2013

The Estimation and Evaluation of Optimal Thresholds for Two Sequential Testing Strategies

Amber R. Wilk
Virginia Commonwealth University

www.manaraa.com

THE ESTIMATION AND EVALUATION OF OPTIMAL THRESHOLDS FOR TWO

SEQUENTIAL TESTING STRATEGIES

A Dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy at Virginia Commonwealth University.

by

AMBER R. WILK

B.S. Mathematics and Statistics Composite Secondary Education,

Utah State University, 2007

Advisor: DONNA K MCCLISH, PH.D.

PROFESSOR, DEPARTMENT OF BIOSTATISTICS

Virginia Commonwealth University

Richmond, Virginia

July 2013

Acknowledgements

This dissertation would not have been possible without the help of so many people in so many ways. First and foremost, my advisor Dr. Donna McClish was instrumental in the development, progression, and completion of this research. Without your unwavering support, kind words, guidance and persistent help, willingness to let me complain, encouragement to keep going, and meeting at a last minutes' notice to talk both professionally and personally, this dissertation could not have been completed. Your passion for your work inspires me. Thank you to the rest of my dissertation committee: Dr. Christine Schubert-Kabban, Dr. Leroy Thacker, Dr. Jessica Ketchum, and Dr. Edmund (Trey) Wickham, whose ideas and comments are all greatly appreciated. Particular thanks goes to Dr. Jessica Ketchum. I have had your constant endorsement and encouragement throughout the past five years. Thank you for the constructive feedback, the opportunity to learn and develop as a researcher, and motivation to keep going when I wanted to quit. Thank you for being a mentor and my friend.

I am thankful to all of my supervisors from various positions I have held as a research assistant over the course of my graduate career. Thank you to Joanne Einsmann and John Matro for giving me time away from work when I needed to study, and for always supporting me during the two years I was part of your group and beyond. Thank you to the team I worked with at Merck & Co., Inc., for the unique opportunity to gain experience working in a pharmaceutical environment while realizing I was still a student.

ii

Finally, thank you to my current supervisor, Dr. Cathy Bradley, for never failing to ask how my work was progressing.

There were key people in the department who helped me get through school. I owe a very important debt to Yvonne Hargrove, Gayle Spivey, Russ Boyle, Dr. Roy Sabo, and Dr. Chris Gennings. Additionally, I want to thank my cohort of peers for the system of support we created: Adam Sima, Jiayi Hou, and especially Caroline Carrico, who shared an office / cubicle with me the entire five years and was always there to laugh, complain, gossip, and talk with. Thank you all for being excited with me and for me as we all graduate and become Biostatisticians.

I am grateful to my friends and family, near and far, for the constant inspiritment, comfort, and encouragement. I will never be able to surely express how appreciative I am. A very special thank you goes to two of my best friends, Kimberly Parker and Claire Russell. We have shared many laughs and tears, dissertation work parties and breaks from studies, vacations and road trips, and so much more. We started graduate school together, and now we finish it together. As we go down our own separate paths, I know that I have found two lifelong friends, and I could not be more grateful and thankful to the both of you. Erich Fritz, you have been my friend through the good times and the not so good times, and I could not have finished this dissertation without your invariable encouragement. Thank you for helping me realize and reach my potential.

I owe my deepest gratitude to my parents, Annalisa and Ed, for molding me into the person I am today and motivating me with their own examples of hard work. Thank you for your indefinite unconditional love, support, and confidence, and for believing in

me even when I didn't believe in myself. From a young age you taught me to always strive to be my best no matter what. Thank you for cheering me on and reminding me when I needed to take a break. Thank you for your endless words of encouragement through phone calls, text messages, emails, and visits. Thank you for always being there; I definitely could not have done this without you, and I can't wait to share the next chapter of my life.

Table of Contents

List of Tables

List of Figures

Abbreviations

AUC	Area Under the Curve
BN	Believe the Negative
BP	Believe the Positive
CDF	Cumulative Distribution Function
CR	Confidence Region
FPR	False Positive Rate
GYI	Generalized Youden Index
NR	Newton-Raphson
OLS	Ordinary Least Squares
OOP	Optimal Operating Point
RMSE	Root Mean Square Error
ROC	Receiver Operator Characteristic
SCI	Simultaneous Confidence Interval
Se	Sensitivity
Sp	Specificity
TPR	True Positive Rate
TSS	Total Sample Size
YI	Youden Index

Abstract

THE ESTIMATION AND EVALUATION OF OPTIMAL THRESHOLDS FOR TWO

SEQUENTIAL TESTING STRATEGIES

By Amber R. Wilk, Ph.D.

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor

of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2013

Advisor:

Donna K. McClish

Professor, Department of Biostatistics

Many continuous medical tests often rely on a threshold for diagnosis. There are two

sequential testing strategies of interest: Believe the Positive (BP) and Believe the Negative

(BN). BP classifies a patient positive if either the first test is greater than a threshold θ_1 or

negative on the first test and greater than θ_2 on the second test. BN classifies a patient

positive if the first test is greater than a threshold θ_3 and greater than θ_4 on the second test.

Threshold pairs $\boldsymbol{\theta} = (\theta_1, \theta_2)$ or (θ_3, θ_4), depending on strategy, are defined as optimal if

they maximized GYI = Se + r(Sp − 1). Of interest is to determine if these optimal

threshold, or optimal operating point (OOP), estimates are "good" when calculated from a

sample. The methods proposed in this dissertation derive formulae to estimate $\hat{\boldsymbol{\theta}}$ assuming

tests follow a binormal distribution, using the Newton-Raphson algorithm with ridging. A

simulation study is performed assessing bias, root mean square error, percentage of over

estimation of Se/Sp, and coverage of simultaneous confidence intervals and confidence

regions for sets of population parameters and sample sizes. Additionally, OOPs are

compared to the traditional empirical approach estimates. Bootstrapping is used to estimate

the variance of each optimal threshold pair estimate. The study shows that parameters such

as the area under the curve, ratio of standard deviations of disease classification groups

within tests, correlation between tests within a disease classification, total sample size, and

allocation of sample size to each disease classification group were all influential on OOP

estimation. Additionally, the study shows that this method is an improvement over the

empirical estimate. Equations for researchers to use in estimating total sample size and SCI

width are also developed. Although the models did not produce high coefficients of

determination, they are a good starting point for researchers when designing a study. A

pancreatic cancer dataset is used to illustrate the OOP estimation methodology for

sequential tests.

1 Introduction

1.1. Introduction

In a clinical setting, determining whether or not a patient has a disease of interest is a common task among physicians; however, the method to determine what test to give and how disease status is determined varies. It is becoming increasingly common that rather than being given a single test, patients may be given a series of tests before diagnosis. Performing multiple tests on patients can be costly, both monetarily and physically. Instead of subjecting all patients to multiple tests, strategies have been derived using logic rules to combine tests so that only a subset of patients receive multiple tests.

Combining diagnostic tests may improve the overall accuracy of diagnosis as opposed to relying on a single test (Shen, 2008). One possible testing strategy that may improve accuracy and decrease cost includes using a combination of diagnostic tests. Give the patient a test (consider this to be test 1). Depending on the result of test 1, a second test may or may not be given. For example, consider body mass index (BMI), which is a number calculated from a person's weight and height as $\frac{weight(lbs)}{height(ins)^2} \times 703$; this is an easy measurement to calculate. An elevated BMI is considered to be a risk factor for diabetes. However, a high BMI alone may not be conclusive evidence to

1

diagnose a patient with diabetes. Another test, for example a blood glucose measurement, may be given before diagnosis is confirmed, at least for some people. Note that, in general, a less expensive or burdensome test is usually used as the first test.

1.2. Methods of Combining Tests

There are several methods in the literature describing ways to combine diagnostic tests; such methods generally fall under two categories: non-sequential and sequential. Non-sequential testing consists of deciding a priori what diagnostic tests will be given, and administering all of them to the patient regardless of test outcomes. Once all the tests have been performed, a diagnosis can be made. One strategy of combining information from the tests is to create a linear combination of the two test results (Shen, 2008; Pepe & Thompson, 2000; Su & Liu 1993; Thompson, 2003; Etzioni, 2008), with the goal of choosing the combination that maximizes accuracy of diagnosis (Pepe & Thompson, 2000). There are two strategies that are often used in non-sequential testing situations: "believe the positive" (BP) and "believe the negative" (BN) (Seto, 2005; Polister, 1980).

Sequential testing, on the other hand, consists of selecting a sequence of tests but only administering the subsequent test(s) dependent on the results of the previous test(s). The focus of this research is on sequential testing strategies. BP and BN approaches can also be applied to sequential testing strategies (Thompson, 2003; Oxley *et al.*, 2011). The logic rules behind these two strategies will be detailed in the next chapter. For both BP and BN, consider two tests; all patients receive the first test. Each of these strategies will either: a) diagnose the patient as positive for disease, b) diagnose the patient as disease-

free, or c) send the patient on to receive the second test; the choice of which decision is selected is dependent on a strategy-specific threshold.

Oftentimes the test results are measured on a continuous scale so that the results need to be dichotomized in order to be interpreted as indicative of non-disease or disease (Halpern *et al.*, 1996). This allows physicians to confirm the participant as healthy, investigate further with an additional test, or start a specific treatment, i.e., diagnosis the patient with disease (Jund *et al.*, 2005). We will assume that higher values of a test suggest disease is present, and vice versa (Leeflang *et al.*, 2008), although tests exist in which low values indicate disease (such as using hemoglobin to screen for anemia). Dichotomizing the continuous outcome is done by means of comparison to a threshold (Skaltsa, Jover, & Carrasco, 2010); methods of analyzing these data as well as determining this threshold have been recently studied throughout the biomedical literature.

1.3. ROC Curves and Related Measurements

Sensitivity (Se) is a measure of the true positive rate (TPR; proportion of correct classification of unhealthy individuals) and specificity (Sp) is a measure of the true negative rate (proportion of correct classification of healthy individuals). Oftentimes, we are more interested in $1 - Sp$, the probability of incorrectly classifying a healthy individual, or the false positive rate (FPR) (Perkins & Schisterman, 2005; Schisterman & Perkins, 2007), rather than Sp itself, as this represents false positive classification.

The ROC curve plots the Se (vertical axis) versus 1-Sp (horizontal axis) across all possible choices of thresholds (Metz, 1978; Faraggi, 2003; Fluss, Faraggi, & Reiser, 2005; Shen, 2008; Skaltsa, Jover, & Carrasco, 2010). ROC curve analysis, although initially developed for electronic signal detection and mainly used in psychophysical-type studies (Hanley, 1989; Schisterman & Perkins, 2007; Pepe, 2012; Green & Swets, 1966), has proven very useful in the biomedical field (fully developed in this area by Swets and Pickett in 1982 (Hanley, 1989)), as it graphically displays a diagnostic test's ability to correctly identify true positive and true negative patients over varied choices of thresholds (Perkins & Schisterman, 2005; Leeflang *et al.*, 2008; Shen, 2008).

The chance line is the line connecting the points (0, 0) and (1, 1), and is used to determine whether or not the diagnostic test(s) is better (good) or worse (bad) than merely selecting a diagnosis by chance. Perfect classification occurs at the (0,1), or where both Se and Sp are equal to 1, indicating that all patients with the disease were correctly classified. Thus it follows that ROC curves falling closer to the (0,1) point are associated with tests that are more accurate in diagnosis. ROC curves can be drawn empirically or smoothed (Hanley, 1989), and are used to select a good, or optimal, threshold; that is, a threshold that detects true positives and true negatives most accurately (Halpern *et al.*, 1996).

1.3.1 Empirical ROC Curves

Empirical ROC curves make no distributional assumption on the data. Because of this, the ROC curve appears as a step curve, where the size of the steps is dependent on the

sample size. This may not show the true relationship between test(s) and the disease of interest because the true shape of the ROC curve has not been exposed. Smoothed ROC curves can also be constructed in a non-parametric way with kernel estimation without making distributional assumptions (Zou *et al.*, 2013).

1.3.2   Binormal ROC Curves

The most common distributional assumption in diagnostic medicine is that both the healthy and diseased patients follow independent normal distributions (Hanley, 1988; Metz, 1980; Swets, 1979). This is referred to as a binormal distributional. If the data being analyzed are not binormal, a Box-Cox transformation may be applied to the data to satisfy this assumption. Placing this assumption on the data allows us to fit a smooth curve, which may better unveil the relationship between the disease of interest and the test(s) (Zhou, Obuchowski, & McClish, 2011).

1.3.3   Measures of Accuracy

In order to determine sufficient discrimination between healthy and disease groups for various thresholds, there are a number of different measures of test accuracy. The choice of measure depends on the method of optimization of the diagnostic accuracy (Pepe & Thompson, 2000). The Se and Sp are common measures of accuracy. Additional measures that combine the Se and Sp are the ratios of Se and Sp such as the diagnostic likelihood ratio (Pepe, 2003), the diagnostic odds ratio (Böhning, Holling, & Patilea,

2010), and the Youden Index (YI) (Youden, 1950; Faraggi, 2003; Fluss, Faraggi, &
Resier, 2005).

1.4. Optimal Operating Point(s)

The focus of this research involves continuous diagnostic tests (tests that could,
theoretically, take on any value), which implies a need for a threshold such that each test
can dichotomize the corresponding results as either healthy and diseased. There are,
theoretically, an infinite number of choices of thresholds, θ, from which to select
(Schisterman *et al.*, 2005). The choice of the optimal operating point (OOP) defines the
best threshold value to discriminate between healthy and disease states. We want to
choose an OOP that in some way optimizes accuracy. For reasons detailed in the next
chapter, the YI is a popular measure (Fluss, Faraggi, & Reiser, 2005; Perkins &
Schisterman, 2005; Ruopp *et al.*, 2008; Schisterman *et al.*, 2008), and is the focus of this
research.

The YI, initially defined as a function of the sum of the Se and Sp-1 and
calculated from the ROC curve, can be used to compare threshold candidates to select the
OOP. A definition of the optimal operating point (OOP) is the threshold that will
maximize the sum of the Se and Sp simultaneously. This measure can be expanded to
weight the Se and Sp, incorporating costs of error rates and disease prevalence – this is
known as the Generalized Youden Index (GYI).

1.5. Motivation and Goals

Given a single diagnostic test, and an assumption of binormal data, formulae have been derived to determine the OOP. However, as previously outlined, sometimes multiple tests are given to enhance the classification of disease. Not much is known about the OOP of these strategies in terms of methodology to locate the OOP or its properties. The goal of this research is to not only develop a method to select an OOP but also evaluate its properties for each of the previously mentioned sequential testing strategies (BP and BN). Specifically, for each of these two sequential testing strategies, we will derive the OOP corresponding to the YI, evaluate how close the estimates are to the true values, and determine under what conditions the estimated OOP performs best.

By using these derived OOPs as opposed to those calculated empirically, our goal is to improve the bias and other associated measurements in order to provide accurate and viable thresholds that maximize correct classification of patients. Various summary measures will be used to evaluate the choice of the optimal threshold pair. These methods will be evaluated using simulated data, and demonstrated with a relevant real-life dataset. Additionally, it is of interest to use the results of the simulation study for each sequential testing strategy to derive formulae for use in sample size estimation as well as widths of confidence intervals. Clinicians will benefit from this research by quickly being able to determine disease classification of patients and/or determine whether or not additional tests may be necessary to maximize correct classification. Additionally, they should be able to select the best testing strategy according to the patient data available.

1.6. Prospectus

This research is organized as follows: chapter 2 outlines a review of current strategies in the literature for determining OOPs. In chapter 3, methodological details will be provided for estimating the OOP and its variance for sequential testing strategies. The summary measurements necessary to describe the accuracy of the OOP will also be calculated. Chapter 4 is devoted to the results of the simulation study, as well as related sub-analyses. Chapter 5 applies the methodology to real data. Discussion and direction for future work is discussed in chapter 6.

2 Review of Relevant Statistical Topics

2.1. Introduction

In clinical practice, medical tests that generate continuous outcomes are often conducted in patients to categorize an individual patient as diseased or non-diseased. However, in order to use this classification technique, a threshold needs to be defined. The optimal value for this threshold depends on what is important to the researcher or clinician. In some cases, it may be more important to correctly classify those with disease vs. non-disease or vice versa.

2.2. Receiver Operator Characteristic (ROC) Curves

The ROC curve is a graphical way to summarize a diagnostic test's ability to correctly classify patients (Perkins & Schisterman, 2005). This curve is a visual plot of the Se (or the TPR) on the horizontal axis versus 1-Sp (or the FPR) on the vertical axis across all possible threshold values. Use of ROC curve analysis can aid in the selection of a good, or optimal, threshold (OOP); that is, a threshold that detects true positives and true negatives most accurately (Halpern *et al.*, 1996). This is done by selecting pairs of Se and Sp off of the curve and considering the thresholds connected to these as possible optimal thresholds. How these selected pairs of Se and Sp would be associated to the thresholds connected to them would have to be done via measures described in section section 2.3.

Fluss, Faraggi, & Reiser (2005) indicate the need to estimate the Se and Sp through the cumulative distribution function (CDF). There are a number of ways to

approximate the CDF comprising of non-parametric and parametric estimation, leading to both empirical and smooth ROC curves (Hanley, 1989). Non-parametric methods include the use of an empirical CDF and a kernel smoothing function; the former provides a step function while the latter is a smooth function. Parametric estimation assumes a distribution of the data, most commonly a binormal distribution.

Empirical estimates of the OOP can be estimated from all possible data points. In real life, the OOP has usually been developed in a data-driven, or empirical, way; the data tends to dictate the OOP rather than theory (Leeflang *et al.*, 2008). As defined by Zhou, Obuchowski, & McClish (2011), let n_H represent the number of patients without disease, n_D represent the number of patients with the disease, and $\overline{F}_{n_H}(c_i)$ and $\overline{F}_{n_D}(c_i)$ represent one minus the CDF, or the survival functions. For the single test case, the empirical ROC curve is comprised of the plotted pairs (FPR, TPR) defined as

$$\left(\overline{F}_{n_H}(c_i) = \frac{1}{n_H} \sum_{j=1}^{n_H} I_{\left[T_{Hj} > c_i \right]}, \overline{F}_{n_D}(c_i) = \frac{1}{n_D} \sum_{j=1}^{n_D} I_{\left[T_{Dj} > c_i \right]} \right) \tag{2.1}$$

where the indicator variable I is equal to one if $T_{Hj} > c_i$ (or $T_{Dj} > c_i$) and zero otherwise. Note that these pairs are constructed to correspond to each possible threshold value from the sample; i ranges from 1, …, TSS, where $TSS = n_H + n_D$. The first coordinate represents the FPR, while the second coordinate represents the TPR. The resulting curve is a step curve. The OOP is chosen as the observation from the ROC curve that maximizes YI. This method makes no distributional assumption on the two groups, nor does it say anything about the form of the plot (Zou, O'Malley, & Mauri, 2007).

Alternatively, the Kernel method (Zou, Hall, & Shapiro, 1998; Fluss, Faraggi, & Reiser, 2005) can be used to estimate a somewhat smooth curve from the empirical CDFs. The points on the ROC curve are defined as

$$\overline{F}_i(\theta) = \sum_{j=1}^{n_i} \int_\theta^\infty \frac{1}{n_i h_i} k\left(\frac{x - T_{ij}}{h_i}\right) dx, \ i = 0,1 \tag{2.2}$$

where k is the kernel function, h_i is the bandwidth, T_{ij} is the continuous observed result of the test of the *j*th subject with or without the condition ($k = 1$ or $k = 0$ respectively). The Gaussian kernel function, $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, can be used to rewrite equation (2.2) as

$$\overline{F}_i(\theta) = \sum_{j=1}^{n_i} \phi\left(\frac{T_{ij} - \theta}{h_i}\right), \ i = 0,1, \tag{2.3}$$

and iterative numerical methods are used to select the OOP. This method isn't stable at the extremes of the ROC curve (Lasko *et al.*, 2005).

A common method of estimating a smooth ROC curve is to assume a parametric distribution for the data. This is usually done by assuming a binormal form. Binormal ROC curves are based on a key assumption regarding the data. It is assumed that the data within the disease (D) and healthy (H) groups each independently follow normal distributions, which together follow a bivariate normal distribution. For example, the healthy group follows a bivariate normal distribution (using the symbol ~ for "is distributed as") as $H \sim BVN(\mu_{1H}, \mu_{2H}, \sigma_{1H}^2, \sigma_{2H}^2, \rho_H)$, where μ_{iH} represents the respective means, σ_{iH}^2 represents the respective variances, and ρ represents the correlation between tests. If the data are not normally distributed, it is often assumed that a transformation, such as a Box-Cox transformation, can be applied to the data in an

attempt to produce a normal distribution (Hanley, 1988; Hanley, 1989). The Box-Cox transformation is defined as

$$y_i^\lambda = \begin{cases} \dfrac{y_i^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log y_i, & \lambda = 0 \end{cases}, \qquad (2.4)$$

where y is the variable of interest and λ is the transformation parameter. The ROC curve for binormal data is defined as

$$ROC(t) = \Phi(a + b\Phi^{-1}(t)) \qquad (2.5)$$

where $a = \dfrac{\mu_D - \mu_H}{\sigma_D}$, $b = \dfrac{\sigma_H}{\sigma_D}$, and Φ refers to the standard normal cumulative distribution function (CDF) (Pepe, 2003).

2.3. Useful Measures to Select an OOP

According to Zou *et al.* (2013), the goal for deriving an optimal threshold should include aspects of accuracy, agreement, best point, and information. There are a couple of measures that summarize the ROC curve into a single value (Fluss, Faraggi, & Reiser, 2005) that can be used to identify an "optimal" threshold. The goal is that these measures minimize classification error (Jund *et al.*, 2005). Measures used to select the OOP can include, but are not limited to the diagnostic odds ratio (Böhning, Holling, & Patilea, 2010), the (0,1) criteria or the Euclidean Distance (ED) (Zou *et al.*, 2013), and the Youden Index (YI) (Faraggi, 2003; Fluss, Faraggi, & Resier, 2005). Zou *et al.* (2013) also mention the measures percent correct diagnosis (PCdx), kappa (κ), and mutual information (MI); the former two maximize agreement between the reference standard

(RS) and the diagnosis (dx) based on tests, while the latter maximizes the information shared between RS and dx.

The diagnostic odds ratio (DOR), as cited by Böhning, Holling, & Patilea (2010) is defined as

$$DOR = \frac{Se}{1 - Se} \times \frac{Sp}{1 - Sp}.$$ (2.6)

An OOP can be defined as the value that maximizes the DOR. However, problems arise with convexity when the variances between the healthy and the diseased groups are equal; values that maximize the DOR are the values on the boundary of the parameter range, which are useless OOPs. Böhning, Holling, & Patilea (2010) also suggest that although this property does not generalize to the unequal variance case, other patterns arise that still rule out the DOR as a useful measure for determining OOPs (these patterns are not explained in the literature; a statement is made by Böhning, Holling, & Patilea (2010) that "Other patterns than convexity arise, which are neither suitable for using the DOR as an optimizing criterion. So it's not worthwhile looking for a generalization of the convexity property – it does not generalize.").

The (0,1) criteria, also known as the geometric method or the ED method, refers to the selection of a threshold such that the point is the closest (in distance) on the ROC curve to the point (0,1) on the axes. Perkins & Schisterman (2005) note that this method more specifically involves locating the threshold corresponding to the shortest radius from the point (0,1) and the ROC curve; where this radius intersects with the curve is then considered to be the OOP. Formally, the OOP using this measure would be determined by

$$\theta^* = \min_\theta \{(1 - Se(\theta))^2 + (1 - Sp(\theta))^2\}, \qquad (2.7)$$

where θ represents all possible threshold values and θ^* is the OOP selected from amongst all θ. However, this method does not provide the opportunity to control for cost (monetary, patient burden, or misclassification) and disease prevalence if desired for overall misclassification rates (Perkins & Schisterman, 2005).

Another common measure is the Youden Index (YI), which is the focus of this research. The YI is the maximum of the sum between the Se and Sp – 1 (Skaltsa *et al.*, 2010; Leeflang *et al.*, 2008; Perkins & Schisterman, 2005; Halpern *et al.*, 1996), and is visualized as the vertical distance between the ROC curve and the chance line. Figure 2.1 depicts this relationship.

Figure 2.1: Youden Index
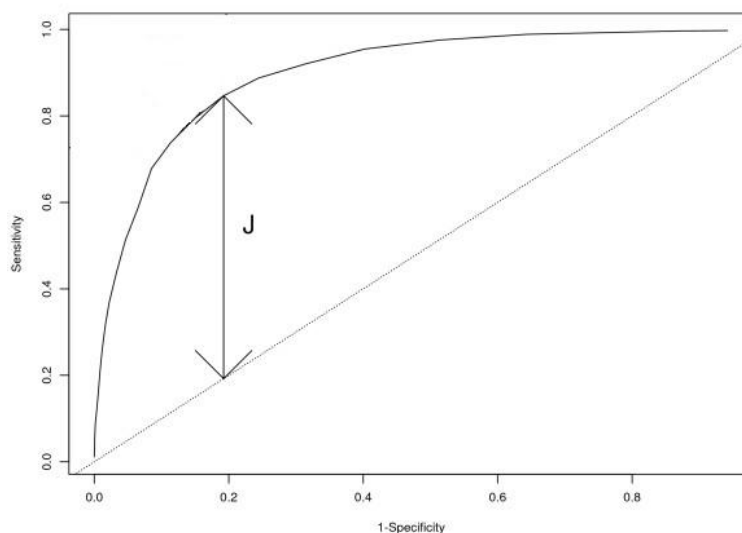
The YI, oftentimes referred to as *J* in the literature, is formally defined as

$$YI = \max_\theta \{Se(\theta) + Sp(\theta) - 1\}, \qquad (2.8)$$

where θ is taken over all possible choices of thresholds. $\theta*$ is the value that maximizes this measure. This may be rewritten as YI = TPR(θ) – FPR(θ). This index ranges between -1 and 1, with the subset [-1, 0) having the opposite interpretation as the subset (0, 1]; the latter is the usual range. Values closer to one represent a more accurate test (Perkins & Schisterman, 2005). The YI does not have the boundary (convexity) issues that the DOR has (Böhning, Holling, & Patilea, 2010), making it a more popular measure (Fluss, Faraggi, & Reiser, 2005; Perkins & Schisterman, 2005; Ruopp *et al.*, 2008; Schisterman *et al.*, 2008).

Fluss, Faraggi, & Resier (2005) note that when the two populations (healthy and diseased) are completely distinguishable, YI = 1, while complete overlap of the two populations yields YI = 0, as shown below in Figure 2.2. Note that this measure is independent of population size.

Figure 2.2: Small, Medium, and Large Population Overlaps (Respectively)
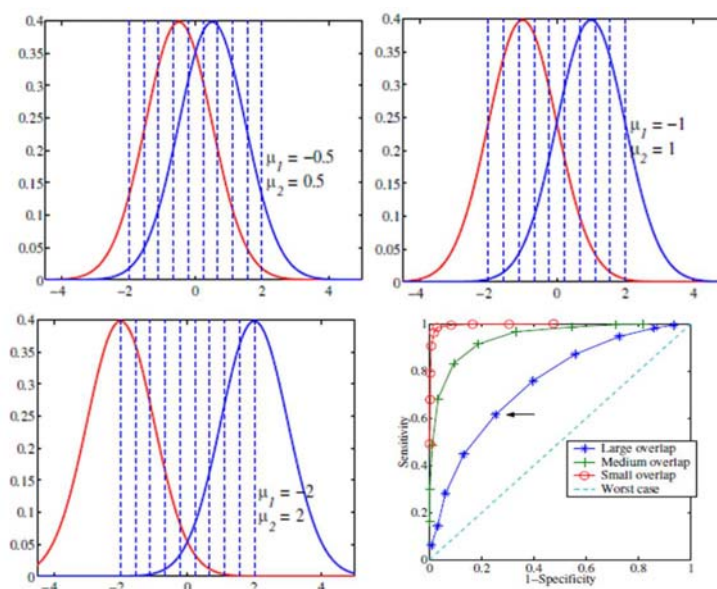
The YI can also be extended to what is known as the Generalized Youden Index (GYI). This extends the YI to include cost and prevalence, and is defined as

$$GYI = \max_\theta \{Se(\theta) + r[Sp(\theta) - 1]\} = \max_\theta \{TPR(\theta) - rFPR(\theta)\}, \qquad (2.9)$$

where r is defined as $\dfrac{(1-p)}{p} \dfrac{(C_{FP} - C_{TN})}{(C_{FN} - C_{TP})}$. The value C represents the cost, p represents the prevalence, FP represents the false positive group, TN represents the true negative group, FN represents the false negative group, and TP represents the true positive group. Equivalently, r can also be notated as $\dfrac{1-\pi}{a\pi}$, where π represents the prevalence of disease and $a = \dfrac{C_{FN}}{C_{FP}}$, which refers to the misclassification cost of a FN as opposed to a FP classification (Perkins & Schisterman, 2005) and C_{TN} and C_{TP} are taken to be zero. Note that when $r = 1$, GYI = YI. Since we can weight the Se and Sp in the YI equation to produce the GYI, this is an improvement over the (0, 1) criterion (Perkins & Schisterman, 2005), again, making the YI/GYI a more popular measure.

There are other possible measures to derive optimal thresholds. The kappa is defined as

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)}, \qquad (2.10)$$

where $P(A)$ is the proportion of times the tests agree and $P(E)$ is the proportion of times the tests are expected to agree by chance. We also have the PCdx and the MI, which is defined as

$$MI_\theta = MI(D_\theta, T) = H(D_\theta) + H(T) - H(D_\theta, T), \qquad (2.11)$$

where θ represents any threshold, *D* represents the diagnosis, *T* represents the gold standard, and *H* are entropies. However, as stated by Zou *et al.* (2013), the YI and ED appear to be robust and stable for the underlying set of ROC parameters (including AUC, correlation, *b*, etc). However, the results based on PCdx, κ, and MI tend to differ, depending on prevalence.

2.4. Area Under the Curve (AUC) and its Influence on the OOP

The AUC is a common measure of overall test performance (Westin, 2001; Greiner, Pfeiffer, & Smith, 2000). This measure summarizes the accuracy over the entire ROC curve (Grenier, Pfeiffer, & Smith, 2001) rather than just at a particular point on or portion of the ROC curve. This measure is defined as

$$AUC = \int_0^1 ROC(t)\,dt \tag{2.12}$$

or

$$AUC = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right) \tag{2.13}$$

assuming the binormal form of the ROC curve.

Recall from Chapter 1 that the chance line refers to the diagonal line connecting the point (0, 0) and (1, 1). An uninformative test (Pepe, 2003) diagnoses the patient no better than guessing, and has an AUC of 0.5. A perfect test would be reflected in an ROC curve passing through the point (0, 1), and hence the corresponding AUC would be equal to 1. Greiner, Pfeiffer, & Smith (2001) cite a guideline suggested by Swets in 1998 that contextualizes the values of AUC: uninformative (AUC = 0.5), less accurate

(0.5 < AUC ≤ 0.7), moderately accurate (0.7 < AUC ≤ 0.9), highly accurate

(0.9 < AUC < 1), and perfect classification (AUC = 1). Thus, a higher AUC reflects a

better test accuracy.

While this measure provides a summary of overall accuracy of the test, it does not

aid in the selection of an OOP for diagnostic purposes. It does, however, influence the

OOP by influencing the shape of the ROC curve. Larger AUC values represent more

accurate diagnostic tests, while lower AUC values indicate less accurate tests.

2.5. OOP for the Univariate Case using the YI and Binormal Assumption

As described above, empirical ROC curves can be used to aid in the selection of the

OOP. Empirically derived Se and Sp pairs would be taken from the curve and the data

point that is associated with them would be assessed via the YI. The data point that yields

the highest YI would then be considered as the OOP. However, Leeflang (2008) and

Fluss, Faraggi, & Reiser (2005) have shown that parametric methods perform better in

the selection of an OOP. Although a kernel smoothing method could be used, in a

univariate diagnostic test case, the formula for determining the OOP via the Youden

Index has been well documented when a binormal model is assumed.

Let X be a test's measure for truly diseased and Y be a test's measure for truly

healthy. Assume X and Y each independently follow a normal distribution such that

$$X_i \sim N(\mu_D, \sigma_D^2) \text{ and } Y_j \sim N(\mu_H, \sigma_H^2), \qquad (2.14)$$

where $i = 1, \ldots, n_D$ and $j = 1, \ldots, n_H$, and assume that $\mu_D > \mu_H$ (this inequality can be reversed if you switch the cases and controls in the subsequent analysis). The sensitivity and specificity are

$$Se(\theta) = \Phi\left(\frac{\mu_D - \theta}{\sigma_D} \right) \tag{2.15}$$

and

$$Sp(\theta) = \Phi\left(\frac{\theta - \mu_H}{\sigma_H} \right), \tag{2.16}$$

where, again, Φ represents the standard normal CDF. To obtain the OOP, $\theta*$, across all values of θ, substitute equations (2.15) and (2.16) into equation (2.8), set equal to zero and solve. It has been shown (Halpern *et al.*, 1996; Fluss, Faraggi, & Reiser, 2005) that the result from the quadratic yields a maximum, or OOP, of

$$\theta* = \frac{\left(\mu_D \sigma_H^2 - \mu_H \sigma_D^2 \right) - \sigma_H \sigma_D \sqrt{\left(\mu_H - \mu_D \right)^2 + \left(\sigma_H^2 - \sigma_D^2 \right) \log\left(\frac{\sigma_H^2}{\sigma_D^2} \right)}}{\left(\sigma_H^2 - \sigma_D^2 \right)}. \tag{2.17}$$

This simplifies when the variance between the healthy and disease groups is equal to

$$\theta* = \frac{\mu_D + \mu_H}{2}. \tag{2.18}$$

To obtain the Youden Index θ^* is inserted into equation (2.8); this value should be the highest YI available, thus validating the choice of the OOP. The same approach could be used for the GYI in equation (2.9) (Jund *et al.*, 2005; Skaltsa *et al.*, 2010).

While Jund *et al.* (2005) and Leeflang (2008) describe situations in which the data are lognormal, these situations can be transformed to normal and the methodology

described above can be applied. Schisterman & Perkins (2007) show that, rather than assuming a normal distribution, one could assume the two groups each follow a gamma distribution such that

$$X_i \sim \Gamma(\alpha_x, \beta_x) \text{ and } Y_j \sim \Gamma(\alpha_y, \beta_y), \tag{2.19}$$

where i and j are defined as in the normal case, Γ is the gamma function, and α and β are the shape and scale parameters respectively. This is an important distinction from the lognormal case because gamma data cannot be transformed to normal; hence new methodology needs to be derived. A closed form solution can be found when the shape or scale parameters are equal across groups; otherwise numerical approximation methods must be used. Then YI and $\theta*$ can be estimated by

$$YI = \overline{F}(\theta) + G(\theta) - 1, \tag{2.20}$$

where F and G are the gamma CDFs with $\overline{F} = 1 - F$. The OOP, $\theta*$, will be the value that yields the largest YI.

2.6. Use of More than One Diagnostic Test (Sequential Testing)

While much is known about the derivation and properties of an OOP for a single diagnostic test strategy (Shen, 2008), not as much is known regarding the same for sequential diagnostic test strategies. As an example, consider two tests. The first test, Test 1, is given to all patients. The second test, Test 2, is given to a subset of these patients pending the results of Test 1. A sequential testing strategy may be considered due to cost of the tests, invasiveness of the tests, and/or for improved diagnostic accuracy (Shen, 2008). These two tests may either be two different tests taken at two different time points,

two tests taken at the same time (for example, low-density lipoprotein and high-density lipoprotein), or the same test taken longitudinally.

There are two popular strategies currently published and used in the literature. These are referred to as Believe the Positive (BP) and Believe the Negative (BN). Other methods to combine the tests exist as well, such as those previously discussed in Chapter 1, but are not the focus of this research and will not be discussed further here. Diagnoses for both testing strategies in this dissertation are made based on Test 1 as to whether the patient is healthy/diseased or needs further testing, and this decision varies based on the strategy employed. Descriptions of both sequential methods are found in the following subsections. A visual summary of each can be seen in Figure 2.3.

2.6.1 Believe the Positive (BP)

For the BP strategy, consider the following: Test 1 is administered to each subject (with continuous test result X_1). If the test result is high (i.e., $X_1 > \theta_1$), then stop. The subject will be considered positive for the disease. Do not do the second test. If the test result is low (i.e., $X_1 \leq \theta_1$) then administer the second test (with continuous test results X_2). If X_2 is high ($X_2 > \theta_2$) then the subject will be considered positive for disease. If X_2 is low ($X_2 \leq \theta_2$) then the subject will be considered negative for the disease. In this case, BP diagnoses patients as positive if either Test 1 or Test 2 comes back positive (Shen, 2008).

2.6.2   Believe the Negative (BN)

Next, consider the BN strategy. Test 1 is measured on each subject (with continuous test result X_1). If the test result is low (i.e., $X_1 \le \theta_3$), then stop. The subject will be considered negative for the disease. Do not administer Test 2. If the test result is high (i.e., $X_1 > \theta_3$) then administer Test 2 (with continuous test results X_2). If X_2 is high ($X_2 > \theta_4$) then the subject will be considered positive for disease. If X_2 is low ($X_2 \le \theta_4$) then the subject will be considered negative for the disease.

Figure 2.3: BP and BN Diagnoses
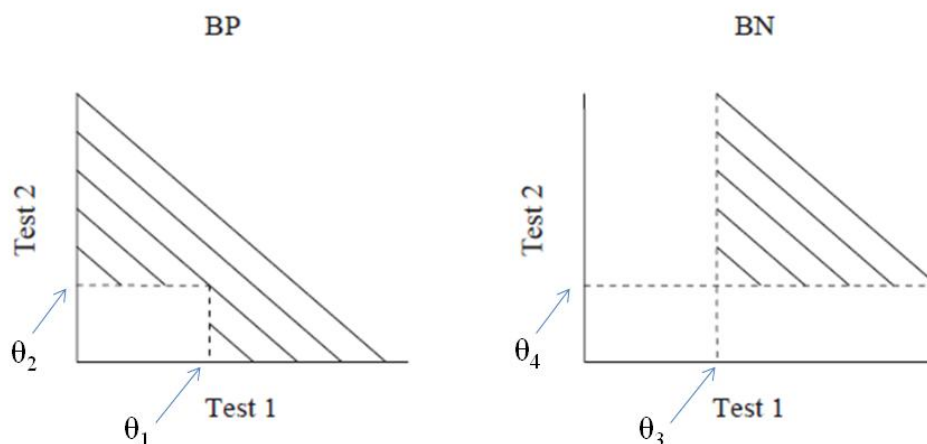
2.6.3   Visual Representation of BP and BN

Refer to Figure 2.3; this illustrates the BP and BN strategies. The BP strategy claims a positive diagnosis when either test is positive, while the BN strategy claims a positive diagnosis only when both tests are positive (Shen, 2008; Marshall, 1989). While Figure 2.3 portrays the rules of the combination of tests, it does not portray the OOP (Shen, 2008).

2.7. Univariate Cases in the Literature

The estimation of OOP for YI has been examined in recent literature for the single test case. Jund *et al.* (2005) looked at normal and log-normal data with equal variances across groups and suggested that selecting the OOP by means of the value that maximizes Se + Sp*R works best when the difference between the two means (one for each group) and sample size are high and when the weighting parameter (*R*) for the Sp is close to one. This was found to be "asymptotically non-biased" and have acceptable coverage as long as the sample size in each group is greater than 50.

Schisterman & Perkins (2007) looked at finding a confidence interval for the OOP as well as the YI using several bootstrapping methods. While both may be calculated and may come to different conclusions, the OOP CI is the result of most interest for this research. The derived CIs had less appropriate coverage and wider intervals than those generated via the delta method. However, these CIs rely heavily on distributional assumptions. If the assumptions are not met, non-parametric methods are more robust (such as bootstrapping).

Schisterman *et al.* (2005) extended the current methods for calculation of the YI and an OOP from a single sample to a pooled sample, suggesting that pooling samples can save cost and is robust in regards to the estimates of the optimal threshold and corresponding YI. Skaltsa *et al.* (2010) looked at varying the cost parameter in the GYI, and found that when *r* was high the estimators had more bias; *r* was also found to affect the estimation of the Se and Sp, selecting estimation values it otherwise would not for the

$r = 1$ case. Additionally, they found that parametric estimators had less variability than empirical estimators.

Fluss, Faraggi, & Reiser (2005) compared methods of obtaining optimal thresholds from four different methods: normally distributed data, transformed to normal data using a Box-Cox transformation, empirical, and kernel smoothing in terms of bias and root mean square error (RMSE). They found the empirical estimate to be the least accurate method (especially for smaller sample sizes), while the transformed to normal appeared perform equally as well as any others so long as the data could be adequately transformed, and works just as well as the kernel method when sample sizes is greater than or equal to 50.

2.8. BP and BN Cases in the Literature

This research focuses on the diagnostic accuracy of OOPs. The ultimate goal is to obtain an OOP for a sequence of tests and determine its accuracy, comparing the effect of different parameters that effect the estimation and as compared to a simple empirical approach, while limiting the bias and obtaining accurate coverage. Two tests, as opposed to one, may lead to a more complex clinical interpretation, but may be more reliable and "improve diagnostic reliability" (Marshall, 1989). Thompson (2003) mentions that, with these sequential testing strategies, the second test could be the same test as Test 1, just administered later in time. The latter, according to Politser (1980), may be especially true if the initial test is unreliable and needs to "be repeated".

Marshall (1989) discusses the predictive values of diagnostic tests. He states that the rules for combining the two tests "do not necessarily have better predictive values than a single test; whether they do may depend on the association between test outcomes." If tests are associated, this may alter the decision of when to conclude that patients are diseased or not. The predictive values also shift when more than two tests are combined sequentially.

Another issue to address for these strategies is identifying which set of tests are best to determine optimality of diagnosis. Researchers may have a host of tests to select from. Ahmed, McClish, & Schubert (2011) looked at accuracy of diagnosis of sequential tests while taking the cost into account. Shen (2008) suggests choosing tests "that have opposite correlation signs between the diseased and non-diseased populations." She suggests that this difference adds another dimension to the distinction between the two populations.

While several studies have been published using sequential strategies, no study mentions the selection and evaluation of an optimal threshold for these strategies. Zöller, Burkard, & Schäfer (1991) utilized a modified BP strategy when examining the effects of using western blot to diagnose Lyme Borreliosis. For this example, bands on the immunoblots were inspected and a positive diagnosis was made when both bands were present and one band exceeded the respective optical density (OD) critical level. This was done for various cutoff values. Se values were compared for Sp values above 90%, and "the couple of critical OD levels for the two bands was chosen to obtain the maximum Se under the restricted Sp in the sample."

Baccouche *et al.* (2009) also utilized the BP strategy to assess the diagnostic synergy of cardiovascular magnetic resonance and endomyocardial biopsy in patients with positive acute chest pain but without significant coronary artery disease. The approach for this data example was as follows: the patient was considered being troponin-positive if at least one of the procedures could establish a diagnosis; if neither procedure yielded a diagnosis, or if they did not yield the same diagnosis, the patient was considered not troponin-positive.

Dreher *et al.* (2004) looked at the validity of $p27^{kip1}$ and Gleason scores for biopsies as predictive tests with both the BP and BN strategies for the Gleason scores of the surgical specimen for the diagnosis of prostate cancer. Note that these were binary tests, not continuous tests. Rather than using thresholds for each of the tests, prediction was used, but details were not given. Results, however, showed that the combination of predictors was not any more reliable than the Gleason score alone.

Cury *et al.* (2006) used BP rules to combine stress first-pass perfusion magnetic resonance imaging and delayed enhancement magnetic resonance imaging to assess the diagnosis of coronary artery disease. They found a slightly better diagnostic accuracy for the combination of tests than for just a single test (first-pass perfusion magnetic resonance imaging).

Ahmed, McClish, & Schubert (2011) looked at BP and BN sequential strategies, accounting for cost, to diagnose diabetes from BMI and fasting glucose levels in the Pima Indian population. It was found that the BN strategy had a lower FPR than the BP strategy for their data example. However, optimal thresholds were calculated by means of

a grid search rather than methodologically. Leeflang (2008) noted that when the optimal threshold is selected in a data-driven method, using the YI, the corresponding Se and Sp estimates are overestimated.

Thompson (2003) used a prostate cancer example from the CARET study involving BP rules in which two scenarios were analyzed for repeated tests: the same threshold used for everyone at each testing point, but the threshold can change at each follow-up (non-adaptive) and allowing the threshold corresponding with the conditional percentiles to differ across individuals and/or follow-ups (adaptive). Although these methods of calculating thresholds were mentioned, no guidance was given as to which one to choose.

