	209	164	151	119	109
	0.4	187	177	136	128	100	95	236	215	172	158	128	115
0.25	0.5	212	202	158	150	117	111	270	245	201	184	148	135
	0.6	259	243	203	194	156	147	328	299	256	231	197	178
	0.8	505	482	437	412	357	333	644	584	550	501	452	402
	0.2	174	166	125	119	94	90	222	201	159	146	119	110
	0.4	176	168	134	128	101	96	225	207	169	155	127	116
0.5	0.5	206	194	156	148	117	111	259	235	198	181	149	135
	0.6	248	236	195	185	150	142	316	287	249	225	191	172
	0.8	518	482	429	402	345	324	656	599	541	490	434	391
	0.2	174	165	133	129	102	97	221	202	169	156	130	118
	0.4	176	166	140	133	106	101	222	204	178	163	135	123
0.75	0.5	199	190	161	153	126	119	254	232	204	187	160	144
	0.6	247	234	206	194	158	149	312	285	260	237	200	181
	0.8	520	492	471	445	357	337	659	597	594	534	452	407
Five-group sequential design (J=5)													
	0.2	170	167	120	118	89	87	210	193	147	135	109	100
	0.4	177	174	127	125	91	91	218	201	156	145	114	106
0	0.5	200	198	147	145	110	109	246	226	179	169	136	127
	0.6	236	236	186	185	142	140	293	270	228	209	174	161
	0.8	475	468	408	410	322	321	586	545	500	458	393	365
	0.2	163	160	118	116	86	85	201	187	144	132	106	98
	0.4	169	166	124	119	91	91	209	191	151	140	112	103
0.25	0.5	194	188	144	142	106	105	237	216	177	164	130	122
	0.6	235	231	183	180	141	140	287	264	224	204	172	159
	0.8	459	457	397	392	323	319	565	527	486	452	392	363
	0.2	159	154	113	113	85	84	195	182	141	130	106	98
	0.4	160	158	122	122	91	91	197	183	149	138	113	104
0.5	0.5	187	183	141	141	107	107	229	211	176	162	132	123
	0.6	226	225	177	175	137	136	277	256	219	200	166	155
	0.8	471	465	387	386	312	312	573	527	479	444	382	354
	0.2	158	156	121	121	92	92	196	178	150	138	114	107
	0.4	159	157	127	127	96	97	197	181	156	144	118	111
0.75	0.5	182	181	147	146	114	113	225	206	180	167	139	130
	0.6	224	224	187	182	143	142	274	254	230	214	175	164
	0.8	474	466	425	420	324	326	582	534	522	482	399	375

Table 4.13: Expected sample sizes using GSD with $\alpha = 0.05$, case II

		Power=80%						**Power=90%**					
		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	t	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock
						Two-group sequential design (J=2)							
	0.2	417	400	299	282	214	204	527	489	378	344	269	245
	0.4	426	405	300	286	212	203	541	493	380	346	271	248
0	0.5	474	448	343	326	251	238	597	547	435	395	318	289
	0.6	552	527	429	407	321	305	705	646	543	494	404	369
	0.8	1049	1005	885	842	687	647	1333	1221	1122	1016	865	789
	0.2	404	388	288	275	209	199	512	474	367	338	261	241
	0.4	409	389	293	279	213	202	521	475	368	339	270	246
0.25	0.5	456	433	338	319	244	231	579	529	427	386	306	282
	0.6	553	525	425	401	319	300	696	635	535	485	403	366
	0.8	1049	988	891	837	691	646	1322	1203	1117	1013	873	790
	0.2	395	374	276	266	203	193	494	456	352	322	256	237
	0.4	387	367	290	275	214	204	489	452	369	330	270	246
0.5	0.5	448	420	333	317	241	228	560	511	420	389	305	278
	0.6	532	499	406	386	311	290	672	612	511	465	393	358
	0.8	1061	990	861	818	692	652	1342	1206	1099	999	874	789
	0.2	388	366	293	279	222	212	488	445	373	343	279	257
	0.4	378	362	302	289	222	209	481	442	384	352	280	255
0.75	0.5	427	404	336	322	261	248	543	498	430	393	328	301
	0.6	519	492	427	407	318	303	656	603	538	488	403	368
	0.8	1057	1009	952	889	714	673	1339	1224	1194	1080	908	821
						Five-group sequential design (J=5)							
	0.2	383	372	270	266	193	191	468	436	333	305	237	219
	0.4	386	380	274	267	195	194	477	441	336	307	238	221
0	0.5	429	423	311	308	229	227	529	489	383	360	280	260
	0.6	506	501	386	384	290	287	617	570	471	435	355	328
	0.8	957	947	806	799	622	625	1178	1096	989	914	767	701
	0.2	369	363	262	257	189	187	452	419	322	300	233	217
	0.4	371	370	265	262	193	193	458	415	327	299	236	222
0.25	0.5	417	410	306	299	222	220	511	463	376	346	272	254
	0.6	497	488	384	375	288	289	613	561	467	429	352	327
	0.8	952	946	804	792	629	616	1161	1081	985	911	758	706
	0.2	355	353	253	252	184	184	438	397	310	287	227	214
	0.4	350	348	263	260	193	193	433	403	325	296	237	222
0.5	0.5	405	396	301	297	219	219	497	454	370	342	270	251
	0.6	482	476	366	364	280	278	595	545	448	421	342	320
	0.8	953	933	782	777	621	621	1175	1094	967	887	760	704
	0.2	353	344	267	267	201	199	430	397	329	305	247	229
	0.4	341	341	274	270	200	200	423	391	337	308	249	231
0.75	0.5	386	384	307	304	238	234	479	445	378	350	291	269
	0.6	474	462	387	383	288	286	580	534	476	440	357	333
	0.8	966	942	857	846	647	638	1190	1093	1044	959	796	737

Table 4.14: Expected sample sizes using GSD with $\alpha = 0.05$, case III

		Power=80%						**Power=90%**					
		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	t	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock
Two-group sequential design (J=2)													
	0.2	1659	1563	1198	1151	838	800	2099	1918	1523	1374	1063	966
	0.4	1628	1554	1163	1106	821	774	2073	1901	1468	1351	1035	953
0	0.5	1802	1724	1280	1223	932	884	2269	2082	1633	1483	1179	1076
	0.6	2080	1976	1568	1495	1165	1097	2640	2439	1992	1811	1466	1333
	0.8	3712	3559	3166	2983	2420	2287	4694	4307	3988	3658	3043	2799
	0.2	1654	1574	1169	1105	836	794	2080	1903	1478	1348	1052	960
	0.4	1644	1540	1138	1078	823	776	2079	1909	1426	1312	1048	946
0.25	0.5	1797	1682	1293	1217	917	867	2255	2051	1641	1485	1160	1056
	0.6	2151	2027	1638	1530	1170	1102	2694	2481	2076	1868	1469	1346
	0.8	3925	3742	3272	3073	2486	2353	4951	4505	4118	3721	3146	2863
	0.2	1576	1493	1107	1057	791	756	2004	1825	1412	1288	1010	923
	0.4	1526	1453	1132	1070	806	768	1915	1752	1428	1303	1023	929
0.5	0.5	1705	1613	1284	1207	916	867	2166	1963	1621	1474	1148	1051
	0.6	2067	1941	1537	1451	1156	1099	2605	2367	1931	1762	1458	1323
	0.8	3943	3653	3178	3004	2498	2335	4982	4533	3995	3635	3122	2848
	0.2	1526	1440	1146	1084	858	820	1904	1757	1458	1335	1092	1005
	0.4	1442	1368	1153	1094	827	783	1810	1677	1454	1325	1049	962
0.75	0.5	1618	1525	1265	1213	968	918	2043	1840	1611	1463	1224	1123
	0.6	1967	1863	1575	1490	1176	1122	2507	2277	1991	1831	1489	1368
	0.8	3885	3687	3378	3156	2544	2418	4925	4515	4267	3872	3224	2940
Five-group sequential design (J=5)													
	0.2	1502	1482	1078	1060	765	749	1834	1716	1330	1217	942	871
	0.4	1481	1454	1059	1035	744	742	1836	1700	1304	1206	918	843
0	0.5	1643	1611	1170	1153	848	837	2035	1881	1434	1327	1049	955
	0.6	1898	1881	1428	1395	1052	1038	2347	2157	1757	1607	1293	1185
	0.8	3412	3372	2877	2859	2189	2174	4215	3934	3533	3274	2705	2467
	0.2	1502	1463	1057	1036	756	738	1846	1682	1304	1172	925	853
	0.4	1488	1442	1036	1009	744	733	1832	1650	1271	1150	916	847
0.25	0.5	1624	1596	1166	1156	832	826	1986	1819	1440	1312	1020	941
	0.6	1947	1908	1479	1432	1056	1041	2382	2151	1802	1628	1298	1190
	0.8	3559	3511	2955	2912	2263	2198	4353	4002	3621	3283	2763	2513
	0.2	1423	1394	1005	995	719	716	1746	1597	1242	1151	886	825
	0.4	1375	1362	1024	1000	737	727	1696	1563	1252	1144	899	822
0.5	0.5	1547	1513	1157	1144	829	820	1904	1759	1428	1301	1016	931
	0.6	1868	1829	1379	1362	1048	1023	2275	2079	1700	1572	1281	1173
	0.8	3582	3465	2877	2842	2243	2198	4354	3938	3535	3208	2756	2509
	0.2	1375	1340	1041	1041	781	779	1691	1557	1285	1181	966	886
	0.4	1303	1294	1046	1035	750	747	1614	1495	1283	1170	925	863
0.75	0.5	1468	1447	1153	1138	880	870	1806	1649	1417	1318	1081	1001
	0.6	1795	1734	1433	1412	1064	1064	2213	2006	1752	1612	1304	1223
	0.8	3534	3427	3070	2999	2304	2301	4351	3999	3765	3449	2837	2603