Lastly, Shen (2008) defined and implemented both a BP and BN strategy on a dementia Alzheimer's disease in African-American patients. The two tests are continuous neuropsychological tests (Word List Learning and Word List Delayed Recall and Recognition), and the outcome was dementia. Again, thresholds were calculated by a grid search rather than a methodological approach. She also suggests that the choice between BP and BN may depend on the underlying distribution of the data as well as the signs of the correlation between the two groups.

2.9. Summary

In this chapter, the choice of an OOP for the univariate (single test) scenario as well as the sequential testing strategies were defined and explained. According to Shen (2008), some literature on detecting an OOP based on the AUC is available. However, as stated

previously, little is known regarding the selection of an OOP for the case when two or more continuous tests are available and are chosen to be used in tandem. The research in this paper will include the development and evaluation of an OOP for each of the two sequential testing strategies detailed above.

3   Methodology

In this chapter, methods for determining the optimal operating point (OOP) in a

sequential testing strategy scenario will be examined by means of simulation studies.

Two sequential testing strategies will be considered: "Believe the Positive" (BP) and

"Believe the Negative" (BN). To describe the properties of the OOP and situations in

which the method of choosing the OOP performs best, evaluation measures must be

considered. For this simulation study, bias, root mean square error (RMSE), percent

overestimation of Se and Sp, 95% simultaneous confidence intervals (SCIs), and

confidence regions (CRs) will be evaluated. Additionally, comparison to the empirical

approach for determining an optimal threshold will be done by calculating the bias of the

empirical OOPs, corresponding Se and Sp, and overestimation of Se and Sp. The

simulated study will be performed in SAS (SAS Institute, Cary, NC).

3.1. Basic Notation

Consider two groups of patients: those without disease (healthy), referred to as H, and

those with disease (disease), referred to as D. Let X be a test's measure for truly diseased

and Y be a test's measure for truly healthy. Assume X and Y each independently follow a

bivariate normal distribution, or can be so transformed such that

$X_i \sim BVN(\mu_{1D}, \mu_{2D}, \sigma_{1D}^2, \sigma_{1D}^2, \rho_D)$ and $Y_j \sim BVN(\mu_{1H}, \mu_{2H}, \sigma_{1H}^2, \sigma_{2H}^2, \rho_H)$, where μ

represents the means associated with each test for each group, σ^2 represents the variance

for each test for each group, and ρ represents the correlation between the two tests within

each group. The numerical subscripts represent the corresponding test with $i = 1, \ldots,$

$n_D, j = 1, \ldots, n_H$. Within tests, assume each group independently follows a normal

distribution, such that the healthy patients follow a standard normal distribution for each

test $\mu_{1H_j} = 0, \mu_{2H_j} = 0, \sigma_{1H_j} = 1, \sigma_{2H_j} = 1$. The entire set of simulated parameters will be

denoted as $\underline{\gamma} = (\mu_{1H}, \mu_{2H}, \mu_{1D}, \mu_{2D}, \sigma_{1H}, \sigma_{2H}, \sigma_{1D}, \sigma_{2D}, \rho_H, \rho_D)$.

There were 144 combinations of parameters available for the simulation studies

for each sequential testing strategy, which will be defined in the next chapter. Varied

parameters include values for the AUC for each test (a function of the population means

and standard deviations), the ratio of standard deviations between the two groups (b),

correlation between the two tests for each group, total sample size (TSS), and the

allocation (k) of H and D within the TSS. The total sample size is $TSS = n_H + n_D$, where

$$n_D = kn_H \tag{3.1}$$

The ratio of standard deviations is denoted as $b_i = \dfrac{\sigma_{iH}}{\sigma_{iD}}$, where i represents the respective

test number (1 or 2). Under the normality assumption, the area under the ROC curve

(AUC) for the tests, are

$$AUC_i = \Phi\left(\frac{a_i}{\sqrt{1+b_i^2}}\right) \tag{3.2}$$

AUC_1 and AUC_2 for the first (1) and second (2) tests respectively, where

$a_i = (\mu_{iD} - \mu_{iH}) / \sigma_{iD}$ (i represents the respective test number).

Note that, for simplicity, we assume that the healthy group follows a standard normal distribution for each test, or that the means and standard deviations for the multivariate distribution for both tests were zero and one respectively. In order to calculate the means and standard deviations for the disease group, we must use the AUC and the assumed mean values. The mean for the disease group for test i is calculated as

$$\mu_{iD} = \sigma_{iD}\sqrt{1+b_i^2}\,\Phi^{-1}(AUC_i) + \mu_{iH} \,. \tag{3.3}$$

The standard deviations for the disease groups for each test is calculated based on the desired values for b_i such that $\sigma_{iD} = \dfrac{\sigma_{iH}}{b_i}$.

The optimal threshold for BP will be notated as (θ_1, θ_2) while the optimal threshold for BN will be notated as (θ_3, θ_4).

3.2. Sequential Testing Formulae

Recall that we are trying to maximize the GYI, so our formulae will be based upon this. The derivation of each set of formulae will be described in the following subsections. Let the GYI be the objective function from which we derive the OOP. The GYI is formally defined as

$$GYI = \max\{Se(c) + r(Sp(c)-1)\} = \max\{TPR(c) - rFPR(c)\}, \tag{3.4}$$

where the constant r is a function of the misclassification cost and prevalence in order to weight the Se and Sp; note that when $r = 1$, GYI=YI.

3.2.1 BP System of Equations

Ahmed, McClish, & Schubert (2011) show that the FPR and the TPR for the BP strategy are given by

$$FPR^{BP}(\theta_1, \theta_2) = 1 - F_{X1H, X2H}(\theta_1, \theta_2) \tag{3.5}$$

and

$$TPR^{BP}(\theta_1, \theta_2) = 1 - F_{X1D, X2D}(\theta_1, \theta_2), \tag{3.6}$$

where $F_{X1H, X2H}(\theta_1, \theta_2)$ and $F_{X1D, X2D}(\theta_1, \theta_2)$ are the cumulative distribution functions. In this research we will assume they represent the bivariate normal distribution. The subscripts 1H and 2H represent test 1 and test 2 for those without disease (H), and subscripts 1D and 2D represent test 1 and test 2 for those with disease (D). Additionally, θ_1 represents the threshold associated with test 1, while θ_2 represents the threshold associated with test 2. For the BP case, we then have

$$GYI^{BP} = [1 - F_{X1D, X2D}(\theta_1, \theta_2)] - r[1 - F_{X1H, X2H}(\theta_1, \theta_2)]. \tag{3.7}$$

3.2.2 BN System of Equations

Ahmed, McClish, & Schubert (2011) show that the FPR and the TPR for the BN strategy are given by

$$FPR^{BN}(\theta_3, \theta_4) = 1 - F_{X1H}(\theta_3) - F_{X2H}(\theta_4) + F_{X1H, X2H}(\theta_3, \theta_4) \tag{3.8}$$

and

$$TPR^{BN}(\theta_3, \theta_4) = 1 - F_{X1D}(\theta_3) - F_{X2D}(\theta_4) + F_{X1D, X2D}(\theta_3, \theta_4). \tag{3.9}$$

The subscripts represent the same tests and disease groups as they did for the BP strategy, θ_3 represents the threshold associated with test 1, and θ_4 represents the threshold associated with test 2. The GYI for the BN strategy is then

$$
\begin{aligned}
GYI^{BN} = &[1 - F_{X1D}(\theta_3) - F_{X2D}(\theta_4) + F_{X1D,X2D}(\theta_3,\theta_4)] \\
&- r[1 - F_{X1H}(\theta_3) - F_{X2H}(\theta_4) + F_{X1H,X2H}(\theta_3,\theta_4)]
\end{aligned}
\tag{3.10}
$$

3.3. Newton-Raphson (NR)

Generally, a maximum can be found by taking derivatives of equations, setting the equations equal to zero, and solving for the necessary parameter(s). For both GYI^{BP} and GYI^{BN}, this process yields a set of two complex nonlinear functions with two unknown parameters. Iterative procedures are necessary since an explicit solution cannot be attained for the unknown parameters. The Newton-Raphson (NR) method is an iterative method to find roots of functions utilizing both first and second derivatives, also known as the gradient (**f**) and the Hessian (**H**) matrices respectively, for real valued functions. This method does so by using the Taylor Series of the function, of which it is based upon. Suppose you have two equations, $f(x,y)$ and $g(x,y)$. An approximation of the Taylor Series expansion for $f(x,y)$ is given by:

$$
\begin{aligned}
f(x + \Delta x, y + \Delta y) = f(x^{(0)}, y^{(0)}) + &\left.\frac{df(x,y)}{dx}\right|_{[x^{(0)},y^{(0)}]} \Delta x \\
+ &\left.\frac{df(x,y)}{dy}\right|_{[x^{(0)},y^{(0)}]} \Delta y + O(\Delta^2),
\end{aligned}
\tag{3.11}
$$

where $\mathbf{\Delta} = \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$, O refers to the remainder due to the approximation (truncation) of

the Taylor Series polynomial, and (0) refers to the starting value. The expansion for

$g(x, y)$ is similar. Note that in equation (3.11), the partial derivative of $f(x, y)$ with

respect to x represents the variation due to x, while the partial derivative of $f(x, y)$ with

respect to y represents the variation due to y. Also $\begin{pmatrix} x^{(0)} \\ y^{(0)} \end{pmatrix} + \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \mathbf{r}$, and $\mathbf{f(r)} = \mathbf{0}$, so we

can set equation (3.11) and its counterpart for $g(x, y)$ equal to zero. Then equation (3.11)

becomes

$$f(x + \Delta x, y + \Delta y) = f(x^{(0)}, y^{(0)}) + \left.\frac{df(x, y)}{dx}\right|_{[x^{(0)}, y^{(0)}]} \Delta x + \left.\frac{df(x, y)}{dy}\right|_{[x^{(0)}, y^{(0)}]} \Delta y. \quad (3.12)$$

Because equation (3.12) (and it counterpart expansion $g(x, y)$) are two complex

linear functions with two unknowns, a closed form solution does not exist. In order to use

the NR algorithm, a gradient and a Hessian matrix must be defined. For our purposes, the

gradient is comprised of $f(x, y)$ and $g(x, y)$, referred so as our sequential testing system

of equations. Taking partial derivatives of the gradient to obtain the Hessian yields a total

of four partial derivates, defined as

$$\mathbf{H} = \begin{pmatrix} \dfrac{df(x, y)}{dx} & \dfrac{df(x, y)}{dy} \\ \dfrac{dg(x, y)}{dx} & \dfrac{dg(x, y)}{dy} \end{pmatrix}. \quad (3.13)$$

The NR algorithm is then

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} f(x, y) \\ g(x, y) \end{bmatrix} + \mathbf{H} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}, \quad (3.14)$$

or reduced,

$$0 = \mathbf{f}^{(0)} + \mathbf{H}^{(0)}\boldsymbol{\Delta}, \qquad (3.15)$$

where $\mathbf{f} = \begin{bmatrix} f(x,y) \\ g(x,y) \end{bmatrix}$. This then implies that

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{H}(\mathbf{x}_k)^{-1}\mathbf{f}(\mathbf{x}_k), \qquad (3.16)$$

where $\mathbf{x}_k = \begin{bmatrix} x \\ y \end{bmatrix}$, and $k = 0$, 1, …, k refers to the iteration number. The left hand side

represents the next "step" in the algorithm, while the right hand side represents the

computations and the previous solution. This algorithm iterates ($\mathbf{H}$ is calculated at every

iteration using the previous estimate and used in solving the system of equations) k times

until we achieve convergence. More specifically, we iterate until the difference between

the current estimate $\mathbf{x}_{k+1}$ and the previous estimate $\mathbf{x}_k$ is less than an *a priori* set

convergence criterion.

3.4. Newton-Raphson (NR) with Ridging

One requirement for the use of the NR formula is that the Hessian matrix, or the matrix of

second derivatives, must be positive definite. This means that the matrix is symmetric

with only positive eigenvalues. In other words, a square matrix $\mathbf{A}$ is positive definite if

the quadratic form $\mathbf{x'Ax}$ is greater than zero for all non-zero vectors of $\mathbf{x}$. If this condition

does not hold, then the inverse cannot be calculated. Positive definiteness can be

calculated in ways including, but not limited to, calculating all the eigenvalues and

checking to see if they are all positive, calculating the determinants of the 2 upper left matrices, or by using a Cholesky decomposition.

If the Hessian matrix is not positive definite, or even nearly so, then the usual NR method cannot be used. It became apparent while setting up this research that the NR method was not converging as often as it should (compared against a grid search). There were several instances when the number of times we found an OOP was significantly less for the NR method as compared to a grid search. These problems seemed to disappear once the algorithm was switched to the Newton-Raphson with ridging method, leading to the conclusion that the Hessian related to the GYI may not satisfy the positive-definite requirement of the NR method.

A related alternative method is Newton-Raphson ridge optimization which works around this problem by adding a multiple of the identity matrix to the Hessian. This adds a positive value (called the ridge parameter) to the diagonal of the Hessian, in turn making it positive definite. The larger the ridge, the more optimization problems there are (SAS 9 Documentation). The updated NR algorithm would then look like

$$\mathbf{x}_{k+1} = \mathbf{x}_k - [\mathbf{H}(\mathbf{x}_k) + \alpha\mathbf{I}]^{-1}\mathbf{f}(\mathbf{x}_k), \qquad (3.17)$$

where α represents the ridge parameter and is large enough to make the matrix positive definite and $\mathbf{I}$ represents the identity matrix.

3.4.1   Non-Linear Programming using SAS PROC NLP

Simulation parameters were generated using PROC NLP in SAS. This procedure uses optimization techniques to minimize/maximize one or more continuous nonlinear

functions. Using Newton-Raphson with ridging, OOPs were determined by maximizing the GYI. Because the multivariate NR is very sensitive to the starting values, our starting value was determined by using a coarse grid search on the interval [-4, 4] by increments of 0.1.

A grid search is considered to be a brute force method in which the specified parameter space is covered by a specified precision, or interval. The statistics of interest are calculated for every step in every n-dimensional direction across the grid space. This is not necessarily efficient or precise, but does yield a general overview of the objective function. In this case, since we are maximizing the GYI, the OOP from the grid search method would be the pair of thetas that yields the largest GYI. This process can be very time consuming and computationally expensive depending on how large the parameter space is or how fine the search increment is.

PROC NLP was used to select a starting value using a grid search rather than programming a specific value. According to the SAS documentation, providing the grid search as a starting point allows the procedure to compute the objective function at each grid point and select the "best", or most feasible, point as a starting value for the optimization technique by means of a rough grid of the parameter space.

For this study, the gradient and Hessian matrices are estimated using finite-difference approximations rather than being explicitly computed to save computation time, although the advantage of PROC NLP is that you can select either the analytic solution or supply the exact derivatives. See Appendix A for the explicit gradient and Hessian formulae for each sequential testing strategy. Additionally, PROC NLP selects

the best ridge parameter automatically. This was chosen rather than comparing models with different, *a priori* selected ridging values to decrease the simulation combination count.

3.5. Bias and Root Mean Square Error (RMSE)

To assess the accuracy of the simulation study in regards to the estimation of the OOPs, bias and root mean square error (RMSE) will be computed by comparing the true threshold to the estimated OOPs.

Bias is a measure of the difference between the true and estimated value of a parameter, or in other words, a measure of accuracy. This specifically measures the magnitude in which the estimated OOP $(\hat{\theta}_1, \hat{\theta}_2)$ does not match the true OOP (θ_1, θ_2). Bias will be calculated as

$$bias((\hat{\theta}_1, \hat{\theta}_2)) = (\hat{\theta}_1 - \theta_1, \hat{\theta}_2 - \theta_2). \tag{3.18}$$

where bias is calculated separately for the thresholds associated with each test. Bias is calculated similarly for the BN and BP strategies. Positive bias indicates that the estimate tends to be larger than the true value, while negative bias indicates that the estimate tends to be smaller than the true value. In this dissertation, we are interested in the bias of the OOP as well as the corresponding Se and Sp. For the simulation study median bias will be reported.

Another common measure of differences between estimates and true values is the RMSE, which incorporates bias and variance simultaneously. RMSE, which is the square root of the MSE, will be calculated as

$$RMSE((\hat{\theta}_1, \hat{\theta}_2)) = \sqrt{\sigma^2_{(\hat{\theta}_1, \hat{\theta}_2)} + bias((\hat{\theta}_1, \hat{\theta}_2))^2} \ . \tag{3.19}$$

The RMSE is sometimes preferred over the MSE for scenarios with many leading zeros as it is easier to understand while also retaining the units.

3.6. Simultaneous Confidence Intervals (SCI) and Confidence Regions (CR)

To assess the precision of an estimate, simultaneous confidence intervals (SCIs) and confidence regions (CRs) will be used. In the multivariate case, estimated parameters may be correlated; that is, the estimation of one threshold may depend on the estimation of another. In the case of the OOP, the increase/decrease in the threshold for test one may affect the increase/decrease in the threshold for test two. Additionally, as the number of parameters, p, increases, the probability that each of the univariate, or "one-at-a-time", confidence intervals accurately covers the space decreases from the set confidence level.

Bonferroni SCIs are set up like Wald-type confidence intervals (Estimate ± z*(Standard Deviation)), where z refers to the confidence level based on a standard normal distribution. We alter z in this case such that $z^* = z_{1-\frac{\alpha}{2p}}$, where α represents the confidence level of interest and p represents the number of parameters we are estimating; this is known as the Bonferroni correction. The Bonferroni SCIs are calculated with limits

$$SCI = \hat{\boldsymbol{\theta}} \pm z^* \sqrt{\text{var}(\hat{\boldsymbol{\theta}})} \ , \tag{3.20}$$

where $\hat{\boldsymbol{\theta}}$ represents the vector of estimated threshold components (for both BP and BN, the dimensions of this vector are 2×1).

This Bonferroni correction adjusts for the multiplicity in the interval estimation. In this dissertation, $p = 2$, as the OOP for each sequential strategy has two components, one for each test. Typically, as p gets larger, the SCIs are more conservative. Widths of these SCIs are calculated by subtracting the lower bound from the upper bound of a particular set. It is worth noting that there are other methods of constructing SCIs such as Scheffé (all possible comparisons) or Tukey (all pairwise comparisons); as we are only doing a small number of *a priori* set comparisons, Bonferroni was the chosen method. Selection of a method depends ultimately on what one is actually interesting in capturing with the SCI.

A better way to describe confidence in a multivariate setting is the use of a confidence region, or a confidence ellipse. Used commonly in multivariate statistics, a parametric $100(1-\alpha)\%$ confidence region for a vector of estimates $\hat{\boldsymbol{\theta}}$ is determined by all $\hat{\boldsymbol{\theta}}$ such that

$$n(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'S^{-1}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \leq F_{p,n-p}(\alpha), \tag{3.21}$$

where n represents the total sample size, S represents the variance-covariance matrix, and F represents the CDF for the F distribution.

Coverage for both the SCIs and CRs is calculated as the percentage of times the true parameter value falls within the respective confidence bounds. The closer your coverage is to your set confidence level the more accurate your estimate is. When the coverage is larger than the set confidence level, the interval/region is considered conservative; likewise, when the coverage is less than the set confidence level, the interval/region is considered permissive (Dodge *et al.*, 2003).

3.7. Estimating a Variance (and Covariance)

Once we obtain estimates for the threshold pairs, or the OOPs, variance estimates are needed to obtain simultaneous confidence intervals, confidence regions, and RMSE estimates. As the BP/BN equations are too complex to obtain a closed solution for the thresholds, the same holds true for calculating a variance estimate.

A common method to estimate variance is the bootstrapping technique. Bootstrapping was first described by Efron & Tibshirani in 1979. This method is commonly used when the distribution of the estimate is unknown. Within each simulated dataset, we will sample B replicates, or bootstrap samples, separately for the H and D groups. This can be done using several sampling techniques. We will implement bootstrapping with replacement using unrestricted random sampling, so some observations may be included more than once, while some observations may not be selected at all.

Within each bootstrap sample, we can calculate the B OOPs, and then calculate a variance for each theta component of the OOP as well as the covariance between the two. Schisterman and Perkins (2007) suggest that this method performs no worse than the delta method for estimation of the optimal threshold when a single test is used, while being similar to or outperforming other method, such as the percentile method or BCa bootstrapping. While this method involves using the mean OOP values across the B samples, we altered it slightly to use the actual simulated OOPs. The percentile method involves taking all B bootstrap sample, calculating the parameter of interest, ranking the results numerically, and finding the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ percentiles; this represents your

confidence intervals. The bias-corrected and accelerated (BCa) bootstrapping method improves the percentile method by correcting for bias and for rate of change of the standard error with the estimated parameter value "with respect to the true parameter value" (Zhou, Obuchowski, & McClish, 2011; pg. 479).

3.8. Assessing Normality

There are several methods available to assess the normality of a random variable. One simple thing is to compare the mean and the median and see how close they are. This is informative because, if the data are normally distributed, the mean should equal the median. The further the data are from this assumption, the further apart these statistics will be. Skewness, or the measure of symmetry,, is defined as

$$\frac{\mathrm{E}\left[(X-\mu)^3\right]}{\left(\mathrm{E}\left[(X-\mu)^2\right]\right)^{3/2}}. \tag{3.22}$$

Kurtosis, or the measure of "peakedness" of a distribution, is defined as

$$\frac{\mathrm{E}\left[(\theta-\mu)^4\right]}{\left(\mathrm{E}\left[(\theta-\mu)^2\right]\right)^2}. \tag{3.23}$$

The closer these values are to zero, the more normal a distribution appears. To view the distribution of a variable, we can construct a histogram. If the data are normally distributed, the histogram should be symmetric about the average, or center. Additionally, a normal quantile plot, which plots the normal quantiles against the the quantiles of the distribution of the variable, should follow the line $y = x$ if the data are normally distributed.

Several formal statistical tests are also available to assess normality. For this research, we will consider the Kolmogrov-Smirnov and Shapiro-Wilk tests for normality. Kolmogrov-Smirnov test statistic is defined as

$$\sup_\theta \left| G_n(\theta) - G(\theta) \right|, \tag{3.24}$$

where G represents the empirical distribution function (EDF) and n represents the number of independent observations (Massey Jr., 1951). The Shapiro-Wilk test statistic is defined as

$$\frac{\left(\sum_{i=1}^{n} a_i \theta_{(i)} \right)^2}{\sum_{i=1}^{n} \left(\theta_i - \bar{\theta} \right)^2}, \tag{3.25}$$

where $\theta_{(i)}$ represents the ordered values from smallest to largest and a_i represents the constants generated from the means, variances, and covariances of the order statistic of an n-sized sample from a normal distribution (Shapiro & Wilk, 1965).

To assess for normality of the theta components of the OOP for a subset of parameters, all measures and methods described in this section will be examined, and an overall conclusion will be made on the normality of the variable.

3.9. Empirical Estimates

The OOP for the sequential testing strategies will also be estimated empirically. This method is based off of the available data, so that the corresponding ROC curve will be more of a step curve than a smooth curve. Assume continuous data, patients are considered negative (healthy) if test 1 and test 2 both yield a negative result for BP.

Creating indicator variables for both tests, the indicator portion of equation 2.1 then becomes

$$\left(\sum_{j=1}^{n_H} I_{\left[T_{Hj1} < c_{i1} \right]} I_{\left[T_{Hj2} < c_{i2} \right]}, \sum_{j=1}^{n_D} I_{\left[T_{Dj1} < c_{i1} \right]} I_{\left[T_{Dj2} < c_{i2} \right]} \right). \qquad (3.26)$$

Expanding these to create pairs to be plotted, it is computationally easier to count those with disease rather than those that are healthy. Counting the number of patients that are positive (with disease), we get

$$\left(\overline{F}_{n_H}(c_{i1}, c_{i2}), \overline{F}_{n_D}(c_{i1}, c_{i2}) \right) =$$
$$\left(\frac{n_H - \sum_{j=1}^{n_H} I_{\left[T_{Hj1} < c_{i1} \right]} I_{\left[T_{Hj2} < c_{i2} \right]}}{n_H}, \frac{n_D - \sum_{j=1}^{n_D} I_{\left[T_{Dj1} < c_{i1} \right]} I_{\left[T_{Dj2} < c_{i2} \right]}}{n_D} \right), \qquad (3.27)$$

which simplifies to

$$\left(\overline{F}_{n_H}(c_{i1}, c_{i2}), \overline{F}_{n_D}(c_{i1}, c_{i2}) \right) =$$
$$\left(1 - \frac{1}{n_H} \sum_{j=1}^{n_H} I_{\left[T_{Hj1} < c_{i1} \right]} I_{\left[T_{Hj2} < c_{i2} \right]}, 1 - \frac{1}{n_D} \sum_{j=1}^{n_D} I_{\left[T_{Dj1} < c_{i1} \right]} I_{\left[T_{Dj2} < c_{i2} \right]} \right). \qquad (3.28)$$

Looking at the BN strategy, patients are considered positive (with disease) only when both tests are positive. To count these, the expanded indicator portion of equation 2.1 becomes

$$\left(\sum_{j=1}^{n_H} I_{\left[T_{Hj1} > c_{i1} \right]} I_{\left[T_{Hj2} > c_{i2} \right]}, \sum_{j=1}^{n_D} I_{\left[T_{Dj1} > c_{i1} \right]} I_{\left[T_{Dj2} > c_{i2} \right]} \right). \qquad (3.29)$$

Expanding these to create pairs to be plotted, we get

$$\left(\overline{F}_{n_H}(c_{i1}, c_{i2}), \overline{F}_{n_D}(c_{i1}, c_{i2})\right) =$$

$$\left(\frac{1}{n_H}\sum_{j=1}^{n_H} I_{\left[T_{Hj1}>c_{i1}\right]} I_{\left[T_{Hj2}>c_{i2}\right]}, \frac{1}{n_D}\sum_{j=1}^{n_D} I_{\left[T_{Dj1}>c_{i1}\right]} I_{\left[T_{Dj2}>c_{i2}\right]}\right). \tag{3.30}$$

For both the BP and BN empirical formulae, the OOP would be selected as it was for the univariate case (i.e. the threshold yielding the highest YI is selected as the OOP).

While no distributional assumptions are placed on the data, this method is obviously very cumbersome. This method does not extrapolate outside of the dataset as a smooth ROC could in finding the OOP. Consequences of this could be over/underestimation of the Se, Sp, or both, as well as inaccuracy of the actual OOP due to a limited range of values available for the threshold calculations and corresponding diagnostic statistics. Additionally, Schisterman & Perkins (2007) mention that when selecting an optimal threshold for the single test case, it is possible that more than one threshold could be selected based on the YI alone. Since this is possible when calculating OOPs from the data available, the OOP with the largest Se was chosen as the optimal threshold.

3.10.   Determining Legitimate BP/BN Strategies

During this research, it became obvious that in solving for the OOP, there were scenarios in which only one test was necessary. For example, the threshold for the second test might be so low that essentially everybody would be classified based on the result of the first test. Due to this, before the simulation study could be performed, valid BN strategies needed to be identified based on the population parameters. The method described above

using PROC NLP was used to determine valid strategies, with starting values being a coarse grid search from -4 to 4 by 0.2. For both BP and BN, a strategy will be considered "legitimate" if the threshold value for test 1 falls within

$(\mu_{1H} - 1.645 \times \sigma_{1H}, \mu_{1D} + 1.645 \times \sigma_{1D})$ and the threshold value for test 2 falls

within $(\mu_{2H} - 1.645 \times \sigma_{2H}, \mu_{2D} + 1.645 \times \sigma_{2D})$. These constraints are based on the normal distribution and identify values that are extreme for that distribution. When allocation of patients across the H and D groups favors D, we would be excluding more people.

3.11.   Determining Valid OOPS

Similar to the manner in which we determined legitimate strategies, bounds were employed to determine valid OOPs. For both BP and BN, an OOP will be considered valid if the threshold value for test 1 falls within $(\mu_{1H} - 1.96 \times \sigma_{1H}, \mu_{1D} + 1.96 \times \sigma_{1D})$ and the threshold value for test 2 falls within $(\mu_{2H} - 1.96 \times \sigma_{2H}, \mu_{2D} + 1.96 \times \sigma_{2D})$. Note that within each legitimate strategy, we were less conservative when specifying bounds for the OOP, allowing more OOPs estimated from simulated data to be accepted.

3.12.   Prediction of Total Sample Size and SCI Widths

A common question in research studies from investigators is "how large of a sample do I need?" For this research, someone who wants to design a study to determine the OOP for a given sequential testing strategy may want guidance for selecting an appropriate sample size. While a derived formula exists for estimating a sample size for the single test case (Skaltsa *et al.*, 2010), nothing is available for the sequential testing strategy scenario. We

do not attempt try to derive an exact formula in this research, however we aim to try and provide guidelines based upon the simulation data.

For each strategy (BP and BN), it was of interest to predict both an overall sample size and the SCI widths for each test using the data from the simulations. SCI widths were considered rather than the CR areas because of the interpretability; for clinicians, widths are more relevant and applicable than areas due to the reduced dimensionality. Because of the abundance of data, we split the data equally into two parts: an estimation data set and a validation data set. The first 500 runs within a specified parameter set were used as the estimation dataset, while the latter 500 runs were assigned to the validation dataset. Recall from Section 3.2 that we have 144 parameter combinations for each testing strategy. With 1,000 runs in each, we have a maximum 72,000 valid observations in both the estimation and validation datasets for both BP and BN.

The first outcome of interest is SCI width, with a separate regression model for each OOP component (i.e. separate models for the width of the SCI for the first and second components for the OOP, which is equivalent to the widths for each test). To predict SCI width for each test and strategy, ordinary least squares (OLS) regression was performed using three different sets of predictor variables:

a) M1: TSS, k, and simulated values for

$$\underline{\gamma} = (\mu_{1H}, \mu_{2H}, \mu_{1D}, \mu_{2D}, \sigma_{1H}, \sigma_{2H}, \sigma_{1D}, \sigma_{2D}, \rho_H, \rho_D);$$

b) M2: TSS, k, and simulated values for AUCs, b, and correlations; and

c) M3: TSS, k, and simulated values for correlations, a's and b's used to calculate the AUC (defined in Chapter 2).

Note that the variables in M2 and M3 are actually just transformations of those in M1. Models including interactions between TSS and k as well as AUC and b were also included where appropriate. Comparisons were made on the basis of R^2 across models; since the sample size was so large, all parameters were significant, so an increase in R^2 was a more reliable measure of whether interactions are important.

Similarly, regression equations for predicting TSS were constructed. Again, OLS regressions were performed in SAS PROC REG with a different prediction equation for each set of predictor variables:

a) M1: SCI width, k, and simulated values for

$$\underline{\gamma} = (\mu_{1H}, \mu_{2H}, \mu_{1D}, \mu_{2D}, \sigma_{1H}, \sigma_{2H}, \sigma_{1D}, \sigma_{2D}, \rho_H, \rho_D);$$

b) M2: SCI width k, and simulated values for AUCs, b, and correlations; and

c) M3: SCI width, k, and simulated values for correlations, a's and b's used to calculate the AUC (defined in Chapter 2).

To further assess the models obtained from the estimation dataset, we used the parameter estimates obtained from the models as regression coefficients for the validation dataset. This technique allows for a generalization of the model to an independent dataset. The coefficient of determination, R^2, was used to determine which model performs best within each set of OLS regressions for the SCI widths, as well as the OLS regressions for the TSS equations.

4 Simulations

This research aims to identify the optimal operating point to distinguish between two populations (healthy (H) and disease (D)) under two sequential testing strategies scenarios. In this chapter, the OOP will be identified and properties will be described in terms of bias, root mean squared error, over/under estimation of the OOP as well as the corresponding Se and Sp, simultaneous confidence interval widths, confidence region area, percent coverage of both the simultaneous confidence intervals and confidence regions, and a comparison to the empirical approach in terms of bias and over/under estimation of the OOP as well as the Se and Sp. The effects of simulation parameters on these measures are examined; we also compare the two strategies.

4.1. Strategy Properties

In this section, some parameters that could affect the estimation of OOP for each of the testing strategies are defined. We assume that the healthy patients follow a standard normal distribution. Upon setting this assumption and selecting varied values for simulation parameters identified in Table 4.1, calculations for additional parameters can be performed, such as the means for the diseased groups for each test as described in chapter 3.

Table 4.1: Population Parameters [1,2,3]

Parameter	Values
b	(0.5, 0.5), (1.0, 1.0), (0.5, 1.0), (1.0, 0.5), (1.5, 1.5), (1.0, 1.5), (1.5, 1.0)
AUC	(0.7, 0.7), (0.75, 0.85), (0.9, 0.9)
ρ	(0.0, 0.0), (0.5, 0.5), (0.25, 0.5), (0.5, 0.25)
r	0.5, 1, 2

[1] $b_i = \dfrac{\sigma_{iH}}{\sigma_{iD}}$

[2] ρ represents the correlation between the tests within a disease classification group

[3] r represents the weighting factor in the GYI

The AUC was selected as follows: (0.7, 0.7) and (0.9, 0.9) were chosen such that both tests had fair/moderate accuracy or high accuracy, and that both tests were equivalent in this accuracy. An AUC of (0.75, 0.85) was chosen such that the second test was slightly more accurate than the first test but so that the differences in the accuracies were not large This parameter is important as it influences the shape of the ROC curve of the individual tests. Formulae to obtain the values for μ_{1D} and μ_{2D} written in terms of the AUC and *b*, are defined in section 3.1.

The ratio of standard deviations, *b*, was selected as this parameter also influences the shape of the ROC curve irrespective of the AUC. Two curves could have an identical AUC but look very different if this ratio varies. The choice of *b* = 1 assumes the spread of the test results for the disease and healthy groups are identical; *b* < 1 assumes that the test results for the disease group has more variability than those for the healthy group; lastly, *b* > 1 assumes that the test results for the disease group has more variability than those for the healthy group. In order to have one value from each group, *b* of 0.5, 1, and 1.5 was

chosen. Figure 4.1 shows three different curves, one for each value of *b*; all three

curves have the same area (AUC = 0.8).

Figure 4.1: ROC Curves and *b*
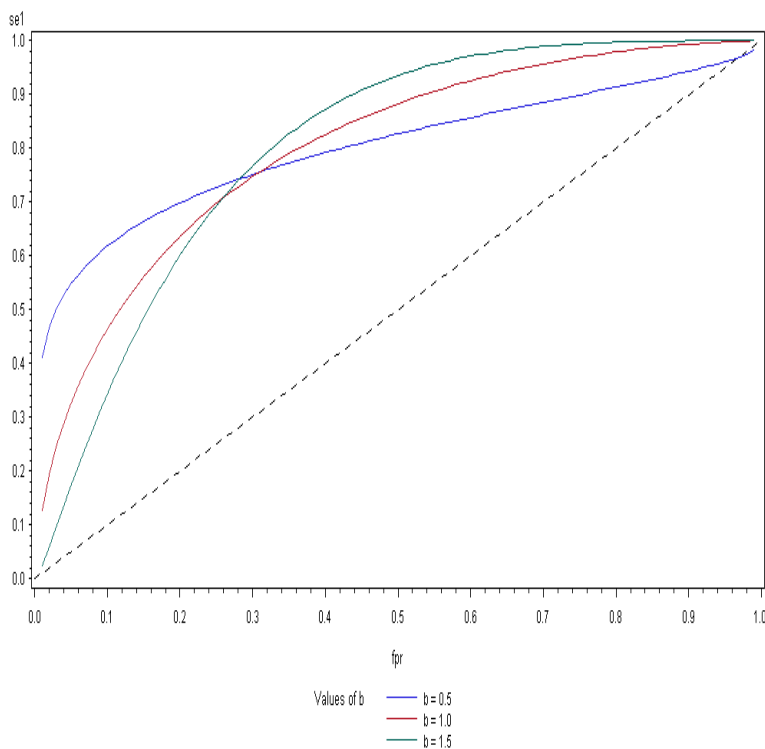
Correlation was varied to either assume independence between the two tests, for

both the healthy and disease groups, by letting $\rho = (0.0, 0.0)$, or allow some correlation

($\rho = (0.25, 0.5)$), ($\rho = (0.5, 0.25)$), or ($\rho = (0.5, 0.5)$)). Additionally, values of *r*, or the

weighting parameter, used in the calculation of the GYI, were varied for the population

parameter tables. The parameter *r* specifies a summary of cost and prevalence to weight

the sum of the TPR and the Sp of the GYI. Values of $r > 1$ result in a higher weight for

specificity versus sensitivity in the GYI, while values of $r < 1$ result in lower weight.

4.2. Population Parameters Results

As mentioned in section 3.8, during this research it became obvious that while solving for the OOP, there were scenarios in which only one test was necessary to diagnose disease. For example, the threshold for the second test might be so low that essentially everybody would be classified based on the result of the first test. This was verified with the use of a grid search, or a brute force method in which OOP was assessed by calculating the GYI for many test values spanning the entire range of possible values. Due to this, before the simulation study could be performed, legitimate strategies needed to be identified based on the population parameters.

4.2.1 Identifying Legitimate Strategies

Consider the BP strategy. Strategies are considered "legitimate" if θ_1 falls within ($\mu_{1H} - 1.645 \times \sigma_{1H}$, $\mu_{1D} + 1.645 \times \sigma_{1D}$) and θ_2 falls within ($\mu_{2H} - 1.645 \times \sigma_{2H}$, $\mu_{2D} + 1.645 \times \sigma_{2D}$). These constraints are based on the normal distribution and identify values that are extreme for that distribution. The same was done for the BN strategy. Recall that the objective is to maximize the GYI, so if a single test performs better than a sequential testing strategy in terms of the GYI, this would be more appropriate. For times where both strategies are legitimate, we assess the values of the GYI and select the strategy that yields the largest GYI.

From the 252 combinations in Table 4.1 for the BP strategy, 181 (72%) were determined valid by the bounds set as mentioned above, and can be seen in Tables 8.1 – 8.3. Illegitimate strategies are shown in italics. Values of *b* that involved 1.5 were,

generally, not appropriate for the BP strategy, except for the largest AUC values. This was true regardless of r or correlation values. Also, illegitimate BP strategies occur more often when the first tests' correlation was less than the second tests'. Regarding r, more problems seem to arise when $r = 2$, but this trend fades as the AUC increases.

Legitimate BN strategies were also examined for 252 combinations of parameters detailed in Table 4.1. These results can be seen in Tables 8.16 – 8.18. Illegitimate strategies are shown in italics. From the 252 combinations in Table 4.1, 172 (68%) were determined valid by the bounds set. Values of b that involved 0.5 were, generally, not appropriate for the BN strategy. This was true regardless of AUC, r, or correlation values. Also, illegitimate BN strategies occur more often when the first tests' correlation is higher or equal to the second tests'. This trend fades as the AUC increases, and appears to only occur for $r = 1$ or $r = 2$.

4.2.2   Comparing BP and BN

BP and BN strategies were compared based on results in Tables 8.1 – 8.3 and 8.16 – 8.18. First, AUC = (0.70, 0.70) was examined. For low values of b (i.e. pairs involving $b = 0.5$), BP is clearly superior to BN in terms of having more legitimate strategies. However, as r increases, the need for sequential tests versus a single test in terms of maximizing the GYI dwindles. For $b = (1.0, 1.0)$, the strategies are approximately equal in terms of the number of legitimate strategies and the GYI. For larger values of b (pairs involving $b = 1.5$), BN is the superior strategy in terms of more legitimate strategies for the specified value. Overall, instances where both strategies are legitimate have an

approximately equivalent GYI. For BP, there were 31 instances where the strategy was not legitimate, while BN had 32. There were three instances in which these overlapped, indicating that the single test was the only relevant strategy. For times when only BP was illegitimate while BN was legitimate, BN yielded an equivalent or slightly higher GYI than using only a single test. The same was true when BN was illegitimate and BP was legitimate, again indicating that the sequential test was better, or at the worst equivalent, to the single test.

Second, AUC = (0.75, 0.85) was examined. Again, low values of b appear to favor BP while high values appear to favor BN, in terms of legitimate strategies. For values of b = (1.0, 1.0), higher values of correlation favor BP, while lower values favor BN. No correlation tended to yield valid strategies for both, with the GYI approximately equal across the two. This is similar to what we found for AUC = (0.70, 0.70). For BP, there were 37 invalid instances for this AUC, while BN had 35. For times when these didn't overlap, the sequential strategy always yielded a higher GYI than a single test. There were four times in which neither BP nor BN was a legitimate strategy; for these combinations, the single test is the only option.

Lastly, AUC = (0.90, 0.90) was examined; BP had six illegitimate strategies (7%) while BN had 13 (15%). For the times where both BP and BN were legitimate strategies (77%), BN performed slightly better for combinations including b = 1.5, while BP was superior for combinations including b = 0.5. For b = (1.0, 1.0), neither BP nor BN had any illegitimate strategies, and were approximately equal in terms of GYI when the correlation was (0.0, 0.0) or (0.50, 0.50). BP saw a slightly higher GYI for

$\rho = (0.25, 0.50)$, while the GYI for BN was slightly higher for $\rho = (0.50, 0.25)$. For the six times BP was in invalid strategy, the GYI associated with BN was always higher, suggesting that the sequential strategy performs better than a single test. Comparing the GYI values even though one strategies is invalid provides insight into whether a single test versus a sequential test strategy is better.

4.2.3   Effect of Population Parameters on Sensitivity, Specificity and GYI

Consider Tables 8.1 – 8.3 for the BP strategy. As r increases, the Se decreases while the Sp increases; the GYI also decreases. We also see that the highest GYI occurs when the two tests are uncorrelated. If one can choose tests to consider for the sequential testing strategy, uncorrelated tests will give a higher GYI regardless of the AUC values. These same observations hold true for the BN strategy and can be seen in Tables 8.16 – 8.18.

4.3. Simulating Data

A subset of parameters was chosen from Table 4.1 to vary during the simulation study. Looking at the population parameter tables (Tables 8.1 – 8.3 and 8.16 – 8.18), we see that b pairs involving 1.5 are not appropriate for BP; likewise for b pairs involving 0.5 for BN. These could be considered the "extreme" cases for the values presented in Table 4.1. In order to make the amount of simulations more manageable, we limited the choice of correlation and r. For correlation, we consider no correlation (0.0, 0.0) or moderate correlation (0.5, 0.5), which can again be thought of as the "extremes" of the four cases presented. Although the value of $r = 1$ will be used in the main part of the simulation

studies, interesting cases of other values of *r* for a subset of the parameters are also considered.

To obtain a variety of total sample sizes (TSS), values of 40, 60, 100, and 200 were used, and allocations (*k*) of the TSS to both the H and D groups were allowed to vary as ½, 1, and 2. Higher values of $k = n_D/n_H$ indicate the number of patients in the healthy group is smaller than the number of patients in the disease group (in this case, a ratio of 1:2) and have better Se, while lower values suggest the opposite (in this case, a ratio of 2:1) and have better Se. A summary of simulation parameters can be seen below in Table 4.2.

Table 4.2: Simulation Parameters

	BP	BN
Simulation Parameters		
b	0.5, 1	1, 1.5
TSS	40, 60, 100, 200	
k	0.5, 1, 2	
AUC	(0.7, 0.7), (0.75, 0.85), (0.9, 0.9)	
ρ	(0.0, 0.0), (0.5, 0.5)	
r	0.5, 1, 2	

All possible combinations of variables from Table 4.2 (with *r* = 1) yields 144 unique parameter combinations, labeled $\psi_1, ..., \psi_{144}$, that we will analyze via the simulation study,. The methods used to simulate the datasets will are described as follows (all analyses were performed using SAS):

- Validity of each of the sequential testing strategies for each set of population parameters was determined;

- For each valid strategy, a true OOP and corresponding Se, Sp, and GYI is calculated from the specified values in Table 4.1;

- For each of the 144 parameter sets $\underline{\psi} = (AUC_1, AUC_2, \rho_H, \rho_D, b_1, b_2, TSS, k)$ shown in Table 4.2, 1,000 datasets for the H and D groups are simulated from a bivariate normal distribution;

- OOPs from the sample are estimated, as well as the corresponding Se, Sp, and GYI;

- For OOPs falling within pre-set specified bounds for the estimation, a variance (and covariance) estimate is calculated by bootstrapping. $B = 1,000$ bootstrap samples for each simulated dataset. Note that the bounds to compare the OOP with are different than the bounds set to determine legitimate strategies (sections 3.10 and 3.11);

- Bias, RMSE, overestimation of the Se and Sp corresponding to the OOP, and 95% SCIs, CRs, and their respective coverages are computed;

- Empirical OOP estimates are calculated from the data using methodology presented in chapter 3, as well as the corresponding Se, Sp, and GYI;

- Bias of the OOP, Se, and Sp and overestimation of the Se and Sp for the empirical estimates are also computed; and

- For a subset of parameters in Table 4.2, the coefficient r will be examined for its effect on the results.

Comparisons across strategies are also made; for valid strategies, the selection of the more appropriate testing strategy will be evaluated on the basis of the GYI. In the case where only BP or BN is a valid strategy for a specified parameter set, the GYI will still be used to determine whether or not a single test is more appropriate than a sequential strategy (i.e., it may be more appropriate at times to perform only a single test rather than a set of sequential tests).

Additionally, prediction equations for SCI widths and TSS are also compared using R^2 both within and across modeling types.

4.4. Believe the Positive (BP) Simulation Results

Simulations were performed for valid BP strategies to estimate the OOP for various scenarios. Note that there are various total sample size, allocation, and AUC combinations used for BP as shown in Table 4.2. In Table 8.4 we see that for each of the 144 combinations, there were times out of 1,000 that the simulation OOPs did not fall within the specified bounds, did not converge, or appeared infinite. Smaller values of b tend to lend themselves to values of estimated OOPs lying within the specified bounds. When the AUC for each test are different from one another, there are more times there was no valid answer as a BP strategy than when the AUC for each test is equal. Additionally, as sample size increases, there are less times there are problems with the method.

Median bias was assessed for both the estimation of the OOPs as well as the corresponding Se and Sp (see Tables 8.5 and 8.7). Note that the values are much higher

for the column $b = 1$, AUC $= (0.75, 0.85)$, and $\rho = (0.5, 0.5)$. This reflects poor estimation, and may indicate that this combination of parameters may have had a true OOP value falling just within the set bounds to be included as a legitimate BP strategy. Although this is true for median bias in the OOP, the same trend is not found when looking at bias in Se and Sp; these measures do not appear to be affected. Median bias in the OOP appears to be reduced when the AUC is increased. Correlation does not appear to play a huge role here, but lower values of b show smaller bias in OOP estimation. This trend appears to hold sometimes as the sample size increases, although the presence of k makes this an inconsistent finding. The allocation parameter appears to be more important for smaller sample sizes as compared to larger sample sizes. This finding appears to hold for both thresholds comprising the OOP. Median bias in Se tends to be reduced as the sample size increases, $b < 1$, and correlation is not present between the two tests. AUC appears to reduce the median bias in Se and Sp as it gets larger, with the exception of when the AUC is different for each test. For median bias in Sp, correlation does not appear to play as large of a role as it does for the median bias in Se. Additionally, median bias in Sp appears to be large in most cases than it is for Se.

Median RMSE was also assessed for the estimated OOPs and the corresponding Se and Sp (see Tables 8.6 and 8.8). Again, as the sample size increased, the median RMSE in the OOPs decreased. Lower b generally yielded a lower median RMSE, and AUC appeared to play more of a role for $b = 0.5$ than $b = 1$. Correlation did not appear to be an important factor, nor did k. However, for smaller AUC, $k = 0.5$ generally led to larger median RMSE.

Median RMSE for Se decreased as sample size increased, correlation was absent between the 2 tests, and AUC increased. The ratio of standard deviations, b, does not appear as important here as it does for other measures. Se appears to have a lower RMSE than Sp when $b = 1$, but the opposite is true when $b = 0.5$. This difference decreases as AUC increases.

Percent of overestimation of Se and Sp (Table 8.9) improves for lower b, as sample size increases. Correlation and k do not appear to have an obvious pattern. Percent overestimation of Se appears to decrease as AUC increases for $b = 1$, but the same cannot be said for Sp. For $b = 0.5$, AUC = (0.75, 0.85) actually yields the best percent overestimation of both Se and Sp, but this finding is not true for $b = 1$.

Since we are producing 95% SCIs and CRs, coverage was expected to be near 95%. Coverage of the confidence intervals and regions (Table 8.10) was not as good as we would have expected, especially for the parameter combination involving AUC = (0.75, 0.85), $\rho = (0.5, 0.5)$, $b = 1$. This may reflect a legitimate strategy that fell just inside the specified bounds. Once data are simulated on an already questionable strategy, simulated values may fall outside of the bounds, corresponding with poor or inaccurate estimation of Se and Sp. Coverage improves as sample size is increased, more so for $b = 1$ than for $b = 0.5$. There appears to be no relationship between coverage and k and correlation between tests, with a slight relationship between AUC and coverage. Lastly, there does not appear to be a difference between coverage of the 95% SCIs versus the 95% CRs; (a formal statistical test needs to be performed to verify this observational

result). This is unexpected, as the CRs are a more appropriate measure of confidence due to the nature of the data.

SCI widths are summarized in Table 8.11. In general, a higher b value yielded a narrower SCI over all other parameters. As sample size increases, widths decrease. Within each value of b, when the AUC for each test was equal, widths of the SCIs decreased as AUCs increased; however, when the AUC for each test was different, the widths were actually wider than they were for scenarios with lower, equivalent AUCs. Correlation played a larger role in scenarios with different AUCs than equivalent AUCs. There is not much difference between the first and second components of the OOP in terms of SCI width. Allocation of $k = 2$ generally yielded the widest SCIs. CR areas are larger when there is correlation between groups, sample size is smaller. For $b = 0.5$, AUC does not matter as much in terms of reducing the CR area as it does when $b = 1$. Along these lines, $k = 2$ produces the widest SCI widths for $b = 0.5$, but this is not always true for $b = 1$. These results are summarized in Table 8.12.

Empirical estimates were compared to estimates derived from the methodology in chapter 3 in terms of median bias of the OOP and corresponding Se, and Sp, as well as percent overestimation of Se and Sp. These are summarized in Tables 8.13 – 8.15. There is no apparent pattern; bias changes from positive to negative, and does not seem to improve with sample size. Additionally, the percentage of overestimation of Se and Sp shows no consistent patterns. The empirical estimates seem very unstable and unpredictable.

4.4.1 Formal Regressions to Determine Relationships

OLS regressions were performed for each measure of interest (bias, RMSE, etc) with the following set as predictor variables: AUCpair, $b$pair, ρpair, TSS, and k. For the multivariate model, only main effects were considered. The variables with "pair" in them refer to the fact that these are technically categorical variables that are simulated based on pairs. The predictor set refers to the simulation parameters from Table 4.2. Note that r is not included in the table below, because these simulations only used $r = 1$. Table 4.3 below shows which variables are significant for each measure of interest; the ● in the box indicates that the variable was a significant predictor at the

$\alpha = 0.05$ level.

Overall, b and TSS were found to be significant for almost every measure, while k was significant the least. AUC and correlation were also fairly significant. Area under the curve was significantly related to all measures except the bias of the OOP estimates and the coverage of the confidence intervals/regions. All measures were related to TSS with the exception of bias of the OOP. The ratio of standard deviations, b, was related to a majority of the measures, as was correlation.

Allocation, k, appeared to have no effect on RMSE of the OOP or SCI widths. Not all relations depicted in Table 4.3 were found using an observational approach, but it was common between the two approaches that k was not as important as the remaining four parameters.

Table 4.3: Formal Regression Analysis to Determine Relationships, BP

Outcome	AUCpair	$b$pair	ρpair	TSS	k
		Predictors			
Median Bias θ_1		●	●		
Median Bias θ_2		●			●
RMSE θ_1		●	●	●	
RMSE θ_2	●	●		●	
Median Bias Se	●	●		●	●
Median Bias Sp	●	●	●	●	●
RMSE Se	●		●	●	●
RMSE Sp	●	●	●	●	
Percent Overestimation Se	●	●	●	●	
Percent Overestimation Sp	●	●	●	●	●
Median SCI Width θ_1	●	●	●	●	
Median SCI Width θ_2	●	●		●	
Median CR Area	●	●	●	●	
Median SCI Coverage		●	●	●	●
Median CR Coverage		●	●	●	●

4.5. Believe the Negative (BN) Simulation Results

Simulations were performed for valid BN strategies to estimate the OOP for various

scenarios under the BN strategy, similar to those for BP. In Table 8.19 we see that for

each of the 144 combinations, there were times out of 1,000 that the simulated OOPs

were not considered valid based on the set bounds. As b gets larger, we have less

instances of this and when the AUC for each test are different from one another, there are

more times there was no valid answer. Additionally, as sample size increases, there are

less times there are problems with the method.

Median bias was assessed for both the estimation of the OOPs as well as the

corresponding Se and Sp (see Tables 8.20 and 8.22). Note that the values are much higher

for the column $b = 1$, AUC = (0.75, 0.85), and ρ = (0.5, 0.5), similar to that for BP.

Although this is true for median bias in the OOP, the same trend is not found when

looking at bias in Se and Sp; these values appear unaffected. Median bias in the OOP appears to be reduced when the AUC is increased, there is no correlation between tests for the H and D groups, and *b* is further in magnitude from zero. This trend appears to hold sometimes as the sample size increases, although the presence of *k* makes this an inconsistent finding. This finding appears to hold for both thresholds comprising the OOP. Median bias in Se tends to be reduced as the sample size increases, *b* is further in magnitude from zero, and there is no correlation between the H and D groups. AUC appears to only play a slight role; this is magnified as *b* is further from zero in magnitude. For median bias in Sp, the bias decreases when the sample size is increased, the AUC is increased, and as *b* increased. Interestingly, for *b* = (1, 1), Sp had a higher median bias when correlation between groups was present, while the opposite is true for *b* = (1.5, 1.5).

Median RMSE was also assessed for the estimated OOPs and the corresponding Se and Sp (see Tables 8.21 and 8.23). Again, as the sample size increased, the median RMSE in the OOPs decreased. Additionally, median RMSE decreases when *b* increases in magnitude from zero, AUC is increased, correlation is not present between tests for either the D and H groups. For *b* = 1.5, *k* = 0.5 yielded the highest median RMSE for both components of the OOP, irrespective of correlation. Conversely, when *b* = 1, *k* = 1 always usually the lowest RMSE for both components of the OOP for lower values of AUC and higher sample sizes. A higher median RMSE for θ_3 as opposed to θ_4 can be attributed to this OOP component lying close to the acceptable bounds for a valid BN strategy.

Median RMSE for Se decreased as sample size increased, as b moved further in magnitude from zero, correlation was absent between the H and D groups, and AUC increased. Se appears to have a higher RMSE than Sp when $b = 1$, but the opposite is true when $b = 1.5$. This differences decreases as AUC increases. When $b = 1$, $k = 0.5$ led to a higher RMSE in both Se and Sp, but this pattern does not always hold for $b = 1.5$. For the larger b, the relationship appears to be entangled with sample size, correlation, and AUC. Similar patterns were found for the percent of overestimation of Se and Sp (see Table 8.24).

Again, coverage of the confidence intervals and regions (Table 8.25) was not as good as we would have expected. Coverage improves as sample size is increased and when $b = 1.5$ rather than $b = 1$. There appears to be no relationship between coverage and k, AUC, and correlation between tests. Lastly, there does not appear to be a difference between coverage of the 95% SCIs versus the 95% CRs; a formal statistical test needs to be performed to verify this observational result). This is unexpected, as the CRs are a more appropriate measure of confidence due to the nature of the data.

SCI widths are summarized in Table 8.26. In general, a higher b value yielded a narrower width over all other parameters. As sample size increases, widths decrease, and when correlation is introduced, the widths become wider. Within each value of b, when the AUC for each test was equal, widths of the SCIs decreased as AUCs increased; however, when the AUC for each test was different, the widths were actually wider than they were for scenarios with lower, equivalent AUCs. When $b = 1.5$, the SCI widths for $k = ½$ are larger than for the other values of k. Widths for the threshold associated with the

second test appear to be much narrower than the width associated for the first test when the AUC for each test is different. CR areas are smaller when there is no correlation between groups, the AUC is larger or the AUC between tests is different, sample size increases, there is equal allocation between groups, and $b = 1.5$ rather than $b = 1$. This is summarized in Table 8.27.

Empirical estimates were compared to estimates derived from the methodology in chapter 3 in terms of median bias of the OOP and corresponding Se, and Sp, as well as percent overestimation of Se and Sp. These are summarized in Tables 8.28 – 8.30. As with BP, there is no apparent pattern, and the empirical estimates seem very unstable and unpredictable.

4.5.1 Formal Regressions to Determine Relationships

As was done for the BP strategy, OLS regressions were performed for each measure of interest with the following set as predictor variables: AUCpair, $b$pair, ρpair, TSS, and k. For the multivariate model, only main effects were considered. Table 4.4 shown on the next page depict which variables are significant for each measure of interest; the ● in the box indicates that the variable was a significant predictor at the $\alpha = 0.05$ level.

Overall, b was very important as it was significant for every measure calculated; TSS was significant for every measure except median bias for the first component of the BN OOP. AUC and correlation appear to be important predictors, while k is important only in about 50% of the measures of interest. Area under the curve appears unrelated to bias of the OOP as well as coverage. TSS and b are related to all measures, while

correlation is unrelated to percentage of overestimation of Se and Sp and k is unrelated to RMSE of the OOPs and SCI widths. Again, it was common between the observational approach and the formal approach that k was not as important of a parameter as AUC, b, correlation, and TSS; otherwise, not all relationships discovered using formal regressions were seen using the observational approach.

If we compare Tables 4.3 and 4.4, we see similar patterns of significance for the parameters AUC, b, TSS, k, and correlation across BP and BN.

Table 4.4: Formal Regression Analysis to Determine Relationships, BN

	Predictors				
Outcome	AUCpair	*b*pair	ρpair	TSS	*k*
Median Bias θ_3		●	●		
Median Bias θ_4		●	●	●	●
RMSE θ_3	●	●	●	●	
RMSE θ_4	●	●		●	
Median Bias Se	●	●		●	●
Median Bias Sp	●	●	●	●	●
RMSE Se	●	●	●	●	
RMSE Sp	●	●	●	●	●
Percent Overestimation Se	●	●		●	●
Percent Overestimation Sp	●	●		●	●
Median SCI Width θ_3	●	●	●	●	
Median SCI Width θ_4	●	●		●	
Median CR Area	●	●	●	●	
Median SCI Coverage		●	●	●	●
Median CR Coverage		●	●	●	●

4.6. Assessing the Normality of the OOPs for a Subset of Simulation Sets

Normality of the theta components of the OOPs was assessed using all the methods described in section 3.8. This is important for the measures we are calculating such as the SCIs, CRs, and coverage, as these are based on the normal distribution. Although the simulated data follows a normal distribution, the OOPs estimated from this data may not

follow the same distribution. A random subset of $n = 14$ was selected from the 144

simulation parameters using PROC SURVEYSELECT in SAS 9.2. This represented

approximately 10% of the parameter sets from the simulation study. Table 4.2 contains

the parameters of the sets chosen.

The parameter sets as well as the results of the normality assessment can be seen

in Tables 8.31 – 8.32. For the BP strategy, almost all selected parameter sets appeared to

satisfy the normality assumption, while BN had more that did not satisfy the assumption.

Figures 4.2 – 4.4 below are three examples of the histograms and QQPlots: one that

obviously satisfies the normality assumption, one that obviously does not satisfy the

normality assumption, and one that is a judgment call when combined with the other

measures in Tables 8.31 – 8.32.

Figure 4.2: BP, Normally Distributed
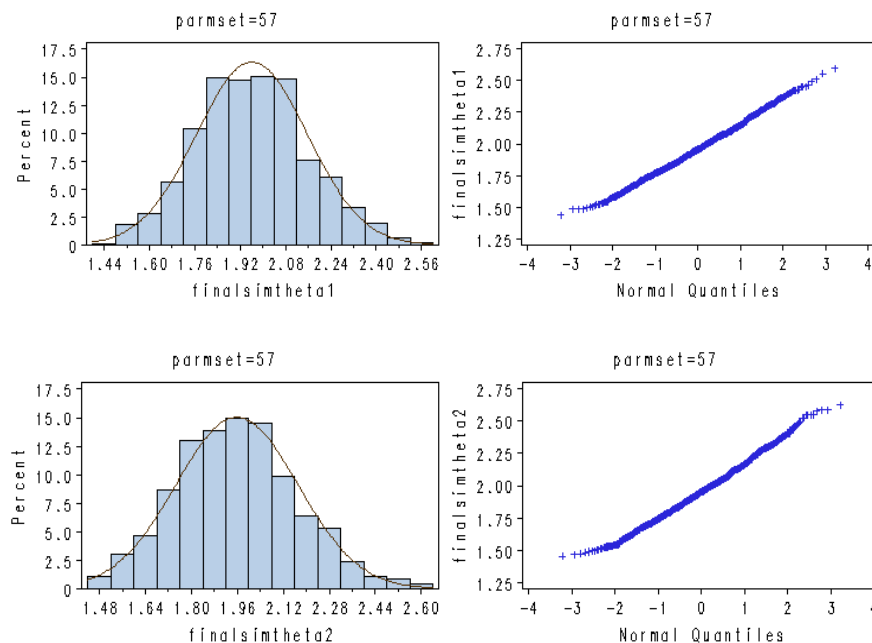
Figure 4.3: BN, Not Normally Distributed
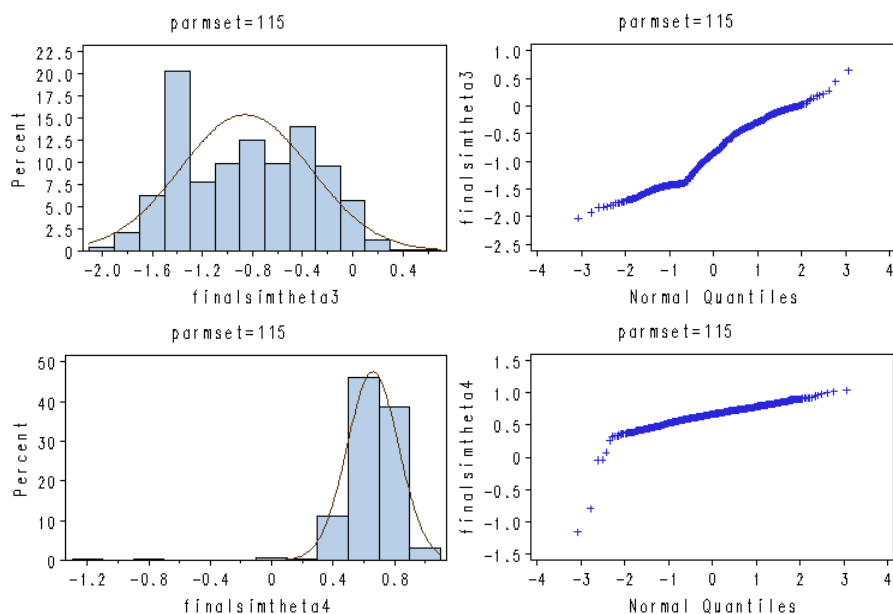
Figure 4.4: BN, Normality Assumption Questionable
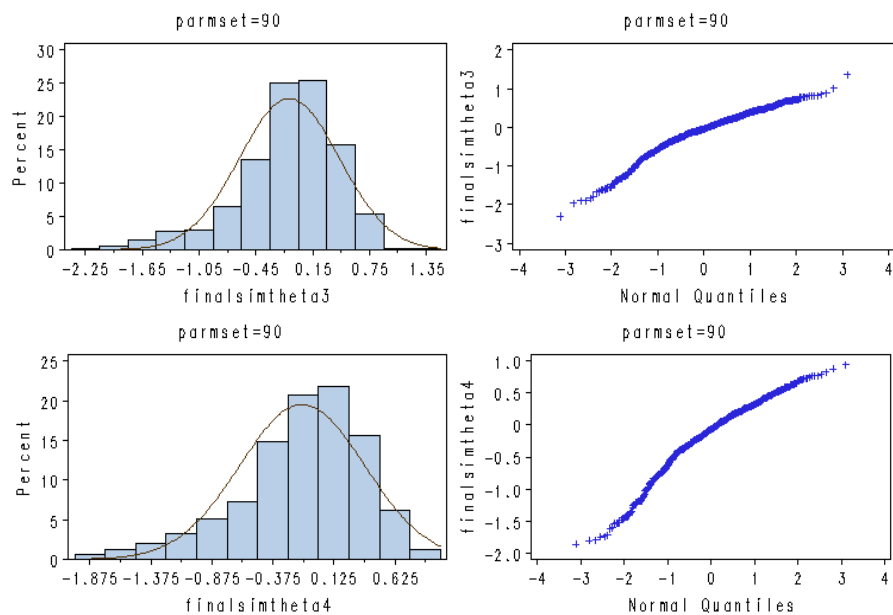
4.7. Varying *r* for a Subset of Simulation Parameter Sets

To assess the effect of *r*, or the weight in the GYI formula (function of prevalence and/or cost of the disease), a subset of parameters from Table 4.2 was looked at. Setting $\rho = (0.0, 0.0)$, TSS = 200, and $k = 1$, values of AUC and *b* were varied over all choices, and *r* was varied as 0.5 and 2. This yields 12 parameter combinations for each sequential testing strategy (values of $r = 1$ are included in simulation tables for comparisons). All calculations looked at for the BP and BN strategies above were assessed for these cases with the exception of the empirical estimates, and are summarized in Tables 8.33 and 8.34.

For BP, the number of times there are invalid OOP estimates remains the same or improves as compared to $r = 1$ regardless of which of the two values of $r \neq 1$. Compared to $r = 1$, bias of the OOP is only decreased for AUC = (0.9, 0.9) for both values of $r \neq 1$, RMSE is only improved for AUC = (0.7, 0.7) for $r = 0.5$, $b = 0.5$ but improves for AUC= (0.7, 0.7) and AUC = (0.75, 0.85) regardless of *r* for $b = 1$. Bias and RMSE of Se and Sp worsens for $r = 2$, but improves for $r = 0.5$ (especially as AUC increases for RMSE). Percent overestimation of Se and Sp improves for $r = 2$ but not for $r = 0.5$; this balances out as AUC increases for the latter and coverage improves slightly for $r = 0.5$ but not $r = 2$. Both SCI widths and CR areas decrease as AUC increases for $r = 0.5$, $b = 0.5$ as compared to $r = 1$ while the opposite is true for $b = 1$; $r = 2$ improves only for $b = 1$.

For BN, the number of times no convergence was improved for $r = 2$ but worsened for $r = 0.5$ for $b = 1$, and remained unaffected for $b = 1.5$. For $b = 1$, $r = 0.5$

improves the bias of the OOP for lower AUC values, while $b = 1.5$ performed best when $r = 1$. RMSE improved for $b = 1$, $r = 0.5$, but worsened for $r = 2$; $b = 1.5$ again performed best for $r = 1$. For $b = 1$, median bias in Se improves for lower AUC values and levels out, while bias in Sp worsens for lower AUC values and then levels out for $r = 0.5$; for $r = 2$, Sp improves while Se worsens. For $b = 1.5$, Se remains similar while Sp improves for $r = 0.5$ and worsens for $r = 2$; RMSE of Se and Sp is similar. Percent overestimation of Se and Sp appears to remain somewhat stable. SCI widths improve for smaller AUC values for $r = 0.5$ and get larger for $r = 2$ for $b = 1$, but are always larger for all AUC values regardless of r for $b = 1.5$. CR areas are always smaller for $r = 1$.

4.8. Predicting SCI Widths and Total Sample Size

A researcher interested in conducting a study to estimate an OOP for a given set of tests might want to design the study such that the OOP is estimated with a certain level of precision. This could include calculating a sample size to use given a certain confidence interval width, or estimating a confidence interval width given a specified available sample size. SCI width, rather than CR area, was used because this value is more attainable and intuitive to a researcher.

With 144 parameter combination sets and 1,000 runs per set, the maximum final estimation count for both BP and BN is 144,000 observations each. Using OLS regression to predict SCI widths/TSS, only valid estimations, or those falling within the specified bounds, will be considered in the regression estimation. Because the first 500 runs for each parameter set are allocated to the estimation dataset and the second 500

runs allocated to the validation dataset for each strategy, we expect the number of invalid observations across datasets to be approximately equivalent. Additionally, splitting the overall dataset in half allows for a maximum of 72,000 useable observations in each dataset. For BP, the estimation dataset yielded 64,684, while the validation dataset yielded 64,720. For BN, the estimation dataset yielded 64,459, while the validation dataset yielded 64,555. A subset of prediction models for SCI Widths and TSS can be seen in Appendix C.

First we will look at the main effects models. OLS regression was performed with the SCI widths as the outcome measure, using PROC REG in SAS. All variables were treated as continuous. To compare the different models available, R^2 was used, where a higher value means that the model is a better fit (i.e., more variance is accounted for). These results are presented in Tables 4.5. M2 is the model most researchers will have the information to carry out. Unfortunately, it is not the best model in terms of R^2. In general, it appears that M1 is the best model. For BP, Table 4.5 shows that M2 performs slightly worse than M1 or M3. With BN, M2 and M3 appear to perform very similarly in terms of R^2. Note that the values of R^2 are in the mid-range ($R^2 = 0.57 - 0.75$).

When the outcome measure was TSS for the main effects models, OLS regression was performed using PROC REG in SAS; results can be seen in Table 4.6. The OLS regression was performed as it was for the SCI widths. R^2 was used to compare models within each strategy for the OLS regressions. When predicting TSS, M1 had the best performance in terms of R^2, although it was only marginally superior to the other models for the BP strategy; it was clearly superior for BN. M2 was the worst performing model

for BP and about the same as M3 for BN. No model accounted for more than 50% of the variance ($R^2 = 0.34 - 0.46$), and all the validation dataset values were higher than the estimation values, meaning the model translates well to other datasets.

Table 4.5: SCI Widths – OLS Regression, R^2

	R-Square			
	BP		BN	
	Estimation	Validation	Estimation	Validation
NO INTERACTIONS				
Outcome = SCI Width for θ_1 or θ_3				
M1: TSS, k, $\hat{\gamma}$	0.6441	0.6421	0.7530	0.7535
M2: TSS, k, $\hat{AUC}_1, \hat{AUC}_2, \hat{b}_1, \hat{b}_2, \hat{p}_H, \hat{p}_D$	0.5789	0.5750	0.6561	0.6576
M3: TSS, k, $\hat{a}_1, \hat{a}_2, \hat{b}_1, \hat{b}_2, \hat{p}_H, \hat{p}_D$	0.6074	0.6041	0.6549	0.6573
Outcome = SCI Width for θ_2 or θ_4				
M1: TSS, k, $\hat{\gamma}$	0.6117	0.6156	0.7408	0.7414
M2: TSS, k, $\hat{AUC}_1, \hat{AUC}_2, \hat{b}_1, \hat{b}_2, \hat{p}_H, \hat{p}_D$	0.5797	0.5833	0.6569	0.6571
M3: TSS, k, $\hat{a}_1, \hat{a}_2, \hat{b}_1, \hat{b}_2, \hat{p}_H, \hat{p}_D$	0.6179	0.6197	0.6406	0.6412
WITH INTERACTIONS				
Outcome = SCI Width for θ_1 or θ_3				
M1 + TSS*k	0.6441	0.6421	0.7531	0.7537
M2 + TSS*k	0.5789	0.5750	0.6561	0.6577
M2 + AUC*b	0.5961	0.5915	0.6640	0.6648
M2 + TSS*k + AUC*b	0.5961	0.5915	0.6640	0.6649
M3 + TSS*k	0.6074	0.6041	0.6549	0.6573
Outcome = SCI Width for θ_2 or θ_4				
M1 + TSS*k	0.6117	0.6156	0.7409	0.7416
M2 + TSS*k	0.5797	0.5833	0.6570	0.6571
M2 + AUC*b	0.6138	0.6126	0.6749	0.6752
M2 + TSS*k + AUC*b	0.6138	0.6126	0.6749	0.6753
M3 + TSS*k	0.6179	0.6197	0.6406	0.6412

Where appropriate (for predicting both SCI Widths and TSS), interactions between AUC and b and/or TSS and k were included; the results can be seen in Tables 4.5 and 4.6. These were estimated in SAS PROC GENMOD and PROC GLM. Because sample sizes were so high, rather than looking at the significance of individual parameters, R^2 was examined for an increase. For predicting SCI widths, the addition of TSS*k alone did not improve R^2; AUC*b appears to slightly improve the model (by two hundredths), but appears to perform equally as well as adding both the TSS*k and the AUC*b interactions. This was true for both BP and BN. For predicting TSS, only the addition of AUC*b was appropriate for M2 since this model includes the AUC; the addition of this interaction appears to improve the R^2 of the model for both BP and BN. Since M2 will most likely be the model used by researches due to the information necessary, the inclusion of this interaction term may be important since it explains more variance.

Table 4.6: Total Sample Size – OLS Regression, R^2

	R-Square			
	BP		**BN**	
	Estimation	**Validation**	**Estimation**	**Validation**
NO INTERACTIONS				
Outcome = TSS				
M1: SCI_wTest1, k, $\hat{\gamma}$	0.4155	0.4604	0.4083	0.4548
M1: SCI_wTest2, k, $\hat{\gamma}$	0.3940	0.4453	0.3969	0.4471
M2: SCI_wTest1, k, $\hat{AUC_1}, \hat{AUC_2}, \hat{b}_1, \hat{b}_2, \hat{p}_H, \hat{p}_D$	0.3593	0.4092	0.3434	0.3989
M2: SCI_wTest2, k, $\hat{AUC_1}, \hat{AUC_2}, \hat{b}_1, \hat{b}_2, \hat{p}_H, \hat{p}_D$	0.3577	0.4134	0.3397	0.3976
M3: SCI_wTest1, k, $\hat{a}_1, \hat{a}_2, \hat{b}_1, \hat{b}_2, \hat{p}_H, \hat{p}_D$	0.3783	0.4282	0.3408	0.3980
M3: SCI_wTest2, k, $\hat{a}_1, \hat{a}_2, \hat{b}_1, \hat{b}_2, \hat{p}_H, \hat{p}_D$	0.3855	0.4362	0.3369	0.3939
WITH INTERACTIONS				
Predictor of Width θ_1 or θ_3 M2 + AUC*b	0.3748	0.4237	0.3544	0.4088
Predictor of Width θ_2 or θ_4 M2 + AUC*b	0.3808	0.4298	0.3538	0.4110

5   Application

This chapter will apply the methodology developed in chapter 3 to two sets of collected data. Each data application section will provide a brief introduction, descriptive statistics on the original distributions, and comparisons between the two groups. Descriptive statistics will include N per group, means, standard deviations, medians, minimums, maximums, as well as graphics; comparisons will be made using two-sample t-tests or Wilcoxon non-parametric tests. As assumed for the methodology developed in this research, normality of the groups will be assessed using measures described in chapter 3, and if necessary, transformations will be applied. These transformations will be determined using the Box-Cox method. Using means and standard deviations of each group, as well as the correlation between tests within each group, the OOP will be estimated, using the YI (i.e., assuming $r = 1$) as well as the corresponding Se and Sp. A variance estimate is estimated using bootstrapping and 95% SCIs and CRs are calculated.

5.1. Wieand Cancer Example

Data were taken from Wieand *et al.*, 1989. According to the authors, a case-control study was conducted at the Mayo Clinic in Rochester, MN, in which blood serum was taken from 141 patients (51 controls with pancreatitis but without pancreatic cancer, 90 cases with pancreatic cancer) to study two antigens: CA125, a cancer antigen, and CA19-9, a carbohydrate antigen. These antigens were measured using an immunoradiometric assay (Bast *et al.*, 1983; Del Villano *et al.*, 1983). A dichotomous indicator variable specifying

the gold standard diagnosis of pancreatic cancer (1 if cancer is present, 0 else) was also included in the dataset. It was not indicated how the diagnosis/gold standard was determined.

For the purpose of the sequential testing strategies that this dissertation describes, the markers must be ordered as a first and second test. Figures 5.1 depicts the empirical ROC curves for each test. AUC can be seen in Table 5.1.

Figure 5.1: Empirical ROC Curves of Tests
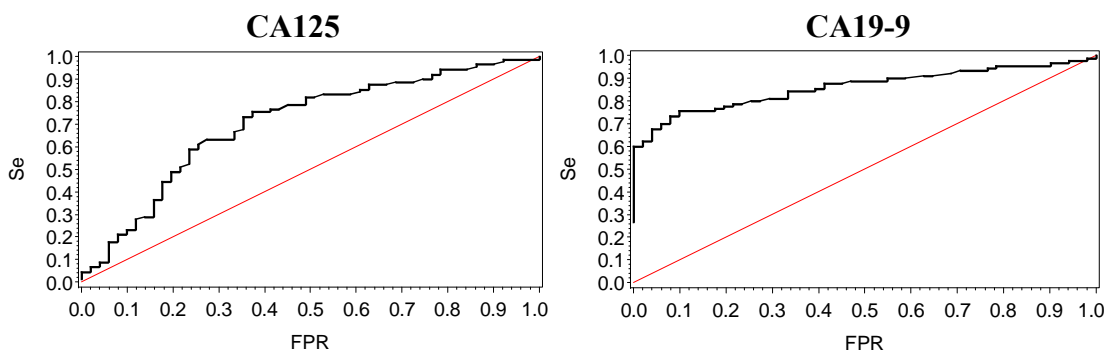
Because the AUC for CA125 is smaller, indicating a less accurate test, than CA19-9, CA125 will be considered Test 1 (X_1); CA19-9 will then be Test 2 (X_2).

Descriptive statistics of the entire sample can be found below in Table 5.1. Table 5.2 includes the descriptive statistics by cancer classification group. As can be seen, the mean for the CA19-9 is much larger than for CA125. For the non-cancer group, CA19-9 ranged from 3.4 to 107.9 U/ml and CA125 ranged from 5.5 to 179.0 U/ml; those with pancreatic cancer had ranges of 2.4 to 24,000.0 U/ml for CA19-9 and 3.7 to 1,024.0 U/ml for CA125. Overall, the correlation between the two variables is 0.23. By group, it is -0.13 for those without pancreatic cancer, and 0.20 for those with pancreatic cancer. There were no missing values in this data.

Table 5.1: Wieand Cancer Sample Descriptive Statistics[1], Overall ($N = 141$)

	Untransformed		Transformed			AUC [2]	
	Mean	SD	Mean	SD	b	Non-Parametric	Parametric
Overall							
CA125	43.0	113.30	1.5	0.19	0.94	0.70	0.70
CA19-9	1,101.5	3,047.00	2.4	0.79	0.62	0.86	0.88

[1] U/ml

[2] Non-parametric AUC is the same for both the untransformed and transformed data, while Parametric AUC is only appropriate for transformed (normally distributed) data

Table 5.2: Cancer Sample Descriptive Statistics[1], by Group
($N_D = 90$, $N_H = 51$, $k = 1.8$ [2])

	Untransformed				Transformed			
	Mean	SD	Median	ρ	Mean	SD	Median	ρ
Healthy				-0.13				-0.13
CA125	21.8	30.30	11.4		1.4	0.17	1.4	
CA19-9	18.0	20.82	10.0		1.8	0.44	1.8	
Disease				0.20				0.04
CA125	55.0	138.83	21.8		1.6	0.18	1.6	
CA19-9	1,715.4	3,681.47	249.3		2.8	0.71	3.0	

[1] U/ml , [2] $k = N_D/N_H$, ρ = Pearson Correlation

From Table 5.2 we see that, as expected, patients that are diagnosed with pancreatic cancer tend to have higher values of CA19-9 than those with just pancreatitis; the same pattern holds for CA125. A Wilcoxon rank sum test verified significant differences in both CA125 ($Z = -4.05$, p-value < 0.001) and CA19-9 ($Z = -7.12$, p-value < 0.0001) between those with and without pancreatic cancer.

Before the methodology of selecting an OOP can be applied, normality of the data needs to be assessed. As can be seen from Table 5.2, the means and the medians are not very similar for either group. Figure 5.2 can be used to assess the normality via histograms for each test within each group, with the top row representing the healthy

group and the bottom row representing disease. The same setup is employed for the

QQPlots in Figure 5.3.

Figure 5.2: Histograms of CA125 and CA19-9 by Cancer Group

Controls

Cases
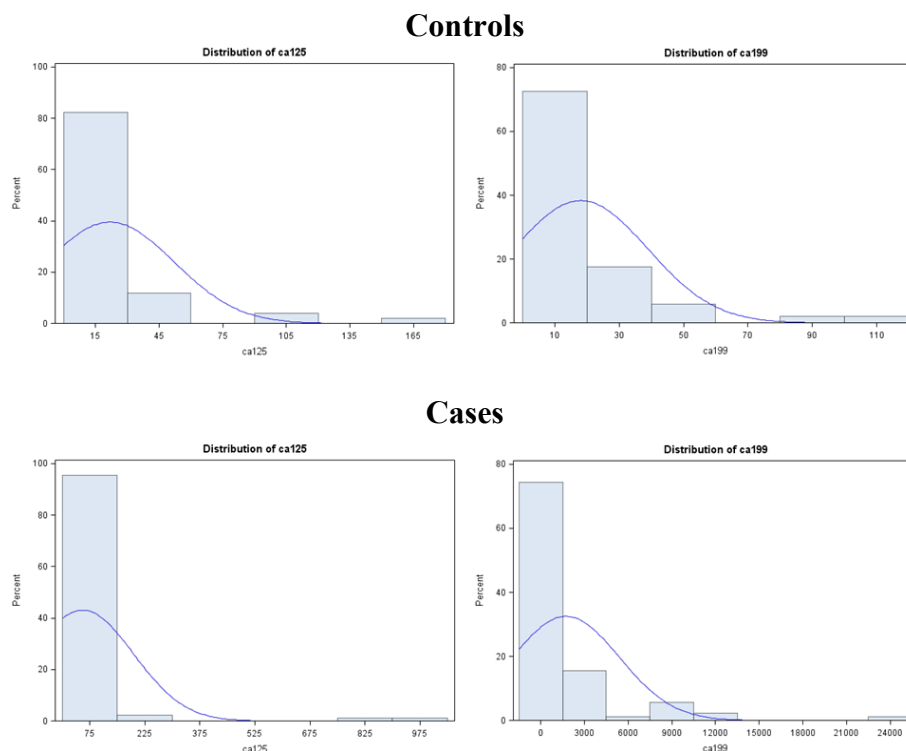
These histograms further demonstrated that those with pancreatic cancer tend to have

higher values of CA19-9 as well as CA125 than non-cancerous patients. None of the four

plots appear to follow a normal distribution, and this is confirmed by all Shapiro-Wilk

tests having p-values smaller than 0.0001 and all Kolmogorov-Smirnov tests having p-

values less than 0.0100. Additionally, the skewness and kurtosis for CA125 and CA19-9

within each group also deviate from zero, representing deviation from symmetry and

even-tailedness. For the healthy group, the skewness for CA19-9 was 2.69 and the

kurtosis was 8.17, and the skewness for CA125 was 3.70 with a kurtosis of 15.58. For the

cancer group, CA125 had a skewness of 6.03 and kurtosis of 37.98, while CA19-9 had a

skewness of 3.59 and a kurtosis of 15.83. The QQPlots below also show obvious

deviation from normality. The blue line in each picture represents the estimated normal

distribution for that data. If the data were normally distributed, the data points would fall

along this line.

Figure 5.3: QQPlots of CA125 and CA19-9 by Cancer Group

Controls
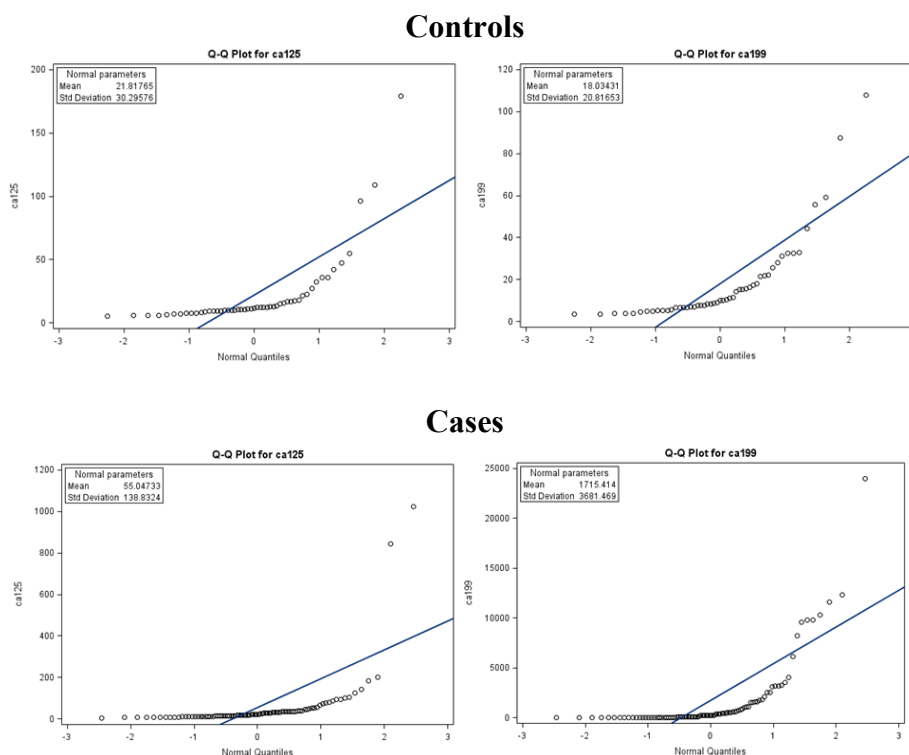
Cases

To transform the data to normality, a Box Cox transformation was employed.

Using the formula

$$X_j^{TRANS} = \begin{cases} \dfrac{X_j^{\lambda_j} - 1}{\lambda_j}, \lambda_j \neq 0 \\ \log(X_j), \lambda_j = 0 \end{cases}, \tag{5.1}$$

where j represents the test number, transformation parameters λ_j were selected using the

best possible option from PROC TRANSREG in SAS 9.2. The transformation parameter

for CA125 was $\lambda_1 = -0.5$, while the transformation parameter for CA19-9 was $\lambda_2 = -0.25$. After transformation, the data appeared to satisfy the normality assumption (see Figures 5.4 and 5.5); the tests will now be referred to as TCA125 and TCA19-9.

Figure 5.4: Histograms of Transformed CA125 and CA19-9 by Cancer Group
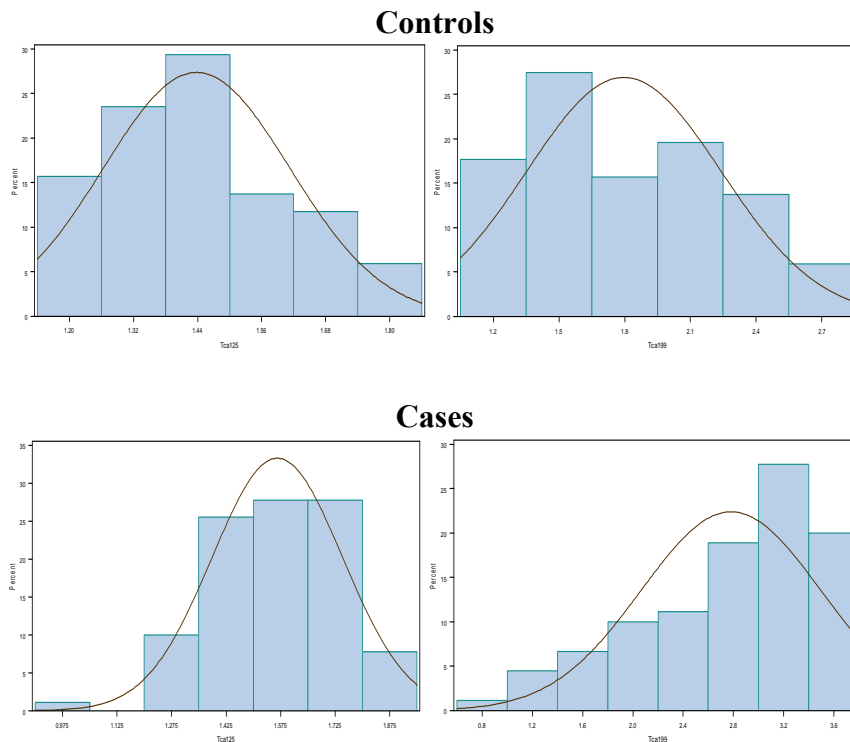
The means and standard deviations for these transformed variables by group can be found in Table 5.2. Medians for TCA125 were 1.4 and 1.6 U/ml for the healthy and cancer groups respectively; they were 1.8 and 3.0 U/ml respectively for TCA19-9. For the healthy group, TCA125 had a skewness of 0.53 and kurtosis of -0.24 while TCA19-9 had a skewness of 0.30 and a kurtosis of -0.78. For the cancer group, TCA125 had a skewness of -0.41 and a kurtosis of 0.37, while TCA19-9 had a skewness of -0.91 and a kurtosis of -0.07. All combinations meet the *p*-value requirements of the Shapiro-Wilk and Kolmogorov-Smirnov tests for normality except TCA19-9 for the cancer group.