Table 4.15: GSD design sample sizes (maximum) with $\alpha = 0.05$, case I

		Power=80%						**Power=90%**					
		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	t	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock
Two-group sequential design (J=2)													
	0.2	208	229	146	161	108	118	278	303	195	213	144	157
	0.4	216	238	153	169	110	121	289	316	205	224	147	161
0	0.5	242	267	178	196	132	146	324	354	238	260	177	193
	0.6	286	315	225	248	172	189	382	417	301	329	229	251
	0.8	574	632	496	546	385	424	767	838	663	725	514	562
	0.2	200	220	143	157	103	114	267	291	191	209	138	150
	0.4	206	227	149	164	109	120	276	301	199	217	146	159
0.25	0.5	234	258	174	192	127	139	313	342	233	255	169	185
	0.6	285	314	223	246	170	187	381	416	298	326	227	248
	0.8	555	611	477	526	387	426	742	810	638	697	518	565
	0.2	194	214	137	150	102	112	259	283	183	199	136	149
	0.4	195	215	148	163	110	121	261	285	198	216	146	160
0.5	0.5	227	250	172	189	127	140	303	331	230	251	170	186
	0.6	275	303	214	236	163	179	367	401	287	313	217	237
	0.8	573	631	466	513	371	408	766	837	623	681	496	542
	0.2	193	212	146	161	111	122	258	282	195	213	149	162
	0.4	194	213	154	169	115	126	259	283	205	224	153	167
0.75	0.5	221	243	176	193	136	150	295	322	235	256	182	199
	0.6	272	299	226	249	172	189	363	397	303	330	230	251
	0.8	575	633	519	572	387	426	768	839	694	758	518	565
Five-group sequential design (J=5)													
	0.2	212	253	149	178	110	131	283	333	199	234	146	172
	0.4	220	263	156	187	112	134	294	346	209	245	150	176
0	0.5	247	295	182	217	135	161	330	388	242	285	180	212
	0.6	291	348	230	275	175	209	389	458	307	361	234	275
	0.8	585	699	506	605	392	469	782	919	676	795	524	616
	0.2	203	243	146	174	105	126	272	320	195	229	140	165
	0.4	210	251	152	181	111	133	281	330	203	238	148	175
0.25	0.5	239	286	178	213	129	154	319	375	237	279	172	203
	0.6	291	348	228	272	173	207	388	457	304	358	232	272
	0.8	566	676	487	582	395	472	756	889	650	765	527	620
	0.2	198	237	139	166	104	124	264	311	186	219	139	163
	0.4	199	238	151	180	112	134	266	312	201	237	149	175
0.5	0.5	231	276	175	210	130	155	309	363	234	276	173	204
	0.6	280	335	219	261	166	198	374	440	292	343	221	260
	0.8	584	698	475	568	378	452	781	918	635	747	505	594
	0.2	197	235	149	178	113	136	263	309	199	234	151	178
	0.4	197	236	157	187	117	140	264	310	209	246	156	183
0.75	0.5	225	269	179	214	139	166	301	353	239	281	186	218
	0.6	277	331	231	276	175	210	370	435	308	363	234	275
	0.8	586	701	529	633	395	472	783	921	707	832	527	620

Table 4.16: GSD design sample sizes (maximum) with $\alpha = 0.05$, case II

		Power=80%						**Power=90%**					
		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	t	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock
						Two-group sequential design (J=2)							
	0.2	464	511	331	364	234	257	620	678	442	483	313	341
	0.4	472	519	332	366	233	257	631	689	444	485	311	340
0	0.5	524	577	378	416	275	303	700	765	505	552	368	402
	0.6	612	674	472	520	353	389	818	894	631	689	473	516
	0.8	1158	1275	979	1078	753	829	1548	1691	1310	1431	1006	1099
	0.2	451	497	319	351	228	251	603	659	427	466	305	333
	0.4	454	500	323	355	233	256	607	663	432	471	311	340
0.25	0.5	508	560	373	411	266	292	680	742	499	545	355	388
	0.6	609	671	471	519	352	387	815	890	630	688	470	513
	0.8	1151	1267	982	1081	758	834	1539	1681	1313	1435	1013	1107
	0.2	438	482	304	334	221	243	586	640	406	443	295	322
	0.4	429	472	320	353	233	257	573	626	428	468	312	340
0.5	0.5	495	545	369	406	263	290	662	723	493	539	352	384
	0.6	588	647	445	490	339	373	786	858	595	650	453	495
	0.8	1174	1292	950	1046	754	831	1570	1714	1271	1388	1009	1102
	0.2	430	474	323	356	242	266	575	628	432	472	323	353
	0.4	419	462	334	368	241	265	561	613	447	488	321	351
0.75	0.5	471	519	371	409	284	313	630	689	496	542	380	415
	0.6	573	631	469	517	347	382	766	837	627	685	464	507
	0.8	1177	1296	1047	1153	783	862	1574	1720	1400	1529	1047	1143
						Five-group sequential design (J=5)							
	0.2	473	566	337	403	238	285	632	744	450	529	318	375
	0.4	481	575	339	405	238	284	642	756	453	532	317	373
0	0.5	534	639	385	461	281	336	714	839	515	606	375	441
	0.6	624	746	481	575	360	431	833	980	643	756	481	566
	0.8	1181	1412	999	1194	768	918	1578	1856	1334	1570	1025	1206
	0.2	460	550	325	389	233	278	615	723	435	511	311	365
	0.4	463	553	329	393	237	284	618	727	440	517	317	373
0.25	0.5	518	620	381	455	271	324	693	815	508	598	362	425
	0.6	622	743	481	574	359	429	830	977	642	755	479	563
	0.8	1174	1403	1002	1197	773	924	1568	1844	1338	1574	1032	1214
	0.2	447	534	310	370	225	269	597	702	414	486	301	354
	0.4	437	523	327	390	238	284	584	687	436	513	318	374
0.5	0.5	505	604	376	450	268	321	675	794	503	591	358	422
	0.6	599	716	454	543	346	413	801	942	606	713	462	543
	0.8	1197	1431	969	1158	769	920	1599	1881	1295	1523	1028	1209
	0.2	439	524	330	394	246	294	586	689	440	518	329	387
	0.4	428	511	341	407	245	293	571	672	455	535	328	385
0.75	0.5	481	575	379	452	290	347	642	755	506	595	387	456
	0.6	584	698	478	572	354	423	780	918	639	752	473	556
	0.8	1201	1435	1068	1276	798	954	1604	1887	1426	1678	1067	1255

Table 4.17: GSD design sample sizes (maximum) with $\alpha = 0.05$, case III

		Power=80%						**Power=90%**					
		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	t	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock
Two-group sequential design (J=2)													
	0.2	1826	2011	1335	1470	931	1025	2442	2667	1785	1950	1244	1359
	0.4	1804	1986	1290	1420	904	995	2412	2635	1725	1884	1209	1320
0	0.5	1997	2199	1418	1562	1034	1138	2671	2917	1897	2072	1382	1510
	0.6	2303	2536	1738	1914	1289	1419	3080	3365	2324	2539	1723	1882
	0.8	4068	4480	3506	3860	2671	2941	5441	5943	4688	5121	3571	3901
	0.2	1845	2032	1301	1433	924	1018	2467	2695	1740	1901	1236	1350
	0.4	1835	2021	1261	1389	912	1004	2454	2681	1686	1842	1220	1332
0.25	0.5	2000	2203	1436	1581	1011	1113	2675	2922	1920	2098	1352	1477
	0.6	2402	2645	1840	2026	1304	1436	3212	3509	2460	2688	1743	1904
	0.8	4368	4810	3646	4015	2782	3064	5842	6382	4876	5326	3721	4064
	0.2	1765	1943	1228	1352	871	959	2360	2578	1641	1793	1164	1272
	0.4	1698	1870	1254	1381	897	988	2271	2481	1677	1832	1199	1310
0.5	0.5	1912	2105	1434	1579	1006	1108	2556	2792	1918	2095	1346	1470
	0.6	2305	2538	1703	1876	1287	1418	3083	3367	2278	2488	1721	1880
	0.8	4417	4863	3527	3884	2763	3043	5906	6452	4717	5153	3695	4036
	0.2	1695	1867	1271	1400	947	1043	2267	2476	1700	1857	1267	1384
	0.4	1592	1753	1274	1402	907	999	2129	2325	1703	1860	1213	1325
0.75	0.5	1785	1966	1399	1541	1069	1177	2388	2608	1871	2044	1430	1562
	0.6	2192	2413	1747	1924	1293	1424	2931	3201	2336	2552	1729	1889
	0.8	4327	4765	3777	4159	2809	3094	5787	6321	5051	5517	3757	4104
Five-group sequential design (J=5)													
	0.2	1862	2226	1361	1627	949	1134	2488	2927	1818	2139	1268	1491
	0.4	1839	2199	1316	1573	922	1102	2458	2891	1758	2068	1231	1449
0	0.5	2037	2435	1447	1729	1054	1260	2721	3201	1933	2274	1408	1657
	0.6	2349	2808	1772	2119	1314	1571	3138	3692	2368	2786	1756	2066
	0.8	4149	4960	3575	4274	2724	3256	5543	6521	4777	5619	3639	4281
	0.2	1881	2249	1327	1586	943	1127	2514	2957	1773	2086	1260	1482
	0.4	1871	2237	1286	1537	930	1112	2500	2941	1718	2021	1243	1462
0.25	0.5	2040	2439	1465	1751	1031	1232	2726	3206	1957	2302	1377	1620
	0.6	2450	2929	1876	2243	1330	1590	3273	3850	2507	2949	1776	2090
	0.8	4455	5326	3718	4445	2837	3392	5952	7002	4968	5844	3791	4459
	0.2	1800	2151	1252	1497	888	1062	2404	2828	1672	1967	1186	1396
	0.4	1732	2071	1279	1529	915	1093	2314	2722	1708	2010	1222	1437
0.5	0.5	1950	2331	1463	1749	1026	1227	2605	3064	1954	2299	1371	1613
	0.6	2351	2810	1737	2077	1313	1569	3141	3695	2321	2730	1754	2063
	0.8	4504	5385	3597	4301	2818	3369	6018	7079	4806	5654	3765	4429
	0.2	1729	2067	1296	1549	966	1155	2310	2717	1732	2037	1291	1518
	0.4	1623	1941	1299	1553	925	1106	2169	2551	1735	2041	1236	1454
0.75	0.5	1821	2177	1427	1706	1090	1304	2433	2862	1906	2242	1457	1714
	0.6	2235	2672	1782	2130	1319	1577	2986	3513	2380	2800	1762	2073
	0.8	4413	5276	3852	4605	2865	3425	5896	6936	5146	6054	3828	4503

Table 4.18: Fixed sample design sample sizes with $\alpha = 0.05$, case I

		Power=80%			**Power=90%**		
$\boldsymbol{\rho}$	t	$\boldsymbol{\lambda} = 0$	$\lambda = 0.5$	$\lambda = 0.75$	$\boldsymbol{\lambda} = 0$	$\lambda = 0.5$	$\lambda = 0.75$
	0.2	206	145	107	276	194	143
	0.4	214	152	109	287	203	146
0	0.5	240	177	131	322	236	175
	0.6	283	224	170	379	299	228
	0.8	569	492	382	762	659	511
	0.2	198	142	102	265	190	137
	0.4	205	148	108	274	198	145
0.25	0.5	232	173	126	311	231	168
	0.6	283	221	169	379	296	226
	0.8	550	474	384	737	634	514
	0.2	193	136	101	258	181	135
	0.4	194	147	109	259	196	145
0.5	0.5	225	171	126	301	228	169
	0.6	273	213	161	365	285	216
	0.8	568	462	368	761	619	492
	0.2	191	145	110	256	194	148
	0.4	192	153	114	257	204	152
0.75	0.5	219	174	135	293	233	181
	0.6	270	225	171	361	300	228
	0.8	570	515	384	763	689	514

Table 4.19: Fixed sample design sample sizes with $\alpha = 0.05$, case II

		Power=80%			**Power=90%**		
$\boldsymbol{\rho}$	t	$\boldsymbol{\lambda} = 0$	$\lambda = 0.5$	$\lambda = 0.75$	$\boldsymbol{\lambda} = 0$	$\lambda = 0.5$	$\lambda = 0.75$
	0.2	460	328	232	616	439	310
	0.4	468	330	231	626	441	309
0	0.5	520	375	273	696	502	365
	0.6	607	468	351	812	627	469
	0.8	1149	972	747	1538	1301	999
	0.2	448	317	226	599	424	303
	0.4	450	320	231	603	429	309
0.25	0.5	504	370	263	675	495	353
	0.6	605	468	349	809	626	467
	0.8	1142	974	752	1528	1304	1006
	0.2	434	301	219	581	403	293
	0.4	425	318	231	569	425	310
0.5	0.5	491	366	261	658	490	349
	0.6	583	442	337	780	591	450
	0.8	1164	943	748	1559	1262	1002
	0.2	427	321	240	571	429	321
	0.4	416	331	239	557	443	319
0.75	0.5	468	368	282	626	493	378
	0.6	568	465	344	761	623	461
	0.8	1168	1039	777	1563	1390	1040

Table 4.20: Fixed sample design sample sizes with $\alpha = 0.05$, case III

		Power=80%			**Power=90%**		
$\boldsymbol{\rho}$	t	$\boldsymbol{\lambda} = 0$	$\lambda = 0.5$	$\lambda = 0.75$	$\boldsymbol{\lambda} = 0$	$\lambda = 0.5$	$\lambda = 0.75$
	0.2	1812	1324	923	2425	1772	1236
	0.4	1789	1280	897	2395	1713	1200
0	0.5	1981	1407	1025	2652	1884	1373
	0.6	2285	1724	1279	3059	2308	1711
	0.8	4036	3478	2649	5403	4655	3547
	0.2	1830	1291	917	2450	1728	1228
	0.4	1821	1251	905	2437	1675	1211
0.25	0.5	1984	1425	1003	2656	1907	1342
	0.6	2383	1825	1293	3190	2443	1731
	0.8	4334	3617	2760	5801	4842	3695
	0.2	1751	1218	864	2343	1630	1156
	0.4	1685	1244	890	2256	1665	1191
0.5	0.5	1896	1423	998	2539	1905	1336
	0.6	2287	1690	1277	3061	2262	1709
	0.8	4382	3499	2741	5865	4684	3669
	0.2	1682	1261	940	2251	1688	1258
	0.4	1579	1264	900	2114	1691	1205
0.75	0.5	1771	1388	1061	2371	1858	1420
	0.6	2174	1733	1283	2910	2320	1717
	0.8	4293	3747	2787	5747	5016	3731