Given all information together, the healthy group appears to follow a normal

distribution while the disease group is questionable for TCA19-9; for purposes of this

example we will assume it meets the requirement.

Figure 5.5: QQPlots of Tranformed CA125 and CA19-9 by Cancer Group
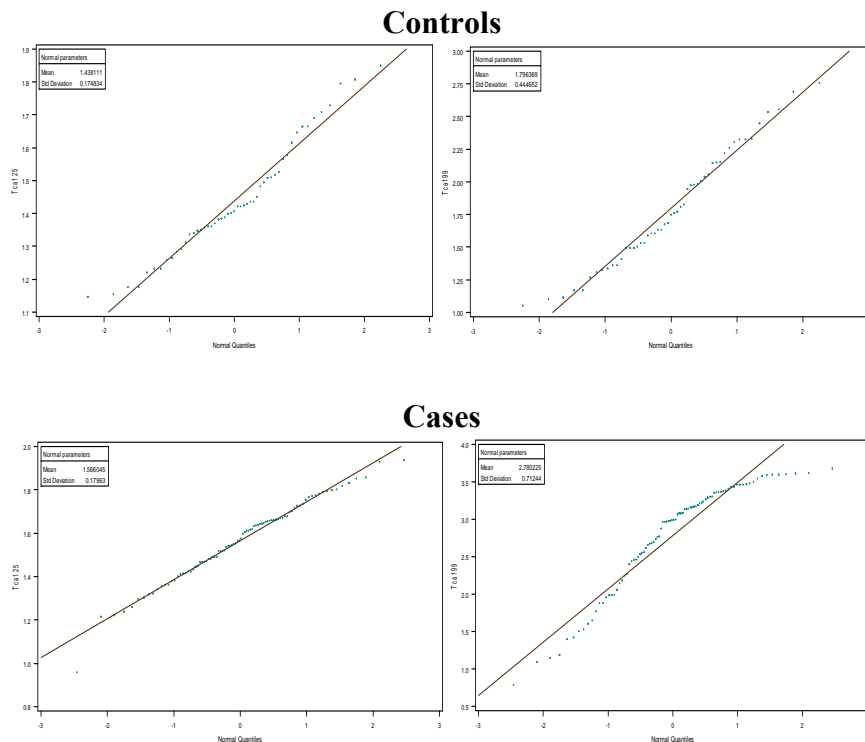
Figure 5.6 shows the scatterplot of the transformed data. Note that a value of

cancer = 0 represents pancreatitis only, while cancer = 1 represents pancreatic cancer.

While we see some separation between groups, there is still overlap between the cases

and controls with regard to both tests.

Figure 5.6: Scatterplot of Transformed CA125 and CA19-9, by Cancer Group
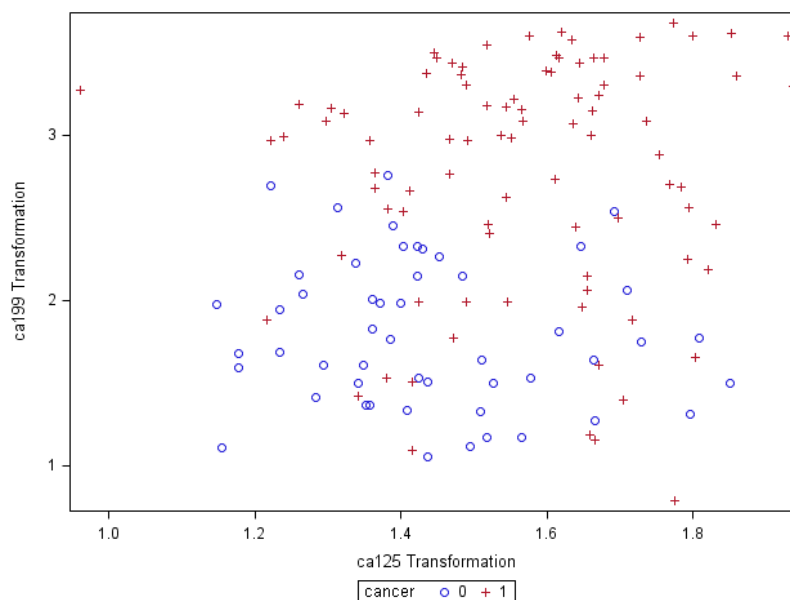
Now that the data have been transformed to normality, OOPs can be estimated assuming a BP and BN sequential testing strategy. Along with these estimates, 95% SCIs and CRs are also estimated. These estimates are compared to 1) the univariate threshold for each test using equation 2.17 and 2) an empirical estimate using equations 3.28 and 3.30. Since both TCA125 and TCA19-9 were transformed from their original units, the results are not directly interpretable and will need to be back-transformed to interpret the results in the original units. The equation

$$X_j = \left(X_j^{TRANS} \lambda_j + 1 \right)^{\frac{1}{\lambda_j}} \tag{5.2}$$

was used to back-transform the data, where j represents the test number.

Table 5.3 contains the results for both univariate tests and sequential testing OOP estimates. These results are shown in the back-transformed, original units for clinical interpretation. These results show that if we were to use both CA125 and CA19-9 as

univariate diagnostic tests, the optimal thresholds would be 16.51 U/ml and 31.85

U/ml respectively. However, using a BP strategy, the thresholds for CA125 and CA19-9

are 111.73 U/ml and 33.51 U/ml. For BN, the thresholds for CA125 and CA19-9 are 5.37

U/ml and 30.85 U/ml. The CIs for the univariate test setting and the SCIs for the

sequential test setting are also shown in Table 5.3. Note that since we are selecting the

threshold with the largest YI, there is a possibility that more than one data point could

yield the same value, thus allowing for more than one optimal empirical estimate (seen in

Table 5.3). We see that the OOPs using the methodology are vastly different from those

calculated empirically, especially for the sequential testing strategies.

Table 5.3: Wieand Cancer Example, OOPs and 95% CIs/SCIs
for each OOP Component (U/ml) [1]

	Single Test		Sequential Test Strategies	
	CA125	**CA19-9**	**(CA125, CA19-9)**	**(CA125, CA19-9)**
			BP	**BN**
Empirical				
OOP			(79.1, 2.4)	
	12.6	32.9	(3.7, 900.0)	(5.6, 3.65)
			(11.7, 3.6)	
Normality Assumption				
OOP	16.51	31.85	(111.73, 33.51)	(5.37, 30.85)
95% CI/SCI			(33.68, 3,499.96)	(4.07, 7.43)
	(13.49, 20.68)	(25.20, 40.84)	&	&
			(23.45, 49.57)	(23.58, 41.15)
Se / Sp / GYI	0.63 / 0.66 / 0.28	0.74 / 0.88 / 0.62	0.75 / 0.87 / 0.63	0.74 / 0.88 / 0.62

[1] In the Original (Back-Transformed) Units

The 95% CRs are shown in Figures 5.7 and 5.8. The first two show the

transformed units for each strategy while the second two show the original, more

appropriate units. It is interesting to note that although the transformed units are elliptical

in shape, once each point comprising the circle is back transformed to original units, the

CR loses the elliptical shape, which applies to normal distributed data. The triangle in these figures represents the OOP for each strategy.

Figure 5.7: BP CR Ellipse, Transformed and Original Units Respectively
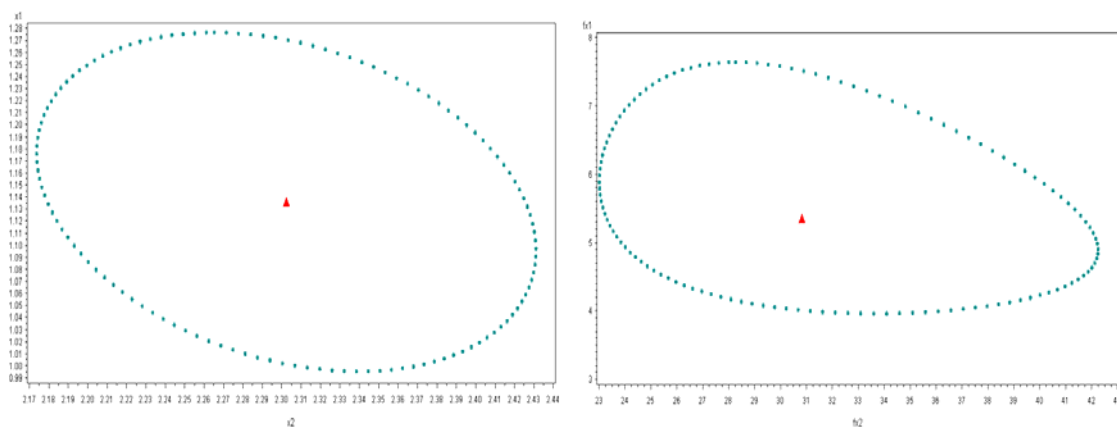
Figure 5.8: BN CR Ellipse, Transformed and Original Units Respectively
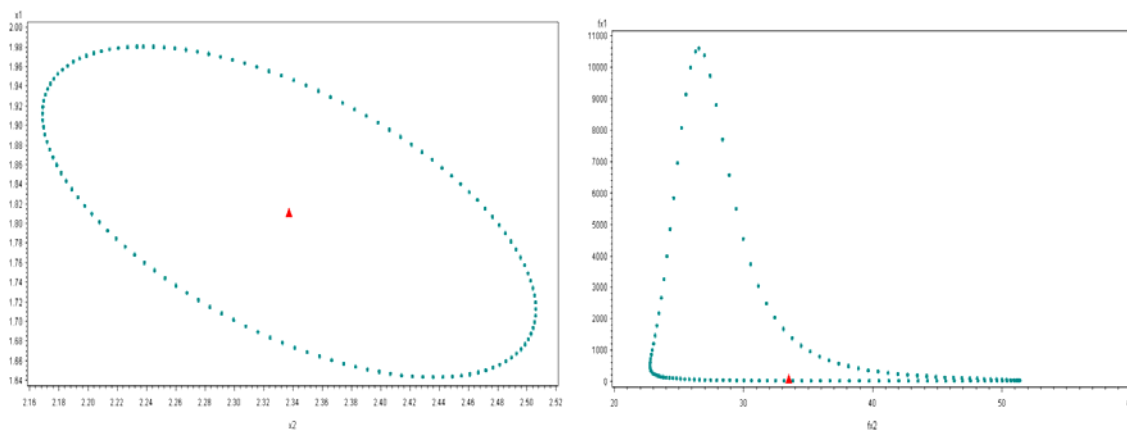
It is noteworthy that from Table 5.3 we see no increase in GYI from the use of a single test (CA19-9) to either sequential testing strategy. This is probably attributed to the fact that the AUC for CA19-9 is so high indicating that is is a fairly accurate test for diagnosing pancreatic cancer. We see a very slight gain in GYI from this test to the BP strategy (+0.01). Another interesting result is that the OOP for CA125 across the single and sequential tests vary greatly while the OOP for CA19-9 appears to stay fairly

constant when assuming normality. If we look at the OOP values for the sequential tests for CA125, we see that it is very high for BP and very low for BN; this indicates that you have have a very extreme value of CA125 to be diagnosed based on the first test, and that most patients would be likely to receive the second, more accurate test.

According to a study by Haglund (1986), less than half of the patients diagnosed with pancreatic cancer had elevated CA125 levels; elevated CA125 was defined as values greater than 35 U/ml. The Pancreatic Cancer Action Organization states that elevated levels of CA19-9 are values greater than 37 U/ml. Haglund (1986) also states that "combination of the CA125 with the CA19-9 test increases the sensitivity only 6% as compared to the CA19-9 assay alone," where combination (either one test or both tests) refers to the use of both test as opposed to a single test. This shows that, while the CA125 assay may not be as accurate in terms of pancreatic cancer diagnosis as CA19-9, the combination of the two tests does increase the Se, or the proportion of patients with disease that test positive for disease. Therefore, sequential testing is very applicable to this dataset.

It may still be worthwhile to do a sequential test in this case even though CA19-9 performs equally as well as a single test depending on the cost of CA19-9, both monetarily or patient burden. The monetary cost of the CA125 test is cheaper than the monetary cost for the CA19-9 test; if not everyone receives the latter, in theory the overall testing cost could be lowered if there is enough of a cost difference between tests. However, upon examining the data we see that for the BP strategy only seven individuals would be diagnosed based solely on the first test, while for the BN strategy only one

person would be diagnosed based solely on the first test. Thus, unless the second test is very expensive, it is more beneficial to just perform the single test of CA19-9 for this specific example.

6   Discussion and Future Work

6.1. Conclusions

In this dissertation, methods to estimate OOPs for two sequential testing strategies are described. The need for such a method is evident upon a literature review, as only methodology to estimate the OOP for a single, or univariate, tests exists. To demonstrate the potential usefulness of the method, we presented an example using a pancreatic cancer dataset.

Methods to estimate an OOP for each of two sequential testing strategies were first examined after identifying legitimate sequential strategies. These were determined by imposing a set of bounds, based on the normal distribution, on an estimated OOP based on the population parameters for that strategy. If a strategy was not deemed legitimate, the sequential test performs no better than a univariate test, i.e., it provides no extra information. For legitimate strategies, OOPs were estimated by maximizing the GYI using the NR with ridging algorithm. These OOPs were evaluated in terms of bias, RMSE, percent overestimation of Se and Sp corresponding to the OOP, 95% SCIs and CRs, coverage of SCIs and CRs, and comparison to empirical methods.

The simulations demonstrated that b, ρ, and TSS all play a large role in the estimation quality, while AUC and k only sometimes influences the estimation for the BP strategy. For the BN strategy, b and TSS were the most influential parameters affecting OOP estimation. Under both sequential testing strategies, an increase in TSS and AUC appear to produce more valid estimations of the OOP; the same can be said for values of b not equal to one. This makes sense since as we increase the sample size, we expect the estimates to be more

accurate, especially under the assumption of normality. Additionally, a higher AUC indicates a more accurate test, which means the OOP associated with it should also be more accurately estimated.

For a subset of parameter sets for each strategy, r was varied to determine the influence of test and/or misclassification cost and well as disease prevalence on OOP estimation; this parameter appears to affect estimation but is dependent on the AUC and ratio of standard deviations, b. The inclusion of this parameter may be beneficial to estimation, and it is worthwhile to expanded the subset we looked at to include other levels of TSS, k, and ρ. It was shown that BP and BN were comparable strategies when both were determined to be legitimate. In cases where only one strategy was legitimate for a given set of parameters, the other sequential strategy still outperformed the univariate test based on the GYI calculation.

Lastly, prediction models were estimated for SCI widths for each component of the OOP as well as the TSS. Although these models do not appear great based on values of R^2, there are currently no guidelines in the literature for determining sample size under a sequential testing strategy setting. Models developed provide a starting point for researchers to design studies to estimate OOPs. We saw that M1 was the model that appeared to perform the best, while M2 and M3 were somewhat similar in terms of R^2. Recall that these models just use transformations of the variables in M1. Although M2 was not the best performing model, it is likely to be the most used model due to the fact that the parameters needed to estimate the model will be the parameters most readily available to researchers. We also saw that the inclusion of an interaction term improved the model.

After characterizing the approach, we applied and interpreted the methodology using real data. Specifically, a pancreatic cancer dataset from Wieand *et al.* (1989) was used to predict those with pancreatic cancer from two blood antigens: CA125 (Test 1) and CA19-9 (Test 2). We saw that the sequential OOP estimates were vastly different from the univariate OOPs for CA125, and that for both tests, the OOPs assuming normality were both different than the empirical estimate. Note that for BP, there were multiple data points that appear to be optimal based on the empirical YI calculation. From this example we saw the importance of test accuracy, and that even though a single test is just as accurate in terms of GYI as a sequential test, cost of the tests may affect testing strategy decisions.

6.2. Limitations

Although the simulation study involved simulating data from a bivariate normal distribution, the estimated OOPs themselves may not follow the same. However, the construction of SCIs and CRs rely on the normality assumption. We randomly selected a subset of population parameter sets (~10%) for which to analyze the normality of the OOPs, but a more extensive analysis should be performed to assess the normality of the OOP for all population parameter sets. Since the SCIs, CRs, and coverage calculations are based on this assumption, OOPs may need to be transformed to satisfy this assumption and make these measures more reliable.

The evaluation of the method detailed in this dissertation was for the most part limited to the assumption that $r = 1$, implying the use of the YI. A limited set of

parameters were analyzed varying r, but this set should be expanded to include various TSS, k, and ρ.

The basic NR algorithm appeared to not always find an OOP when a coarse grid search appeared to find one. The substitution of the NR with ridging algorithm appeared to solve this problem. However, the underlying problem with the positive-definiteness of the Hessian matrix may still be an issue and should be investigated further.

When characterizing the results of the simulations for the BP and BN using formal regressions, note that the prediction models were multivariate main effects models; however, when looking at all the results in Tables 8.1 − 8.30, we see that sometimes the results of a predictor may depend on another. A main effects model would not pick up the interactions between predictors.

OLS regressions were used to develop models for the SCI widths as well as the TSS. However, inspection of the TSS involved only four levels. Investigation of other modeling types and/or interaction terms should be further explored to see if an increase in R^2 is obtained. Additional simulations including more values of TSS would also be helpful.

For the Wieand cancer application, we used the methodology developed in this dissertation regardless of the fact that the second test did not appear to fully satisfy the normality assumption. This assumption is important to the functioning of the methodology. Additionally, one of the single tests in this case appears to perform just as well as a sequential testing strategy in terms of the YI; more research can be done looking

at what happens if we factor in the prevalence of pancreatic cancer and/or cost of the two tests.

6.3. Future Work

The definition of the OOP in this dissertation was the set of theta values, corresponding to each test, that maximize the GYI. There are several other measures besides the GYI that could be used to define the OOP (and are mentioned in chapter 2), such as the (0,1)/ED/geometric method, kappa, MI, PCdx, or the DOR.

The NR algorithm is very sensitive to the selection of an initial value, has an increased computation time, and may not always converge to the global maximum. Other numerical estimation methods, such as the expectation-maximization (EM) algorithm or the Levenberg-Marquardt algorithm, should be compared to NR.

Formal regressions to summarize the effect of the five predictors on each measure of interest (as described in Tables 4.3 and 4.4) were determined using only a main effects model. In order to see how the predictors vary with one another, a higher level model including interaction terms is necessary. With a sample size of only 144, a full-factorial model may not be feasible given the levels of various predictors, but perhaps a model with four-way interactions could be estimated. From this, we could eliminate combinations of predictors which do not appear important until we have a significant model.

Variance estimates were obtained by bootstrapping due to its computational simplicity. The delta method requires more derivations as well as covariance estimates of

parameters; however, a more exact estimate may be obtained by employing this method. The delta method might also provide a method that would be less computationally intensive than bootstrapping. Other bootstrap methods, some of which are described in chapter 2, could be considered. Additionally, other methods of constructing confidence intervals, such as using exact methods, the Fiducial or Clopper-Pearson intervals, could be considered. Lastly, it may be appropriate to compare the coverage between the CIs constructed in this dissertation to those calculated with the mean of the OOP components as the point estimate rather than the actual simulated value (for example, mean $\theta_1 \pm 1.96(\mathrm{var}(\theta_1))$).

It may be interesting, and useful, to develop parallel methodology for different distributions of test results. The gamma distribution may be a good alternative for data that cannot be transformed to normality. Schisterman & Perkins (2007) develop a formula for the univariate test case using the GYI assuming a gamma distribution on the data. Additionally, other non-parametric methods, such as the kernel smoothing method described in chapter 2, should be explored in addition to the empirical estimate.

The inclusion of a third sequential testing strategy, most recently named Believe the Extreme (BE) by Ahmed, McClish, & Schubert (2011) should be examined. This strategy combines the BP and BN strategies by adding a third dimension to the OOP estimate (i.e. there are three threshold estimates rather than two). For this strategy, Test 2 is only administered only to those whose Test 1 values fall between the first two thresholds, or components in the OOP estimate; in other words, they are not positive or negative for disease based on Test 1.

6.4. Summary

This dissertation has described methods that can be employed to estimate and evaluate OOPs for sequential testing strategies. It was shown that oftentimes, it is more beneficial to perform a sequential test rather than a single test, as well as that this method is an improvement of the current empirical estimate. Parameters such as the AUC, TSS, b, ρ, and k were analyzed for influence on the estimation, and for the most part were found to be important. Additionally, sample size estimation formulae for a variety of parameters were developed and may be used as a guideline for researchers wishing to design sequential testing studies.

7 List of References

Ahmed, A.E., McClish, D.K., & Schubert, C.M. (2011). Accuracy and cost in medical testing using sequential testing strategies. *Statistics in Medicine* 30(29), 3416-3430.

Baccouche, H. *et al.* (2009). Diagnostic synergy of non-invasive cardiovascular magnetic resonance and invasive endomyocardial biopsy in troponin-positive patients without coronary artery disease. *European Heart Journal* 30(23), 2869-2879.

Bast, R.C. *et al.* (1983). A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. *New England Journal of Medicine* 309(15), 883-887.

Böhning, D., Holling, H., & Patilea, V. (2010). A limitation of the diagnostic-odds ratio in determining an optimal cut-off value for a continuous diagnostic test. *Statistical Methods in Medical Research* 00, 1-10.

Cury, R.C. *et al.* (2006). Diagnostic performance of stress perfusion and delayed-enhancement MR imaging in patients with coronary artery disease. *Radiology* 240, 39-45.

Del Villano, B.C. *et al.* (1983). *Radioimmunometric assay for a monoclonal antibody-defined tumor marker, CA 19-9. Clinical Chemistry* 29(3), 549-552.

Dodge *et al.* (2003). *The Oxford Dictionary of Statistical Terms*. Oxford University
Press, Inc., New York.

Dreher *et al.* (2004). Reduction of PTEN and p27^{kip1} expression correlates with tumor
grade in prostate cancer. Analysis in radical prostatectomy specimens and needle
biopsies. *Virchows Archiv* 444(6), 509-517.

Efron, B. & Tibshirani, R.J. (1979). *An Introduction to the Bootstrap.* Chapman &
Hall.

Etzioni, R. (2003). Combining biomarkers to detect disease with application to prostate
cancer. *Biostatistics*, 4, 523-538.

Farragi, D. (2003). Adjusting receiver operating characteristic curves and related indices
for covariates. *The Statistician* 52(2), 179-192.

Fluss, R., Faraggi, D., & Resier, B. (2005). Estimation of the Youden Index and its
Associated Cutoff Point. *Biometrical Journal* 47, 458-472.

Greiner, M., Pfeiffer, D., & Smith, R.D. (2000). Principles and practical application of
the receiver-operating characteristic analysis for diagnostic tests. *Preventative
Veterinary Medicine* 45, 23-41.

Green, D.M. & Swets, J.A. (1966). *Signal Detection Theory and Psychophysics*. John
Wiley & Sons, Inc.

Haglund, C. (1986). Tumour marker antigen CA125 in pancreatic cancer: A comparison
with CA19-9 and CEA. *British Journal of Cancer* 54(6), 897-901.

Halpern, E.J. *et al.* (1996). Comparison of Receiver Operating Characteristic Curves on
the Basis of Optimal Operating Points. *Statistics for Radiologists* 3, 245-253.

Hanley, J.A. (1988). The Robustness of the "Binormal" Assumptions Used in Fitting ROC Curves. *Medical Decision Making* 8(3), 197-203.

Hanley, J.A. (1989). Receiver Operating Characteristic (ROC) Methodology: The State of the Art. *Critical Reviews in Diagnostic Imaging* 29(3), 307-335.

Jund, J. *et al.* (2005). Methods to Estimate the Optimal Threshold for Normally or Log-Normally Distributed Biological Tests. *Medical Decision Making*, 406-415.

Lasko *et al.* (2005). The use of receiver operating characteristic curves in biomedical informatics. *Journal of Biomedical Informatics* 38, 404-415.

Leeflang, M.M.G. *et al.* (2008). Bias in Sensitivity and Specificity Caused by Data-Driven Selection of Optimal Cutoff Values: Mechanisms, Magnitude, and Solutions. *Clinical Chemistry* 54(4), 729-737.

Marshall, R.J. (1989). The Predictive Value of Simple Rules for Combining Two Diagnostic Tests. *Biometrics* 45(4), 1213-1222.

Massey Jr., F.J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association* 46(253), 68-78.

Metz, C.E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine* 8(4), 283-298.

Metz, C.E. & Kronman, H.B. (1980). Statistical significance tests for Binormal ROC curves. *Journal of Mathematical Psychology* 22(3), 218-243.

Oxley *et al.* (2011). Sequential Fusion. *Proc. SPIE 8050, Signal Processing, Sensor Fusion, and Target Recognition* XX, 80500V.

Pepe, M. (2003). *The Statistical evaluation of medical tests for classification and prediction*. New York: Oxford University Press.

Pepe, M.S. (2012). Receiver Operating Characteristic Methodology. *Journal of the American Statistical Association* 95(449), 308-311.

Pepe, M.S. & Thompson, M.L. (2000). Combining diagnostic test results to increase accuracy. *Biostatistics* 1(2), 123-140.

Perkins, N.J. & Schisterman, E.F. (2005). The Youden Index and the Optimal Cut-Point Corrected for Measurement Error. *Biometrical Journal* 47(4), 428-441.

Polister, P. (1980). Computer-Based Consultation for Repeated Diagnostic Tests. *Proc Annu Symp Comput Appl Med Care* 5(1), 572–581.

Ruopp, M.D. (2008). Youden Index and Optimal Cut-Point Estimated from Observations Affected by a Lower Limit of Detection. Biometrical Journal 50(3), 419–430.

SAS Institute, Inc, SAS 9.2 & 9.3, SAS Institute Inc., Cary, NC, 2000–2008.

Schisterman, E.F. & Perkins, N. (2007). Confidence Intervals for the Youden Index and Corresponding Optimal Cut-Point. Communications in Statistics – Simulation and Computation 36, 549-563.

Schisterman, E.F. *et al.* (2005). Optimal Cut-point and Its Corresponding Youden Index to Discriminate Individuals Using Pooled Blood Samples. *Epidemiology* 16(1), 73-81.

Seto, M.C. (2005). Is More Better? Combining Actuarial Risk Scales to Predict Recidivism Among Adult Sex Offenders. *Psychological Assessment* 17(2), 156-167.

Shapiro, S.S. & Wilk, M.B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* 3(4), 591-611.

Shen, C. (2008). UCI On the Principles of Believe the Positive and Believe the Negative for Diagnosis Using Two Continuous Tests. *Journal of Data Science* 6, 189-205.

Skaltsa, K., Jover, L., & Carrasco, J.L. (2010). Estimation of the diagnostic threshold accounting for decision costs and sampling uncertainty. *Biometrical Journal* 52(5), 676-697.

Su, J.Q. & Lu, J.S. (1993). Linear Combinations of Multiple Diagnostic Markers. *Journal of the American Statistical Association* 88(424), 1350-1355.

Swets, J.A. (1979). ROC Analysis Applied to the Evaluation of Medical Imaging Techniques. Investigative Radiology 14(2), 109-121.

Thompson, M.L. (2003). Assessing the diagnostic accuracy of a sequence of tests. *Biostatistics* 4(3), 341-351.

Westin, L.K. (2001). Receiver operating characteristic (ROC) analysis: Evaluating discriminance effects among decision support systems. Tech. Rep. UMINF 01.18, ISSN-0348-0542, Department of Computer Science, Umea University, SE-90187, Umea, Sweden, http://www.cs.umu.se/research/reports/2001/018/part1.pdf.

Wieand, S. *et al.* (1989) A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika* 76(3), 585-592.

Youden, W.J. Index for Rating Diagnostic Tests. *Cancer* 3(1), 32-35.

Zhou, X-H., Obuchowski, N.A., & McClish, D.K. (2011). *Statistical Methods in Diagnostic Medicine*. Wiley Series in Probability and Statistics.

Zöller, L., Burkard, S., & Schäfer, H. (1991). Validity of western immunoblot band patterns in the serodiagnosis of Lyme borreliosis. *Journal of Clinical Microbiology* 29(1), 174-182.

Zou , K.H. *et al.* (2013). Optimal Thresholds by Maximizing or Minimizing Various Metrics via ROC-Type Analysis. *Academic Radiology* 20 (7), 807-815.

Zou, K.H., Hall, W.J., & Shapiro, D.E. (1997). Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Statistics in Medicine* 16(19), 2143-2156.

Zou, K.H., O'Malley, A.J, & Mauri, L. (2007). Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. *Circulation* 115(5), 654-657

8   Appendices

8.1. Appendix A: Exact Gradient and Hessian Equations for BP and BN

In order to maximize our objective function, the GYI, we must first define **f** and **H** for each sequential testing strategy. First consider BP. We must take the partial derivatives $\frac{dGYI}{d\theta_1}$ and $\frac{dGYI}{d\theta_2}$ to comprise the gradient matrix, set them equal to zero, and solve the system. Taking the respective derivatives, our BP system of equations is as follows:

$$\frac{dGYI^{BP}}{d\theta_1} = \left[\frac{r}{\sigma_{1H}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\theta_1 - \mu_{1H}}{\sigma_{1H}}\right)^2} \cdot \Phi\left(\frac{\sigma_{1H}\theta_2 - \sigma_{1H}\mu_{2H} - \rho_H(\theta_1 - \mu_{1H})\sigma_{2H}}{\sigma_{1H}\sigma_{2H}\sqrt{1 - \rho_H^2}} \right) \right]$$
$$- \left[\frac{1}{\sigma_{1D}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\theta_1 - \mu_{1D}}{\sigma_{1D}}\right)^2} \cdot \Phi\left(\frac{\sigma_{1D}\theta_2 - \sigma_{1D}\mu_{2D} - \rho_D(\theta_1 - \mu_{1D})\sigma_{2D}}{\sigma_{1D}\sigma_{2D}\sqrt{1 - \rho_D^2}} \right) \right] \quad (8.1)$$

and

$$\frac{dGYI^{BP}}{d\theta_2} = \left[\frac{r}{\sigma_{2H}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\theta_2 - \mu_{2H}}{\sigma_{2H}}\right)^2} \cdot \Phi\left(\frac{\sigma_{2H}\theta_1 - \sigma_{2H}\mu_{1H} - \rho_H(\theta_2 - \mu_{2H})\sigma_{1H}}{\sigma_{1H}\sigma_{2H}\sqrt{1 - \rho_H^2}} \right) \right]$$
$$- \left[\frac{1}{\sigma_{2D}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\theta_2 - \mu_{2D}}{\sigma_{2D}}\right)^2} \cdot \Phi\left(\frac{\sigma_{2D}\theta_1 - \sigma_{2D}\mu_{1D} - \rho_D(\theta_2 - \mu_{2D})\sigma_{1D}}{\sigma_{1D}\sigma_{2D}\sqrt{1 - \rho_D^2}} \right) \right] \quad (8.2)$$

We also need the Hessian matrix of second derivatives; each entry is shown below.

$$\frac{d}{d\theta_1}\left[\frac{dGYI^{BP}}{d\theta_1}\right]=$$

$$\frac{-re^{-\frac{1}{2}\left(\frac{\theta_1-\mu_{1H}}{\sigma_{1H}}\right)^2}}{\sigma_{1H}^2\sqrt{2\pi}}\left(\frac{\rho_H}{2\pi\sqrt{1-\rho_H^2}}e^{-\frac{1}{2}\left(\frac{\sigma_{1H}\theta_2-\sigma_{1H}\mu_{2H}-\rho_H(\theta_1-\mu_{1H})\sigma_{2H}}{\sigma_{1H}\sigma_{2H}\sqrt{1-\rho_H^2}}\right)^2}+\left(\frac{\theta_1-\mu_{1H}}{\sigma_{1H}}\right)\Phi\left(\frac{\sigma_{1H}\theta_2-\sigma_{1H}\mu_{2H}-\rho_H(\theta_1-\mu_{1H})\sigma_{2H}}{\sigma_{1H}\sigma_{2H}\sqrt{1-\rho_H^2}}\right)\right)$$

$$-\frac{e^{-\frac{1}{2}\left(\frac{\theta_1-\mu_{1D}}{\sigma_{1D}}\right)^2}}{\sigma_{1D}^2\sqrt{2\pi}}\left(\frac{\rho_D}{2\pi\sqrt{1-\rho_D^2}}e^{-\frac{1}{2}\left(\frac{\sigma_{1D}\theta_2-\sigma_{1D}\mu_{2D}-\rho_D(\theta_1-\mu_{1D})\sigma_{2D}}{\sigma_{1D}\sigma_{2D}\sqrt{1-\rho_D^2}}\right)^2}+\left(\frac{\theta_1-\mu_{1D}}{\sigma_{1D}}\right)\Phi\left(\frac{\sigma_{1D}\theta_2-\sigma_{1D}\mu_{2D}-\rho_D(\theta_1-\mu_{1D})\sigma_{2D}}{\sigma_{1D}\sigma_{2D}\sqrt{1-\rho_D^2}}\right)\right) \quad (8.3)$$

$$\frac{d}{d\theta_1}\left[\frac{dGYI^{BP}}{d\theta_2}\right]=$$

$$r\left[\frac{1}{\sigma_{1H}\sigma_{2H}\sqrt{2\pi}\sqrt{1-\rho_H^2}}e^{-\frac{1}{2}\left(\frac{\theta_2-\mu_{2H}}{\sigma_{2H}}\right)^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\sigma_{2H}\theta_1-\sigma_{2H}\mu_{1H}-\rho_H(\theta_2-\mu_{2H})\sigma_{1H}}{\sigma_{1H}\sigma_{2H}\sqrt{1-\rho_H^2}}\right)^2}\right] \quad (8.4)$$

$$-\left[\frac{1}{\sigma_{1D}\sigma_{2D}\sqrt{2\pi}\sqrt{1-\rho_D^2}}e^{-\frac{1}{2}\left(\frac{\theta_2-\mu_{2D}}{\sigma_{2D}}\right)^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\sigma_{2D}\theta_1-\sigma_{2D}\mu_{1D}-\rho_D(\theta_2-\mu_{2D})\sigma_{1D}}{\sigma_{1D}\sigma_{2D}\sqrt{1-\rho_D^2}}\right)^2}\right]$$

$$\frac{d}{d\theta_2}\left[\frac{dGYI^{BP}}{d\theta_1}\right]=$$

$$r\left[\frac{1}{\sigma_{1H}\sigma_{2H}\sqrt{2\pi}\sqrt{1-\rho_H^2}}e^{-\frac{1}{2}\left(\frac{\theta_2-\mu_{1H}}{\sigma_{1H}}\right)^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\sigma_{1H}\theta_2-\sigma_{1H}\mu_{2H}-\rho_H(\theta_1-\mu_{1H})\sigma_{2H}}{\sigma_{1H}\sigma_{2H}\sqrt{1-\rho_H^2}}\right)^2}\right] \quad (8.5)$$

$$-\left[\frac{1}{\sigma_{1D}\sigma_{2D}\sqrt{2\pi}\sqrt{1-\rho_D^2}}e^{-\frac{1}{2}\left(\frac{\theta_1-\mu_{1D}}{\sigma_{1D}}\right)^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\sigma_{1D}\theta_2-\sigma_{1D}\mu_{2D}-\rho_D(\theta_1-\mu_{1D})\sigma_{2D}}{\sigma_{1D}\sigma_{2D}\sqrt{1-\rho_D^2}}\right)^2}\right]$$

$$\frac{d}{d\theta_2}\left[\frac{dGYI^{BP}}{d\theta_2}\right] =$$

$$\frac{-re^{-\frac{1}{2}\left(\frac{\theta_2-\mu_{2H}}{\sigma_{2H}}\right)^2}}{\sigma_{2H}^2\sqrt{2\pi}}\left(\begin{array}{l}\frac{\rho_H}{2\pi\sqrt{1-\rho_H^2}}e^{-\frac{1}{2}\left(\frac{\sigma_{2H}\theta_1-\sigma_{2H}\mu_{1H}-\rho_H(\theta_2-\mu_{2H})\sigma_{1H}}{\sigma_{1H}\sigma_{2H}\sqrt{1-\rho_H^2}}\right)^2} \\ +\left(\frac{\theta_2-\mu_{2H}}{\sigma_{2H}}\right)\Phi\left(\frac{\sigma_{2H}\theta_1-\sigma_{2H}\mu_{1H}-\rho_H(\theta_2-\mu_{2H})\sigma_{1H}}{\sigma_{1H}\sigma_{2H}\sqrt{1-\rho_H^2}}\right)\end{array}\right)$$

$$\frac{-e^{-\frac{1}{2}\left(\frac{\theta_2-\mu_{2D}}{\sigma_{2D}}\right)^2}}{\sigma_{2D}^2\sqrt{2\pi}}\left(\begin{array}{l}\frac{\rho_D}{2\pi\sqrt{1-\rho_D^2}}e^{-\frac{1}{2}\left(\frac{\sigma_{2D}\theta_1-\sigma_{2D}\mu_{1D}-\rho_D(\theta_2-\mu_{2D})\sigma_{1D}}{\sigma_{1D}\sigma_{2D}\sqrt{1-\rho_D^2}}\right)^2} \\ +\left(\frac{\theta_2-\mu_{2D}}{\sigma_{2D}}\right)\Phi\left(\frac{\sigma_{2D}\theta_1-\sigma_{2D}\mu_{1D}-\rho_D(\theta_2-\mu_{2D})\sigma_{1D}}{\sigma_{1D}\sigma_{2D}\sqrt{1-\rho_D^2}}\right)\end{array}\right) \quad (8.6)$$

These equations and derivatives are substituted into the NR algorithm (equation 3.16 or 3.17). This iterative process to simultaneously solve for (θ_1,θ_2) while both of the following criterion are met: the difference between the previous and current estimate of θ_1 differ by more than a set convergence criterion or the difference between the previous and current estimate of θ_2 differ by more than a set convergence criterion, and the number of iterations is less than the maximum amount set.

Next, consider the BN strategy. Like the BP, we must first derive the gradient matrix:

$$\frac{dGYI^{BN}}{d\theta_1} =$$

$$\left[\frac{1}{\sigma_{1D}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\theta_3 - \mu_{1D}}{\sigma_{1D}}\right)^2} \cdot \Phi\left(\frac{\sigma_{1D}\theta_4 - \sigma_{1D}\mu_{2D} - \rho_D(\theta_3 - \mu_{1D})\sigma_{2D}}{\sigma_{1D}\sigma_{2D}\sqrt{1-\rho_D^2}} \right) \right]$$

$$- \left[\frac{r}{\sigma_{1H}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\theta_3 - \mu_{1H}}{\sigma_{1H}}\right)^2} \cdot \Phi\left(\frac{\sigma_{1H}\theta_4 - \sigma_{1H}\mu_{2H} - \rho_H(\theta_3 - \mu_{1H})\sigma_{2H}}{\sigma_{1H}\sigma_{2H}\sqrt{1-\rho_H^2}} \right) \right] \quad (8.7)$$

$$+ \frac{r}{\sigma_{1H}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mu_{1H} - \theta_3}{\sigma_{1H}}\right)^2} - \frac{1}{\sigma_{1D}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mu_{1D} - \theta_3}{\sigma_{1D}}\right)^2}$$

and

$$\frac{dGYI^{BN}}{d\theta_2} =$$

$$\left[\frac{1}{\sigma_{2D}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\theta_4 - \mu_{2D}}{\sigma_{2D}}\right)^2} \cdot \Phi\left(\frac{\sigma_{2D}\theta_3 - \sigma_{3D}\mu_{1D} - \rho_D(\theta_4 - \mu_{2D})\sigma_{1D}}{\sigma_{1D}\sigma_{2D}\sqrt{1-\rho_D^2}} \right) \right]$$

$$- \left[\frac{r}{\sigma_{2H}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\theta_4 - \mu_{2H}}{\sigma_{2H}}\right)^2} \cdot \Phi\left(\frac{\sigma_{2H}\theta_3 - \sigma_{2H}\mu_{1H} - \rho_H(\theta_4 - \mu_{2H})\sigma_{1H}}{\sigma_{1H}\sigma_{2H}\sqrt{1-\rho_H^2}} \right) \right] \quad (8.8)$$

$$+ \frac{r}{\sigma_{2H}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mu_{2H} - \theta_4}{\sigma_{2H}}\right)^2} - \frac{1}{\sigma_{2D}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mu_{2D} - \theta_4}{\sigma_{2D}}\right)^2}$$

Again, to maximize the GYI using the NR algorithm for this strategy, a Hessian matrix of second derivatives is needed. Like they are listed for the BP strategy, below are the four components for the Hessian matrix for the BN strategy:

$$\frac{d}{d\theta_3}\left[\frac{dGYI^{BN}}{d\theta_3}\right]=$$

$$-\frac{e^{-\frac{1}{2}\left(\frac{\theta_3-\mu_{1D}}{\sigma_{1D}}\right)^2}}{\sigma_{1D}^2\sqrt{2\pi}}\left(\frac{\rho_D}{2\pi\sqrt{1-\rho_D^2}}e^{-\frac{1}{2}\left(\frac{\sigma_{1D}\theta_4-\sigma_{1D}\mu_{2D}-\rho_D(\theta_3-\mu_{1D})\sigma_{2D}}{\sigma_{1D}\sigma_{2D}\sqrt{1-\rho_D^2}}\right)^2}\right.$$
$$\left.+\left(\frac{\theta_3-\mu_{1D}}{\sigma_{1D}}\right)\Phi\left(\frac{\sigma_{1D}\theta_4-\sigma_{1D}\mu_{2D}-\rho_D(\theta_3-\mu_{1D})\sigma_{2D}}{\sigma_{1D}\sigma_{2D}\sqrt{1-\rho_D^2}}\right)\right)$$

$$-\frac{-re^{-\frac{1}{2}\left(\frac{\theta_3-\mu_{1H}}{\sigma_{1H}}\right)^2}}{\sigma_{1H}^2\sqrt{2\pi}}\left(\frac{\rho_H}{2\pi\sqrt{1-\rho_H^2}}e^{-\frac{1}{2}\left(\frac{\sigma_{1H}\theta_4-\sigma_{1H}\mu_{2H}-\rho_H(\theta_3-\mu_{1H})\sigma_{2H}}{\sigma_{1H}\sigma_{2H}\sqrt{1-\rho_H^2}}\right)^2}\right.$$
$$\left.+\left(\frac{\theta_3-\mu_{1H}}{\sigma_{1H}}\right)\Phi\left(\frac{\sigma_{1H}\theta_4-\sigma_{1H}\mu_{2H}-\rho_H(\theta_3-\mu_{1H})\sigma_{2H}}{\sigma_{1H}\sigma_{2H}\sqrt{1-\rho_H^2}}\right)\right)$$

$$+r\frac{(\mu_{1H}-\theta_3)}{\sigma_{1H}^3\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\mu_{1H}-\theta_3}{\sigma_{1H}}\right)^2}-\frac{(\mu_{1D}-\theta_3)}{\sigma_{1D}^3\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\mu_{1D}-\theta_3}{\sigma_{1D}}\right)^2}\tag{8.9}$$

$$\frac{d}{d\theta_3}\left[\frac{dGYI^{BN}}{d\theta_4}\right]=$$

$$\left[\frac{1}{\sigma_{1D}\sigma_{2D}\sqrt{2\pi}\sqrt{1-\rho_D^2}}e^{-\frac{1}{2}\left(\frac{\theta_3-\mu_{1D}}{\sigma_{1D}}\right)^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\sigma_{1D}\theta_4-\sigma_{1D}\mu_{2D}-\rho_D(\theta_3-\mu_{1D})\sigma_{2D}}{\sigma_{1D}\sigma_{2D}\sqrt{1-\rho_D^2}}\right)^2}\right]\tag{8.10}$$

$$-r\left[\frac{1}{\sigma_{1H}\sigma_{2H}\sqrt{2\pi}\sqrt{1-\rho_H^2}}e^{-\frac{1}{2}\left(\frac{\theta_3-\mu_{1H}}{\sigma_{1H}}\right)^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\sigma_{1H}\theta_4-\sigma_{1H}\mu_{2H}-\rho_H(\theta_3-\mu_{1H})\sigma_{2H}}{\sigma_{1H}\sigma_{2H}\sqrt{1-\rho_H^2}}\right)^2}\right]$$

$$\frac{d}{d\theta_4}\left[\frac{dGYI^{BN}}{d\theta_3}\right]=$$

$$\left[\frac{1}{\sigma_{1D}\sigma_{2D}\sqrt{2\pi}\sqrt{1-\rho_D^2}}e^{-\frac{1}{2}\left(\frac{\theta_4-\mu_{2D}}{\sigma_{2D}}\right)^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\sigma_{2D}\theta_3-\sigma_{2D}\mu_{1D}-\rho_D(\theta_4-\mu_{2D})\sigma_{1D}}{\sigma_{1D}\sigma_{2D}\sqrt{1-\rho_D^2}}\right)^2}\right] \qquad (8.11)$$

$$-r\left[\frac{1}{\sigma_{1H}\sigma_{2H}\sqrt{2\pi}\sqrt{1-\rho_H^2}}e^{-\frac{1}{2}\left(\frac{\theta_4-\mu_{2H}}{\sigma_{2H}}\right)^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\sigma_{2H}\theta_3-\sigma_{2H}\mu_{1H}-\rho_H(\theta_4-\mu_{2H})\sigma_{1H}}{\sigma_{1H}\sigma_{2H}\sqrt{1-\rho_H^2}}\right)^2}\right]$$

$$\frac{d}{d\theta_4}\left[\frac{dGYI^{BN}}{d\theta_4}\right]=$$

$$-\frac{e^{-\frac{1}{2}\left(\frac{\theta_4-\mu_{2D}}{\sigma_{2D}}\right)^2}}{\sigma_{2D}^2\sqrt{2\pi}}\left(\frac{\rho_D}{2\pi\sqrt{1-\rho_D^2}}e^{-\frac{1}{2}\left(\frac{\sigma_{2D}\theta_3-\sigma_{2D}\mu_{1D}-\rho_D(\theta_4-\mu_{2D})\sigma_{1D}}{\sigma_{1D}\sigma_{2D}\sqrt{1-\rho_D^2}}\right)^2} + \left(\frac{\theta_4-\mu_{2D}}{\sigma_{2D}}\right)\Phi\left(\frac{\sigma_{2D}\theta_3-\sigma_{2D}\mu_{1D}-\rho_D(\theta_4-\mu_{2D})\sigma_{1D}}{\sigma_{1D}\sigma_{2D}\sqrt{1-\rho_D^2}}\right)\right)$$

$$\frac{-re^{-\frac{1}{2}\left(\frac{\theta_4-\mu_{2H}}{\sigma_{2H}}\right)^2}}{\sigma_{2H}^2\sqrt{2\pi}}\left(\frac{\rho_H}{2\pi\sqrt{1-\rho_H^2}}e^{-\frac{1}{2}\left(\frac{\sigma_{2H}\theta_3-\sigma_{2H}\mu_{1H}-\rho_H(\theta_4-\mu_{2H})\sigma_{1H}}{\sigma_{1H}\sigma_{2H}\sqrt{1-\rho_H^2}}\right)^2} + \left(\frac{\theta_4-\mu_{2H}}{\sigma_{2H}}\right)\Phi\left(\frac{\sigma_{2H}\theta_3-\sigma_{2H}\mu_{1H}-\rho_H(\theta_4-\mu_{2H})\sigma_{1H}}{\sigma_{1H}\sigma_{2H}\sqrt{1-\rho_H^2}}\right)\right)$$

$$+r\frac{(\mu_{2H}-\theta_4)}{\sigma_{2H}^3\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\mu_{2H}-\theta_4}{\sigma_{2H}}\right)^2} - \frac{(\mu_{2D}-\theta_4)}{\sigma_{2D}^3\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\mu_{2D}-\theta_4}{\sigma_{2D}}\right)^2} \qquad (8.12)$$

These pieces are substituted into the NR algorithm (equation 3.16 or 3.17). This iterative process to simultaneously solve for (θ_3,θ_4) while both of the following criterion are met:

the difference between the previous and current estimate of θ_3 differ by more than a

set convergence criterion or the difference between the previous and current estimate of

θ_4 differ by more than a set convergence criterion, and the number of iterations is less

than the maximum amount set.