Table 4.21: AUC, power(%) using the O'Brien-Fleming GSD with $\alpha = 0.05$, case I

	Power=80%						**Power=90%**					
	$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN
	Two-group sequential design (J=2)											
0	79.2	79.1	80.3	80.2	80.4	80.4	89.1	89.7	90.2	89.5	91.0	90.5
0.25	80.2	79.9	78.9	80.1	81.1	79.5	89.7	89.3	89.8	89.5	90.0	90.4
0.5	80.4	79.8	78.8	79.8	81.1	80.9	89.8	89.9	90.3	89.8	90.2	90.2
0.75	80.5	79.3	79.4	78.3	79.4	79.6	89.6	90.5	90.0	89.7	89.7	89.8
	Five-group sequential design (J=5)											
0	79.9	79.4	79.0	79.5	80.4	81.4	89.6	89.9	89.9	90.1	90.2	90.4
0.25	80.0	79.4	80.3	80.1	79.5	80.7	89.7	89.2	89.5	89.4	89.9	90.2
0.5	78.9	79.6	79.4	80.4	80.8	80.4	89.1	90.1	89.2	88.9	90.3	89.9
0.75	80.0	81.2	79.5	79.5	80.3	80.0	90.0	90.5	89.6	89.4	90.3	90.1

For statistical studies comparing AUCs of two clustered ROC, we get the following results using same approach. Instead of analyzing on particular FPR, here we study the summary measurement, AUC, of the investigational ROC curves. We follow the same steps as discussed earlier conducting the study, which includes sample size calculation, GSD size determination, simulation for powers, and calculation of expected sample size.

The simulation results in Tables 4.21 - 4.35 shows that the simulated powers are close to the expected values, 80% or 90%, with sample sizes calculated using power approach. In each case we find that the power goals are closely met for both OBrien-Fleming and Pocock methods with different number of interim looks and also for different ρ, λ combinations. Not surprisingly, we have similar results for lognormal and normal data. Furthermore, Pocock method tend to have larger GSD design size and smaller expected sample size than OBrien-Fleming method.

4.4 Example of Glaucomatous Deterioration Detection

In this section, we illustrate the GSD in a glaucomatous deterioration detection diagnostic trial. Glaucoma is a progressive optic neuropathy. The related symptoms include loss of retinal ganglion cells, and morphological changes to the optic nerve and retinal nerve fiber

Table 4.22: AUC, power(%) using the Pocock GSD with $\alpha = 0.05$, case I

	Power=80%						Power=90%					
	$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN
						Two-group sequential design (J=2)						
0	79.3	80.4	79.8	80.6	80.8	80.7	89.7	89.2	89.6	89.7	90.8	91.2
0.25	79.5	80.1	79.9	79.8	80.5	80.8	89.8	89.4	88.9	89.3	89.8	90.7
0.5	79.2	79.8	81.7	81.1	80.5	79.2	89.9	90.7	89.6	89.2	90.0	89.2
0.75	79.4	80.0	80.9	79.9	79.7	79.9	90.0	90.5	89.5	89.6	89.3	89.2
						Five-group sequential design (J=5)						
0	79.4	81.0	81.1	80.1	80.9	80.7	89.4	90.4	91.0	90.5	90.9	91.5
0.25	79.4	79.4	79.4	78.9	80.3	80.1	89.9	88.8	90.1	89.3	90.2	88.8
0.5	80.1	79.9	81.3	80.1	81.5	80.6	89.8	90.0	89.6	90.1	90.2	89.7
0.75	80.9	81.1	80.2	80.3	80.5	80.0	89.8	89.9	90.0	90.2	90.5	90.5

Table 4.23: AUC, power(%) using the O'Brien-Fleming GSD with $\alpha = 0.05$, case II

	Power=80%						Power=90%					
	$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN
						Two-group sequential design (J=2)						
0	79.5	80.0	79.8	79.4	80.9	81.1	89.5	89.0	89.3	89.2	90.4	90.9
0.25	79.0	79.7	79.5	80.6	79.4	80.4	89.4	89.5	90.4	89.7	89.9	88.9
0.5	80.2	80.8	80.2	79.1	79.8	80.1	89.3	90.0	90.3	89.8	89.9	90.1
0.75	80.2	79.3	78.8	79.5	79.7	79.5	90.4	89.5	89.9	89.5	90.3	89.7
						Five-group sequential design (J=5)						
0	80.5	79.5	80.4	78.8	81.0	80.2	89.6	89.3	90.2	90.0	90.4	90.9
0.25	79.1	79.6	79.0	78.9	78.7	79.7	90.2	89.4	90.0	89.8	89.8	90.4
0.5	80.9	80.2	79.1	78.7	80.1	79.4	89.7	89.0	88.5	89.9	89.4	90.2
0.75	80.1	80.1	78.6	80.1	79.7	79.2	90.3	89.3	90.2	88.8	89.7	90.3

Table 4.24: AUC, power(%) using the Pocock GSD with $\alpha = 0.05$, case II

	Power=80%						**Power=90%**					
	$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN
							Two-group sequential design (J=2)					
0	79.7	78.7	79.9	79.8	80.7	80.1	88.3	89.8	89.7	89.8	90.0	90.2
0.25	79.2	79.7	79.2	79.5	79.3	79.6	89.9	89.3	89.5	89.4	89.7	90.0
0.5	80.2	80.3	79.3	79.1	80.3	80.4	90.5	89.5	89.8	88.8	89.6	89.4
0.75	79.6	80.8	79.0	80.4	80.4	79.7	89.9	90.3	89.0	89.8	89.3	90.0
							Five-group sequential design (J=5)					
0	78.5	79.8	80.0	80.2	80.5	80.4	90.1	88.6	90.0	90.2	89.3	90.3
0.25	79.7	79.6	80.1	79.4	80.3	79.0	89.8	89.5	89.4	89.7	89.5	89.9
0.5	80.7	78.7	79.0	79.7	79.9	81.4	90.2	89.9	89.9	89.8	90.4	89.2
0.75	80.5	79.3	78.8	79.4	81.3	80.7	89.1	89.0	89.3	89.7	89.5	89.8

Table 4.25: AUC, power(%) using the O'Brien-Fleming GSD with $\alpha = 0.05$, case III

	Power=80%						**Power=90%**					
	$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN
							Two-group sequential design (J=2)					
0	79.0	79.4	79.9	79.1	80.0	79.9	89.0	89.3	89.6	89.4	89.6	88.5
0.25	80.2	81.3	79.5	79.1	79.6	79.3	90.2	90.0	89.5	89.6	89.5	89.2
0.5	80.8	79.8	80.2	79.0	80.5	81.4	90.9	90.1	89.5	89.8	90.5	90.0
0.75	80.5	80.2	79.7	78.6	79.2	79.7	90.6	90.1	89.0	90.0	89.8	89.6
							Five-group sequential design (J=5)					
0	78.7	78.8	79.3	78.1	79.2	80.1	89.2	89.0	89.5	89.4	90.7	90.0
0.25	80.6	80.8	79.9	79.3	79.1	80.2	90.3	90.4	90.1	90.1	89.9	89.6
0.5	80.3	80.5	80.0	80.3	79.7	81.1	89.3	90.2	89.9	89.8	90.2	90.6
0.75	80.3	79.2	78.7	78.9	80.0	79.8	89.8	90.3	89.3	89.2	90.3	89.5

Table 4.26: AUC, power(%) using the Pocock GSD with $\alpha = 0.05$, case III

	Power=80%						**Power=90%**					
	$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN	Normal	LogN
						Two-group sequential design (J=2)						
0	78.8	78.1	79.0	79.0	78.7	79.8	88.0	89.2	89.4	89.0	89.6	89.3
0.25	80.4	79.4	79.9	79.4	79.8	79.1	90.6	90.1	89.7	90.3	90.0	89.7
0.5	80.5	80.7	80.4	80.3	79.9	80.7	90.6	90.6	89.4	90.1	90.2	90.3
0.75	79.8	80.1	79.4	78.8	79.9	79.4	90.2	89.3	89.1	89.9	89.8	89.1
						Five-group sequential design (J=5)						
0	79.1	78.5	79.0	79.9	78.3	79.0	89.7	89.7	89.8	90.0	89.3	89.4
0.25	80.7	81.5	79.6	79.5	80.2	79.9	90.3	90.6	90.4	90.5	90.0	90.4
0.5	80.7	81.4	80.2	80.6	81.0	81.0	90.4	90.3	90.2	90.4	89.2	90.1
0.75	80.0	80.5	79.5	79.4	79.7	79.7	90.0	89.5	89.8	89.7	90.2	90.5

Table 4.27: AUC, expected sample sizes using GSD with $\alpha = 0.05$, case I

	Power=80%						**Power=90%**					
	$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock
						Two-group sequential design (J=2)						
0	123	118	69	65	37	35	157	142	87	79	47	43
0.25	114	108	63	60	35	33	145	132	80	74	44	41
0.5	104	100	60	57	34	32	131	121	76	70	43	39
0.75	95	91	61	58	34	33	121	111	77	71	44	40
						Five-group sequential design (J=5)						
0	111	109	62	60	34	32	137	124	76	68	41	37
0.25	104	101	57	56	32	30	127	116	71	64	38	35
0.5	94	91	54	52	31	29	117	107	67	61	37	33
0.75	87	83	54	53	31	30	107	95	67	61	39	35

Table 4.28: AUC, expected sample sizes using GSD with $\alpha = 0.05$, case II

	Power=80%						**Power=90%**					
	$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock
						Two-group sequential design (J=2)						
0	273	256	150	141	81	76	346	317	190	173	102	92
0.25	255	243	140	133	75	71	322	293	175	161	95	87
0.5	232	221	132	125	73	69	294	267	167	154	91	83
0.75	209	196	132	124	74	70	263	241	164	153	93	87
						Five-group sequential design (J=5)						
0	247	242	135	134	72	70	304	278	167	153	89	82
0.25	231	225	127	124	68	66	284	258	156	143	83	77
0.5	210	205	120	117	65	64	260	235	148	133	80	72
0.75	189	182	119	116	67	65	233	211	145	132	82	74

Table 4.29: AUC, expected sample sizes using GSD with $\alpha = 0.05$, case III

	Power=80%						**Power=90%**					
	$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock
						Two-group sequential design (J=2)						
0	1075	1023	585	560	310	293	1356	1242	738	678	391	358
0.25	1022	962	556	528	292	276	1289	1182	698	637	368	336
0.5	925	874	520	489	279	263	1166	1072	661	598	352	324
0.75	806	763	499	476	282	266	1013	932	630	578	354	325
						Five-group sequential design (J=5)						
0	975	954	531	529	282	278	1200	1102	651	600	343	313
0.25	929	908	504	495	266	257	1136	1043	616	559	325	298
0.5	838	825	471	457	253	245	1027	941	578	530	310	287
0.75	725	712	454	440	254	248	894	830	558	512	313	283

Table 4.30: AUC, GSD design sample sizes (maximum) with $\alpha = 0.05$, case I

	Power=80%						**Power=90%**					
	$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock
						Two-group sequential design (J=2)						
0	138	152	77	85	42	46	185	202	103	112	56	61
0.25	128	141	71	79	39	43	171	187	95	104	52	57
0.5	118	130	68	75	38	42	157	172	91	99	51	55
0.75	108	119	69	76	39	43	145	158	92	100	53	57
						Five-group sequential design (J=5)						
0	141	168	78	94	43	51	188	221	105	123	57	67
0.25	131	156	73	87	40	47	174	205	97	114	53	62
0.5	120	144	69	83	39	46	160	189	92	109	51	60
0.75	111	132	70	84	40	48	148	173	93	110	54	63

Table 4.31: AUC, GSD design sample sizes (maximum) with $\alpha = 0.05$, case II

	Power=80%						**Power=90%**					
	$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock
						Two-group sequential design (J=2)						
0	305	336	168	185	90	99	408	446	225	245	120	131
0.25	285	314	157	172	84	92	381	417	209	228	112	122
0.5	261	287	148	163	81	89	348	381	198	216	108	118
0.75	235	258	147	162	83	92	313	342	196	214	111	122
						Five-group sequential design (J=5)						
0	311	372	171	205	91	109	416	489	229	269	122	143
0.25	291	348	160	191	86	102	389	457	213	251	114	134
0.5	266	318	151	181	83	99	355	418	202	237	110	129
0.75	239	286	150	179	85	102	319	376	200	235	113	133

Table 4.32: AUC, GSD design sample sizes (maximum) with $\alpha = 0.05$, case III

	Power=80%						**Power=90%**					
	$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$		$\boldsymbol{\lambda} = 0$		$\lambda = 0.5$		$\lambda = 0.75$	
$\boldsymbol{\rho}$	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock	OF	Pocock
	Two-group sequential design (J=2)											
0	1192	1313	654	720	345	380	1595	1742	875	956	461	504
0.25	1147	1263	621	684	326	359	1534	1676	830	907	436	477
0.5	1040	1145	582	640	313	344	1391	1519	778	849	418	457
0.75	900	990	558	614	314	346	1203	1314	746	815	420	459
	Five-group sequential design (J=5)											
0	1216	1454	667	798	352	420	1625	1911	891	1049	470	553
0.25	1170	1399	633	757	333	398	1563	1839	846	995	445	523
0.5	1061	1268	593	709	319	381	1417	1667	792	932	426	501
0.75	917	1097	569	680	320	383	1226	1442	760	894	428	503

Table 4.33: AUC, fixed sample design sample sizes with $\alpha = 0.05$, case I

	Power=80%			**Power=90%**		
$\boldsymbol{\rho}$	$\boldsymbol{\lambda} = 0$	$\lambda = 0.5$	$\lambda = 0.75$	$\boldsymbol{\lambda} = 0$	$\lambda = 0.5$	$\lambda = 0.75$
0	137	76	41	183	102	55
0.25	127	71	39	170	95	52
0.5	117	67	38	156	90	50
0.75	108	68	39	144	91	52

Table 4.34: AUC, fixed sample design sample sizes with $\alpha = 0.05$, case II

	Power=80%			**Power=90%**		
$\boldsymbol{\rho}$	$\boldsymbol{\lambda} = 0$	$\lambda = 0.5$	$\lambda = 0.75$	$\boldsymbol{\lambda} = 0$	$\lambda = 0.5$	$\lambda = 0.75$
0	303	167	89	405	223	119
0.25	283	155	83	379	208	111
0.5	259	147	80	346	197	107
0.75	233	146	83	311	195	111

Table 4.35: AUC, fixed sample design sample sizes with $\alpha = 0.05$, case III

	Power=80%			Power=90%		
$\boldsymbol{\rho}$	$\boldsymbol{\lambda} = 0$	$\lambda = 0.5$	$\lambda = 0.75$	$\boldsymbol{\lambda} = 0$	$\lambda = 0.5$	$\lambda = 0.75$
0	1183	649	342	1583	869	458
0.25	1138	616	324	1523	824	433
0.5	1032	577	310	1381	772	415
0.75	892	553	312	1194	741	417

layer. The global prevalence of glaucoma for population aged 40-80 years is 3.54%. The number of people with glaucoma worldwide was estimated to be 64.3 million in 2013, which will increase to 76.0 million in 2020 and 111.8 million in 2040 (Tham et al. 2014). If glaucoma is not diagnosed and treated, damage can progress and cause a loss of peripheral vision and may eventually lead to complete sight loss. In fact, glaucoma is one of the leading causes of global preventable blindness. Glaucoma does not cause symptoms in early stages, which makes it hard to be diagnosed, but an eye exam might detect the signs of glaucoma. The visual field deterioration due to glaucoma can be tested using imaging techniques. But it is challenging to accurately identify the progressive eyes in glaucoma patients since the structure of the image data is complex and it is very difficult to detect the changes. Li and Zhou (2008) studied the accuracy of probability scores generated from two Bayesian hierarchical models on classifying the stable and progressive eyes. The study includes 171 patients and visual field tests were given to these patients over 8 years of follow-up study. Some patients were measured on both eyes and others were measured only on one eye. Because some data are from both eyes of the same patient, test scores from the hierarchical models calculated from both eyes of the same patients are cluster-correlated. We applied the previously mentioned method to the dataset to generate empirical ROC curves for both models. Our study covers the GSD methods for the trial both on a point estimate of the ROC curve and on the AUC differences. The empirical ROC curves of the two biomarkers are shown in Figure 4.6.