8.2. Appendix B: Simulation Results

Below are tables with the results from the simulation study detailed in chapter 4. Tables

8.1 – 8.15 refer to BP, Tables 8.16 – 8.30 refer to BN, Tables 8.31 and 8.32 detail a

summary of the normality assessed in section 4.6, and Tables 8.33 and 8.34 refer to the

results for varying r detailed in section 4.7.

Table 8.1: BP, Population Parameters' Optimal Points, listed as Se / Sp / GYI, AUC = (0.7, 0.7)

	ρ	b = (0.5, 0.5)	b = (1.0, 1.0)	b = (0.5, 1.0)	b = (1.0, 0.5)	b = (1.5, 1.5)	b = (1.0, 1.5)	b = (1.5, 1.0)
r = 0.5	(0.00, 0.00)	0.77 / 0.74 / 0.64	0.89 / 0.38 / 0.58	0.85 / 0.51 / 0.61	0.85 / 0.51 / 0.61	*0.94 / 0.35 / 0.61*	0.93 / 0.36 / 0.62	0.93 / 0.36 / 0.62
	(0.25, 0.50)	0.70 / 0.72 / 0.56	*0.90 / 0.29 / 0.55*	0.87 / 0.37 / 0.56	0.87 / 0.37 / 0.56	*0.94 / 0.35 / 0.61*	*0.94 / 0.35 / 0.61*	*0.94 / 0.35 / 0.61*
	(0.50, 0.25)	0.76 / 0.72 / 0.62	0.90 / 0.39 / 0.59	0.85 / 0.49 / 0.60	0.86 / 0.49 / 0.60	0.94 / 0.35 / 0.62	0.93 / 0.37 / 0.62	0.93 / 0.37 / 0.62
	(0.50, 0.50)	0.72 / 0.71 / 0.57	0.90 / 0.33 / 0.56	0.86 / 0.41 / 0.57	0.86 / 0.41 / 0.57	*0.94 / 0.35 / 0.61*	*0.94 / 0.35 / 0.61*	*0.94 / 0.35 / 0.61*
r = 1	(0.00, 0.00)	0.67 / 0.88 / 0.55	0.69 / 0.66 / 0.35	0.65 / 0.79 / 0.45	0.65 / 0.79 / 0.45	*0.82 / 0.51 / 0.33*	0.81 / 0.53 / 0.34	0.81 / 0.53 / 0.34
	(0.25, 0.50)	0.59 / 0.89 / 0.48	*0.65 / 0.64 / 0.29*	0.53 / 0.85 / 0.38	0.53 / 0.85 / 0.38	*0.82 / 0.51 / 0.33*	*0.82 / 0.51 / 0.33*	*0.82 / 0.51 / 0.33*
	(0.50, 0.25)	0.65 / 0.88 / 0.53	0.71 / 0.65 / 0.36	0.64 / 0.79 / 0.43	0.64 / 0.79 / 0.43	0.84 / 0.50 / 0.33	0.81 / 0.54 / 0.35	0.81 / 0.54 / 0.35
	(0.50, 0.50)	0.60 / 0.88 / 0.48	0.66 / 0.65 / 0.31	0.58 / 0.82 / 0.39	0.58 / 0.82 / 0.39	*0.82 / 0.51 / 0.33*	0.82 / 0.51 / 0.33	0.82 / 0.51 / 0.33
r = 2	(0.00, 0.00)	0.58 / 0.95 / 0.47	0.38 / 0.88 / 0.14	0.46 / 0.93 / 0.33	0.46 / 0.93 / 0.33	*0.00 / 1.00 / 0.00*	*0.29 / 0.90 / 0.09*	*0.29 / 0.90 / 0.09*
	(0.25, 0.50)	0.50 / 0.95 / 0.40	0.29 / 0.91 / 0.10	*0.40 / 0.95 / 0.31*	*0.40 / 0.95 / 0.31*	*0.00 / 1.00 / 0.00*	*0.29 / 0.90 / 0.09*	*0.29 / 0.90 / 0.09*
	(0.50, 0.25)	0.55 / 0.95 / 0.45	0.39 / 0.88 / 0.14	0.44 / 0.94 / 0.32	0.44 / 0.94 / 0.32	*0.00 / 1.00 / 0.00*	*0.29 / 0.90 / 0.09*	*0.29 / 0.90 / 0.09*
	(0.50, 0.50)	0.51 / 0.95 / 0.41	0.33 / 0.89 / 0.12	*0.40 / 0.95 / 0.31*	*0.40 / 0.95 / 0.31*	*0.00 / 1.00 / 0.00*	*0.29 / 0.90 / 0.09*	*0.29 / 0.90 / 0.09*

Note: Red italicized cells refer to illegitimate BP strategies (combination of values were not valid)

Table 8.2: BP, Population Parameters' Optimal Points, listed as Se / Sp / GYI, AUC = (0.75, 0.85)

	ρ	b = (0.5, 0.5)	b = (1.0, 1.0)	b = (0.5, 1.0)	b = (1.0, 0.5)	b = (1.5, 1.5)	b = (1.0, 1.5)	b = (1.5, 1.0)
r = 0.5	(0.00, 0.00)	0.85 / 0.83 / 0.77	0.90 / 0.62 / 0.71	0.89 / 0.71 / 0.74	0.86 / 0.73 / 0.72	*0.93 / 0.60 / 0.73*	0.93 / 0.60 / 0.74	0.92 / 0.55 / 0.69
	(0.25, 0.50)	0.80 / 0.81 / 0.70	*0.89 / 0.60 / 0.69*	0.88 / 0.64 / 0.70	*0.78 / 0.78 / 0.67*	*0.93 / 0.60 / 0.73*	*0.93 / 0.60 / 0.73*	*0.89 / 0.60 / 0.69*
	(0.50, 0.25)	0.84 / 0.82 / 0.75	0.90 / 0.61 / 0.70	0.89 / 0.69 / 0.73	0.85 / 0.71 / 0.71	*0.93 / 0.60 / 0.73*	0.93 / 0.60 / 0.74	0.92 / 0.56 / 0.70
	(0.50, 0.50)	0.81 / 0.80 / 0.71	0.89 / 0.60 / 0.69	0.88 / 0.65 / 0.70	0.81 / 0.73 / 0.68	*0.93 / 0.60 / 0.73*	*0.93 / 0.60 / 0.73*	*0.89 / 0.60 / 0.69*
r = 1	(0.00, 0.00)	0.80 / 0.91 / 0.71	0.79 / 0.77 / 0.56	0.80 / 0.83 / 0.63	0.76 / 0.87 / 0.63	*0.85 / 0.71 / 0.56*	0.85 / 0.71 / 0.57	*0.77 / 0.77 / 0.54*
	(0.25, 0.50)	0.73 / 0.91 / 0.64	*0.77 / 0.77 / 0.54*	0.76 / 0.81 / 0.57	*0.70 / 0.90 / 0.60*	*0.85 / 0.71 / 0.56*	*0.85 / 0.71 / 0.56*	*0.77 / 0.77 / 0.54*
	(0.50, 0.25)	0.77 / 0.91 / 0.68	0.79 / 0.77 / 0.55	0.79 / 0.83 / 0.61	0.74 / 0.88 / 0.62	*0.85 / 0.71 / 0.56*	0.86 / 0.71 / 0.57	*0.77 / 0.77 / 0.54*
	(0.50, 0.50)	0.74 / 0.91 / 0.64	0.77 / 0.77 / 0.54	0.76 / 0.81 / 0.58	*0.70 / 0.90 / 0.60*	*0.85 / 0.71 / 0.56*	*0.85 / 0.71 / 0.56*	*0.77 / 0.77 / 0.54*
r = 2	(0.00, 0.00)	0.74 / 0.96 / 0.65	0.63 / 0.88 / 0.40	0.68 / 0.92 / 0.52	0.66 / 0.95 / 0.55	*0.69 / 0.82 / 0.33*	0.70 / 0.82 / 0.34	*0.60 / 0.89 / 0.37*
	(0.25, 0.50)	0.66 / 0.96 / 0.58	*0.60 / 0.89 / 0.37*	0.61 / 0.92 / 0.44	*0.62 / 0.95 / 0.53*	*0.69 / 0.82 / 0.33*	*0.69 / 0.82 / 0.33*	*0.60 / 0.89 / 0.37*
	(0.50, 0.25)	0.71 / 0.96 / 0.62	0.63 / 0.88 / 0.39	0.66 / 0.92 / 0.49	0.64 / 0.95 / 0.54	*0.69 / 0.82 / 0.33*	0.70 / 0.82 / 0.34	*0.60 / 0.89 / 0.37*
	(0.50, 0.50)	0.67 / 0.96 / 0.58	*0.60 / 0.89 / 0.37*	0.62 / 0.92 / 0.45	*0.62 / 0.95 / 0.53*	*0.69 / 0.82 / 0.33*	*0.69 / 0.82 / 0.33*	*0.60 / 0.89 / 0.37*

Note: Red italicized cells refer to illegitimate BP strategies (combination of values were not valid)

Table 8.3: BP, Population Parameters' Optimal Points, listed as Se / Sp / GYI, AUC = (0.9, 0.9)

	ρ	b = (0.5, 0.5)	b = (1.0, 1.0)	b = (0.5, 1.0)	b = (1.0, 0.5)	b = (1.5, 1.5)	b = (1.0, 1.5)	b = (1.5, 1.0)
r = 0.5	(0.00, 0.00)	0.92 / 0.91 / 0.88	0.93 / 0.78 / 0.82	0.93 / 0.84 / 0.85	0.93 / 0.84 / 0.85	0.95 / 0.72 / 0.81	0.95 / 0.75 / 0.82	0.95 / 0.75 / 0.82
	(0.25, 0.50)	0.88 / 0.88 / 0.82	0.91 / 0.73 / 0.77	0.89 / 0.80 / 0.79	0.89 / 0.80 / 0.79	*0.94 / 0.69 / 0.79*	0.94 / 0.70 / 0.79	0.94 / 0.70 / 0.79
	(0.50, 0.25)	0.91 / 0.89 / 0.85	0.93 / 0.76 / 0.81	0.92 / 0.83 / 0.83	0.92 / 0.83 / 0.83	0.95 / 0.72 / 0.81	0.94 / 0.73 / 0.81	0.94 / 0.73 / 0.81
	(0.50, 0.50)	0.88 / 0.88 / 0.82	0.91 / 0.74 / 0.78	0.90 / 0.80 / 0.80	0.90 / 0.80 / 0.80	0.94 / 0.69 / 0.79	0.94 / 0.71 / 0.79	0.94 / 0.71 / 0.79
r = 1	(0.00, 0.00)	0.90 / 0.95 / 0.84	0.88 / 0.85 / 0.73	0.88 / 0.91 / 0.79	0.88 / 0.91 / 0.79	0.90 / 0.79 / 0.69	0.89 / 0.82 / 0.71	0.89 / 0.82 / 0.71
	(0.25, 0.50)	0.84 / 0.94 / 0.78	0.83 / 0.83 / 0.67	0.82 / 0.90 / 0.72	0.82 / 0.90 / 0.72	*0.88 / 0.78 / 0.66*	0.87 / 0.78 / 0.66	0.87 / 0.78 / 0.66
	(0.50, 0.25)	0.87 / 0.94 / 0.82	0.87 / 0.85 / 0.71	0.86 / 0.90 / 0.76	0.86 / 0.90 / 0.76	0.90 / 0.79 / 0.69	0.89 / 0.81 / 0.70	0.89 / 0.81 / 0.70
	(0.50, 0.50)	0.84 / 0.94 / 0.78	0.84 / 0.84 / 0.68	0.83 / 0.90 / 0.73	0.83 / 0.90 / 0.73	0.88 / 0.77 / 0.66	0.87 / 0.80 / 0.67	0.87 / 0.80 / 0.67
r = 2	(0.00, 0.00)	0.86 / 0.97 / 0.81	0.79 / 0.91 / 0.62	0.82 / 0.95 / 0.72	0.82 / 0.95 / 0.72	0.81 / 0.86 / 0.52	0.80 / 0.88 / 0.57	0.80 / 0.88 / 0.57
	(0.25, 0.50)	0.80 / 0.97 / 0.73	0.73 / 0.91 / 0.54	0.75 / 0.95 / 0.65	0.75 / 0.95 / 0.65	*0.77 / 0.85 / 0.48*	*0.70 / 0.90 / 0.50*	*0.70 / 0.90 / 0.50*
	(0.50, 0.25)	0.84 / 0.97 / 0.78	0.78 / 0.91 / 0.60	0.80 / 0.95 / 0.69	0.80 / 0.95 / 0.65	0.80 / 0.86 / 0.51	0.79 / 0.88 / 0.55	0.79 / 0.88 / 0.55
	(0.50, 0.50)	0.80 / 0.97 / 0.74	0.74 / 0.91 / 0.56	0.76 / 0.95 / 0.66	0.76 / 0.95 / 0.66	*0.77 / 0.85 / 0.48*	0.76 / 0.88 / 0.51	0.76 / 0.88 / 0.51

Note: Red italicized cells refer to illegitimate BP strategies (combination of values were not valid)

Table 8.4: BP, Number of Invalid OOPs, listed as θ_1 / θ_2 / θ_1 & θ_2

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$
b=½	40	½	8 / 8 / 0	5 / 6 / 0	5 / 0 / 0	15 / 0 / 0	0 / 0 / 0	2 / 0 / 0
	40	1	3 / 3 / 0	3 / 2 / 0	5 / 0 / 0	7 / 0 / 0	0 / 0 / 0	1 / 0 / 0
	40	2	1 / 1 / 0	2 / 0 / 0	2 / 0 / 0	6 / 0 / 0	0 / 1 / 0	0 / 0 / 0
	60	½	1 / 2 / 0	2 / 0 / 0	2 / 0 / 0	2 / 0 / 0	0 / 0 / 0	1 / 0 / 0
	60	1	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	1 / 0 / 0	1 / 0 / 0	0 / 1 / 0
	60	2	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	2 / 0 / 0	0 / 0 / 0	0 / 0 / 0
	100	½	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	2 / 0 / 0	0 / 0 / 0	0 / 0 / 0
	100	1	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	1 / 0 / 0	0 / 0 / 0	0 / 0 / 0
	100	2	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0
	200	½	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0
	200	1	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0
	200	2	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0
b=1	40	½	172 / 180 / 1	209 / 215 / 0	284 / 16 / 0	488 / 1 / 0	34 / 43 / 0	71 / 68 / 0
	40	1	167 / 143 / 0	181 / 208 / 0	240 / 10 / 0	468 / 3 / 0	29 / 28 / 0	45 / 62 / 0
	40	2	142 / 149 / 0	204 / 191 / 0	248 / 12 / 0	417 / 8 / 0	30 / 34 / 0	71 / 52 / 0
	60	½	132 / 141 / 0	169 / 190 / 0	229 / 3 / 0	483 / 3 / 0	12 / 11 / 0	38 / 32 / 0
	60	1	137 / 116 / 0	174 / 182 / 0	197 / 3 / 0	441 / 0 / 0	9 / 10 / 0	34 / 36 / 0
	60	2	134 / 125 / 0	179 / 159 / 0	211 / 6 / 0	424 / 1 / 0	15 / 15 / 0	44 / 36 / 0
	100	½	89 / 98 / 0	150 / 146 / 0	171 / 0 / 0	467 / 0 / 0	3 / 2 / 0	8 / 20 / 0
	100	1	77 / 84 / 0	120 / 144 / 0	138 / 0 / 0	432 / 0 / 0	3 / 1 / 0	16 / 16 / 0
	100	2	80 / 86 / 0	153 / 143 / 0	140 / 1 / 0	417 / 1 / 0	5 / 3 / 0	16 / 16 / 0
	200	½	61 / 34 / 0	87 / 77 / 0	93 / 0 / 0	447 / 0 / 0	0 / 0 / 0	3 / 2 / 0
	200	1	24 / 23 / 0	68 / 81 / 0	79 / 0 / 0	399 / 0 / 0	0 / 0 / 0	3 / 2 / 0
	200	2	34 / 24 / 0	76 / 82 / 0	92 / 0 / 0	389 / 0 / 0	0 / 1 / 0	2 / 1 / 0

Table 8.5: BP, Median Bias in (θ_1, θ_2), listed as θ_1 / θ_2

	N	k	AUC = (0.7, 0.7) ρ = (0.0, 0.0)	AUC = (0.7, 0.7) ρ = (0.5, 0.5)	AUC = (0.75, 0.85) ρ = (0.0, 0.0)	AUC = (0.75, 0.85) ρ = (0.5, 0.5)	AUC = (0.9, 0.9) ρ = (0.0, 0.0)	AUC = (0.9, 0.9) ρ = (0.5, 0.5)
b=½	40	½	-0.020 / -0.036	-0.009 / 0.003	0.010 / -0.026	0.002 / -0.010	0.023 / -0.001	0.014 / 0.006
	40	1	-0.035 / -0.026	-0.070 / -0.033	-0.018 / -0.019	0.005 / -0.017	-0.039 / 0.005	-0.005 / 0.003
	40	2	-0.051 / -0.031	-0.050 / -0.032	-0.039 / -0.012	-0.028 / -0.046	-0.060 / -0.025	-0.007 / -0.025
	60	½	-0.013 / -0.021	-0.019 / -0.016	0.006 / -0.007	0.018 / -0.008	0.025 / -0.004	0.015 / 0.015
	60	1	-0.011 / -0.024	-0.016 / -0.011	-0.023 / -0.020	-0.011 / -0.013	-0.008 / -0.007	0.010 / -0.010
	60	2	-0.050 / -0.019	-0.016 / -0.027	-0.012 / -0.008	0.010 / -0.010	-0.011 / -0.030	-0.027 / -0.018
	100	½	-0.002 / -0.017	0.009 / -0.022	0.013 / -0.009	0.015 / -0.016	0.003 / -0.002	0.006 / 0.000
	100	1	-0.003 / -0.026	-0.011 / -0.007	-0.013 / -0.009	-0.010 / -0.003	-0.000 / -0.003	0.002 / 0.008
	100	2	-0.025 / -0.012	-0.030 / -0.033	-0.004 / -0.010	0.012 / -0.007	-0.030 / -0.007	0.007 / -0.021
	200	½	-0.008 / -0.005	-0.002 / -0.000	0.002 / 0.003	0.010 / -0.000	0.002 / -0.006	0.013 / -0.004
	200	1	-0.003 / -0.009	-0.009 / -0.008	0.003 / -0.002	0.001 / -0.005	-0.010 / -0.009	0.002 / 0.009
	200	2	0.001 / -0.012	-0.010 / -0.006	-0.026 / 0.005	0.003 / -0.011	-0.005 / -0.016	0.007 / 0.002
b=1	40	½	0.024 / -0.011	0.038 / -0.002	-0.270 / 0.122	-0.978 / 0.150	0.033 / 0.006	-0.002 / 0.065
	40	1	0.006 / -0.007	-0.007 / -0.000	-0.236 / 0.100	-0.999 / 0.107	-0.016 / 0.010	-0.012 / -0.002
	40	2	-0.012 / -0.019	0.006 / -0.033	-0.275 / 0.098	-1.114 / 0.114	0.004 / -0.026	-0.059 / -0.028
	60	½	0.042 / -0.014	0.061 / -0.025	-0.129 / 0.055	-0.830 / 0.120	0.016 / 0.004	0.018 / 0.044
	60	1	0.026 / -0.017	0.049 / 0.008	-0.198 / 0.075	-0.904 / 0.097	-0.018 / 0.002	-0.016 / 0.013
	60	2	-0.034 / -0.015	-0.012 / 0.017	0.193 / 0.065	-1.032 / 0.087	-0.008 / -0.019	-0.009 / -0.024
	100	½	0.008 / 0.008	0.070 / -0.002	-0.066 / 0.054	-0.825 / 0.087	0.024 / -0.017	-0.002 / 0.001
	100	1	0.020 / -0.019	0.033 / -0.009	-0.092 / 0.047	-0.769 / 0.057	0.011 / -0.022	-0.011 / 0.003
	100	2	-0.003 / -0.024	0.031 / -0.026	-0.108 / 0.041	-0.822 / 0.067	0.007 / -0.015	0.003 / -0.019
	200	½	-0.003 / 0.008	0.012 / 0.006	-0.034 / 0.016	-0.616 / 0.045	0.001 / 0.001	0.011 / -0.010
	200	1	0.002 / -0.005	0.048 / -0.010	-0.023 / 0.013	-0.625 / 0.043	-0.000 / 0.000	0.015 / -0.022
	200	2	-0.016 / 0.003	-0.008 / 0.012	-0.070 / 0.017	-0.709 / 0.038	-0.004 / -0.011	0.003 / -0.010

112

www.manaraa.com

Table 8.6: BP, Median RMSE in (θ₁, θ₂), listed as θ₁ / θ₂

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$
b=½	40	½	0.451 / 0.441	0.472 / 0.469	0.490 / 0.378	0.564 / 0.352	0.446 / 0.441	0.470 / 0.477
	40	1	0.439 / 0.440	0.451 / 0.464	0.501 / 0.385	0.565 / 0.343	0.441 / 0.442	0.450 / 0.448
	40	2	0.512 / 0.504	0.504 / 0.506	0.557 / 0.436	0.614 / 0.416	0.511 / 0.515	0.489 / 0.508
	60	½	0.356 / 0.365	0.386 / 0.380	0.405 / 0.305	0.473 / 0.277	0.353 / 0.343	0.362 / 0.364
	60	1	0.356 / 0.361	0.363 / 0.372	0.408 / 0.302	0.445 / 0.284	0.354 / 0.354	0.361 / 0.359
	60	2	0.399 / 0.410	0.417 / 0.405	0.442 / 0.354	0.509 / 0.315	0.415 / 0.412	0.399 / 0.400
	100	½	0.266 / 0.266	0.281 / 0.274	0.302 / 0.224	0.341 / 0.203	0.256 / 0.261	0.256 / 0.258
	100	1	0.268 / 0.275	0.281 / 0.277	0.301 / 0.234	0.331 / 0.210	0.266 / 0.266	0.267 / 0.265
	100	2	0.313 / 0.312	0.313 / 0.315	0.349 / 0.268	0.382 / 0.254	0.315 / 0.313	0.312 / 0.302
	200	½	0.184 / 0.181	0.185 / 0.184	0.202 / 0.155	0.223 0.144	0.180 / 0.177	0.181 / 0.177
	200	1	0.187 / 0.193	0.192 / 0.187	0.210 / 0.161	0.231 / 0.148	0.189 / 0.189	0.184 / 0.183
	200	2	0.218 / 0.221	0.217 / 0.217	0.240 / 0.191	0.259 / 0.175	0.219 / 0.222	0.209 / 0.211
b=1	40	½	0.529 / 0.529	0.584 / 0.589	0.597 / 0.390	1.130 / 0.352	0.485 / 0.471	0.572 / 0.600
	40	1	0.539 / 0.521	0.582 / 0.594	0.602 / 0.378	1.143 / 0.325	0.469 / 0.478	0.545 / 0.552
	40	2	0.558 / 0.554	0.582 / 0.567	0.631 / 0.394	1.244 / 0.361	0.518 / 0.517	0.557 / 0.562
	60	½	0.507 / 0.499	0.564 / 0.544	0.565 / 0.330	0.980 / 0.288	0.413 / 0.407	0.515 / 0.557
	60	1	0.496 / 0.497	0.507 / 0.510	0.576 / 0.319	1.033 / 0.286	0.402 / 0.411	0.489 / 0.500
	60	2	0.502 / 0.514	0.524 / 0.531	0.577 / 0.337	1.151 / 0.300	0.437 / 0.432	0.507 / 0.511
	100	½	0.477 / 0.464	0.511 / 0.506	0.497 / 0.250	0.964 / 0.218	0.331 / 0.317	0.443 / 0.446
	100	1	0.457 / 0.459	0.480 / 0.489	0.483 / 0.251	0.899 / 0.194	0.324 / 0.322	0.416 / 0.425
	100	2	0.475 / 0.477	0.490 / 0.490	0.514 / 0.264	0.958 / 0.211	0.346 / 0.355	0.435 / 0.428
	200	½	0.393 / 0.396	0.459 / 0.464	0.396 / 0.176	0.760 / 0.137	0.228 / 0.228	0.320 / 0.316
	200	1	0.385 / 0.395	0.446 / 0.441	0.396 / 0.171	0.779 / 0.124	0.218 / 0.219	0.295 / 0.288
	200	2	0.396 / 0.406	0.429 / 0.437	0.419 / 0.185	0.849 / 0.137	0.252 / 0.257	0.304 / 0.297

Table 8.7: BP, Median Bias in Sensitivity and Specificity for (θ₁, θ₂), listed as Se / Sp

			AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
	N	k	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$
b=½	40	½	0.021 / 0.002	0.018 / 0.004	0.017 / 0.005	0.013 / 0.007	0.014 / 0.003	0.014 / 0.005
	40	1	0.018 / 0.003	0.023 / 0.006	0.015 / 0.006	0.012 / 0.004	0.005 / 0.002	0.011 / 0.005
	40	2	0.019 / 0.007	0.018 / 0.006	0.012 / 0.006	0.016 / 0.010	0.011 / 0.005	0.013 / 0.008
	60	½	0.010 / 0.001	0.016 / 0.002	0.013 / 0.003	0.011 / 0.004	0.010 / 0.002	0.013 / 0.003
	60	1	0.014 / 0.002	0.012 / 0.002	0.005 / 0.002	0.010 / 0.004	0.007 / 0.002	0.007 / 0.004
	60	2	0.014 / 0.004	0.010 / 0.004	0.008 / 0.003	0.011 / 0.005	0.005 / 0.002	0.008 / 0.006
	100	½	0.007 / 0.001	0.010 / 0.001	0.004 / 0.001	0.007 / 0.002	0.004 / 0.000	0.006 / 0.002
	100	1	0.008 / 0.001	0.006 / 0.001	0.007 / 0.002	0.007 / 0.003	0.003 / -0.000	0.005 / 0.003
	100	2	0.007 / 0.002	0.012 / 0.004	0.005 / 0.002	0.005 / 0.003	0.004 / 0.002	0.003 / 0.003
	200	½	0.004 / 0.001	-0.000 / 0.000	0.005 / 0.002	0.003 / 0.002	0.003 / 0.000	-0.001 / 0.000
	200	1	0.005 / 0.000	0.006 / 0.001	0.004 / 0.001	0.003 / 0.002	0.003 / 0.001	0.001 / 0.001
	200	2	0.004 / 0.000	0.004 / 0.001	0.002 / 0.001	0.003 / 0.001	0.002 / 0.000	0.001 / 0.002
b=1	40	½	0.063 / 0.029	0.081 / 0.024	0.036 / 0.017	0.035 / 0.020	0.019 / 0.011	0.027 / 0.016
	40	1	0.057 / 0.032	0.063 / 0.028	0.028 / 0.015	0.023 / 0.027	0.016 / 0.015	0.020 / 0.016
	40	2	0.047 / 0.043	0.060 / 0.041	0.030 / 0.027	0.022 / 0.021	0.017 / 0.017	0.021 / 0.026
	60	½	0.047 / 0.010	0.057 / 0.017	0.024 / 0.011	0.026 / 0.013	0.015 / 0.007	0.018 / 0.009
	60	1	0.033 / 0.015	0.034 / 0.031	0.021 / 0.008	0.016 / 0.013	0.010 / 0.009	0.017 / 0.014
	60	2	0.034 / 0.027	0.035 / 0.038	0.016 / 0.014	0.011 / 0.014	0.011 / 0.013	0.014 / 0.019
	100	½	0.030 / 0.006	0.037 / 0.016	0.017 / 0.006	0.012 / 0.008	0.008 / 0.005	0.011 / 0.005
	100	1	0.023 / 0.017	0.022 / 0.018	0.011 / 0.006	0.012 / 0.008	0.006 / 0.005	0.011 / 0.009
	100	2	0.019 / 0.019	0.029 / 0.023	0.011 / 0.008	0.013 / 0.006	0.007 / 0.007	0.010 / 0.009
	200	½	0.018 / 0.004	0.012 / 0.004	0.007 / 0.004	0.007 / 0.004	0.004 / 0.003	0.005 / 0.001
	200	1	0.012 / 0.005	0.010 / 0.009	0.005 / 0.006	0.005 / 0.005	0.003 / 0.002	0.005 / 0.003
	200	2	0.012 / 0.006	0.014 / 0.013	0.007 / 0.005	0.005 / 0.005	0.002 / 0.003	0.007 / 0.004

Table 8.8: BP, Median RMSE in Sensitivity and Specificity for (θ_1, θ_2), listed as Se / Sp

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	ρ = (0.0, 0.0)	ρ = (0.5, 0.5)
b=½	40	½	0.110 / 0.043	0.131 / 0.048	0.083 / 0.033	0.101 / 0.037	0.052 / 0.023	0.076 / 0.028
	40	1	0.097 / 0.039	0.116 / 0.042	0.075 / 0.031	0.090 / 0.034	0.048 / 0.021	0.067 / 0.026
	40	2	0.093 / 0.040	0.105 / 0.043	0.070 / 0.033	0.084 / 0.037	0.045 / 0.022	0.061 / 0.027
	60	½	0.094 / 0.034	0.109 / 0.038	0.071 / 0.027	0.089 / 0.029	0.046 / 0.019	0.064 / 0.022
	60	1	0.082 / 0.031	0.094 / 0.034	0.061 / 0.025	0.075 / 0.028	0.040 / 0.017	0.056 / 0.021
	60	2	0.076 / 0.032	0.087 / 0.035	0.057 / 0.027	0.071 / 0.031	0.038 / 0.018	0.051 / 0.022
	100	½	0.072 / 0.025	0.084 / 0.027	0.056 / 0.020	0.067 / 0.022	0.036 / 0.014	0.050 / 0.017
	100	1	0.061 / 0.024	0.071 / 0.025	0.047 / 0.019	0.058 / 0.021	0.031 / 0.014	0.043 / 0.017
	100	2	0.058 / 0.025	0.068 / 0.027	0.045 / 0.020	0.054 / 0.024	0.030 / 0.014	0.041 / 0.018
	200	½	0.051 / 0.017	0.059 / 0.019	0.039 / 0.014	0.049 / 0.015	0.026 / 0.010	0.036 / 0.012
	200	1	0.043 / 0.016	0.051 / 0.018	0.034 / 0.014	0.042 / 0.015	0.023 / 0.010	0.031 / 0.012
	200	2	0.041 / 0.017	0.046 / 0.019	0.032 / 0.015	0.039 / 0.016	0.021 / 0.010	0.029 / 0.013
b=1	40	½	0.110 / 0.112	0.131 / 0.124	0.076 / 0.076	0.089 / 0.079	0.050 / 0.053	0.064 / 0.061
	40	1	0.105 / 0.112	0.124 / 0.134	0.070 / 0.075	0.082 / 0.078	0.047 / 0.055	0.059 / 0.063
	40	2	0.102 / 0.122	0.121 / 0.142	0.070 / 0.084	0.081 / 0.090	0.046 / 0.059	0.060 / 0.067
	60	½	0.093 / 0.097	0.114 / 0.106	0.063 / 0.063	0.074 / 0.065	0.043 / 0.045	0.056 / 0.052
	60	1	0.084 / 0.096	0.100 / 0.107	0.056 / 0.063	0.067 / 0.066	0.038 / 0.045	0.050 / 0.052
	60	2	0.083 / 0.101	0.102 / 0.118	0.055 / 0.069	0.065 / 0.073	0.038 / 0.050	0.048 / 0.058
	100	½	0.075 / 0.076	0.092 / 0.086	0.050 / 0.049	0.058 / 0.052	0.034 / 0.034	0.043 / 0.040
	100	1	0.067 / 0.076	0.082 / 0.087	0.044 / 0.050	0.053 / 0.052	0.031 / 0.035	0.039 / 0.041
	100	2	0.066 / 0.081	0.085 / 0.096	0.044 / 0.056	0.052 / 0.059	0.030 / 0.040	0.039 / 0.045
	200	½	0.054 / 0.055	0.067 / 0.064	0.036 / 0.035	0.043 / 0.037	0.024 / 0.024	0.031 / 0.028
	200	1	0.047 / 0.053	0.059 / 0.063	0.032 / 0.035	0.037 / 0.038	0.021 / 0.025	0.028 / 0.029
	200	2	0.047 / 0.059	0.060 / 0.067	0.031 / 0.040	0.036 / 0.043	0.021 / 0.028	0.027 / 0.032

Table 8.9: BP, Percent of Overestimation of Sensitivity and Specificity, listed as % Se / % Sp

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$
b=½	40	½	58.54 / 52.95	57.53 / 54.60	58.19 / 56.78	57.56 / 59.90	64.50 / 56.00	59.42 / 58.12
	40	1	57.65 / 54.33	57.69 / 57.19	58.79 / 59.40	56.19 / 56.39	55.80 / 53.60	56.66 / 57.06
	40	2	59.92 / 58.62	59.72 / 57.21	57.41 / 57.92	58.45 / 62.07	60.06 / 59.06	60.60 / 61.80
	60	½	55.07 / 51.05	57.21 / 53.01	58.82 / 55.31	55.51 / 56.01	60.90 / 55.20	58.36 / 57.26
	60	1	56.70 / 52.90	55.00 / 53.00	54.20 / 54.00	56.86 / 55.46	57.86 / 55.86	54.55 / 58.16
	60	2	60.50 / 56.00	56.50 / 55.60	58.00 / 57.60	57.41 / 57.41	56.30 / 55.70	57.40 / 61.70
	100	½	54.20 / 52.40	56.10 / 53.20	53.60 / 52.90	54.71 / 52.81	55.60 / 51.20	55.30 / 57.00
	100	1	57.30 / 53.50	54.60 / 51.70	57.20 / 55.60	56.16 / 57.06	56.40 / 48.80	55.10 / 58.80
	100	2	57.60 / 55.60	58.10 / 57.10	55.20 / 55.00	55.10 / 54.50	57.10 / 56.00	53.50 / 59.60
	200	½	53.80 / 52.20	49.90 / 51.30	56.10 / 54.30	52.30 / 56.00	56.30 / 51.20	48.90 / 52.20
	200	1	55.20 / 50.50	55.20 / 51.90	54.60 / 55.50	54.00 / 56.40	56.00 / 52.50	51.20 / 55.70
	200	2	55.90 / 50.60	54.90 / 53.90	52.90 / 53.20	54.20 / 54.00	54.50 / 50.60	52.30 / 57.40
b=1	40	½	76.27 / 61.63	77.95 / 57.47	69.43 / 60.14	67.32 / 63.01	65.66 / 61.00	69.57 / 63.18
	40	1	73.77 / 63.48	72.83 / 59.90	68.53 / 60.27	63.52 / 64.46	64.79 / 62.57	65.29 / 62.71
	40	2	70.38 / 67.70	72.07 / 62.81	71.49 / 65.95	64.52 / 61.57	64.74 / 63.46	65.91 / 67.62
	60	½	72.08 / 55.85	71.45 / 56.79	67.19 / 57.55	64.79 / 60.70	64.69 / 57.01	63.87 / 56.56
	60	1	69.61 / 58.23	68.32 / 64.29	66.75 / 56.63	59.57 / 61.90	61.06 / 60.86	66.67 / 61.29
	60	2	70.45 / 62.48	65.86 / 62.99	64.50 / 59.39	60.52 / 57.04	63.20 / 63.09	62.72 / 65.22
	100	½	68.88 / 53.75	68.75 / 57.95	65.50 / 57.42	59.47 / 57.60	60.60 / 57.19	61.52 / 56.69
	100	1	66.51 / 59.48	63.59 / 60.05	64.39 / 55.68	60.04 / 58.27	56.73 / 56.73	62.29 / 59.19
	100	2	66.55 / 60.79	66.05 / 62.93	63.91 / 58.32	61.51 / 55.67	60.69 / 58.06	60.95 / 58.88
	200	½	66.74 / 53.37	59.93 / 52.75	61.52 / 55.24	56.06 / 54.25	58.30 / 54.40	57.19 / 51.66
	200	1	62.85 / 53.52	57.34 / 56.52	56.03 / 58.85	58.90 / 55.57	55.90 / 52.70	57.99 / 55.38
	200	2	62.31 / 57.22	58.43 / 59.50	61.01 / 55.40	58.59 / 55.32	55.96 / 54.45	60.98 / 57.17

Table 8.10: BP, Coverage for SCIs and CRs for (θ_1, θ_2), listed as SCI / CR

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$
b=½	40	½	90.55 / 91.36	91.71 / 92.72	92.16 / 93.57	91.27 / 92.39	93.30 / 94.80	93.98 / 94.29
	40	1	88.53 / 89.74	88.54 / 89.25	89.15 / 90.95	91.24 / 91.84	91.20 / 91.80	90.59 / 91.09
	40	2	84.87 / 85.77	86.57 / 87.47	87.47 / 87.78	85.01 / 84.91	86.99 / 87.99	88.40 / 89.40
	60	½	92.98 / 93.98	94.69 / 94.49	91.58 / 92.59	92.89 / 93.69	94.40 / 94.90	95.70 / 96.20
	60	1	90.90 / 91.36	91.90 / 92.80	91.60 / 92.80	90.59 / 91.79	92.29 / 93.09	91.19 / 92.39
	60	2	89.00 / 89.00	89.70 / 90.40	90.40 / 90.20	90.08 / 90.38	88.30 / 88.50	90.80 / 90.30
	100	½	93.50 / 93.70	93.80 / 94.10	92.30 / 93.30	93.89 / 94.39	94.70 / 95.20	95.60 / 96.90
	100	1	93.70 / 93.60	94.10 / 93.50	92.40 / 92.00	93.49 / 94.09	94.30 / 94.60	94.30 / 93.60
	100	2	90.90 / 91.10	90.10 / 91.00	91.90 / 91.70	90.50 / 92.00	91.50 / 90.90	92.50 / 93.20
	200	½	95.50 / 94.90	94.90 / 95.30	94.40 / 94.50	95.30 / 95.50	94.00 / 94.00	95.60 / 95.20
	200	1	94.10 / 93.60	94.10 / 94.20	93.70 / 92.70	95.10 / 94.80	93.60 / 93.00	94.50 / 93.70
	200	2	92.60 / 92.70	93.80 / 94.10	93.80 / 94.30	93.40 / 92.30	93.20 / 93.10	94.60 / 95.40
b=1	40	½	83.51 / 85.82	83.85 / 85.42	84.43 / 87.57	61.25 / 66.14	92.63 / 95.12	89.08 / 90.01
	40	1	85.36 / 87.39	84.94 / 88.54	84.67 / 85.07	63.89 / 67.67	89.50 / 92.05	88.80 / 91.60
	40	2	85.61 / 88.29	83.97 / 85.62	82.43 / 84.05	59.13 / 61.39	87.93 / 89.74	85.18 / 89.05
	60	½	86.93 / 89.13	85.49 / 88.61	88.02 / 87.76	65.18 / 66.73	90.38 / 92.73	91.51 / 91.94
	60	1	88.49 / 89.42	89.13 / 91.30	87.63 / 87.38	62.79 / 63.15	92.25 / 92.86	91.51 / 93.33
	60	2	88.80 / 89.61	87.16 / 88.52	87.61 / 87.36	59.48 / 59.13	88.87 / 91.13	86.96 / 89.67
	100	½	91.64 / 92.99	87.50 / 90.91	91.31 / 90.95	67.92 / 68.67	94.47 / 94.47	91.46 / 93.52
	100	1	89.87 / 91.42	88.59 / 90.22	91.53 / 92.23	70.25 / 69.19	93.98 / 94.48	91.12 / 91.63
	100	2	87.53 / 90.05	86.51 / 87.78	89.99 / 89.41	67.35 / 66.49	92.44 / 92.24	89.77 / 91.32
	200	½	92.38 / 93.04	90.19 / 90.43	93.38 / 92.61	75.95 / 76.13	94.40 / 94.90	94.27 / 94.97
	200	1	91.82 / 92.86	91.19 / 91.42	92.73 / 92.29	74.04 / 74.04	94.40 / 93.90	93.27 / 94.27
	200	2	91.61 / 92.68	89.43 / 90.02	92.18 / 91.41	72.18 / 69.56	93.29 / 93.69	92.08 / 92.38