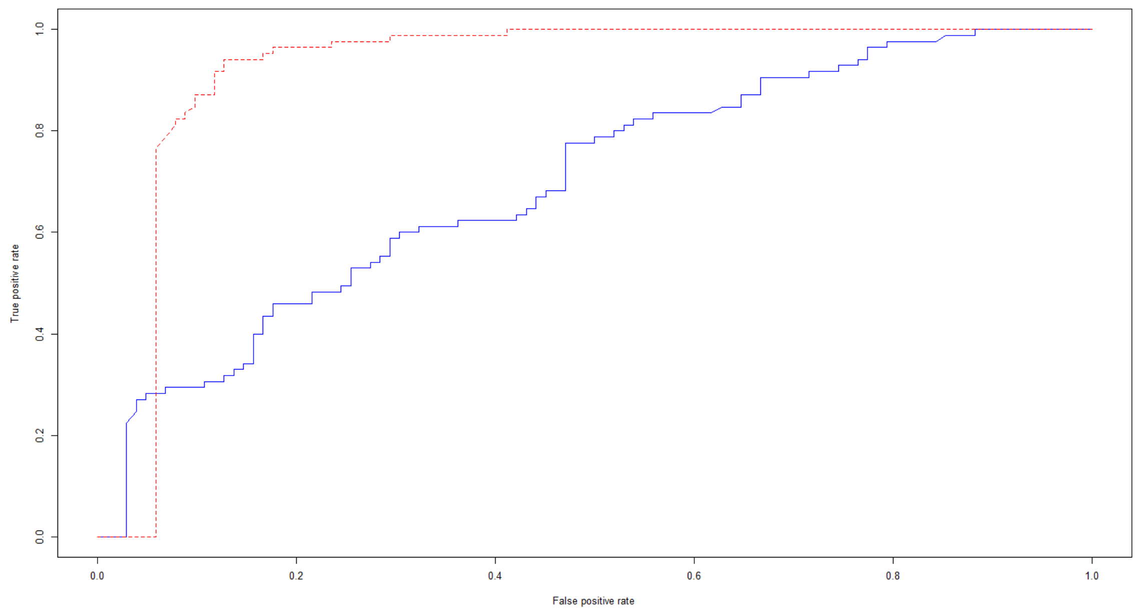

Figure 4.6: Empirical ROC curves of two models for glaucoma deterioration detection

Consider testing the null hypothesis of $\Delta(t) = 0$ for t={0.2,0.4,0.5,0.6,0.8}. The Glaucoma example is a possible case under the alternative hypothesis condition, with $\Delta(t) = \{0.506, 0.365, 0.212, 0.165, 0.024\}$ for t={0.2,0.4,0.5,0.6,0.8} respectively. With empirical ROC estimates and Bootstrap method for the variance estimation, in Table 4.36 we show the interim looks of one run with statistics and corresponding boundaries displayed at the bottom for O'Brien-Fleming GSD method with $J = 5$. The sequential empirical ROCs at the interim analysis point and the final step are calculated and displayed in Figure 4.8, where the last graph is identical to Figure 4.6. Similarly, the sequential empirical ROCs for $J = 2$ are shown in Figure 4.7.

Table 4.36: Interim test statistics of the glaucomatous deterioration detection example

	Interim Z-Statistic				
FPR	1	2	3	4	5
0.2	1.613	2.324	5.411		
0.4	0.000	2.280	3.756		
0.5	0.000	1.345	2.105	2.673	
0.6	0.000	1.707	2.543	3.262	
0.8	0.000	0.000	0.000	0.461	0.741
Boundaries	±4.56	±3.23	±2.63	±2.28	±2.04

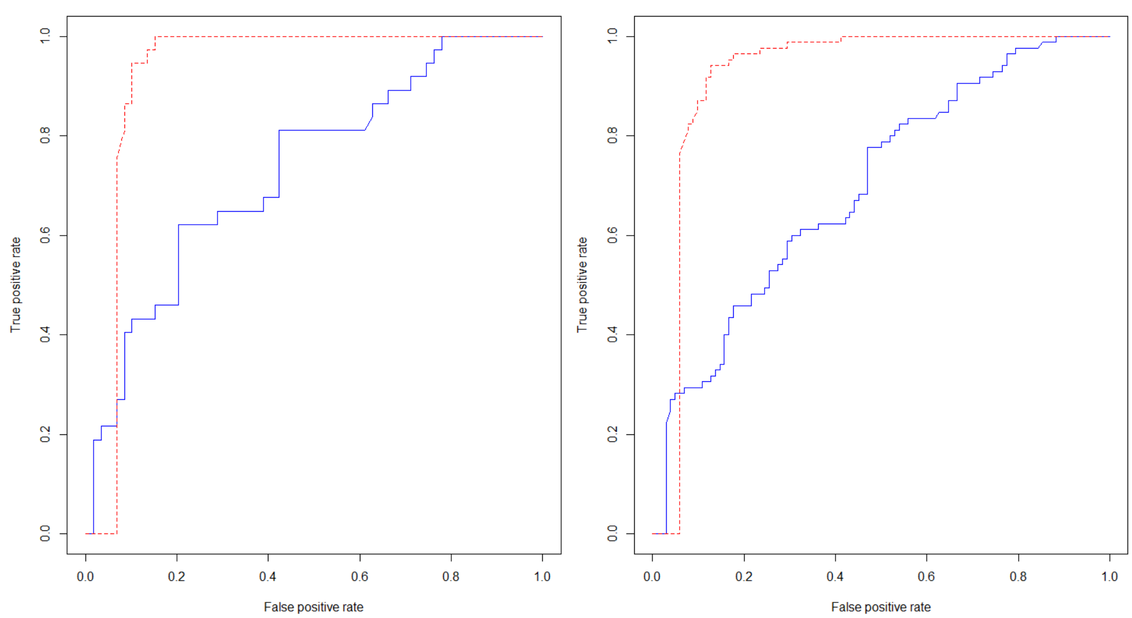

Figure 4.7: Sequential empirical ROC curves at interim analyses for glaucoma deterioration detection (J=2)