Table 8.11: BP, Median SCI Widths for (θ₁, θ₂), listed as θ₁ / θ₂

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	ρ = (0.0, 0.0)	ρ = (0.5, 0.5)
b=½	40	½	1.55 / 1.51	1.71 / 1.66	1.79 / 1.29	2.08 / 1.23	1.58 / 1.57	1.64 / 1.69
	40	1	1.51 / 1.53	1.55 / 1.60	1.73 / 1.31	2.03 / 1.21	1.51 / 1.52	1.55 / 1.54
	40	2	1.69 / 1.66	1.69 / 1.68	1.86 / 1.49	2.06 / 1.34	1.70 / 1.70	1.66 / 1.69
	60	½	1.26 / 1.28	1.33 / 1.34	1.43 / 1.05	1.72 / 0.97	1.24 / 1.23	1.27 / 1.28
	60	1	1.26 / 1.24	1.25 / 1.27	1.41 / 1.06	1.58 / 0.98	1.24 / 1.23	1.25 / 1.25
	60	2	1.36 / 1.39	1.41 / 1.40	1.54 / 1.23	1.75 / 1.10	1.40 / 1.39	1.36 / 1.34
	100	½	0.95 / 0.95	0.99 / 0.96	1.08 / 0.79	1.24 / 0.73	0.94 / 0.92	0.94 / 0.93
	100	1	0.96 / 0.96	0.97 / 0.98	1.07 / 0.83	1.18 / 0.76	0.96 / 0.96	0.95 / 0.94
	100	2	1.08 / 1.11	1.06 / 1.09	1.21 / 0.96	1.35 / 0.86	1.10 / 1.09	1.07 / 1.05
	200	½	0.66 / 0.66	0.67 / 0.66	0.73 / 0.56	0.83 / 0.52	0.66 / 0.65	0.66 / 0.64
	200	1	0.68 / 0.68	0.69 / 0.69	0.75 / 0.59	0.83 / 0.54	0.69 / 0.68	0.67 / 0.68
	200	2	0.79 / 0.79	0.78 / 0.77	0.86 / 0.69	0.95 / 0.62	0.79 / 0.79	0.76 / 0.76
b=1	40	½	1.81 / 1.78	1.98 / 2.06	1.82 / 1.42	2.42 / 1.24	1.71 / 1.67	2.15 / 2.23
	40	1	1.81 / 1.77	2.03 / 2.04	1.86 / 1.34	2.48 / 1.14	1.65 / 1.69	2.04 / 2.06
	40	2	1.88 / 1.89	1.98 / 1.93	1.95 / 1.45	2.50 / 1.31	1.79 / 1.77	1.99 / 2.00
	60	½	1.75 / 1.71	1.94 / 1.95	1.78 / 1.20	2.33 / 0.98	1.44 / 1.47	1.99 / 2.11
	60	1	1.77 / 1.71	1.88 / 1.88	1.78 / 1.14	2.28 / 1.00	1.41 / 1.48	1.82 / 1.85
	60	2	1.78 / 1.77	1.84 / 1.91	1.84 / 1.25	2.34 / 1.05	1.54 / 1.55	1.90 / 1.83
	100	½	1.66 / 1.64	1.87 / 1.83	1.65 / 0.88	2.25 / 0.76	1.18 / 1.13	1.69 / 1.69
	100	1	1.64 / 1.59	1.78 / 1.79	1.64 / 0.89	2.18 / 0.71	1.16 / 1.13	1.53 / 1.58
	100	2	1.68 / 1.66	1.79 / 1.74	1.70 / 0.95	2.21 / 0.75	1.23 / 1.25	1.59 / 1.53
	200	½	1.41 / 1.39	1.69 / 1.71	1.39 / 0.61	1.99 / 0.47	0.80 / 0.81	1.21 / 1.16
	200	1	1.40 / 1.38	1.65 / 1.66	1.40 / 0.60	1.98 / 0.46	0.80 / 0.79	1.09 / 1.03
	200	2	1.42 / 1.43	1.57 / 1.60	1.46 / 0.66	2.03 / 0.50	0.89 / 0.89	1.09 / 1.07

Table 8.12: BP, Median CR Area for (θ₁, θ₂)

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$
b=½	40	½	2.20	2.81	2.20	2.60	2.48	2,94
	40	1	2.19	2.53	2.14	2.42	2.24	2.52
	40	2	2.63	2.80	2.62	2.75	2.69	2.83
	60	½	1.44	1.76	1.39	1.64	1.50	1.67
	60	1	1.39	1.59	1.36	1.53	1.42	1.55
	60	2	1.73	1.94	1.70	1.87	1.75	1.81
	100	½	0.80	0.92	0.77	0.87	0.80	0.85
	100	1	0.82	0.91	0.79	0.87	0.82	0.86
	100	2	1.06	1.12	1.03	1.11	1.05	1.08
	200	½	0.38	0.42	0.36	0.40	0.38	0.39
	200	1	0.40	0.44	0.38	0.42	0.41	0.42
	200	2	0.53	0.57	0.51	0.55	0.54	0.54
b=1	40	½	2.56	3.60	2.07	2.69	2.33	3.80
	40	1	2.39	3.55	1.94	2.57	2.10	3.43
	40	2	2.69	3.31	2.10	2.79	2.33	3.17
	60	½	2.15	3.05	1.57	2.01	1.59	3.10
	60	1	2.07	2.79	1.46	1.87	1.46	2.59
	60	2	2.15	2.82	1.59	2.05	1.64	2.47
	100	½	1.72	2.47	1.02	1.42	0.97	1.96
	100	1	1.58	2.35	0.97	1.28	0.90	1.69
	100	2	1.71	2.31	1.07	1.39	1.01	1.70
	200	½	1.14	1.96	0.56	0.79	0.43	0.98
	200	1	1.06	1.77	0.53	0.74	0.39	0.79
	200	2	1.13	1.72	0.59	0.83	0.49	0.83

Table 8.13: BP, Median Bias in Empirical (θ_1, θ_2), listed as θ_1 / θ_2

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	ρ = (0.0, 0.0)	ρ = (0.5, 0.5)
b=½	40	½	0.517 / -0.469	-0.042 / -0.379	-0.632 / 0.197	-0.338 / -0.161	0.483 / 0.051	1.121 / -0.087
	40	1	-0.324 / -0.422	-0.189 / -0.414	-1.127 / -0.160	-0.046 / -0.310	-0.298 / 0.474	-0.325 / -0.063
	40	2	0.809 / -0.271	-0.405 / 0.361	-0.531 / 0.342	0.667 / 0.348	-0.065 / 0.146	1.099 / -0.502
	60	½	0.217 / 0.099	0.632 / -0.882	-0.064 / 0.005	-0.629 / 0.814	0.933 / -0.480	0.470 / 0.357
	60	1	-0.361 / -0.123	-0.959 / 0.400	-0.246 / -0.123	-0.244 / -0.501	-0.483 / 1.742	0.034 / -0.663
	60	2	-0.209 / -1.060	-0.406 / -0.610	-0.859 / 0.075	0.276 / -0.332	-0.620 / 0.880	-0.752 / 0.033
	100	½	0.196 / 0.248	0.400 / -0.150	-0.894 / 0.230	-0.193 / 0.809	1.179 / -0.179	0.211 / -0.036
	100	1	0.166 / 0.157	0.086 / -0.314	-0.217 / 0.059	-0.305 / 0.227	0.123 / 0.167	-0.756 / 0.364
	100	2	-0.183 / -0.061	0.729 / -0.071	0.739 / -0.081	0.246 / -0.265	-0.677 / -0.171	0.164 / 0.248
	200	½	0.104 / 0.088	0.415 / 0.342	-0.145 / 0.603	-0.313 / -0.352	0.376 / -0.200	0.156 / 0.210
	200	1	-0.430 / 0.069	-0.087 / -0.035	0.148 / 0.068	-0.380 / 0.028	-0.023 / -0.102	-0.113 / 0.257
	200	2	-0.608 / -0.032	-0.169 / 0.139	-0.401 / -0.108	-0.332 / 0.218	-0.053 / 0.487	0.016 / 0.185
b=1	40	½	-0.113 / -0.873	0.614 / -0.239	0.082 / -0.037	-0.979 / -0.199	-0.226 / 0.929	-0.001 / -0.774
	40	1	0.124 / -0.929	-0.407 / 0.494	-0.402 / -0.068	-1.039 / 0.366	0.153 / -0.157	-0.422 / 0.204
	40	2	-0.229 / 0.685	-0.863 / 0.483	-0.240 / -0.318	-1.030 / -0.026	-0.122 / 0.884	-0.821 / -0.367
	60	½	0.526 / -0.134	-0.558 / -0.192	-0.354 / 0.272	-1.197 / -0.247	0.230 / -0.246	0.789 / -0.052
	60	1	-0.023 / 0.232	0.099 / -0.142	-0.572 / 0.677	-0.855 / 0.418	-0.020 / 0.033	-0.001 / -0.774
	60	2	1.040 / -0.114	0.493 / 0.775	-0.331 / 0.730	-0.760 / 0.383	-0.427 / 0.134	-0.377 / 0.032
	100	½	-0.207 / 0.202	0.178 / 0.063	0.390 / -0.184	-0.272 / 0.534	-0.088 / 0.157	0.087 / 0.043
	100	1	-0.649 / 0.532	0.616 / -0.056	0.640 / -0.202	-1.311 / 0.003	0.247 / -0.301	-0.033 / 0.615
	100	2	0.441 / -0.293	-0.306 / 0.169	0.363 / -0.167	-1.003 / 0.217	0.128 / 0.107	-0.184 / 0.495
	200	½	-0.550 / 1.079	0.152 / -0.204	-0.418 / 0.596	-0.760 / 0.017	0.089 / -0.003	-0.048 / -0.016
	200	1	0.442 / 0.080	0.141 / -0.241	-0.071 / -0.125	-0.480 / 0.224	-0.130 / 0.095	-0.410 / 0.326
	200	2	0.501 / -0.046	-0.321 / 0.258	0.189 / -0.396	-0.400 / 0.060	-0.107 / 0.556	-0.221 / 0.639

Table 8.14: BP, Median Bias in Sensitivity and Specificity for Empirical (θ_1, θ_2), listed as Se / Sp

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$
b=½	40	½	0.045 / 0.002	0.043 / 0.039	-0.011 / 0.048	-0.095 / 0.016	-0.039 / 0.013	0.012 / 0.025
	40	1	0.080 / -0.182	0.150 / 0.016	0.103 / -0.064	0.162 / -0.007	0.004 / 0.001	0.055 / 0.014
	40	2	0.182 / 0.041	-0.007 / 0.116	0.018 / 0.048	-0.034 / 0.016	-0.007 / 0.051	-0.030 / 0.064
	60	½	0.130 / 0.018	0.100 / -0.159	0.153 / 0.036	-0.088 / -0.032	0.054 / 0.001	0.005 / 0.039
	60	1	0.064 / -0.016	0.233 / -0.051	-0.064 / -0.014	0.029 / -0.074	0.071 / -0.016	0.088 / -0.036
	60	2	0.130 / -0.182	0.175 / -0.134	0.103 / -0.114	0.037 / -0.007	0.054 / 0.051	0.030 / -0.086
	100	½	-0.052 / 0.057	0.165 / 0.071	-0.003 / -0.050	-0.120 / 0.048	0.016 / -0.010	0.037 / 0.003
	100	1	-0.010 / 0.018	0.040 / 0.056	0.003 / 0.006	-0.158 / 0.053	0.004 / -0.009	0.055 / 0.024
	100	2	0.047 / -0.004	-0.033 / 0.025	-0.036 / 0.026	0.038 / 0.032	-0.000 / 0.021	-0.024 / 0.064
	200	½	-0.102 / 0.035	-0.033 / 0.056	-0.006 / -0.011	-0.006 / -0.057	0.015 / -0.047	0.036 / 0.011
	200	1	0.060 / -0.002	-0.010 / 0.016	-0.007 / 0.006	0.052 / -0.017	-0.006 / -0.019	-0.045 / -0.006
	200	2	0.039 / -0.019	0.101 / 0.040	0.002 / 0.041	0.046 / 0.017	0.037 / 0.021	-0.031 / 0.003
b=1	40	½	0.310 / -0.467	0.052 / 0.117	0.139 / 0.037	0.161 / -0.077	0.053 / -0.047	0.088 / -0.106
	40	1	0.110 / -0.209	0.238 / -0.052	0.111 / 0.030	-0.218 / -0.069	0.075 / -0.055	0.059 / 0.113
	40	2	0.162 / 0.110	0.227 / -0.268	0.100 / 0.153	0.158 / 0.154	-0.060 / 0.068	-0.026 / 0.009
	60	½	0.010 / 0.116	0.338 / -0.102	-0.039 / 0.005	0.132 / -0.144	0.125 / -0.005	0.009 / -0.012
	60	1	0.077 / 0.107	0.071 / 0.048	0.044 / 0.030	-0.001 / 0.165	0.092 / 0.079	0.093 / -0.137
	60	2	0.035 / -0.059	-0.212 / 0.248	-0.164 / 0.180	0.032 / 0.081	0.075 / 0.095	0.009 / 0.113
	100	½	-0.013 / 0.038	0.014 / 0.136	0.034 / -0.013	-0.121 / 0.110	0.125 / 0.024	-0.076 / -0.064
	100	1	0.070 / -0.039	-0.002 / 0.148	0.031 / 0.130	0.032 / -0.029	0.045 / 0.065	-0.041 / 0.063
	100	2	-0.018 / 0.220	0.099 / 0.075	0.017 / -0.013	-0.066 / 0.019	0.020 / -0.006	-0.020 / 0.103
	200	½	0.086 / -0.050	0.054 / 0.077	-0.028 / 0.019	0.053 / 0.043	0.006 / -0.013	0.085 / 0.035
	200	1	-0.020 / 0.051	0.058 / 0.058	0.111 / -0.060	-0.028 / 0.061	0.015 / 0.035	0.079 / -0.017
	200	2	-0.011 / 0.220	0.129 / 0.075	0.076 / -0.058	0.061 / 0.095	0.006 / 0.024	-0.005 / 0.027

Table 8.15: BP, Percent of Overestimation of Empirical Sensitivity and Specificity, listed as % Se / % Sp

	N	k	AUC = (0.7, 0.7) ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	AUC = (0.75, 0.85) ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	AUC = (0.9, 0.9) ρ = (0.0, 0.0)	ρ = (0.5, 0.5)
b=½	40	½	67.8 / 50.3	88.6 / 70.1	37.3 / 62.1	0.0 / 84.0	42.5 / 63.4	57.7 / 85.2
	40	1	67.5 / 17.6	95.3 / 58.7	95.3 / 36.4	99.8 / 0.8	86.2 / 84.7	92.9 / 86.0
	40	2	92.4 / 84.3	9.4 / 96.9	53.3 / 72.3	18.7 / 89.9	31.3 / 91.7	31.5 / 70.0
	60	½	97.8 / 68.6	58.7 / 15.1	100.0 / 88.8	0.1 / 19.6	97.2 / 60.9	72.3 / 78.9
	60	1	83.1 / 34.9	100.0 / 17.7	0.0 / 28.4	53.9 / 29.9	99.1 / 43.7	85.0 / 14.5
	60	2	98.1 / 0.3	87.8 / 20.4	80.1 / 32.4	74.3 / 22.2	99.8 / 99.9	59.4 / 32.6
	100	½	6.0 / 80.7	100.0 / 90.0	48.2 / 20.7	0.2 / 80.9	84.9 / 46.6	94.8 / 54.7
	100	1	43.4 / 57.5	83.7 / 94.2	65.4 / 56.4	0.0 / 84.7	63.4 / 46.1	97.0 / 88.8
	100	2	82.1 / 41.2	20.8 / 57.5	15.0 / 73.8	84.6 / 95.3	41.7 / 63.2	4.9 / 98.5
	200	½	3.9 / 80.2	42.7 / 65.1	44.8 / 48.7	32.9 / 0.7	61.1 / 30.0	97.5 / 73.9
	200	1	85.2 / 48.4	27.5 / 62.4	44.1 / 55.4	98.0 / 29.2	49.4 / 36.2	11.6 / 34.3
	200	2	78.8 / 27.9	100.0 / 55.0	55.8 / 91.6	91.8 / 73.4	96.5 / 89.4	5.8 / 70.3
b=1	40	½	98.2 / 1.5	67.2 / 93.0	90.5 / 71.9	93.6 / 20.3	55.6 / 23.3	94.0 / 6.3
	40	1	77.3 / 20.2	98.8 / 26.9	98.9 / 64.1	2.4 / 24.7	77.1 / 20.2	95.1 / 99.0
	40	2	91.5 / 89.5	100.0 / 0.0	80.2 / 99.9	99.9 / 99.9	21.9 / 70.8	49.1 / 90.0
	60	½	50.4 / 87.7	100.0 / 0.0	33.0 / 57.8	85.5 / 13.2	97.4 / 44.4	81.4 / 40.1
	60	1	88.9 / 87.0	92.0 / 85.1	82.3 / 60.7	15.4 / 96.9	99.8 / 98.5	88.2 / 1.7
	60	2	50.7 / 48.1	5.5 / 94.5	20.0 / 79.2	97.5 / 96.7	99.3 / 100.0	68.6 / 100.0
	100	½	45.8 / 63.0	57.9 / 100.0	94.9 / 49.3	36.5 / 76.9	98.8 / 72.7	47.5 / 26.4
	100	1	90.0 / 12.6	42.9 / 98.2	76.8 / 98.1	70.9 / 39.1	99.4 / 99.4	2.9 / 97.6
	100	2	46.0 / 100.0	96.6 / 95.0	75.6 / 41.4	22.6 / 66.8	65.2 / 37.5	33.1 / 99.1
	200	½	84.7 / 27.0	99.8 / 74.0	31.7 / 70.8	89.1 / 86.7	60.7 / 35.5	100.0 / 92.6
	200	1	44.2 / 66.8	94.2 / 97.2	99.9 / 0.1	32.8 / 92.5	67.2 / 85.0	97.9 / 6.3
	200	2	46.5 / 100.0	97.3 / 91.6	76.9 / 25.7	99.2 / 99.3	55.6 / 70.7	36.0 / 76.6

Table 8.16: BN, Population Parameters' Optimal Points, listed as Se / Sp / GYI, AUC = (0.7, 0.7)

	ρ	b = (0.5, 0.5)	b = (1.0, 1.0)	b = (0.5, 1.0)	b = (1.0, 0.5)	b = (1.5, 1.5)	b = (1.0, 1.5)	b = (1.5, 1.0)
r = 0.5	(0.00, 0.00)	*1.00 / 0.00 / 0.50*	0.88 / 0.38 / 0.57	*0.90 / 0.29 / 0.55*	*0.90 / 0.29 / 0.55*	0.92 / 0.50 / 0.67	0.91 / 0.43 / 0.63	0.91 / 0.43 / 0.63
	(0.25, 0.50)	*1.00 / 0.00 / 0.50*	0.88 / 0.39 / 0.57	*0.90 / 0.29 / 0.55*	*0.90 / 0.29 / 0.55*	0.93 / 0.48 / 0.67	0.91 / 0.42 / 0.62	0.91 / 0.42 / 0.62
	(0.50, 0.25)	*1.00 / 0.00 / 0.50*	0.91 / 0.29 / 0.55	*0.90 / 0.29 / 0.55*	*0.90 / 0.29 / 0.55*	0.93 / 0.42 / 0.67	*0.94 / 0.35 / 0.61*	*0.94 / 0.35 / 0.61*
	(0.50, 0.50)	*1.00 / 0.00 / 0.50*	0.89 / 0.33 / 0.56	*0.90 / 0.29 / 0.55*	*0.90 / 0.29 / 0.55*	0.93 / 0.43 / 0.65	0.93 / 0.36 / 0.61	0.93 / 0.36 / 0.61
r = 1	(0.00, 0.00)	*0.50 / 0.88 / 0.38*	0.66 / 0.69 / 0.35	0.50 / 0.88 / 0.38	0.50 / 0.88 / 0.38	0.82 / 0.65 / 0.46	0.74 / 0.66 / 0.40	0.74 / 0.66 / 0.40
	(0.25, 0.50)	*0.50 / 0.88 / 0.38*	0.65 / 0.71 / 0.36	0.51 / 0.88 / 0.38	0.51 / 0.88 / 0.38	0.81 / 0.63 / 0.45	0.74 / 0.66 / 0.40	0.74 / 0.66 / 0.40
	(0.50, 0.25)	*0.50 / 0.88 / 0.38*	*0.64 / 0.65 / 0.29*	*0.50 / 0.88 / 0.38*	*0.50 / 0.88 / 0.38*	0.83 / 0.57 / 0.39	0.76 / 0.57 / 0.34	0.77 / 0.57 / 0.34
	(0.50, 0.50)	*0.50 / 0.88 / 0.38*	0.65 / 0.66 / 0.31	*0.50 / 0.88 / 0.38*	*0.50 / 0.88 / 0.38*	0.82 / 0.59 / 0.41	0.75 / 0.61 / 0.36	0.75 / 0.61 / 0.36
r = 2	(0.00, 0.00)	*0.40 / 0.95 / 0.31*	0.38 / 0.89 / 0.16	0.40 / 0.95 / 0.31	0.40 / 0.95 / 0.31	0.59 / 0.80 / 0.20	0.45 / 0.87 / 0.19	0.45 / 0.87 / 0.19
	(0.25, 0.50)	*0.40 / 0.95 / 0.31*	0.39 / 0.90 / 0.18	0.40 / 0.95 / 0.31	0.40 / 0.95 / 0.31	0.57 / 0.81 / 0.18	0.44 / 0.87 / 0.18	0.44 / 0.87 / 0.18
	(0.50, 0.25)	*0.40 / 0.95 / 0.31*	*0.29 / 0.90 / 0.09*	*0.40 / 0.95 / 0.31*	*0.40 / 0.95 / 0.31*	0.54 / 0.76 / 0.06	0.32 / 0.89 / 0.10	0.32 / 0.89 / 0.10
	(0.50, 0.50)	*0.40 / 0.95 / 0.31*	0.33 / 0.90 / 0.12	*0.40 / 0.95 / 0.31*	*0.40 / 0.95 / 0.31*	0.53 / 0.78 / 0.10	0.37 / 0.88 / 0.12	0.37 / 0.88 / 0.12

Note: Red italicized cells refer to illegitimate BN strategies (combination of values were not valid)

Table 8.17: BN, Population Parameters' Optimal Points, listed as Se / Sp / GYI, AUC = (0.75, 0.85)

	ρ	b = (0.5, 0.5)	b = (1.0, 1.0)	b = (0.5, 1.0)	b = (1.0, 0.5)	b = (1.5, 1.5)	b = (1.5, 1.0)	b = (1.0, 1.5)
r = 0.5	(0.00, 0.00)	*0.78 / 0.78 / 0.70*	0.88 / 0.63 / 0.70	*0.89 / 0.60 / 0.69*	0.78 / 0.78 / 0.67	0.93 / 0.68 / 0.77	0.90 / 0.66 / 0.73	0.92 / 0.64 / 0.74
	(0.25, 0.50)	*0.78 / 0.78 / 0.67*	0.88 / 0.63 / 0.69	*0.89 / 0.60 / 0.69*	0.78 / 0.78 / 0.67	0.93 / 0.66 / 0.76	0.90 / 0.65 / 0.72	0.93 / 0.62 / 0.74
	(0.50, 0.25)	*0.78 / 0.78 / 0.67*	*0.89 / 0.60 / 0.69*	*0.89 / 0.60 / 0.69*	*0.78 / 0.78 / 0.67*	0.93 / 0.62 / 0.74	0.90 / 0.60 / 0.70	*0.93 / 0.60 / 0.73*
	(0.50, 0.50)	*0.78 / 0.78 / 0.67*	*0.89 / 0.60 / 0.69*	0.89 / 0.60 / 0.69	*0.78 / 0.78 / 0.67*	0.93 / 0.62 / 0.75	0.90 / 0.61 / 0.70	*0.93 / 0.60 / 0.73*
r = 1	(0.00, 0.00)	*0.70 / 0.90 / 0.60*	0.77 / 0.79 / 0.56	*0.77 / 0.77 / 0.54*	0.70 / 0.90 / 0.60	0.87 / 0.78 / 0.64	0.81 / 0.79 / 0.60	0.83 / 0.76 / 0.59
	(0.25, 0.50)	*0.70 / 0.90 / 0.60*	0.77 / 0.79 / 0.55	*0.77 / 0.77 / 0.54*	0.70 / 0.90 / 0.60	0.86 / 0.76 / 0.62	0.80 / 0.79 / 0.59	0.83 / 0.75 / 0.58
	(0.50, 0.25)	*0.70 / 0.90 / 0.60*	*0.77 / 0.77 / 0.54*	*0.77 / 0.77 / 0.54*	*0.70 / 0.90 / 0.60*	0.86 / 0.72 / 0.58	0.78 / 0.76 / 0.55	*0.85 / 0.71 / 0.56*
	(0.50, 0.50)	*0.70 / 0.90 / 0.60*	0.77 / 0.77 / 0.54	*0.77 / 0.77 / 0.54*	*0.70 / 0.90 / 0.60*	0.86 / 0.72 / 0.58	0.79 / 0.77 / 0.56	0.85 / 0.71 / 0.56
r = 2	(0.00, 0.00)	*0.62 / 0.95 / 0.53*	0.62 / 0.90 / 0.41	0.51 / 0.95 / 0.41	0.62 / 0.95 / 0.53	0.75 / 0.86 / 0.47	0.67 / 0.89 / 0.46	0.68 / 0.87 / 0.42
	(0.25, 0.50)	*0.62 / 0.96 / 0.53*	0.61 / 0.90 / 0.41	0.52 / 0.94 / 0.41	0.62 / 0.95 / 0.53	0.73 / 0.85 / 0.43	0.66 / 0.89 / 0.44	0.67 / 0.87 / 0.40
	(0.50, 0.25)	*0.62 / 0.95 / 0.53*	*0.60 / 0.89 / 0.37*	*0.60 / 0.89 / 0.37*	*0.62 / 0.95 / 0.53*	0.71 / 0.82 / 0.36	0.62 / 0.88 / 0.38	*0.69 / 0.82 / 0.33*
	(0.50, 0.50)	*0.62 / 0.95 / 0.53*	0.60 / 0.89 / 0.38	0.48 / 0.95 / 0.38	*0.62 / 0.95 / 0.53*	0.71 / 0.83 / 0.37	0.62 / 0.88 / 0.39	0.67 / 0.84 / 0.34

Note: Red italicized cells refer to illegitimate BN strategies (combination of values were not valid)

Table 8.18: BN, Population Parameters' Optimal Points, listed as Se / Sp / GYI, AUC = (0.9, 0.9)

	ρ	b = (0.5, 0.5)	b = (1.0, 1.0)	b = (0.5, 1.0)	b = (1.0, 0.5)	b = (1.5, 1.5)	b = (1.5, 1.0)	b = (1.0, 1.5)
r = 0.5	(0.00, 0.00)	*0.83 / 0.83 / 0.75*	0.91 / 0.79 / 0.81	0.86 / 0.83 / 0.77	0.86 / 0.83 / 0.77	0.95 / 0.83 / 0.87	0.94 / 0.81 / 0.84	0.94 / 0.81 / 0.84
	(0.25, 0.50)	*0.83 / 0.83 / 0.75*	0.91 / 0.78 / 0.80	0.86 / 0.82 / 0.77	0.86 / 0.82 / 0.77	0.95 / 0.80 / 0.85	0.93 / 0.79 / 0.83	0.93 / 0.78 / 0.83
	(0.50, 0.25)	*0.83 / 0.83 / 0.75*	0.91 / 0.73 / 0.77	*0.90 / 0.70 / 0.75*	*0.90 / 0.70 / 0.75*	0.95 / 0.76 / 0.83	0.93 / 0.74 / 0.80	0.93 / 0.74 / 0.80
	(0.50, 0.50)	*0.83 / 0.83 / 0.75*	0.91 / 0.73 / 0.78	0.84 / 0.82 / 0.75	0.84 / 0.82 / 0.75	0.95 / 0.77 / 0.83	0.93 / 0.75 / 0.81	0.93 / 0.75 / 0.81
r = 1	(0.00, 0.00)	*0.77 / 0.92 / 0.69*	0.85 / 0.88 / 0.73	0.80 / 0.92 / 0.72	0.80 / 0.92 / 0.72	0.92 / 0.88 / 0.80	0.89 / 0.87 / 0.76	0.89 / 0.87 / 0.76
	(0.25, 0.50)	0.76 / 0.93 / 0.69	0.85 / 0.87 / 0.71	0.79 / 0.92 / 0.71	0.79 / 0.92 / 0.71	0.91 / 0.86 / 0.77	0.88 / 0.86 / 0.74	0.88 / 0.86 / 0.74
	(0.50, 0.25)	*0.77 / 0.92 / 0.69*	0.83 / 0.83 / 0.67	0.77 / 0.92 / 0.69	0.77 / 0.92 / 0.69	0.90 / 0.82 / 0.73	0.87 / 0.83 / 0.70	0.87 / 0.83 / 0.70
	(0.50, 0.50)	*0.77 / 0.92 / 0.69*	0.84 / 0.84 / 0.68	0.77 / 0.92 / 0.69	0.77 / 0.92 / 0.69	0.90 / 0.83 / 0.73	0.87 / 0.83 / 0.70	0.87 / 0.83 / 0.70
r = 2	(0.00, 0.00)	0.70 / 0.96 / 0.63	0.78 / 0.93 / 0.64	0.74 / 0.96 / 0.66	0.74 / 0.96 / 0.66	0.86 / 0.92 / 0.70	0.82 / 0.92 / 0.67	0.82 / 0.92 / 0.67
	(0.25, 0.50)	0.71 / 0.96 / 0.64	0.76 / 0.93 / 0.61	0.73 / 0.96 / 0.65	0.73 / 0.96 / 0.65	0.84 / 0.91 / 0.66	0.80 / 0.92 / 0.64	0.80 / 0.92 / 0.64
	(0.50, 0.25)	*0.71 / 0.96 / 0.63*	0.73 / 0.91 / 0.54	*0.71 / 0.96 / 0.63*	*0.71 / 0.96 / 0.63*	0.82 / 0.88 / 0.58	0.77 / 0/90 / 0.56	0.77 / 0.90 / 0.56
	(0.50, 0.50)	*0.71 / 0.96 / 0.63*	0.74 / 0.91 / 0.56	0.71 / 0.96 / 0.63	0.71 / 0.96 / 0.63	0.82 / 0.89 / 0.60	0.77 / 0.90 / 0.58	0.77 / 0.90 / 0.58

Note: Red italicized cells refer to illegitimate BN strategies (combination of values were not valid)

Table 8.19: BN, Number of Times Invalid OOPs, listed as θ_3 / θ_4 / θ_3 & θ_4

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$
b=1	40	½	160 / 149 / 0	199 / 189 / 1	223 / 12 / 0	425 / 9 / 0	33 / 30 / 0	55 / 64 / 0
	40	1	150 / 151 / 0	184 / 198 / 0	247 / 6 / 0	433 / 4 / 0	31 / 26 / 0	48 / 61 / 0
	40	2	183 / 164 / 0	222 / 201 / 0	285 / 24 / 0	461 / 5 / 0	41 / 26 / 0	52 / 69 / 0
	60	½	157 / 112 / 0	180 / 156 / 0	193 / 7 / 0	410 / 2 / 0	15 / 12 / 0	39 / 37 / 0
	60	1	112 / 123 / 0	182 / 168 / 0	207 / 3 / 0	434 / 9 / 0	11 / 11 / 0	31 / 26 / 0
	60	2	128 / 139 / 0	172 / 164 / 0	221 / 4 / 0	450 / 0 / 0	14 / 16 / 0	34 / 36 / 0
	100	½	73 / 70 / 0	135 / 133 / 0	162 / 0 / 0	352 / 1 / 0	4 / 3 / 0	16 / 16 / 0
	100	1	87 / 74 / 0	136 / 143 / 0	144 / 0 / 0	410 / 0 / 0	3 / 1 / 0	13 / 17 / 0
	100	2	104 / 108 / 0	161 / 147 / 0	176 / 1 / 0	431 / 0 / 0	1 / 7 / 0	24 / 7 / 0
	200	½	34 / 34 / 0	78 / 78 / 0	90 / 0 / 0	350 / 0 / 0	0 / 0 / 0	2 / 4 / 0
	200	1	23 / 33 / 0	77 / 58 / 0	71 / 0 / 0	363 / 0 / 0	0 / 0 / 0	0 / 2 / 0
	200	2	34 / 40 / 0	86 / 100 / 0	92 / 0 / 0	402 / 0 / 0	0 / 0 / 0	5 / 0 / 0
b=1½	40	½	17 / 16 / 0	27 / 23 / 0	26 / 3 / 0	57 / 0 / 0	1 / 0 / 0	1 / 1 / 0
	40	1	18 / 14 / 0	27 / 24 / 0	28 / 0 / 0	55 / 1 / 0	1 / 0 / 0	2 / 4 / 0
	40	2	32 / 50 / 0	50 / 44 / 1	41 / 4 / 0	72 / 1 / 0	2 / 5 / 0	6 / 3 / 0
	60	½	2 / 7 / 0	3 / 9 / 0	11 / 0 / 0	28 / 0 / 0	0 / 0 / 0	1 / 0 / 0
	60	1	2 / 9 / 0	10 / 15 / 0	13 / 0 / 0	18 / 0 / 0	0 / 0 / 0	0 / 2 / 0
	60	2	17 / 10 / 0	15 / 15 / 0	20 / 0 / 0	39 / 1 / 0	0 / 0 / 0	2 / 0 / 0
	100	½	0 / 0 / 0	1 / 1 / 0	0 / 0 / 0	9 / 0 / 0	0 / 0 / 0	0 / 0 / 0
	100	1	0 / 0 / 0	1 / 0 / 0	2 / 0 / 0	7 / 0 / 0	0 / 0 / 0	0 / 0 / 0
	100	2	1 / 0 / 0	2 / 0 / 0	1 / 0 / 0	8 / 0 / 0	0 / 0 / 0	0 / 0 / 0
	200	½	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0
	200	1	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	1 / 0 / 0	0 / 0 / 0	0 / 0 / 0
	200	2	0 / 0 / 0	1 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0

Table 8.20: BN, Median Bias in (θ_3, θ_4), listed as θ_3 / θ_4

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	ρ = (0.0, 0.0)	ρ = (0.5, 0.5)
b=1	40	½	0.044 / 0.036	0.012 / -0.030	0.295 / -0.092	0.927 / 0.108	0.015 / 0.039	0.031 / -0.035
	40	1	-0.021 / 0.007	-0.006 / -0.062	0.252 / -0.084	0.860 / -0.102	0.004 / -0.011	-0.004 / -0.013
	40	2	-0.044 / 0.024	-0.088 / -0.032	0.183 / -0.106	0.826 / -0.121	-0.007 / -0.030	-0.029 / -0.017
	60	½	-0.009 / 0.014	-0.023 / 0.053	0.214 / -0.076	0.924 / -0.084	0.029 / 0.004	0.035 / 0.006
	60	1	-0.083 / 0.048	-0.004 / 0.002	0.204 / -0.090	0.817 / -0.083	-0.003 / -0.014	-0.003 / -0.005
	60	2	-0.023 / 0.013	-0.047 / -0.021	0.163 / -0.076	0.724 / -0.114	-0.027 / 0.004	-0.027 / -0.010
	100	½	0.006 / 0.002	-0.002 / 0.011	0.097 / -0.040	0.705 / -0.068	-0.001 / 0.030	0.008 / -0.005
	100	1	-0.019 / 0.020	-0.028 / -0016	0.114 / -0.040	0.562 / -0.052	-0.015 / 0.018	0.003 / -0.001
	100	2	-0.048 / -0.009	-0.044 / -0.022	0.111 / -0.051	0.605 / -0.074	-0.007 / 0.011	-0.016 / -0.005
	200	½	-0.021 / 0.020	-0.004 / 0.005	0.063 / -0.016	0.439 / -0.036	0.010 / 0.007	-0.014 / 0.025
	200	1	-0.023 / 0.016	0.002 / -0.012	0.024 / -0.007	0.420 / -0.033	0.005 / -0.000	-0.010 / 0.016
	200	2	-0.010 / 0.001	0.010 / -0.028	0.029 / -0.031	0.384 / -0.045	-0.002 / 0.006	-0.000 / 0.013
b=1½	40	½	0.024 / 0.027	0.024 / 0.022	0.039 / 0.016	0.009 / -0.008	0.023 / 0.024	0.014 / 0.019
	40	1	0.014 / 0.013	0.025 / -0.002	0.033 / -0.000	0.005 / 0.001	0.008 / -0.019	0.003 / -0.002
	40	2	-0.007 / 0.007	0.023 / -0.004	0.015 / -0.014	0.007 / -0.027	0.002 / -0.007	-0.020 / -0.008
	60	½	0.011 / 0.016	0.017 / 0.011	0.028 / 0.008	0.041 / 0.001	-0.010 / 0.018	0.019 / 0.012
	60	1	0.008 / 0.012	0.013 / 0.003	0.009 / -0.003	0.015 / -0.005	0.003 / 0.014	0.002 / 0.012
	60	2	0.023 / -0.006	0.008 / 0.022	0.000 / -0.010	0.005 / -0.004	-0.004 / 0.014	0.003 / -0.005
	100	½	0.007 / 0.001	-0.001 / 0.029	0.014 / 0.006	0.008 / -0.002	0.001 / 0.022	0.005 / 0.008
	100	1	0.011 / 0.023	0.009 / 0.002	0.011 / 0.002	0.028 / -0.005	0.002 / 0.008	0.005 / 0.014
	100	2	0.006 / -0.005	0.002 / 0.008	0.008 / 0.003	-0.012 / -0.002	0.007 / -0.010	-0.007 / 0.005
	200	½	0.002 / 0.005	-0.004 / 0.005	0.014 / -0.001	0.011 / 0.001	-0.001 / 0.004	-0.008 / 0.006
	200	1	0.000 / 0.008	-0.004 / -0.001	0.005 / 0.003	0.011 / 0.001	0.001 / 0.005	0.009 / 0.000
	200	2	-0.001 / -0.007	0.008 / 0.004	0.012 / -0.003	0.001 / 0.001	-0.002 / 0.007	-0.001 / -0.003

Table 8.21: BN, Median RMSE in (θ₃, θ₄), listed as θ₃ / θ₄

Table 8.21: BN, Median RMSE in (θ_3, θ_4), listed as θ_3 / θ_4

	N	k	AUC = (0.7, 0.7) $\rho = (0.0, 0.0)$	AUC = (0.7, 0.7) $\rho = (0.5, 0.5)$	AUC = (0.75, 0.85) $\rho = (0.0, 0.0)$	AUC = (0.75, 0.85) $\rho = (0.5, 0.5)$	AUC = (0.9, 0.9) $\rho = (0.0, 0.0)$	AUC = (0.9, 0.9) $\rho = (0.5, 0.5)$
b=1	40	½	0.548 / 0.550	0.564 / 0.564	0.632 / 0.397	1.078 / 0.353	0.508 / 0.491	0.549 / 0.570
	40	1	0.543 / 0.533	0.561 / 0.591	0.618 / 0.375	0.999 / 0.319	0.478 / 0.471	0.554 / 0.564
	40	2	0.534 / 0.547	0.613 / 0.606	0.597 / 0.380	1.001 / 0.356	0.481 / 0.469	0.595 / 0.592
	60	½	0.498 / 0.505	0.532 / 0.522	0.577 / 0.334	1.039 / 0.292	0.441 / 0.441	0.494 / 0.498
	60	1	0.509 / 0.489	0.535 / 0.542	0.566 / 0.317	0.961 / 0.280	0.405 / 0.412	0.485 / 0.496
	60	2	0.507 / 0.490	0.572 /0.556	0.535 / 0.327	0.904 / 0.303	0.404 / 0.414	0.538 / 0.522
	100	½	0.477 / 0.476	0.493 / 0.493	0.504 / 0.262	0.864 / 0.218	0.354 / 0.343	0.431 / 0.428
	100	1	0.459 / 0.456	0.497 / 0.494	0.489 / 0.248	0.758 / 0.192	0.329 / 0.326	0.416 / 0.416
	100	2	0.468 / 0.466	0.518 / 0.502	0.480 / 0.256	0.793 / 0.217	0.328 / 0.320	0.438 / 0.446
	200	½	0.407 / 0.407	0.431 / 0.446	0.417 / 0.183	0.643 / 0.137	0.248 / 0.251	0.312 / 0.304
	200	1	0.378 / 0.382	0.423 / 0.415	0.396 / 0.165	0.618 / 0.129	0.221 / 0.224	0.302 / 0.289
	200	2	0.401 / 0.398	0.442 / 0.444	0.399 / 0.177	0.616 / 0.137	0.225 / 0.229	0.317 / 0.319
b=1½	40	½	0.342 / 0.348	0.356 / 0.362	0.383 / 0.265	0.444 / 0.255	0.336 / 0.331	0.338 / 0.339
	40	1	0.320 / 0.326	0.350 / 0.341	0.356 / 0.251	0.420 / 0.223	0.296 / 0.297	0.318 / 0.321
	40	2	0.330 / 0.331	0.355 / 0.370	0.358 / 0.262	0.440 / 0.242	0.299 / 0.310	0.250 / 0.342
	60	½	0.299 / 0.296	0.302 / 0.313	0.335 / 0.230	0.387 / 0.198	0.275 / 0.270	0.284 / 0.274
	60	1	0.275 / 0.273	0.290 / 0.291	0.302 / 0.202	0.363 / 0.177	0.242 / 0.238	0.256 / 0.251
	60	2	0.272 / 0.280	0.304 / 0.304	0.309 / 0.208	0.377 / 0.188	0.238 / 0.240	0.269 / 0.272
	100	½	0.233 / 0.231	0.244 / 0.244	0.260 / 0.172	0.309 / 0.151	0.203 / 0.200	0.210 / 0.209
	100	1	0.217 / 0.211	0.224 / 0.224	0.240 / 0.156	0.285 / 0.134	0.178 / 0.179	0.189 / 0.186
	100	2	0.220 / 0.217	0.238 / 0.229	0.239 / 0.150	0.302 / 0.134	0.179 / 0.182	0.201 / 0.199
	200	½	0.163 / 0.159	0.166 / 0.164	0.183 / 0.119	0.224 / 0.107	0.145 / 0.144	0.145 / 0.144
	200	1	0.146 / 0.148	0.152 / 0.148	0.164 / 0.105	0.205 / 0.095	0.126 / 0.125	0.127 / 0.130
	200	2	0.151 / 0.148	0.156 / 0.155	0.165 / 0.106	0.212 / 0.095	0.124 / 0.120	0.128 / 0.128

Table 8.22: BN, Median Bias in Sensitivity and Specificity for (θ_3, θ_4), listed as Se / Sp

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	ρ = (0.0, 0.0)	ρ = (0.5, 0.5)
b=1	40	½	0.036 / 0.049	0.057 / 0.053	0.033 / 0.031	0.022 / 0.028	0.018 / 0.016	0.023 / 0.021
	40	1	0.034 / 0.052	0.032 / 0.052	0.015 / 0.027	0.019 / 0.028	0.015 / 0.017	0.016 / 0.021
	40	2	0.025 / 0.057	0.030 / 0.077	0.017 / 0.037	0.016 / 0.031	0.014 / 0.020	0.015 / 0.030
	60	½	0.035 / 0.031	0.038 / 0.033	0.015 / 0.018	0.020 / 0.021	0.011 / 0.010	0.016 / 0.013
	60	1	0.024 / 0.029	0.026 / 0.045	0.017 / 0.019	0.006 / 0.016	0.006 / 0.009	0.013 / 0.017
	60	2	0.023 / 0.042	0.025 / 0.055	0.007 / 0.020	0.009 / 0.019	0.006 / 0.013	0.011 / 0.021
	100	½	0.018 / 0.018	0.023 / 0.026	0.012 / 0.009	0.006 / 0.011	0.007 / 0.007	0.008 / 0.007
	100	1	0.011 / 0.021	0.020 / 0.026	0.009 / 0.009	0.005 / 0.010	0.003 / 0.005	0.008 / 0.011
	100	2	0.013 / 0.026	0.010 / 0.032	0.004 / 0.014	0.010 / 0.015	0.005 / 0.008	0.005 / 0.012
	200	½	0.013 / 0.006	0.011 / 0.009	0.008 / 0.006	0.005 / 0.006	0.003 / 0.004	0.004 / 0.004
	200	1	0.008 / 0.011	0.009 / 0.012	0.003 / 0.007	0.001 / 0.002	0.002 / 0.004	0.002 / 0.006
	200	2	0.006 / 0.015	0.008 / 0.014	0.004 / 0.006	0.005 / 0.006	0.002 / 0.005	0.004 / 0.006
b=1½	40	½	0.013 / 0.031	0.016 / 0.029	0.011 / 0.021	0.011 / 0.023	0.010 / 0.014	0.008 / 0.018
	40	1	0.009 / 0.021	0.016 / 0.041	0.006 / 0.024	0.009 / 0.014	0.004 / 0.011	0.008 / 0.013
	40	2	0.006 / 0.035	0.013 / 0.036	0.007 / 0.020	0.010 / 0.031	0.006 / 0.018	0.007 / 0.026
	60	½	0.008 / 0.014	0.020 / 0.019	0.007 / 0.012	0.007 / 0.016	0.004 / 0.007	0.007 / 0.009
	60	1	0.006 / 0.020	0.009 / 0.015	0.005 / 0.012	0.007 / 0.012	0.002 / 0.009	0.003 / 0.011
	60	2	0.004 / 0.023	0.007 / 0.028	0.004 / 0.010	0.007 / 0.018	0.004 / 0.011	0.007 / 0.026
	100	½	0.005 / 0.007	0.007 / 0.011	0.003 / 0.008	0.006 / 0.007	0.003 / 0.006	0.003 / 0.007
	100	1	0.003 / 0.008	0.003 / 0.008	0.002 / 0.009	0.003 / 0.012	0.002 / 0.004	0.002 / 0.006
	100	2	0.003 / 0.010	0.003 / 0.019	0.002 / 0.008	0.003 / 0.008	0.000 / 0.005	0.001 / 0.009
	200	½	0.002 / 0.003	0.003 / 0.004	0.002 / 0.003	0.002 / 0.006	0.001 / 0.002	0.001 / 0.002
	200	1	0.002 / 0.008	0.002 / 0.008	0.002 / 0.004	0.001 / 0.005	0.001 / 0.004	0.001 / 0.003
	200	2	0.002 / 0.004	0.001 / 0.004	0.001 / 0.004	0.002 / 0.005	0.000 / 0.003	0.000 / 0.004