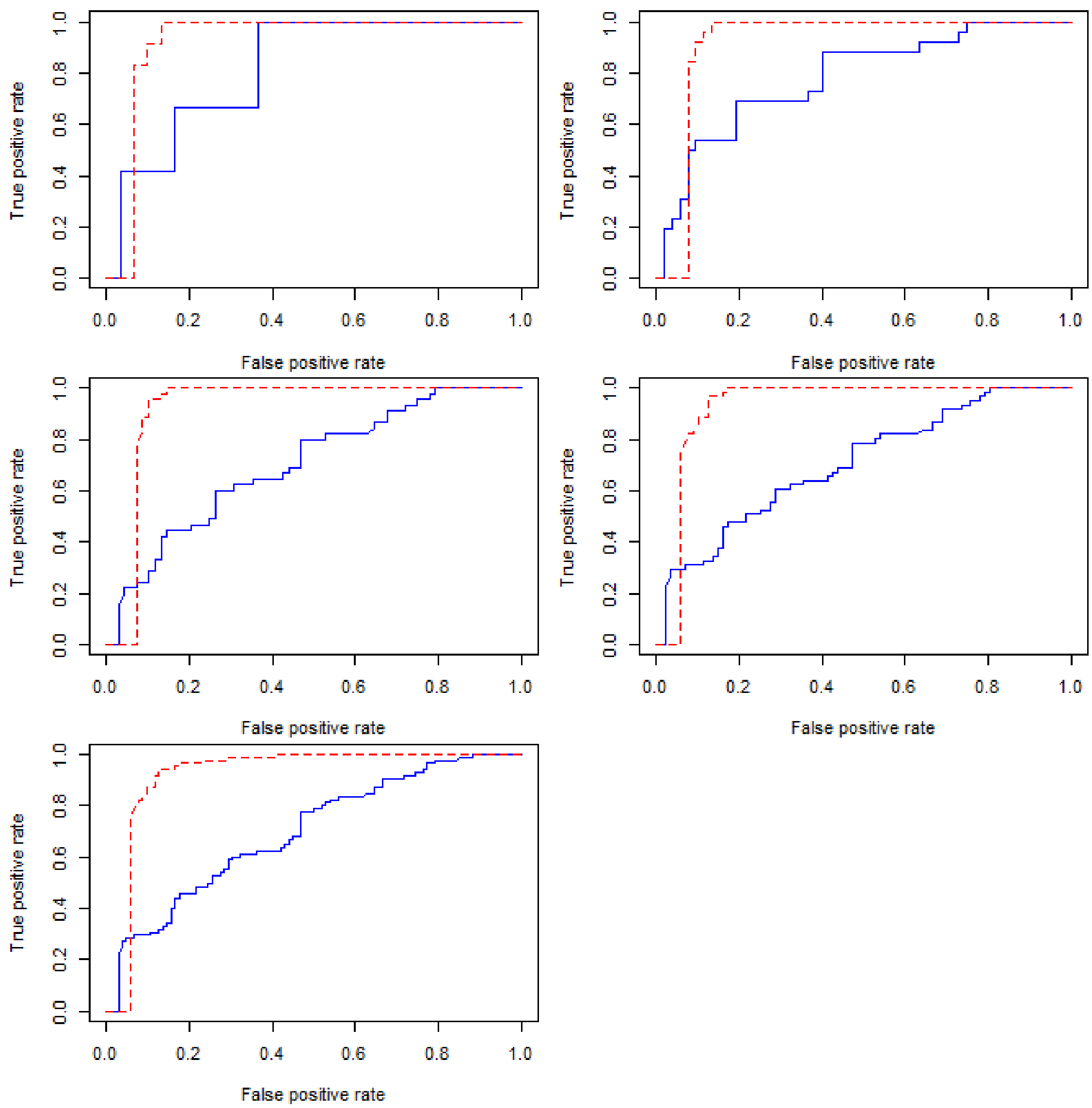

Figure 4.8: Sequential empirical ROC curves at interim analyses for glaucoma deterioration detection (J=5)

Suppose FPR =0.2 and the number of looks for O'Brien-Fleming GSD is 5. At the first endpoint, with 34 subjects' test results become available, the Z-statistic is 1.613, which lies within the rejection boundaries of the hypothesis testing. Thus we fail to reject the null hypothesis, and continue to recruit 34 additional subjects. The difference between the ROC curves at FPR=0.2 and its variance can be estimated using the accrued subjects' data up to this point which is 64 in total. The statistic is calculated to be 2.324 which again is smaller than the boundary 3.23. Again, we fail to reject the null hypothesis, and continue to recruit another 34 subjects. At the third interim analysis with overall 102 subjects' data, we calculate the Z-statistic to be 5.411, which is greater than the boundary 2.63. Therefor, we reject the null hypothesis of $\Delta(0.2) = 0$ at this step, and conclude that the two biomarkers are significantly different in their accuracy at the false positive rate of 0.2.

For testing of AUCs' difference, the AUCs are estimated to be 0.70 and 0.95 for model 1 and model 2 respectively, where AUCs are estimated using Wilcoxon-Mann-Whitney statistics (DeLong et al. 1988). We applied the O'Brien-Fleming GSD with J=2 or 5 to the trial respectively. We found that we can reject the null hypothesis of equal AUCs of the two ROCs at the first (J=2) and the third interim analysis (J=5), which used 85 and 102 subjects respectively. Using the same procedure but with Pocock GSD, we can reject the null hypothesis of equal AUCs at the first (J=2) and the second interim analysis (J=5), which will use 85 and 68 subjects respectively.

4.5 Discussion

The empirical distribution function defined in this chapter puts equal weight on each observation. In future, we can use the following definition which puts equal weight on each subject instead of observations.

$$\hat{F}_{[nt]}(x) = \frac{1}{[nt]} \sum_{i=1}^{[nt]} \frac{1}{m_i} \sum_{j=1}^{m_i} I(X_{ij} \leq x),$$

where t is the percentage of subjects accrued so far at this analysis point. $\hat{F}_{[nt]}(x)$ can be simplified as $\hat{F}_t(x)$.

The sequential empirical process is then defined as

$$n^{-1/2}[nt](\hat{F}_{[nt]}(x) - F(x)).$$

With the new definition, it is also of interest to derive the asymptotic properties of the clustered ROC curves and apply it in group sequential ROC comparison study. We can also remove the requirement that the average of m_i converges to a constant.

Furthermore, the group sequential method we propose can also be extended to comparing multiple clustered-correlated ROC curves.

Chapter 5: Discussion

5.1 Summary

This dissertation covers three issues in the group sequential diagnostic biomarkers' comparison studies. We consider group sequential designs that allow early termination for significant difference. Chapter 2 derived asymptotic theory for correlated ROC curves, which is necessary to apply existing standard group sequential methodology to comparing correlated ROCs. Chapter 3 extended this to the field of correlated PPV and NPV curves, both indexed by the FPR and by the percentil value. Chapter 4 developed the asymptotic theory for clustered ROC curves.

In Chapter 2 we first investigated the asymptotic properties of the sequential empirical difference of two correlated ROC curves. We first extended the work of Koopmeiners and Feng (2011) by showing that the sequential empirical difference of two correlated ROC curves converges to a Gaussian process and show that the sequential empirical estimate of $\Delta(t)$, a point on the ROCs' difference curve, has an independent increments covariance structure.

We then can conduct group sequential comparison studies on two correlated diagnostic biomarkers. The proof of the independent increments allows us to apply existing standard group sequential methodology to correlated diagnostic biomarker comparison studies. We showed the weak convergence of the sequential empirical difference of ROC curves to a Gaussian process, and based on this we derived asymptotic theory. Through integration, this would also allow us to derive asymptotic theory for the sequential empirical summary measure difference of correlated ROC curves. In the thesis we only present results for a point difference on the ROC curves, however it is straight forward to derive distribution theory for other summary measures used to evaluate the performance difference between diagnostic biomarkers. These results provide great flexibility for designing group sequential diagnostic biomarkers comparison studies.

The covariance structure were verified by a simulation study. We also conducted group sequential simulation studies on Type I error rates compared to the nominal value. Another group sequential simulation studies show that actual sample size could be substantially decreased due to early study termination while still maintain the power requirement and α level. We also presented an example on a lung cancer trial with CT and PET comparison study.

In Chapter 3, we studied the sequential difference of empirical correlated PPV and NPV curves, either indexed by FPR or indexed by percentile value. We showed that the sequential empirical difference of correlated PPV and NPV curves converge to a Gaussian process with independent covariance structure.

In Chapter 4, we derived the distribution theory for the sequential empirical ROC difference of two diagnostic biomarkers in a clustered data setting. We further studied the group sequential design for ROCs comparison in this clustered data setting based on the asymptotic properties. We also conducted simulation studies to verify the covariance structure, and group sequential simulation studies on Type I error rates and expected sample sizes.

5.2 Future Work

This dissertation studies the group sequential comparison methods for two correlated or clustered diagnostic biomarkers. Based on the distribution theory we derived for the sequential empirical differences, we can apply the existing group sequential methodology to the comparison study.

In correlated ROCs comparison study, we can use either the variance formula derived or Bootstrap method to estimate the empirical difference's variance, while Bootstrap method is much more computationally intensive. However, using the empirical cumulative distribution functions and Kernel density estimation to estimate the variance has some limitation due to the difficulty in Kernel density estimation. It is desirable if we can develop a non-parametric estimation method for variance without involving density estimation. Currently, we mainly deal with two correlated ROC curves with variance covariance formula developed. We can also apply similar approach to compare multiple ROC curves.

For clustered ROCs comparison study, currently the empirical distribution function defined in the thesis have equal weights on each observation. Another approach would be to put equal weights on each subject instead of observation. The group sequential method we propose can also be extended to PPV and NPV comparison as well as comparing multiple cluster-correlated ROC curves.

Appendix A: R Packages Used

1. **library(gsDesign):** gsDesign is a package that derives group sequential designs and describes their properties. The library is used to calculate the boundaries at interim analysis points for a group sequential design.

2. **library(MASS):** We use the functions provided by the library to generate multivariate normal and lognormal random variables. We used the function mvrnorm() to generate multivariate normal random variable for simulation studies.

3. **library(mvtnorm):** To calculate the theoretical values, we used function pmvnorm() which computes the distribution function of the multivariate normal distribution for arbitrary limits and correlation matrices in Chapter 2.

4. **library(ROCR):** We use the plot() function provided by the package for all ROC curve graphs plotting.

Simulation and example programs are written in R, with some core functions implemented with C.

Bibliography

Armitage, P., C. McPherson, and B. Rowe (1969). Repeated significance tests on accumulating data. *Journal of the Royal Statistical Society. Series A (General)*, 235–244.

Csörgő, M. and B. Szyszkowicz (1998). 21 sequential quantile and bahadur-kiefer processes. *Handbook of Statistics 16*, 631–688.

DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics 44*, 837–845.

Dodd, L. E. and M. S. Pepe (2003). Partial auc estimation and regression. *Biometrics 59*(3), 614–623.

Dvoretzky, A., J. Kiefer, and J. Wolfowitz (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, 642–669.

Emir, B., S. Wieand, S.-H. Jung, and Z. Ying (2000). Comparison of diagnostic markers with repeated measurements: a non-parametric roc curve approach. *Statistics in Medicine 19*(4), 511–523.

Fleming, T. R., D. P. Harrington, and P. C. O'Brien (1984). Designs for group sequential tests. *Controlled Clinical Trials 5*(4), 348–361.

Hsieh, F., B. W. Turnbull, et al. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *The Annals of Statistics 24*(1), 25–40.

Huang, Y., M. Sullivan Pepe, and Z. Feng (2007). Evaluating the predictiveness of a continuous marker. *Biometrics 63*(4), 1181–1188.

Jennison, C. and B. W. Turnbull (2000). *Group Sequential Methods with Applications to Clinical Trials*. New York: Chapman and Hall.

Karr, A. F. (1993). *Probability*. New York: Springer.

Kim, K. and D. L. Demets (1992). Sample size determination for group sequential clinical trials with immediate response. *Statistics in Medicine 11*(10), 1391–1399.

Komlós, J., P. Major, and G. Tusnády (1975). An approximation of partial sums of independent rv'-s, and the sample df. i. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 32*(1-2), 111–131.

Koopmeiners, J. S. and Z. Feng (2011). Asymptotic properties of the sequential empirical roc, ppv and npv curves under case-control sampling. *The Annals of Statistics 39*(6), 3234–3261.

Koopmeiners, J. S., Z. Feng, and M. S. Pepe (2012). Conditional estimation after a two-stage diagnostic biomarker study that allows early termination for futility. *Statistics in Medicine 31*(5), 420–435.

Koopmeiners, J. S. and R. I. Vogel (2013). Early termination of a two-stage study to develop and validate a panel of biomarkers. *Statistics in Medicine 32*(6), 1027–1037.

Lardinois, D., W. Weder, T. F. Hany, E. M. Kamel, S. Korom, B. Seifert, G. K. von Schulthess, and H. C. Steinert (2003). Staging of non–small-cell lung cancer with integrated positron-emission tomography and computed tomography. *New England Journal of Medicine 348*(25), 2500–2507.

Li, G. and K. Zhou (2008). A unified approach to nonparametric comparison of receiver operating characteristic curves for longitudinal and clustered data. *Journal of the American Statistical Association 103*(482), 705–713.

Liu, A., E. F. Schisterman, and C. Wu (2005). Nonparametric estimation and hypothesis testing on the partial area under receiver operating characteristic curves. *Communications in Statistics - Theory and Methods 34*(9-10), 2077–2088.

Liu, A., C. Wu, and E. F. Schisterman (2008). Nonparametric sequential evaluation of diagnostic biomarkers. *Statistics in Medicine 27*(10), 1667–1678.

Massart, P. (1990). The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *The Annals of Probability*, 1269–1283.

Mazumdar, M. (2004). Group sequential design for comparative diagnostic accuracy studies: Implications and guidelines for practitioners. *Medical Decision Making 24*(5), 525–533.

Mazumdar, M. and A. Liu (2003). Group sequential design for diagnostic accuracy studies. *Statistics in Medicine 22*, 727–739.

McNeil, B. J. and S. J. Adelstein (1976). Determining the value of diagnostic and screening tests. *Journal of Nuclear Medicine 17*(6), 439–448.

Moskowitz, C. S. and M. S. Pepe (2004). Quantifying and comparing the predictive accuracy of continuous prognostic factors for binary outcomes. *Biostatistics 5*(1), 113–127.

O'Brien, P. C. and T. R. Fleming (1979). A multiple testing procedure for clinical trials. *Biometrics*, 549–556.

Obuchowski, N. A. (1997). Nonparametric analysis of clustered roc curve data. *Biometrics 53*(2), 567–578.

Obuchowski, N. A. (2005). Roc analysis. *American Journal of Roentgenology 184*(2), 364–372.

Pepe, M., G. Longton, and H. Janes (2009). Estimation and comparison of receiver operating characteristic curves. *The Stata Journal 9*(1), 1.

Pepe, M. S. (2000). Receiver operating characteristic methodology. *Journal of the American Statistical Association 95*(449), 308–311.

Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York: Oxford.

Pepe, M. S., Z. Feng, G. Longton, and J. Koopmeiners (2009). Conditional estimation of sensitivity and specificity from a phase 2 biomarker study allowing early termination for futility. *Statistics in Medicine 28*(5), 762–779.

Pepe, M. S., H. Janes, G. Longton, W. Leisenring, and P. Newcomb (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology 159*(9), 882–890.

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika 64*(2), 191–199.

Rao, J. and A. Scott (1992). A simple method for the analysis of clustered binary data. *Biometrics*, 577–585.

Silvestri, G. A., L. T. Tanoue, M. L. Margolis, J. Barker, and F. Detterbeck (2003). The noninvasive staging of non-small cell lung cancerthe guidelines. *CHEST Journal 123*(1_suppl), 147S–156S.

Sox, H. C., S. Stern, D. Owens, and H. L. Abrams (1989). *Assessment of diagnostic technology in health care: rationale, methods, problems, and directions.* Washington: National Academies Press.

Tang, L., S. S. Emerson, and X.-H. Zhou (2008). Nonparametric and semiparametric group sequential methods for comparing accuracy of diagnostic tests. *Biometrics 64*(4), 1137–1145.

Tang, L. and A. Liu (2010). Sample size recalculation in sequential diagnostic trials. *Biostatistics 11*, 151–163.

Tham, Y.-C., X. Li, T. Y. Wong, H. A. Quigley, T. Aung, and C.-Y. Cheng (2014). Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology 121*(11), 2081–2090.

van der Vaart, A. W. and J. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* New York: Springer.

Wang, S. K. and A. A. Tsiatis (1987). Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics*, 193–199.

Whitehead, J. (1999). A unified theory for sequential clinical trials. *Statistics in Medicine 18*(17-18), 2271–2286.

Wieand, S., M. H. Gail, B. R. James, and K. L. James (1989). A family of non-parametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika 76*, 585–592.

Ye, X. and L. L. Tang (2015). Group sequential methods for comparing correlated receiver operating characteristic curves. *Applied Statistics in Biomedicine and Clinical Trials Design, Springer*, 89–108.

Zheng, Y., T. Cai, M. S. Pepe, and W. C. Levy (2008). Time-dependent predictive values of prognostic biomarkers with failure time outcome. *Journal of the American Statistical Association 103*(481), 362–368.

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

Zhou, X. H., S. M. Li, and C. A. Gatsonis (2008). Wilcoxon-based group sequential designs for comparison of areas under two correlated roc curves. *Statistics in Medicine 27*, 213 – 223.

Zhou, X.-H., N. A. Obuchowski, and D. K. McClish (2011). *Statistical Methods in Diagnostic Medicine*, Volume 712. New York: Wiley.

Zweig, M. H. and G. Campbell (1993). Receiver-operating characteristic (roc) plot: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry 39*, 561–577.

Curriculum Vitae

Xuan received his bachelor's degree from the University of Science and Technology of China, and master of science degree from the University of Illinois at Chicago. He has work experience in the fields of software development and statistics.