Table 8.23: BN, Median RMSE in Sensitivity and Specificity for (θ_3, θ_4), listed as Se / Sp

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	ρ = (0.0, 0.0)	ρ = (0.5, 0.5)
b=1	40	½	0.121 / 0.098	0.135 / 0.122	0.087 / 0.068	0.091 / 0.079	0.059 / 0.046	0.066 / 0.059
	40	1	0.114 / 0.103	0.130 / 0.123	0.076 / 0.069	0.078 / 0.082	0.055 / 0.047	0.064 / 0.060
	40	2	0.113 / 0.109	0.127 / 0.132	0.077 / 0.075	0.083 / 0.090	0.055 / 0.051	0.063 / 0.065
	60	½	0.106 / 0.083	0.122 / 0.102	0.071 / 0.057	0.076 / 0.065	0.051 / 0.038	0.057 / 0.0490
	60	1	0.095 / 0.085	0.109 / 0.101	0.064 / 0.057	0.068 / 0.065	0.045 / 0.039	0.051 / 0.050
	60	2	0.093 / 0.091	0.108 / 0.111	0.063 / 0.062	0.066 / 0.072	0.045 / 0.043	0.052 / 0.056
	100	½	0.081 / 0.066	0.095 / 0.081	0.056 / 0.043	0.058 / 0.052	0.039 / 0.030	0.045 / 0.038
	100	1	0.076 / 0.067	0.087 / 0.081	0.049 / 0.044	0.054 / 0.052	0.035 / 0.031	0.040 / 0.039
	100	2	0.076 / 0.074	0.091 / 0.091	0.050 / 0.050	0.052 / 0.058	0.035 / 0.035	0.040 / 0.045
	200	½	0.059 / 0.049	0.069 / 0.060	0.040 / 0.031	0.042 / 0.037	0.028 / 0.021	0.031 / 0.027
	200	1	0.053 / 0.047	0.061 / 0.059	0.036 / 0.032	0.038 / 0.037	0.025 / 0.022	0.028 / 0.028
	200	2	0.055 / 0.053	0.064 / 0.066	0.036 / 0.036	0.037 / 0.043	0.024 / 0.025	0.028 / 0.032
b=1½	40	½	0.065 / 0.100	0.071 / 0.118	0.050 / 0.072	0.055 / 0.088	0.035 / 0.049	0.042 / 0.066
	40	1	0.063 / 0.104	0.071 / 0.123	0.048 / 0.078	0.052 / 0.091	0.033 / 0.050	0.039 / 0.068
	40	2	0.071 / 0.120	0.076 / 0.140	0.051 / 0.087	0.057 / 0.110	0.034 / 0.057	0.040 / 0.077
	60	½	0.053 / 0.082	0.059 / 0.097	0.042 / 0.060	0.047 / 0.071	0.029 / 0.041	0.034 / 0.053
	60	1	0.051 / 0.085	0.058 / 0.099	0.038 / 0.062	0.043 / 0.076	0.027 / 0.042	0.032 / 0.057
	60	2	0.056 / 0.096	0.062 / 0.118	0.040 / 0.073	0.046 / 0.088	0.028 / 0.048	0.033 / 0.065
	100	½	0.040 / 0.063	0.044 / 0.074	0.032 / 0.048	0.036 / 0.056	0.023 / 0.032	0.027 / 0.042
	100	1	0.039 / 0.065	0.043 / 0.078	0.030 / 0.048	0.033 / 0.059	0.021 / 0.033	0.025 / 0.045
	100	2	0.042 / 0.075	0.047 / 0.089	0.031 / 0.057	0.034 / 0.069	0.022 / 0.038	0.025 / 0.051
	200	½	0.028 / 0.045	0.032 / 0.053	0.023 / 0.033	0.026 / 0.040	0.017 / 0.023	0.019 / 0.030
	200	1	0.027 / 0.047	0.029 / 0.055	0.021 / 0.035	0.024 / 0.042	0.015 / 0.024	0.018 / 0.031
	200	2	0.028 / 0.054	0.032 / 0.061	0.021 / 0.040	0.024 / 0.050	0.015 / 0.027	0.017 / 0.036

Table 8.24: BN, Percent of Overestimation of Sensitivity and Specificity, listed as % Se / % Sp

			AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
	N	k	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$
b=1	40	½	67.00 / 71.49	70.31 / 68.52	64.71 / 72.29	62.54 / 66.08	64.14 / 66.38	67.42 / 67.20
	40	1	63.81 / 61.87	61.33 / 69.26	60.78 / 67.74	65.01 / 66.61	61.40 / 65.54	62.18 / 67.79
	40	2	61.87 / 76.88	62.05 / 73.83	58.03 / 73.52	59.18 / 64.79	62.38 / 66.88	60.86 / 69.97
	60	½	65.80 / 68.40	64.01 / 65.81	60.25 / 65.63	61.22 / 64.80	59.82 / 60.23	62.99 / 61.47
	60	1	62.88 / 66.14	61.69 / 72.62	60.63 / 64.94	54.55 / 62.92	57.26 / 62.58	60.66 / 64.26
	60	2	63.17 / 72.71	59.04 / 69.88	56.00 / 65.03	55.27 / 60.36	56.60 / 64.74	57.53 / 66.67
	100	½	60.44 / 63.36	60.38 / 64.75	59.55 / 61.22	53.94 / 61.05	57.30 / 58.61	57.64 / 60.43
	100	1	56.85 / 65.32	60.33 / 65.19	56.66 / 60.05	54.75 / 59.83	53.71 / 59.24	59.69 / 64.64
	100	2	59.39 / 67.01	57.08 / 68.79	54.07 / 62.45	58.35 / 60.98	57.96 / 61.79	55.11 / 60.17
	200	½	60.94 / 57.08	58.41 / 57.70	59.45 / 57.91	54.46 / 57.08	55.30 / 59.30	55.94 / 57.44
	200	1	57.52 / 60.70	56.76 / 58.73	53.61 / 60.39	52.28 / 53.06	53.90 / 57.00	54.01 / 58.22
	200	2	56.37 / 63.61	55.28 / 62.16	55.29 / 57.60	57.69 / 56.19	53.80 / 57.90	56.38 / 58.59
b=1½	40	½	60.29 / 66.49	60.42 / 60.74	59.63 / 63.75	58.43 / 60.98	64.93 / 63.03	58.92 / 61.32
	40	1	56.51 / 60.23	59.96 / 65.23	57.61 / 63.37	59.11 / 57.42	55.66 / 58.86	59.86 / 58.85
	40	2	54.79 / 63.94	59.43 / 63.29	56.96 / 60.63	58.79 / 62.14	57.20 / 64.15	59.23 / 65.09
	60	½	58.73 / 58.43	58.20 / 58.30	57.33 / 60.16	56.89 / 60.60	57.80 / 58.10	58.86 / 57.86
	60	1	55.41 / 59.86	57.74 / 58.36	56.13 / 57.65	58.45 / 57.94	54.20 / 59.10	55.41 / 58.92
	60	2	54.16 / 62.18	54.85 / 61.24	55.00 / 56.53	57.40 / 58.96	55.50 / 61.70	53.51 / 60.12
	100	½	57.20 / 54.70	58.52 / 57.62	55.60 / 58.20	57.32 / 55.10	55.80 / 60.40	57.10 / 57.00
	100	1	53.60 / 56.70	54.35 / 55.76	52.81 / 57.52	54.98 / 60.02	55.30 / 55.00	54.00 / 56.40
	100	2	53.35 / 55.86	53.41 / 60.82	53.75 / 56.56	55.75 / 55.44	50.00 / 58.50	53.20 / 59.00
	200	½	52.60 / 53.40	53.80 / 55.70	54.90 / 55.60	52.50 / 57.40	52.70 / 54.30	51.50 / 52.70
	200	1	52.30 / 57.10	53.90 / 57.80	55.20 / 54.80	51.45 / 55.36	54.20 / 57.60	53.00 / 54.30
	200	2	53.50 / 53.30	51.55 / 53.45	51.80 / 53.60	53.00 / 54.80	50.30 / 52.90	51.20 / 55.90

Table 8.25: BN, Coverage for SCIs and CRs for (θ₃, θ₄), listed as SCI / CR

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$
b=1	40	½	86.98 / 88.71	86.62 / 89.56	81.96 / 83.92	62.54 / 66.08	87.83 / 89.43	86.72 / 88.65
	40	1	85.55 / 88.70	87.54 / 88.03	85.54 / 86.88	68.21 / 70.34	89.61 / 91.09	90.01 / 91.81
	40	2	85.91 / 88.06	85.27 / 87.18	86.11 / 86.54	71.91 / 75.66	91.43 / 92.71	88.62 / 89.65
	60	½	87.69 / 89.47	86.45 / 86.60	88.63 / 88.25	61.05 / 63.10	88.80 / 90.75	88.64 / 90.15
	60	1	86.80 / 89.28	87.23 / 89.08	86.84 / 87.97	68.98 / 69.52	91.00 / 92.94	90.99 / 92.58
	60	2	88.13 / 89.36	87.20 / 88.70	88.00 / 88.00	70.36 / 72.91	92.16 / 93.81	90.86 / 92.37
	100	½	88.80 / 89.50	85.93 / 89.89	89.62 / 88.31	75.58 / 74.96	91.14 / 91.74	89.05 / 90.70
	100	1	87.96 / 88.92	87.79 / 89.32	89.95 / 90.07	77.46 / 76.27	93.78 / 94.28	91.34 / 92.99
	100	2	90.36 / 91.12	86.56 / 88.01	91.25 / 91.37	75.87 / 76.27	94.56 / 95.26	91.85 / 93.19
	200	½	92.81 / 92.49	89.69 / 91.11	92.97 / 91.98	81.23 / 80.31	92.00 / 92.30	91.15 / 91.95
	200	1	91.42 / 91.95	91.45 / 91.91	92.36 / 92.25	82.73 / 81.79	94.70 / 94.90	92.59 / 92.69
	200	2	92.33 / 93.74	89.43 / 90.54	93.17 / 92.29	85.45 / 84.62	94.70 / 94.30	93.87 / 94.67
b=1½	40	½	88.31 / 88.52	88.42 / 89.05	88.57 / 89.50	86.64 / 87.80	87.27 / 88.28	86.87 / 87.27
	40	1	89.46 / 91.53	89.46 / 90.52	88.48 / 90.02	89.72 / 90.57	91.49 / 92.29	91.45 / 92.45
	40	2	90.31 / 91.83	92.06 / 92.39	92.46 / 94.14	91.91 / 92.99	93.76 / 94.06	93.84 / 94.95
	60	½	89.30 / 89.61	89.88 / 90.38	87.97 / 88.47	87.24 / 87.65	89.30 / 89.30	88.79 / 89.89
	60	1	92.11 / 92.11	91.90 / 92.82	90.58 / 91.69	91.96 / 92.46	91.60 / 92.90	92.59 / 92.89
	60	2	91.57 / 92.50	93.51 / 93.81	91.43 / 91.43	92.50 / 93.02	93.90 / 93.70	94.89 / 96.19
	100	½	91.60 / 91.30	90.08 / 90.98	93.00 / 92.70	90.62 / 90.72	91.40 / 91.20	90.90 / 92.50
	100	1	93.20 / 93.50	93.19 / 93.39	90.98 / 91.68	91.54 / 91.94	94.40 / 93.40	94.40 / 94.80
	100	2	93.69 / 93.59	92.89 / 93.79	94.39 / 94.79	94.46 / 94.86	93.70 / 93.90	94.30 / 94.20
	200	½	92.80 / 92.60	93.90 / 93.20	94.00 / 94.20	92.90 / 93.90	93.20 / 92.50	94.70 / 94.90
	200	1	94.30 / 94.70	95.00 / 95.50	93.70 / 93.30	93.69 / 94.09	94.80 / 94.50	94.00 / 94.40
	200	2	94.90 / 94.20	96.10 / 95.80	93.80 / 93.70	94.30 / 95.30	96.20 / 95.40	95.70 / 95.90

Table 8.26: BN, Median SCI Widths for (θ_3, θ_4), listed as θ_3 / θ_4

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	ρ = (0.0, 0.0)	ρ = (0.5, 0.5)
b=1	40	½	1.80 / 1.78	1.96 / 1.95	1.92 / 1.42	2.45 / 1.29	1.76 / 1.71	1.99 / 2.06
	40	1	1.86 / 1.82	1.96 / 2.04	1.87 / 1.37	2.39 / 1.14	1.63 / 1.66	2.06 / 2.06
	40	2	1.83 / 1.79	2.10 / 2.12	1.88 / 1.42	2.45 / 1.28	1.70 / 1.65	2.24 / 2.23
	60	½	1.74 / 1.77	1.88 / 1.88	1.84 / 1.23	2.26 / 1.03	1.54 / 1.55	1.80 / 1.91
	60	1	1.78 / 1.68	1.90 / 1.88	1.76 / 1.13	2.27 / 1.00	1.42 / 1.47	1.81 / 1.88
	60	2	1.73 / 1.70	2.02 / 1.99	1.75 / 1.18	2.29 / 1.06	1.49 / 1.47	2.06 / 1.98
	100	½	1.66 / 1.65	1.83 / 1.76	1.69 / 0.92	2.21 / 0.78	1.25 / 1.22	1.58 / 1.57
	100	1	1.63 / 1.62	1.79 / 1.79	1.63 / 0.88	2.15 / 0.70	1.17 / 1.16	1.55 / 1.56
	100	2	1.62 / 1.61	1.85 / 1.84	1.62 / 0.91	2.17 / 0.75	1.17 / 1.16	1.72 / 1.70
	200	½	1.45 / 1.45	1.60 / 1.63	1.46 / 0.65	1.94 / 0.49	0.88 / 0.87	1.15 / 1.10
	200	1	1.39 / 1.37	1.57 / 1.59	1.40 / 0.59	1.94 / 0.46	0.79 / 0.80	1.08 / 1.02
	200	2	1.43 / 1.44	1.63 / 1.68	1.41 / 0.61	1.93 / 0.48	0.81 / 0.80	1.18 / 1.18
b=1½	40	½	1.16 / 1.16	1.21 / 1.24	1.30 / 0.91	1.51 / 0.86	1.10 / 1.07	1.17 / 1.14
	40	1	1.10 / 1.11	1.20 / 1.21	1.23 / 0.85	1.49 / 0.78	1.01 / 1.04	1.12 / 1.13
	40	2	1.15 / 1.15	1.29 / 1.35	1.27 / 0.92	1.54 / 0.86	1.06 / 1.07	1.27 / 1.24
	60	½	1.03 / 1.01	1.06 / 1.08	1.13 / 0.77	1.33 / 0.67	0.96 / 0.93	0.95 / 0.92
	60	1	0.96 / 0.95	1.02 / 1.03	1.05 / 0.69	1.30 / 0.62	0.83 / 0.83	0.90 / 0.87
	60	2	0.98 / 0.99	1.08 / 1.08	1.09 / 0.70	1.38 / 0.66	0.83 / 0.84	0.97 / 1.00
	100	½	0.81 / 0.80	0.83 / 0.82	0.92 / 0.61	1.11 / 0.53	0.72 / 0.71	0.72 / 0.72
	100	1	0.76 / 0.74	0.77 / 0.79	0.82 / 0.54	1.05 / 0.48	0.64 / 0.64	0.67 / 0.66
	100	2	0.78 / 0.77	0.83 / 0.82	0.87 / 0.53	1.13 / 0.47	0.63 / 0.64	0.70 / 0.69
	200	½	0.58 / 0.58	0.59 / 0.58	0.66 / 0.44	0.79 / 0.38	0.51 / 0.51	0.51 / 0.51
	200	1	0.52 / 0.52	0.54 / 0.54	0.58 / 0.38	0.74 / 0.34	0.45 / 0.45	0.46 / 0.47
	200	2	0.53 / 0.52	0.55 / 0.55	0.48 / 0.38	0.79 / 0.34	0.44 / 0.44	0.46 / 0.47

Table 8.27: BN, Median CR Area for (θ_3, θ_4)

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$
b=1	40	½	2.49	3.35	2.07	2.68	2.22	3.24
	40	1	2.54	3.37	1.99	2.44	2.13	3.29
	40	2	2.60	3.83	2.11	2.83	2.31	3.92
	60	½	2.12	2.77	1.57	2.00	1.62	2.48
	60	1	2.02	2.84	1.41	1.91	1.47	2.55
	60	2	2.13	3.22	1.52	2.11	1.60	3.01
	100	½	1.70	2.35	1.04	1.42	1.00	1.68
	100	1	1.60	2.36	0.95	1.25	0.89	1.69
	100	2	1.71	2.51	1.06	1.36	0.97	2.04
	200	½	1.13	1.75	0.58	0.79	0.47	0.86
	200	1	1.05	1.67	0.53	0.73	0.39	0.79
	200	2	1.18	1.88	0.57	0.77	0.44	0.98
b=1½	40	½	1.13	1.42	1.03	1.24	1.06	1.28
	40	1	1.04	1.33	0.93	1.12	0.97	1.22
	40	2	1.12	1.60	1.04	1.28	1.08	1.50
	60	½	0.85	1.02	0.74	0.84	0.73	0.83
	60	1	0.75	0.95	0.62	0.76	0.61	0.73
	60	2	0.79	1.05	0.66	0.86	0.64	0.89
	100	½	0.53	0.65	0.45	0.53	0.41	0.48
	100	1	0.44	0.57	0.37	0.46	0.34	0.40
	100	2	0.48	0.61	0.38	0.49	0.35	0.44
	200	½	0.26	0.31	0.22	0.27	0.20	0.23
	200	1	0.21	0.26	0.18	0.23	0.17	0.18
	200	2	0.22	0.27	0.18	0.24	0.16	0.19

Table 8.28: BN, Median Bias in Empirical (θ_3, θ_4), listed as θ_3 / θ_4

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$
b=1	40	½	-1.037 / -0.630	0.560 / 0.346	-0.040 / -0.087	1.246 / -0.305	-0.557 / -0.213	-0.274 / -0.466
	40	1	-0.678 / 0.074	0.270 / -0.619	0.881 / -0.272	0.763 / -1.363	-0.630 / 0.068	-0.420 / -0.485
	40	2	-0.178 / 0.113	-0.362 / -0.928	-0.307 / 0.046	0.623 / -0.269	0.112 / -0.743	-0.346 / -0.658
	60	½	-0.322 / 0.636	0.105 / -1.007	0.011 / 0.386	1.288 / -0.244	0.342 / -0.416	-0.139 / 0.539
	60	1	-0.572 / 0.408	-0.095 / 0.400	0.998 / -1.244	0.697 / -0.367	0.487 / -1.320	-0.543 / -0.138
	60	2	-0.498 / 0.257	-0.039 / -0.018	0.079 / -0.870	0.928 / 0.043	-0.498 / -0.183	0.214 / -0.494
	100	½	-0.285 / 0.232	0.545 / -0.808	0.791 / -0.220	1.617 / -0.111	0.062 / 0.006	-0.106 / -0.121
	100	1	0.367 / -0.575	-0.655 / 0.023	-0.136 / 0.104	0.820 / -0.509	-0.800 / 0.463	-0.408 / -0.276
	100	2	-0.868 / 0.181	0.467 / -0.990	0.240 / -0.582	0.875 / -0.207	-0.248 / 0.152	0.157 / -0.214
	200	½	0.250 / -0.079	-0.057 / -0.459	1.169 / -0.691	0.950 / 0.020	-0.153 / 0.069	0.175 / 0.116
	200	1	-0.154 / 0.563	-0.582 / 0.350	0.009 / -0.067	1.335 / 0.061	0.225 / -0.669	0.177 / 0.016
	200	2	-1.197 / 0.462	0.084 / -0.733	0.186 / -0.200	0.988 / 0.061	-0.259 / 0.060	0.316 / -0.078
b=1½	40	½	-0.268 / 0.405	0.196 / 0.323	0.520 / -0.173	0.012 / -0.290	0.208 / 0.184	0.028 / 0.130
	40	1	-0.302 / 0.386	-0.067 / -0.283	-0.029 / -0.132	-0.196 / 0.280	0.228 / -0.492	0.039 / 0.412
	40	2	0.064 / 0.984	0.413 / 0.347	0.057 / -0.451	0.211 / 0.087	-0.062 / -0.176	-0.410 / -0.145
	60	½	-0.286 / -0.099	-0.135 / 0.332	0.039 / 0.848	0.126 / -0.391	-0.594 / 0.287	-0.143 / 0.127
	60	1	0.272 / -0.558	0.365 / 0.005	0.052 / -0.220	0.066 / 0.109	0.357 / -0.659	0.012 / -0.223
	60	2	-0.180 / -0.037	0.149 / 0.077	0.333 / -0.358	-0.021 / 0.044	0.325 / -0.330	0.209 / -0.118
	100	½	0.001 / 0.375	0.073 / 0.374	-0.102 / 0.007	-0.098 / -0.223	0.082 / 0.145	-0.580 / 0.412
	100	1	-0.195 / -0.027	0.213 / 0.135	-0.148 / 0.137	0.093 / -0.287	0.167 / -0.324	0.341 / 0.049
	100	2	0.429 / -0.302	0.599 / 0.164	0.174 / 0.017	0.027 / -0.115	0.177 / -0.105	0.310 / -0.254
	200	½	0.133 / -0.239	0.123 / -0.011	0.002 / 0.184	-0.235 / 0.035	0.261 / -0.105	-0.084 / 0.321
	200	1	0.410 / -0.371	-0.032 / -0.384	-0.307 / 0.302	-0.082 / -0.141	0.159 / -0.152	-0.131 / -0.245
	200	2	-0.006 / -0.350	-0.307 / 0.155	0.073 / -0.143	0.424 / 0.011	0.069 / 0.034	0.069 / -0.168

Table 8.29: BN, Median Bias in Sensitivity and Specificity for Empirical (θ₃, θ₄), listed as Se / Sp

	N	k	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
			$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$	$\rho = (0.0, 0.0)$	$\rho = (0.5, 0.5)$
b=1	40	½	0.270 / -0.460	-0.008 / 0.221	0.087 / 0.095	0.160 / 0.001	0.074 / 0.048	0.164 / -0.149
	40	1	0.092 / -0.140	0.249 / -0.014	0.080 / 0.161	0.082-0.268	0.095 / -0.075	0.064 / 0.059
	40	2	-0.029 / 0.233	0.238 / -0.202	-0.029 / 0.134	0.158 / 0.155	0.034 / 0.048	0.015 / 0.082
	60	½	-0.008 / 0.160	0.199 / 0.011	-0.070 / 0.036	0.082 / -0.043	0.145 / 0.050	0.014 / 0.059
	60	1	0.042 / 0.110	-0.051 / 0.136	-0.004 / -0.089	0.132 / 0.065	0.012 / 0.025	0.097 / -0.041
	60	2	0.092 / -0.040	-0.001 / 0.036	0.055 / -0.139	0.007 / 0.132	0.070 / 0.075	-0.036 / 0.159
	100	½	0.018 / 0.082	-0.092 / 0.177	0.024 / 0.125	0.026 / 0.080	0.027 / 0.049	-0.013 / -0.008
	100	1	0.002 / 0.050	0.129 / -0.024	0.070 / 0.171	0.092 / -0.088	-0.035 / 0.085	0.064 / -0.061
	100	2	0.028 / 0.128	-0.009 / 0.124	0.081 / -0.001	-0.037 / 0.050	0.056 / -0.057	0.029 / 0.129
	200	½	0.073 / 0.054	0.066 / -0.025	-0.024 / 0.083	0.053 / 0.074	0.011 / -0.025	0.074 / 0.054
	200	1	-0.058 / 0.140	0.049 / 0.056	0.040 / 0.001	-0.018 / 0.092	-0.015 / 0.035	0.044 / -0.011
	200	2	0.065 / 0.113	0.178 / 0.033	0.058 / -0.001	0.030 / 0.126	0.048 / 0.034	-0.001 / 0.008
b=1½	40	½	0.040 / 0.084	0.037 / 0.260	-0.080 / -0.008	0.070 / 0.080	0.010 / 0.084	0.027 / 0.091
	40	1	-0.067 / 0.103	0.030 / -0.036	0.034 / -0.027	-0.009 / 0.172	0.032 / 0.022	-0.002 / 0.118
	40	2	-0.224 / 0.046	-0.079 / 0.337	0.060 / -0.008	-0.118 / 0.195	0.045 / -0.032	0.061 / 0.091
	60	½	0.083 / -0.047	-0.020 / 0.064	-0.066 / 0.173	0.041 / -0.103	-0.018 / 0.022	0.048 / -0.057
	60	1	0.016 / -0.013	0.047 / 0.180	-0.032 / -0.010	-0.092 / 0.006	0.049 / 0.122	0.032 / 0.068
	60	2	0.083 / 0.103	0.105 / 0.114	0.059 / -0.027	0.041 / 0.022	0.007 / 0.072	-0.002 / 0.018
	100	½	-0.082 / 0.126	-0.085 / 0.187	-0.013 / 0.079	0.023 / -0.152	-0.036 / 0.077	0.038 / 0.001
	100	1	0.083 / -0.067	-0.040 / 0.134	0.034 / -0.037	0.021 / -0.068	0.002 / 0.022	-0.002 / 0.108
	100	2	-0.026 / 0.111	-0.193 / 0.141	-0.090 / 0.102	0.022 / 0.121	-0.023 / 0.092	-0.036 / 0.108
	200	½	0.034 / -0.022	0.031 / 0.060	-0.045 / 0.013	-0.023 / 0.039	0.007 / 0.009	-0.021 / 0.048
	200	1	0.003 / 0.053	0.070 / 0.044	0.004 / 0.013	0.021 / 0.002	0.022 / 0.042	0.048 / -0.092
	200	2	0.034 / -0.010	0.038 / 0.050	0.022 / -0.004	-0.023 / 0.045	-0.037 / 0.046	0.009 / 0.001

Table 8.30: BN, Percent of Overestimation of Empirical Sensitivity and Specificity, listed as % Se / % Sp

	N	k	AUC = (0.7, 0.7) ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	AUC = (0.75, 0.85) ρ = (0.0, 0.0)	ρ = (0.5, 0.5)	AUC = (0.9, 0.9) ρ = (0.0, 0.0)	ρ = (0.5, 0.5)
b=1	40	½	85.1 / 4.7	23.6 / 99.7	96.5 / 96.2	98.8 / 53.5	99.3 / 94.7	99.4 / 0.5
	40	1	59.3 / 37.1	88.7 / 41.6	99.2 / 99.8	58.6 / 16.9	98.4 / 13.0	99.6 / 94.8
	40	2	34.1 / 99.4	99.7 / 0.0	41.9 / 98.2	99.8 / 99.7	61.8 / 53.0	57.8 / 88.9
	60	½	10.1 / 92.6	100.0 / 53.9	8.4 / 68.9	87.5 / 12.3	96.1 / 92.8	56.2 / 74.3
	60	1	57.7 / 92.2	11.3 / 98.1	10.6 / 19.2	91.8 / 93.7	74.5 / 52.5	74.2 / 7.5
	60	2	86.7 / 26.2	44.8 / 73.1	76.6 / 5.7	87.7 / 96.8	98.1 / 96.1	17.6 / 99.8
	100	½	63.1 / 84.6	7.1 / 98.0	62.4 / 99.6	56.2 / 99.7	73.4 / 95.9	16.7 / 35.2
	100	1	61.8 / 83.3	65.5 / 43.5	93.8 / 99.9	76.7 / 16.3	29.9 / 95.7	97.1 / 3.9
	100	2	69.4 / 97.4	45.0 / 91.0	82.8 / 48.9	28.7 / 72.7	80.3 / 32.9	64.9 / 97.6
	200	½	98.5 / 89.7	79.6 / 41.2	9.0 / 98.0	99.3 / 99.7	63.5 / 21.7	100.0 / 99.6
	200	1	15.0 / 98.9	78.5 / 94.3	96.8 / 66.2	31.4 / 100.0	13.0 / 98.1	89.7 / 43.4
	200	2	94.6 / 97.5	100.0 / 96.2	93.4 / 47.9	87.2 / 100.0	93.2 / 92.3	40.8 / 70.3
b=1½	40	½	53.9 / 91.1	85.4 / 100.0	7.8 / 46.3	58.9 / 96.6	99.2 / 96.7	88.6 / 98.7
	40	1	22.1 / 76.4	57.7 / 39.7	53.5 / 43.9	15.6 / 98.3	66.6 / 57.6	16.8 / 99.1
	40	2	21.6 / 66.3	1.7 / 99.9	81.4 / 3.8	19.4 / 97.8	56.2 / 37.8	62.8 / 99.9
	60	½	95.1 / 12.4	25.4 / 82.4	39.0 / 94.8	58.9 / 17.8	41.9 / 59.2	77.9 / 2.0
	60	1	65.6 / 49.6	78.7 / 99.7	34.3 / 49.9	16.4 / 57.8	99.1 / 99.7	65.3 / 90.8
	60	2	62.2 / 90.1	95.1 / 99.0	85.5 / 37.2	67.3 / 75.1	65.1 / 90.4	19.2 / 90.3
	100	½	24.3 / 85.0	5.0 / 99.9	36.9 / 95.9	62.2 / 0.4	0.3 / 99.4	70.4 / 61.4
	100	1	84.6 / 16.5	29.4 / 88.6	70.8 / 35.7	78.2 / 17.5	64.9 / 79.3	34.2 / 99.8
	100	2	35.6 / 97.2	23.2 / 71.0	14.2 / 88.6	62.1 / 99.0	20.4 / 99.2	16.7 / 97.4
	200	½	58.4 / 38.7	81.8 / 96.4	35.2 / 55.5	33.7 / 68.2	77.1 / 63.5	26.1 / 89.7
	200	1	50.9 / 87.1	99.2 / 97.3	58.9 / 72.6	81.2 / 59.9	89.0 / 90.8	97.8 / 0.0
	200	2	56.9 / 48.9	89.3 / 100.0	73.0 / 32.3	28.2 / 85.8	5.3 / 93.0	70.4 / 80.5

Table 8.31: BP, Assessing Normality

Parameter Set #	Test	TSS	k	# Valid Estimated OOPs	Mean	Median	Skewness	Kurtosis	Kolmogrov-Smirnov p-value	Shapiro-Wilk p-value	Histogram Look Normal?	QQPlot Look Normal?	Decision on Normality
3	1	40	0.5	984	1.547	1.532	0.631	1.114	<0.0100	<0.0001	y	y	Y
	2				1.551	1.518	0.583	0.801	<0.0100	<0.0001	y	y	Y
15	1	40	0.5	989	1.515	1.484	0.513	0.794	<0.0100	<0.0001	y	y	Y
	2				1.527	1.497	1.191	4.674	<0.0100	<0.0001	y	y	Y
17	1	60	2	1000	1.497	1.477	0.516	0.918	<0.0100	<0.0001	y	y	Y
	2				1.499	1.466	0.935	3.725	<0.0100	<0.0001	y	y	Y
27	1	40	0.5	995	1.862	1.830	0.873	1.915	<0.0100	<0.0001	y	y	Y
	2				1.598	1.585	0.350	0.559	0.0692	<0.0001	y	y	Y
48	1	200	0.5	1000	1.845	1.831	0.620	1.422	<0.0100	<0.0001	y	y	Y
	2				1.470	1.470	0.146	0.395	>0.1500	0.0151	y	y	Y
57	1	100	0.5	1000	1.962	1.957	0.141	-0.113	>0.1500	0.1155	y	y	Y
	2				1.958	1.954	0.244	-0.069	>0.1500	0.001	y	y	Y
59	1	200	2	1000	1.952	1.949	-0.024	0.294	>0.1500	0.3184	y	y	Y
	2				1.939	1.940	0.018	0.235	>0.1500	0.3783	y	y	Y
67	1	100	1	1000	1.823	1.807	0.374	0.480	<0.0100	<0.0001	y	y	Y
	2				1.816	1.813	0.130	-0.012	>0.1500	0.1232	y	y	Y
69	1	100	0.5	1000	1.822	1.812	0.570	1.714	<0.0100	<0.0001	y	y	Y
	2				1.824	1.805	0.509	0.710	<0.0100	<0.0001	y	y	Y
78	1	60	0.5	727	0.989	0.926	0.702	0.530	<0.0100	<0.0001	y	y	Y
	2				0.923	0.870	0.680	0.245	<0.0100	<0.0001	y	y	Y
80	1	100	2	834	0.949	0.881	0.625	0.278	<0.0100	<0.0001	y	y	Y
	2				0.923	0.860	0.824	0.767	<0.0100	<0.0001	y	y	Y
113	1	60	2	575	1.578	1.464	0.307	-1.051	<0.0100	<0.0001	n	n	N
	2				0.868	0.827	1.633	6.971	<0.0100	<0.0001	y	n	N
124	1	60	1	981	1.479	1.414	1.067	1.941	<0.0100	<0.0001	y	y	Y
	2				1.492	1.444	1.076	2.603	<0.0100	<0.0001	y	y	Y
129	1	100	0.5	995	1.477	1.457	0.773	2.084	<0.0100	<0.0001	y	y	Y
	2				1.467	1.424	1.108	2.293	<0.0100	<0.0001	y	y	Y

Table 8.32: BN, Assessing Normality

Parameter Set #	Test	TSS	k	# Valid Estimated OOPs	Mean	Median	Skewness	Kurtosis	Kolmogrov-Smirnov p-value	Shapiro-Wilk p-value	Histogram Look Normal?	QQPlot Look Normal?	Decision on Normality
21	1	100	0.5	998	-0.209	-0.195	-0.620	0.633	<0.0100	<0.0001	y	y	Y
	2				-0.188	-0.165	-0.955	2.350	<0.0100	<0.0001	y	n	Y
24	1	200	0.5	1000	-0.203	-0.198	-0.494	0.729	0.0163	<0.0001	y	y	Y
	2				-0.193	-0.188	-0.359	0.298	0.0236	<0.0001	y	y	Y
29	1	60	2	980	-0.318	-0.295	-0.698	1.115	<0.0100	<0.0001	y	n	Y
	2				0.335	0.341	-0.355	0.706	>0.1500	<0.0001	y	y	Y
36	1	200	0.5	1000	-0.294	-0.280	-0.384	0.462	<0.0100	<0.0001	y	y	Y
	2				0.351	0.350	-0.168	0.088	>0.1500	0.1712	y	y	Y
40	1	60	1	982	-0.365	-0.319	-0.929	1.510	<0.0100	<0.0001	n	n	N
	2				0.443	0.449	-0.338	0.619	0.0248	<0.0001	y	y	Y
82	1	200	1	944	-0.193	-0.157	-0.900	1.666	<0.0100	<0.0001	y	n	Y
	2				-0.174	-0.130	-0.819	1.177	<0.0100	<0.0001	y	y	Y
85	1	40	1	618	-0.112	-0.014	-0.933	1.028	<0.0100	<0.0001	n	n	N
	2				-0.171	-0.069	-0.852	0.768	<0.0100	<0.0001	y	y	Y
86	1	40	2	577	-0.214	-0.096	-0.818	0.514	<0.0100	<0.0001	y	y	Y
	2				-0.129	-0.039	-0.846	1.003	<0.0100	<0.0001	y	y	Y
90	1	60	0.5	664	-0.096	-0.031	-0.957	1.357	<0.0100	<0.0001	y	y	Y
	2				-0.134	-0.060	-0.847	0.715	<0.0100	<0.0001	y	n	Y
96	1	200	0.5	844	-0.116	-0.012	-1.324	2.055	<0.0100	<0.0001	n	n	N
	2				-0.084	-0.003	-1.194	1.949	<0.0100	<0.0001	n	n	N
97	1	40	1	747	-0.446	-0.376	-0.407	-0.219	<0.0100	<0.0001	y	y	Y
	2				0.410	0.481	-1.415	3.773	<0.0100	<0.0001	n	n	N
105	1	100	0.5	838	-0.567	-0.531	-0.347	-0.201	<0.0100	<0.0001	y	y	Y
	2				0.501	0.525	-0.924	1.626	<0.0100	<0.0001	y	n	Y
115	1	100	1	590	-0.855	-0.843	0.082	-1.045	<0.0100	<0.0001	n	n	N
	2				0.660	0.674	-3.644	33.241	<0.0100	<0.0001	n	n	N
139	1	100	1	970	0.439	0.518	-1.880	5.460	<0.0100	<0.0001	n	n	N
	2				0.433	0.514	-1.761	4.985	<0.0100	<0.0001	n	n	N

Table 8.33: BP, Varying *r* (TSS = 200, *k* = 1, *ρ* = (0.0, 0.0))

	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
	b = 0.5	*b* = 1	*b* = 0.5	*b* = 1	*b* = 0.5	*b* = 1
_r_ = 0.5						
Number of Times Invalid OOP Estimates, $\theta_1 / \theta_2 / \theta_1$ & θ_2	0 / 0 / 0	11 / 9 / 0	0 / 0 / 0	29 / 0 / 0	0 / 0 / 0	0 / 0 / 0
Bias of θ_1 / θ_2	-0.008 / -0.007	0.016 / 0.027	-0.010 / 0.010	-0.014 / 0.017	0.007 / -0.006	0.001 / 0.013
RMSE of θ_1 / θ_2	0.190 / 0.192	0.519 / 0.519	0.204 / 0.156	0.441 / 0.201	0.180 / 0.177	0.220 / 0.229
Bias of Se / Sp corresponding to (θ_1, θ_2)	0.004 / -0.001	0.005 / 0.010	0.003 / 0.003	0.003 / 0.006	0.002 / 0.000	0.001 / 0.004
RMSE of Se / Sp corresponding to (θ_1, θ_2)	0.032 / 0.041	0.020 / 0.084	0.025 / 0.029	0.015 / 0.056	0.017 / 0.018	0.012 / 0.038
Percent Overestimation of Se and Sp, % Se / % Sp	56.10 / 49.00	63.37 / 53.47	56.10 / 53.40	60.25 / 56.33	56.00 / 50.70	54.50 / 54.90
Coverage for SCIs / CRs for (θ_1, θ_2)	96.20 / 96.20	93.88 / 94.08	94.80 / 95.70	91.35 / 91.66	94.90 / 94.90	95.90 / 94.50
Median SCI Widths for θ_1 / θ_2	0.69 / 0.69	1.81 / 1.82	0.73 / 0.57	1.55 / 0.69	0.65 / 0.65	0.79 / 0.79
Median CR Area for (θ_1, θ_2)	0.41	1.82	0.36	0.69	0.38	0.41
_r_ = 2.0						
Number of Times Invalid OOP Estimates, $\theta_1 / \theta_2 / \theta_1$ & θ_2	0 / 0 / 0	101 / 102 / 0	0 / 0 / 0	156 / 0 / 0	0 / 0 / 0	0 / 0 / 0
Bias of θ_1 / θ_2	-0.005 / -0.013	0.010 / -0.025	0.011 / 0.006	-0.077 / 0.015	-0.007 / -0.003	0.002 / -0.012
RMSE of θ_1 / θ_2	0.210 / 0.212	0.336 / 0.335	0.231 / 0.181	0.371 / 0.173	0.203 / 0.203	0.234 / 0.242
Bias of Se / Sp corresponding to (θ_1, θ_2)	0.004 / 0.001	0.023 / 0.002	0.001 / 0.000	0.010 / 0.001	0.002 / 0.001	0.006 / 0.002
RMSE of Se / Sp corresponding to (θ_1, θ_2)	0.052 / 0.006	0.082 / 0.020	0.041 / 0.006	0.055 / 0.018	0.029 / 0.005	0.037 / 0.015
Percent Overestimation of Se and Sp, % Se / % Sp	53.50 / 57.10	63.36 / 56.59	51.70 / 52.10	56.99 / 52.25	53.60 / 56.00	58.50 / 56.50
Coverage for SCIs / CRs for (θ_1, θ_2)	93.90 / 93.40	91.22 / 89.96	93.10 / 92.70	92.06 / 90.52	95.00 / 94.90	93.20 / 92.50
Median SCI Widths for θ_1 / θ_2	0.77 / 0.76	1.22 / 1.20	0.83 / 0.66	1.27 / 0.65	0.74 / 0.74	0.84 / 0.83
Median CR Area for (θ_1, θ_2)	0.52	1.12	0.48	0.59	0.48	0.47

Table 8.34: BN, Varying *r* (TSS = 200, *k* = 1, *ρ* = (0.0, 0.0))

	AUC = (0.7, 0.7)		AUC = (0.75, 0.85)		AUC = (0.9, 0.9)	
	b = 1	*b* = 1.5	*b* = 1	*b* = 1.5	*b* = 1	*b* = 1.5
r = 0.5						
Number of Times Invalid OOP Estimates, θ_3 / θ_4 / θ_3 & θ_4	100 / 118 / 3	0 / 0 / 0	151 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0
Bias of θ_3 / θ_4	0.007 / 0.009	0.005 / 0.001	0.080 / -0.029	-0.002 / 0.008	0.014 / 0.000	-0.000 / 0.002
RMSE of θ_3 / θ_4	0.329 / 0.335	0.161 / 0.159	0.366 / 0.175	0.177 / 0.118	0.239 / 0.236	0.134 / 0.135
Bias of Se / Sp corresponding to (θ_3, θ_4)	0.002 / 0.020	0.001 / 0.004	0.002 / 0.007	0.002 / 0.003	0.002 / 0.004	0.001 / 0.002
RMSE of Se / Sp corresponding to (θ_3, θ_4)	0.020 / 0.082	0.009 / 0.060	0.017 / 0.054	0.010 / 0.048	0.015 / 0.037	0.008 / 0.033
Percent Overestimation of Se and Sp, % Se / % Sp	55.80 / 61.53	55.70 / 53.10	56.42 / 56.89	56.90 / 53.70	54.80 / 55.20	52.10 / 52.80
Coverage for SCIs / CRs for (θ_3, θ_4)	89.55 / 89.43	95.30 / 95.10	92.23 / 91.40	93.00 / 92.90	94.10 / 93.50	93.30 / 92.90
Median SCI Widths for θ_3 / θ_4	1.21 / 1.22	0.58 / 0.58	1.26 / 0.66	0.63 / 0.43	0.84 / 0.84	0.49 / 0.48
Median CR Area for (θ_3, θ_4)	1.12	0.29	0.60	0.23	0.47	0.19
r = 2.0						
Number of Times Invalid OOP estimates, θ_3 / θ_4 / θ_3 & θ_4	23 / 15 / 1	0 / 0 / 0	37 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0
Bias of θ_3 / θ_4	-0.028 / -0.017	-0.010 / 0.003	0.022 / -0.024	0.002 / -0.000	-0.006 / -0.000	-0.005 / 0.007
RMSE of θ_3 / θ_4	0.523 / 0.503	0.178 / 0.177	0.439 / 0.204	0.166 / 0.111	0.227 / 0.224	0.122 / 0.123
Bias of Se / Sp corresponding to (θ_3, θ_4)	0.011 / 0.003	0.005 / 0.008	0.005 / 0.004	0.001 / 0.001	0.005 / 0.003	0.002 / 0.003
RMSE of Se / Sp corresponding to (θ_3, θ_4)	0.084 / 0.020	0.070 / 0.029	0.056 / 0.016	0.042 / 0.023	0.039 / 0.012	0.027 / 0.016
Percent Overestimation of Se and Sp, % Se / % Sp	56.39 / 59.61	53.00 / 61.60	54.41 / 60.96	56.90 / 53.80	56.20 / 59.20	54.00 / 58.50
Coverage for SCIs / CRs for (θ_3, θ_4)	93.46 / 93.98	96.90 / 96.50	93.56 / 94.39	94.80 / 94.90	95.50 / 94.80	95.30 / 94.30
Median SCI Widths for θ_3 / θ_4	1.83 / 1.80	0.62 / 0.61	1.56 / 0.69	0.63 / 0.40	0.80 / 0.78	0.44 / 0.44
Median CR Area for (θ_3, θ_4)	1.85	0.30	0.69	0.19	0.41	0.16

8.3. Appendix C: SCI Widths and Total Sample Size Prediction Equations

OLS models to predict SCI widths are shown below. Because M1 is the best performing

model for both sequential testing strategies, it will be shown in (a) – (d). Additionally,

since M2 is the model most researchers will likely have the information for, and the

inclusion of interactions increase the accuracy of the model, M2 + AUC*b will also be

included below in (e) – (h) (the addition of TSS*k did not add anything to R^2, so this

model's equation will not be included). M3 will be omitted since this is the least practical

model.

SCI Widths:

BP, M1:

(a) SCI Width $\theta_1 = \begin{array}{l} 0.83251 - 0.00527*TSS - 0.71403*\rho_H + 1.17412*\rho_D + \\ 0.43870*\mu_{1H} - 0.40471*\mu_{2H} - 0.38568*\mu_{1D} + 0.31489*\mu_{2D} + \\ 2.01130*\sigma_{1H} - 0.03104*\sigma_{1D} - 0.16414*\sigma_{2H} - 0.45470*\sigma_{2D} + \\ 0.05500*k \end{array}$

(b) SCI Width $\theta_2 = \begin{array}{l} 0.64666 - 0.00504*TSS - 0.64798*\rho_H + 0.87311*\rho_D - \\ 0.43776*\mu_{1H} + 0.51615*\mu_{2H} + 0.40555*\mu_{1D} - 0.45528*\mu_{2D} - \\ 0.10578*\sigma_{1H} - 0.41209*\sigma_{1D} + 1.76455*\sigma_{2H} + 0.14255*\sigma_{2D} + \\ 0.04761*k \end{array}$

BN, M1:

(c) SCI Width $\theta_1 = \begin{array}{l} -0.07104 - 0.00435*TSS + 1.19158*\rho_H - 0.70836*\rho_D + \\ 0.56884*\mu_{1H} - 0.47832*\mu_{2H} - 0.64600*\mu_{1D} + 0.49428*\mu_{2D} + \\ 0.34499*\sigma_{1H} + 2.39569*\sigma_{1D} - 0.58781*\sigma_{2H} + 0.15716*\sigma_{2D} + \\ 0.00154*k \end{array}$

(d) SCI Width $\theta_2 =$
$$0.15169 - 0.00424*TSS + 0.90838*\rho_H - 0.66492*\rho_D -$$
$$0.58312*\mu_{1H} + 0.68249*\mu_{2H} + 0.76058*\mu_{1D} - 0.89918*\mu_{2D} -$$
$$0.64907*\sigma_{1H} + 0.04272*\sigma_{1D} + 0.54057*\sigma_{2H} + 2.09272*\sigma_{2D} +$$
$$0.00291*k$$

BP, M2 + AUC*_b_:

(e) SCI Width $\theta_1 =$
$$-0.10571 - 0.00516*TSS - 0.83143*AUC_1 + 2.10710*AUC_2 -$$
$$0.67019*\rho_H + 1.17282*\rho_D + 3.80276*b_1 - 0.64531*b_2 +$$
$$0.04898*k - 3.63939*AUC_1*b_1 + 0.99793*AUC_2*b_2$$

(f) SCI Width $\theta_2 =$
$$0.45704 - 0.00490*TSS + 1.25848*AUC_1 - 0.36435*AUC_2 -$$
$$0.58838*\rho_H + 0.82264*\rho_D - 2.40759*b_1 + 5.05018*b_2 +$$
$$0.04235*k + 3.19752*AUC_1*b_1 - 5.41204*AUC_2*b_2$$

BN, M2 + AUC*_b_:

(g) SCI Width $\theta_1 =$
$$5.55144 - 0.00450*TSS - 5.70396*AUC_1 + 2.90849*AUC_2 +$$
$$1.28358*\rho_H - 0.76890*\rho_D - 2.34340*b_1 - 0.16024*b_2 -$$
$$0.00831*k + 1.80882*AUC_1*b_1 - 0.33338*AUC_2*b_2$$

(h) SCI Width $\theta_2 =$
$$5.11474 - 0.00430*TSS + 5.46691*AUC_1 - 7.97909*AUC_2 +$$
$$0.97419*\rho_H - 0.71025*\rho_D + 0.87370*b_1 - 2.97751*b_2 -$$
$$0.00629*k - 1.54576*AUC_1*b_1 + 2.75218*AUC_2*b_2$$

OLS models to predict TSS are shown below. Again, M1 will be shown (i) – (l) and

M2 + AUC*_b_ will also be included below in (m) – (q). M3 will again be omitted since

this is the least practical model.

OLS TSS:

BP, M1:

(i) $TSS = $
$$101.19614 - 78.00019*SCIWidth\theta_1 - 62.68480*\rho_H + 99.82246*\rho_D + 37.12254*\mu_{1H} - 31.94721*\mu_{2H} - 31.11355*\mu_{1D} + 24.30843*\mu_{2D} + 172.84236*\sigma_{1H} - 3.39221*\sigma_{1D} - 0.58137*\sigma_{2H} - 35.93269*\sigma_{2D} + 4.32768*k$$

(j) $TSS = $
$$87.54881 - 77.24835*SCIWidth\theta_2 - 57.30298*\rho_H + 75.99069*\rho_D - 30.80523*\mu_{1H} + 39.47751*\mu_{2H} + 30.25934*\mu_{1D} - 35.43230*\mu_{2D} + 8.37728*\sigma_{1H} - 32.83994*\sigma_{1D} + 148.98030*\sigma_{2H} + 10.52834*\sigma_{2D} + 3.71734*k$$

BN, M1:

(k) $TSS = $
$$41.10121 - 92.85079*SCIWidth\theta_1 + 118.77820*\rho_H - 71.93783*\rho_D + 58.80122*\mu_{1H} - 44.98378*\mu_{2H} - 62.94795*\mu_{1D} + 45.98943*\mu_{2D} + 32.11888*\sigma_{1H} + 235.65039*\sigma_{1D} - 50.99490*\sigma_{2H} + 16.58264*\sigma_{2D} + 0.34079*k$$

(l) $TSS = $
$$62.65313 - 92.55910*SCIWidth\theta_2 + 92.37382*\rho_H - 67.82936*\rho_D - 47.87442*\mu_{1H} + 62.58796*\mu_{2H} + 67.37496*\mu_{1D} - 83.13018*\mu_{2D} - 59.98950*\sigma_{1H} + 17.41633*\sigma_{1D} + 53.68744*\sigma_{2H} + 195.72953*\sigma_{2D} + 0.47022*k$$

BP, M2 + AUC*b:

(m) $TSS = $
$$21.85069 - 72.01482*SCIWidth\theta_1 - 36.55223*AUC_1 + 175.17757*AUC_2 - 54.19900*\rho_H + 92.36197*\rho_D + 317.14622*b_1 - 24.33313*b_2 + 3.33640*k - 310.95943*AUC_1*b_1 + 44.97021*AUC_2*b_2$$

(n) $TSS = $
$$58.00913 - 77.11872*SCIWidth\theta_2 + 120.15292*AUC_1 - 4.88689*AUC_2 - 51.25369*\rho_H + 71.26731*\rho_D - 142.79182*b_1 + 411.39110*b_2 + 3.07690*k + 198.18581*AUC_1*b_1 - 444.00865*AUC_2*b_2$$

BN, M2 + AUC*_b_:

$$(o)\ \text{TSS} = \begin{aligned} &595.72050 - 76.98047 * SCIWidth\theta_1 - 531.21323 * AUC_1 + \\ &206.78486 * AUC_2 + 110.02441 * \rho_H - 67.68694 * \rho_D - \\ &230.34756 * b_1 - 32.90193 * b_2 - 0.42526 * k + \\ &192.55419 * AUC_1 * b_1 - 5.18954 * AUC_2 * b_2 \end{aligned}$$

$$(p)\ \text{TSS} = \begin{aligned} &579.17550 - 80.28883 * SCIWidth\theta_2 + 346.72896 * AUC_1 - \\ &657.75943 * AUC_2 + 89.44080 * \rho_H - 65.52959 * \rho_D + \\ &20.15155 * b_1 - 259.64551 * b_2 - 0.29103 * k - \\ &70.74960 * AUC_1 * b_1 + 241.46128 * AUC_2 * b_2 \end{aligned}$$

8.4. Appendix D: SAS Programs

Below are SAS programs associated with the simulation study and related sub-analyses

described in chapter 4.

8.4.1   Determining Legitimate BP and BN Strategies

```
libname pop 'C:\Documents and Settings\wilkar\My Documents\Dropbox\Dissertation!';

data values;
      set pop.pop_parms;
run;

%let pi=4*atan(1);
%let m1h=0;   %let s1h=1;   %let m2h=0;   %let s2h=1;
%let iter=1000;        %let converge=0.001;%let func=1000;

data parms;   set values; id=_n_;

/*       Assuming that the healthy group follows a Standard Normal Distribution */
      m1h=&m1h;   s1h=&s1h;   m2h=&m2h;   s2h=&s2h;   cpratio=r;

      s1d=s1h/b1;   s2d=s2h/b2;

/*       b1 = s1h / s1d;       b2 = s2h / s2d;*/

      m1d = ( s1d * Sqrt(1 + (b1)**2) * probit(AUC1) ) + m1h;
```

```
        m2d = ( s2d * Sqrt(1 + (b2)**2) * probit(AUC2) ) + m2h;

        a1 = (m1d - m1h) / s1d;          a2 = (m2d - m2h) / s2d;

        lcutoff1 = m1h - 1.645*s1h;   hcutoff1 = m1d + 1.645*s1d;
        lcutoff2 = m2h - 1.645*s2h;   hcutoff2 = m2d + 1.645*s2d;
run;


/* BN */
proc nlp data=parms noprint out=oop_bn absgconv=&converge maxiter=&iter
            maxfunc=&func;
        by id;
        max GYI;
        parms theta3 = -4 to 4 by 0.2, theta4 = -4 to 4 by 0.2;
        TPR = 1-probnorm((theta3-m1d)/s1d)-probnorm((theta4-m2d)/s2d) + probbnrm(
            ((theta3-m1d)/s1d), ((theta4-m2d)/s2d), pd);
        FPR = 1-probnorm((theta3-m1h)/s1h)-probnorm((theta4-m2h)/s2h) + probbnrm(
            ((theta3-m1h)/s1h), ((theta4-m2h)/s2h), ph);
        Sens = TPR;
        Spec = 1 - FPR;

        GYI = TPR - cpratio*FPR;
run;


data oop_bn2;
        merge parms oop_bn (drop=_obs_ _type_);
        by id;
run;


data oop_bn3;
        set oop_bn2;
        if theta3=. or theta3 > hcutoff1 or theta3 < lcutoff1 then flag1=1;   else flag1=0;
        if theta4=. or theta4 > hcutoff2 or theta4 < lcutoff2 then flag2=1;   else flag2=0;
        if flag1=1 and flag2=1 then flag12=1;else flag12=0;

        if AUC1=0.7 and AUC2=0.7 then AUCpair=1;
        else    if AUC1=0.75 and AUC2=0.85 then AUCpair=2;
        else if AUC1=0.9 and AUC2=0.9 then AUCpair=3;
run;


data bad_bn;  set oop_bn3;  if flag1=1 or flag2=1 or flag12=1;        run;

proc sort data=oop_bn3;       by AUCpair cpratio;   run;
proc print data=oop_bn3;
        var b1 b2 ph pd cpratio sens spec GYI;
        by AUCpair;
        where flag1=0 and flag2=0;
run;
proc sort data=bad_bn;        by AUCpair cpratio;   run;
```

```
proc print data=bad_bn;
        var b1 b2 ph pd cpratio sens spec GYI;
        by AUCpair;
run;

/* BP */
proc nlp data=parms noprint out=oop_bp absgconv=&converge maxiter=&iter
             maxfunc=&func;
        by id;
        max GYI;
        parms theta1 = -4 to 4 by 0.2, theta2 = -4 to 4 by 0.2;
        TPR = 1-probbnrm( ((theta1-m1d)/s1d), ((theta2-m2d)/s2d), pd);
        FPR = 1-probbnrm( ((theta1-m1h)/s1h), ((theta2-m2h)/s2h), ph);
        Sens = TPR;
        Spec = 1 - FPR;
        GYI = TPR - cpratio*FPR;
run;

data oop_bp2;
        merge parms oop_bp (drop=_obs_ _type_);
        by id;
run;

data oop_bp3;
        set oop_bp2;
        if theta1=. or theta1 > hcutoff1 or theta1 < lcutoff1 then flag1=1;    else flag1=0;
        if theta2=. or theta2 > hcutoff2 or theta2 < lcutoff2 then flag2=1;    else flag2=0;
        if flag1=1 and flag2=1 then flag12=1; else flag12=0;

        if AUC1=0.7 and AUC2=0.7 then AUCpair=1;
        else    if AUC1=0.75 and AUC2=0.85 then AUCpair=2;
        else if AUC1=0.9 and AUC2=0.9 then AUCpair=3;
run;

data bad_bp;
        set oop_bp3;
        if flag1=1 or flag2=1 or flag12=1;
run;
proc sort data=oop_bp3;    by AUCpair cpratio;    run;
proc print data=oop_bp3;
        var b1 b2 ph pd cpratio sens spec GYI;
        by AUCpair;
        where flag1=0 and flag2=0;
run;
proc sort data=bad_bp;    by AUCpair cpratio;    run;
proc print data=bad_bp;
        var b1 b2 ph pd cpratio sens spec GYI;
        by AUCpair;
run;
```

8.4.2 BP Simulations and Empirical Estimates

```
options ps=65 ls=116 obs=max formdlim=# nodate nonumber;
libname BP 'C:\Documents and Settings\wilkar\Desktop\Dissertation!\FINAL FROM
CLUSTER - KOTI and MERLOT\BP';

/* Defines what parameter set these datasets belong to */
%let i=1;

data sasparms;
        set BP.sasparms;        if _n_=&i;
        call symput('TSS',TSS);        call symput('k',k);        call symput('AUC1',AUC1);
        call symput('AUC2',AUC2);    call symput('ph',ph);    call symput('pd',pd);
        call symput('s1d',s1d);        call symput('s2d',s2d);
run;

%let pi=4*atan(1);
%let times=1000;        %let rep=1000;        %let p=2;        %let z=1.645;
%let bratio1=s1h/s1d; %let bratio2=s2h/s2d;
%let R=1;
%let N0=floor(&TSS/(&k+1));        %let N1=&TSS - &N0;
%let m1h=0;   %let s1h=1;   %let m2h=0;   %let s2h=1;
%let converge=0.001; %let iter=1000;        %let func=1000;        %let tech=nrridg;

data parms&i;
        R=&R; ph=&ph;        pd=&pd;        AUC1 = &AUC1;        AUC2 = &AUC2;

        /* Assuming that the healthy group follows a Standard Normal Distribution */
        m1h=&m1h; s1h=&s1h;        m2h=&m2h; s2h=&s2h;        s1d=&s1d;
        s2d=&s2d;

        b1=s1h/s1d;   b2=s2h/s2d;

        m1d = ( s1d * Sqrt(1 + (b1)**2) * probit(AUC1) ) + m1h;
        m2d = ( s2d * Sqrt(1 + (b2)**2) * probit(AUC2) ) + m2h;

        Ca1=(m1d-m1h)/s1d; Ca2=(m2d-m2h)/s2d;
        Cauc1 = probnorm( Ca1 / sqrt(1+b1**2) );
        Cauc2 = probnorm( Ca2 / sqrt(1+b2**2) );

        b1 = s1h / s1d;        a1 = (m1d - m1h) / s1d;
        b2 = s2h / s2d;        a2 = (m2d - m2h) / s2d;
run;

proc nlp data=parms&i noprint out=NRTrueValues&i tech=&tech absgconv=&converge
                maxiter=&iter maxfunc=&func;
        max trueGYI;
        parms truetheta1 = -4 to 4 by 0.1, truetheta2 = -4 to 4 by 0.1;
```

```
            trueTPR = 1-probbnrm( ((truetheta1-m1d)/s1d), ((truetheta2-m2d)/s2d), pd);
            trueFPR = 1-probbnrm( ((truetheta1-m1h)/s1h), ((truetheta2-m2h)/s2h), ph);
            trueSens = trueTPR;
            trueSpec = 1 - trueFPR;
            trueGYI = trueTPR - &R*trueFPR;
run;

data sim;
        set NRTrueValues&i;
        call symput('t1',truetheta1);
        call symput('t2',truetheta2);
run;

data Parms_TrueValues&i;
        N_d=&N1;      N_h=&N0;      R=&R;
        merge Parms&i (keep=AUC1 AUC2 b1 b2 ph pd)
        NRTrueValues&i (keep=truetheta1 truetheta2 trueGYI trueSens trueSpec);
run;

/* Simulating Data */
%let seed1=%eval(12345+&i);        %let seed2=%eval(54321+&i);
data disease&i (keep=id test1d test2d nrun h_d_flag) healthy&i (keep=id test1h test2h
nrun h_d_flag);
        set parms&i;
        cd=sqrt(1-pd**2);        ch=sqrt(1-ph**2);
            do k=1 to &times;
                    nrun=k;
                    do l=1 to &N1;
                            id=l;
                            h_d_flag=1;
                                    test1d = rannor(&seed1);
                                    test2d = pd*test1d+cd*rannor(&seed1);
                    test1d = m1d + s1d*test1d;
                    test2d = m2d + s2d*test2d;
                            output disease&i;
                    end;
                    do j=1 to &N0;
                            id=j;
                            h_d_flag=0;
                                    test1h = rannor(&seed2);
                            test2h = ph*test1h+ch*rannor(&seed2);
                            test1h = m1h + s1h*test1h;
                            test2h = m2h + s2h*test2h;
                            output healthy&i;
                    end;
            end;
run;

data combined&i;        merge disease&i healthy&i;    by nrun id;        run;
```

```
/* Finding the Means, Variances, and Correlations for each simulated dataset
sampled from the specified set of parameters */
proc summary data=combined&i;
        class nrun;
        var test1d test2d test1h test2h;
        output out=simstats&i (where=(nrun>.) drop=_freq_ _type_) mean=m1d m2d
m1h m2h std=s1d s2d s1h s2h;
run;
/*Diseased  correlation for each simulated dataset*/
proc corr data=combined&i out=Dcorr&i noprint;
    var test1d test2d;
         by nrun;
run;
data Dcorr&i; keep _TYPE_ _NAME_ test1d;
        set Dcorr&i;
        where _TYPE_='CORR' and _NAME_='test2d';
run;
data Dcorr&i (rename=(test1d=pd)); drop _TYPE_ _NAME_;        set Dcorr&i;    run;
/* Non-diseased correlation for each simulated dataset*/
proc corr data=combined&i out=Hcorr&i noprint;
    var test1h test2h;
         by nrun;
run;
data Hcorr&i; keep _TYPE_ _NAME_ test1h;
        set Hcorr&i;
        where _TYPE_='CORR' and _NAME_='test2h';
run;
data Hcorr&i (rename=(test1h=ph)); drop _TYPE_ _NAME_;        set Hcorr&i;    run;

data SimDataMSDC&i;        merge simstats&i Dcorr&i Hcorr&i;    run;

filename junk dummy;
proc printto  log=junk; run;
%macro sim;
%do m=1 %to &times;
proc nlp data=SimDataMSDC&i noprint out=AllSimDataNLP&m tech=&tech
              absgconv=&converge maxiter=&iter maxfunc=&func;
      where nrun=&m;
      max GYI;
      parms theta1 = &t1, theta2 = &t2;
      nrun=&m;
      TPR = 1-probbnrm( ((theta1-m1d)/s1d), ((theta2-m2d)/s2d), pd);
      FPR = 1-probbnrm( ((theta1-m1h)/s1h), ((theta2-m2h)/s2h), ph);
      Sens = TPR;
      Spec = 1 - FPR;
      GYI = TPR - &R*FPR;
run;
quit;
proc append data=AllSimDataNLP&m base=AllSimData&i; run;
```

```
proc datasets nolist; delete AllSimDataNLP&m; run; quit;
%end;
%mend sim;
%sim;
proc printto;  run;

data AllSimData&i;
        set AllSimData&i;
        lcutoff1 = m1h - (&z*s1h);      lcutoff2 = m2h - (&z*s2h);
        hcutoff1 = m1d + (&z*s1d);     hcutoff2 = m2d + (&z*s2d);
run;

data extreme_sim&i;
        set AllSimData&i;
                if theta1<lcutoff1 or theta2<lcutoff2 or theta1>hcutoff1 or theta2>hcutoff2;
run;
proc sort data=AllSimData&i;          by nrun;          run;
data finalsimdata&i (keep = nrun m1h s1h m2h s2h m1d s1d m2d s2d ph pd lcutoff1
hcutoff1 lcutoff2 hcutoff2 finalsimtheta1 finalsimtheta2 finalsimGYI finalsimSens
finalsimSpec good_sim_thetas prob_sim1 prob_sim2 prob_sim12);
        set AllSimData&i;
        by nrun;
        if lcutoff1 <= theta1 <= hcutoff1 and lcutoff2 <= theta2 <= hcutoff2 then
        good_sim_thetas=1;
                else good_sim_thetas=0;

        if good_sim_thetas=1 then do;
                finalsimtheta1 = theta1*good_sim_thetas;
                finalsimtheta2 = theta2*good_sim_thetas;
                finalsimGYI = GYI*good_sim_thetas;
                finalsimSens = Sens*good_sim_thetas;
                finalsimSpec = Spec*good_sim_thetas;
        end;
        if good_sim_thetas=0 then do;
                finalsimtheta1 = .;
                finalsimtheta2 = .;
                finalsimGYI = .;
                finalsimSens = .;
                finalsimSpec = .;
        end;

        /* Determine if theta1, theta2, or both were bad */
        if good_sim_thetas = 0 then do;
                if theta1 < lcutoff1 or theta1 > hcutoff1 then prob_sim1=1;
                if theta2 < lcutoff2 or theta2 > hcutoff2 then prob_sim2=1;
        end;

        if good_sim_thetas = 1 then do;
                prob_sim1=0;  prob_sim2=0;
```

```
            end;

            if prob_sim1=. then prob_sim1=0;    if prob_sim2=. then prob_sim2=0;

            if prob_sim1=1 and prob_sim2=1 then prob_sim12=1;        else prob_sim12=0;
run;

/* Bootstrapping the Variance of the Theta1 and Theta2 Estimates for each Test for each
Simulated Dataset (j of them)*/

/* Merge the disease and healthy datasets with the final simulated datasets to make sure
that we're only bootstrapping to calculate variances for acceptable simulated theta BP
pairs */
data cd_finalsim&i;
        merge disease&i finalsimdata&i (keep=nrun good_sim_thetas);
        by nrun;
run;
data ch_finalsim&i;
        merge healthy&i finalsimdata&i (keep=nrun good_sim_thetas);
        by nrun;
run;

/* Re-sample with Replacement 1,000 times from each of the j simulated datasets, */
/* with each Re-sample (Replicate) having 100 observations, N1 from diseased and */
/* N0 from healthy */
%let seedval1=%eval(123456789+&i);        %let seedval0=%eval(987654321+&i);

%macro boot;
        %do q=1 %to &times;
                proc surveyselect data=cd_finalsim&i noprint out=bootsim_d&q
                        seed=&seedval1 method=urs samprate=1 outhits rep=&rep;
                        where nrun=&q;
                run;
                proc surveyselect data=ch_finalsim&i noprint out=bootsim_h&q
                        seed=&seedval0 method=urs samprate=1 outhits rep=&rep;
                        where nrun=&q;
                run;

                proc summary data=bootsim_d&q;
                        class nrun replicate;
                        var test1d test2d;
                        output out=bootstats_d&q (where=(replicate ne . and nrun ne .)
                        drop=_freq_ _type_) mean=m1d m2d std=s1d s2d;
                run;
                proc summary data=bootsim_h&q;
                        class nrun replicate;
                        var test1h test2h;
                        output out=bootstats_h&q (where=(replicate ne . and nrun ne .)
                        drop=_freq_ _type_) mean=m1h m2h std=s1h s2h;
```

```
                run;
                data bootsim_means&q; merge bootstats_d&q bootstats_h&q;        run;

                proc sort data=bootsim_d&q; by nrun replicate;        run;
                proc sort data=bootsim_h&q; by nrun replicate;        run;

                proc corr data=bootsim_d&q out=Dcorrboot_&q noprint;
                var test1d test2d;
                        by nrun replicate;
                run;
                data Dcorrboot_&q; keep _TYPE_ _NAME_ test1d;
                        set Dcorrboot_&q;
                        where _TYPE_='CORR' and _NAME_='test2d';
                run;
                data Dcorrboot_&q (rename=(test1d=pd)); drop _TYPE_ _NAME_;
                        set Dcorrboot_&q;        run;
                proc corr data=bootsim_h&q out=Hcorrboot_&q noprint;
                        var test1h test2h;
                        by nrun replicate;
                run;
                data Hcorrboot_&q; keep _TYPE_ _NAME_ test1h;
                        set Hcorrboot_&q;
                        where _TYPE_='CORR' and _NAME_='test2h';
                run;
                data Hcorrboot_&q (rename=(test1h=ph)); drop _TYPE_ _NAME_;
                        set Hcorrboot_&q;        run;

                data BootDataMSDC_&q;        merge bootsim_means&q Dcorrboot_&q
                        Hcorrboot_&q;run;

                proc datasets nolist;    delete bootsim_d&q bootsim_h&q
                        bootsim_comb&q bootsim_means&q Dcorrboot_&q
                        Hcorrboot_&q;run; quit;
                data first&q;    set BootDataMSDC_&q;
                        if 1<=replicate<=333; run;
                data middle&q;         set BootDataMSDC_&q;
                        if 334<=replicate<=667;         run;
                data last&q;    set BootDataMSDC_&q;
                        if 668<=replicate<=1000;        run;

filename junk dummy;
proc printto  log=junk; run;
proc nlp data=first&q noprint out=AllBootDataNLP_1 tech=&tech
        absgconv=&converge maxiter=&iter maxfunc=&func;
        by replicate;
                max GYI;
                parms boottheta1 = &t1, boottheta2 = &t2;
                nrun=&q;
```

```
                  TPR = 1-probbnrm( ((boottheta1-m1d)/s1d), ((boottheta2-
                                m2d)/s2d), pd);
                  FPR = 1-probbnrm( ((boottheta1-m1h)/s1h), ((boottheta2-
                            m2h)/s2h), ph);
                  Sens = TPR;
                  Spec = 1 - FPR;
                  GYI = TPR - &R*FPR;
            run;   quit;

            proc nlp data=middle&q noprint out=AllBootDataNLP_2 tech=&tech
                  absgconv=&converge maxiter=&iter maxfunc=&func;
                  by replicate;
                  max GYI;
                  parms boottheta1 = &t1, boottheta2 = &t2;
                  nrun=&q;
                  TPR = 1-probbnrm( ((boottheta1-m1d)/s1d), ((boottheta2-
                            m2d)/s2d), pd);
                  FPR = 1-probbnrm( ((boottheta1-m1h)/s1h), ((boottheta2-
                        m2h)/s2h), ph);
                  Sens = TPR;
                  Spec = 1 - FPR;
                  GYI = TPR - &R*FPR;
            run;   quit;

            proc nlp data=last&q noprint out=AllBootDataNLP_3 tech=&tech
                  absgconv=&converge maxiter=&iter maxfunc=&func;
                  by replicate;
                  max GYI;
                  parms boottheta1 = &t1, boottheta2 = &t2;
                  nrun=&q;
                  TPR = 1-probbnrm( ((boottheta1-m1d)/s1d), ((boottheta2-
                            m2d)/s2d), pd);
                  FPR = 1-probbnrm( ((boottheta1-m1h)/s1h), ((boottheta2-
                            m2h)/s2h), ph);
                  Sens = TPR;
                  Spec = 1 - FPR;
                  GYI = TPR - &R*FPR;
            run;   quit;

data AllBootDataNLP;
      set AllBootDataNLP_1 AllBootDataNLP_2 AllBootDataNLP_3;
run;

proc datasets nolist;
      append data=AllBootDataNLP base=AllBootData&i;
      delete BootDataMSDC_&q first&q middle&q last&q AllBootDataNLP_1
            AllBootDataNLP_2 AllBootDataNLP_3 AllBootDataNLP;
run; quit;
proc printto;   run;
```

```
%end;
%mend;
%boot;


data AllBootData&i;
       set AllBootData&i (keep = nrun replicate m1h s1h m2h s2h m1d s1d m2d s2d ph
               pd boottheta1 boottheta2 GYI Sens Spec);
               lcutoff1 = m1h - (&z*s1h);      lcutoff2 = m2h - (&z*s2h);
               hcutoff1 = m1d + (&z*s1d);    hcutoff2 = m2d + (&z*s2d);
run;


data extreme&i;
       set allbootdata&i;
               if boottheta1<lcutoff1 or boottheta2<lcutoff2 or boottheta1>hcutoff1 or
               boottheta2>hcutoff2;
run;


data finalbootdata&i (keep = nrun replicate lcutoff1 lcutoff2 hcutoff1 hcutoff2
       finalboottheta1 finalboottheta2 finalbootSens finalbootSpec good_boot_thetas);
       set AllBootData&i;
       by nrun replicate;
       if lcutoff1 <= boottheta1 <= hcutoff1 and lcutoff2 <= boottheta2 <= hcutoff2 then
       good_boot_thetas=1;
               else good_boot_thetas=0;

       if good_boot_thetas=1 then do;
               finalboottheta1 = boottheta1;
               finalboottheta2 = boottheta2;
               finalbootSens = Sens;
               finalbootSpec = Spec;
       end;
       if good_boot_thetas=0 then do;
               finalboottheta1 = .;
               finalboottheta2 = .;
               finalbootSens = .;
               finalbootSpec = .;
       end;
run;

/* Calculate the Variances of  Theta1 and Theta2 for each of the Bootstrapped Datasets'
1,000 Replicates */
proc sort data=finalbootdata&i;        by nrun;        run;
proc corr data = finalbootdata&i cov out=OutCov&i (type=cov) nocorr noprint;
       var finalboottheta1 finalboottheta2;
       by nrun;
       where good_boot_thetas=1;
run;
data OutCov1_&i;
       set OutCov&i;
```

```
        where _TYPE_='COV';
        by nrun;
run;
data S1_&i S2_&i;
        set OutCov1_&i;
        by nrun;
        first=first.nrun;
        last=last.nrun;
        if first.nrun = 1 then output S1_&i;
        if last.nrun = 1 then output S2_&i;
run;
data S1_&i;
        set S1_&i (keep = nrun finalboottheta1 finalboottheta2);
        rename finalboottheta1=VarT1;
        rename finalboottheta2=CovT1T2;
run;
data S2_&i;
        set S2_&i (keep = nrun finalboottheta2);
        rename finalboottheta2=VarT2;
run;
data VarCov&i;
        merge S1_&i S2_&i;
        by nrun;
run;
/* Calculating the Mean Bootstrapped Theta's */
proc means data = finalbootdata&i noprint;
        by nrun;
        var finalboottheta1;
        where good_boot_thetas=1;
        output out = Means1_&i mean = MeanTheta1;
run;
proc means data = finalbootdata&i noprint;
        by nrun;
        var finalboottheta2;
        where good_boot_thetas=1;
        output out = Means2_&i mean = MeanTheta2;
run;
data Means&i;
        merge Means1_&i Means2_&i;
        by nrun;
        drop _type_ _freq_;
run;

proc summary data=finalbootdata&i;
        class nrun;
        var finalbootSens finalbootSpec;
        output out=bootvarSS&i (where=(nrun>.) drop=_freq_ _type_) var=VarSe VarSp;
        where good_boot_thetas=1;
run;
```

```
data Means_VarCov&i;        merge Means&i VarCov&i;      by nrun;        run;
data SimThetaVarCov&i;       merge finalsimdata&i Means_VarCov&i bootvarSS&i;
        by nrun;             run;
```

/* Calculation of the Simultaneous (1-alpha)100% CI's for the Estimated Thresholds */
```
data SimConfInt&i;
        set SimThetaVarCov&i;
        f = finv(0.95, &p, &TSS-&p);
        cv = probit(1-0.05/(2*&p));
        lobon1 = finalsimtheta1 - cv*sqrt(varT1);
        upbon1 = finalsimtheta1 + cv*sqrt(varT1);
        lobon2 = finalsimtheta2 - cv*sqrt(varT2);
        upbon2 = finalsimtheta2 + cv*sqrt(varT2);
run;
```

/* Combine Original "True" Threshold Values with the Estimated Simultaneous CI's to ID if */
/* the True Values lie within the SCIs/CRs.. Done with an Indicator Variable 0 or 1. */
/* This will be used to estimate the Coverage. Additionally the area of each SCI/CR is calculated. */
```
data SCI_CR&i;
        if _n_ = 1 then set Parms_TrueValues&i;
        set SimConfInt&i (keep=nrun good_sim_thetas finalsimtheta1 finalsimtheta2
                finalsimSens finalsimSpec finalsimGYI MeanTheta1 MeanTheta2
                VarT1 VarT2 CovT1T2 VarSe VarSp f lobon1 upbon1 lobon2 upbon2
                prob_sim1 prob_sim2 prob_sim12);

        if good_sim_thetas=0 then do;
                MeanTheta1=.;        MeanTheta2=.;
                VarT1=.;        VarT2=.;        CovT1T2=.;      VarSe=.;        VarSp=.;
        end;

        /* 95% Simultaneous Confidence Regions for each of the Simulated Dataset
        Theta1 and Theta2 Optimal Threshold Values */
        if lobon1 <= truetheta1 <= upbon1 AND lobon2 <= truetheta2 <= upbon2 then
        TrueInSCI = 1;
                else TrueInSCI=0;
        if lobon1 = . or upbon1 = . or lobon2 = . or upbon2 = . then TrueInSCI=.;

        /* Calculates the width of each CI for each test */
        SCI_wTest1 = upbon1-lobon1;        SCI_wTest2 = upbon2-lobon2;

        /* Calculates the area of the rectangular Simultaneous CI region */
        SCI_area = SCI_wTest1 * SCI_wTest2;

        /* 95% Confidence Regions for each of the Simulated Dataset Theta1 and
        Theta2 Optimal Threshold Values */
        /* The Variances and Covariances were estimated using the Bootstrapped
        Datasets from each of the Simulated Datasets */
```

```
if good_sim_thetas=1 then do;
        det = (VarT1*VarT2) - ((CovT1T2)**2);
        matmult = (VarT2*(finalsimtheta1 - truetheta1)**2)
                - (2*CovT1T2*(finalsimtheta1 - truetheta1)*(finalsimtheta2 -
                  truetheta2))  + (VarT1*(finalsimtheta2 - truetheta2)**2);
        LHS = (1/det) * matmult;
        RHS = ( (&p*(&TSS-1)) / (&TSS-&p) ) * f;
end;
if good_sim_thetas=0 then do;
        det=.;   matmult=.;        LHS=.;          RHS=.;
end;

if LHS <= RHS then TrueInConfReg = 1;
if LHS > RHS then TrueInConfReg = 0;
if LHS = . and RHS = . then TrueInConfReg = .;
```
/* Calculate the Area of the CRs */
```
        trace = VarT1 + VarT2;
        lambda1 = ( trace + Sqrt( trace**2 - 4*det ) ) / 2;
        lambda2 = ( trace - Sqrt( trace**2 - 4*det ) ) / 2;
        HLength_lambda1 = Sqrt(lambda1) * Sqrt( RHS );
        HLength_lambda2 = Sqrt(lambda2) * Sqrt( RHS );
        Area_CR = &pi * HLength_lambda1 * Hlength_lambda2;
        correlation = ( CovT1T2 / (sqrt(VarT1*VarT2)) );

run;
data all_SCI_CR&i;
        set SCI_CR&i (drop = f det matmult LHS RHS trace lambda1 lambda2
                HLength_lambda1 HLength_lambda2);
run;

data bias_MSE_SCI_CR&i;   set all_SCI_CR&i;
        biasTheta1 = finalsimtheta1 - truetheta1;
        biasTheta2 = finalsimtheta2 - truetheta2;
        biasSens = finalsimSens - trueSens;
        biasSpec = finalsimSpec - trueSpec;

        rmse_theta1 = sqrt( VarT1 + (biasTheta1)**2 );
        rmse_theta2 = sqrt( VarT2 + (biasTheta2)**2 );
        rmse_Sens = sqrt( VarSe + (biasSens)**2 );
        rmse_Spec = sqrt( VarSp + (biasSpec)**2 );

        if biasSens > 0 then flagOverSens=1;
        if biasSens < 0 then flagOverSens=0;
        if biasSens = . then flagOverSens = .;
        if biasSpec > 0 then flagOverSpec=1;
        if biasSpec < 0 then flagOverSpec=0;
        if biasSpec = . then flagOverSpec = .;
run;
```

```
/* Estimate the Coverage of the Simultaneous CI's and Confidence Regions */
/* Calculates the number of times the sensitivity and specificity from the simulated data
is over/underestimated from the true values */
proc freq data = bias_MSE_SCI_CR&i noprint;
        table good_sim_thetas / list norow nocol nofreq nocum out=num_missing&i;
        table prob_sim1 / list norow nocol nofreq nocum out=prob_T1_&i;
        table prob_sim2 / list norow nocol nofreq nocum out=prob_T2_&i;
        table prob_sim12 / list norow nocol nofreq nocum out=prob_T1T2_&i;
        table TrueInSCI / list norow nocol nofreq nocum out=Coverage_SCI&i;
        table TrueInConfReg / list norow nocol nofreq nocum out=Coverage_CR&i;
        table flagOverSens / list norow nocol nofreq nocum out=Overest_SimSe&i;
        table flagOverSpec / list norow nocol nofreq nocum out=Overest_SimSp&i;
run;


/* Calculates the mean Bias and RMSE for each simulated theta1 and theta2 value from
the true values */
proc means data = bias_MSE_SCI_CR&i n mean median std q1 q3 noprint;
        var biasTheta1 biasTheta2 biasSens biasSpec rmse_theta1 rmse_theta2
                rmse_Sens rmse_Spec;
        output out=bias_sim&i (drop=_type_ _freq_)
        mean=mean_biasT1 mean_biasT2 mean_biasSe mean_biasSp mean_RMSET1
        mean_RMSET2 mean_RMSESe mean_RMSESp
         median=med_biasT1 med_biasT2 med_biasSe med_biasSp med_RMSET1
        med_RMSET2 med_RMSESe med_RMSESp;
run;
/* Calculates the mean/median of the Variances, Covariances, and computed correlation
*/
/* Compute the mean/median of the width of the SCIs, and the area of the CR */
proc summary data=bias_MSE_SCI_CR&i;
        var VarT1 VarT2 CovT1T2 correlation SCI_wTest1 SCI_wTest2 Area_CR;
        output out=VarCovCorrWidthArea_Boot&i (drop=_freq_ _type_)
         mean=mean_VarT1 mean_VarT2 mean_CovT1T2 mean_comp_corr
        meanSCI_width1 meanSCI_width2 meanCR_area
        median=med_VarT1 med_VarT2 med_CovT1T2 med_comp_corr
        medSCI_width1 medSCI_width2 medCR_area;
run;


/* Calculates the correlation between the simulated theta1 and theta2 values */
proc corr data=finalsimdata&i outp=auto_corr_simthetas&i noprint;
        var finalsimtheta1 finalsimtheta2;
        where good_sim_thetas=1;
run;
data auto_corr_simthetas&i; keep _TYPE_ _NAME_ finalsimtheta1;
        set auto_corr_simthetas&i;
        where _TYPE_='CORR' and _NAME_='finalsimtheta2';
run;
data auto_corr_simthetas&i (rename=(finalsimtheta1=auto_corr)); drop _TYPE_
        _NAME_;      set auto_corr_simthetas&i;    run;
```

```
/* Create Parameter Set Data File */
proc sort data=bias_MSE_SCI_CR&i;          by nrun;          run;
proc sort data=finalsimdata&i;          by nrun;          run;
data final_parm_set_nosim&i;
        set bias_MSE_SCI_CR&I;
        by nrun;
run;
data final_parm_set_test&i;
        merge final_parm_set_nosim&i finalsimdata&i (keep = nrun m1h m2h m1d m2d
        s1h s2h s1d s2d ph pd);
        by nrun;
        rename m1h=sim_m1h;        rename m2h=sim_m2h;
        rename m1d=sim_m1d;        rename m2d=sim_m2d;
        rename s1h=sim_s1h;        rename s2h=sim_s2h;
        rename s1d=sim_s1d;        rename s2d=sim_s2d;
        rename ph=sim_ph;    rename pd=sim_pd;
run;


data bp.final_parm_set&i;
        merge final_parm_set_nosim&i (keep=n_d n_h R ph pd AUC1 AUC2 b1 b2)
                final_parm_set_test&i (drop=n_d n_h R AUC1 AUC2 b1 b2);
        TSS = &TSS;  k=&k;
run;

/* Create Parameterization File - Results of Simulation Studies */
data num_missing&i; set num_missing&i (keep=good_sim_thetas count);
        where good_sim_thetas=0;   drop good_sim_thetas;
        rename count=Num_Sim_Missing;   label count = 'Num_Sim_Missing';
run;
data prob_T1_&i;      set prob_T1_&i (keep= prob_sim1 count);
        where prob_sim1=1;  drop prob_sim1;
        rename count=Num_Prob_T1;        label count = 'Num_Prob_T1';
run;
data prob_T2_&i;      set prob_T2_&i (keep= prob_sim2 count);
        where prob_sim2=1;  drop prob_sim2;
        rename count=Num_Prob_T2;        label count = 'Num_Prob_T2';
run;
data prob_T1T2_&i;  set prob_T1T2_&i (keep= prob_sim12 count);
        where prob_sim12=1; drop prob_sim12;
        rename count=Num_Prob_T1T2;    label count = 'Num_Prob_T1T2';
run;
data Coverage_SCI&i;        set Coverage_SCI&i (keep=TrueInSCI percent);
        where TrueInSCI=1;  drop TrueInSCI;
        rename percent=SCI_Coverage;      label percent = 'SCI_Coverage';
run;
data Coverage_CR&i;        set Coverage_CR&i (keep=TrueInConfReg percent);
        where TrueInConfReg=1;      drop TrueInConfReg;
        rename percent=CR_Coverage;      label percent = 'CR_Coverage';
run;
```

```
data Overest_SimSe&i;        set Overest_SimSe&i (keep=flagOverSens percent);

        where flagOverSens=1;        drop flagOverSens;
        rename percent=Perc_Overest_Se;  label percent = 'Perc_Overest_Se';
run;
data Overest_SimSp&i;        set Overest_SimSp&i (keep=flagOverSpec percent);
        where flagOverSpec=1;        drop flagOverSpec;
        rename percent=Perc_Overest_Sp;  label percent = 'Perc_Overest_Sp';
run;

/* Creates Final Parameterization Results - 1 observation for each Parameter set (Wide)
*/
data results&i;
        merge sasparms NRTrueValues&i VarCovCorrWidthArea_Boot&i
        auto_corr_simthetas&i bias_sim&i Num_missing&i
        Coverage_SCI&i Coverage_CR&i Overest_SimSe&i Overest_SimSp&I;
        N_H = &N0;    N_D=&N1;
        if Num_Sim_Missing = . then Num_Sim_Missing=0;
run;
proc append data=results&i base=bp.BPallresults; run;

/* Creates "Problem" file */
data problems&i;
        merge sasparms Num_missing&i prob_T1_&i prob_T2_&i prob_T1T2_&i;
        N_H = &N0;    N_D=&N1;
        /* If the numbers below are missing, there were no missing values so no dataset
        was created. For completeness, force these as 0's */
        if Num_Sim_Missing = . then Num_Sim_Missing=0;
        if Num_Prob_T1 = . then Num_Prob_T1=0;
        if Num_Prob_T2 = . then Num_Prob_T2=0;
        if Num_Prob_T1T2 = . then Num_Prob_T1T2=0;
run;
proc append data=problems&i base=bp.BPallproblems;    run;

/* Empirical */
options ps=65 ls=116 obs=max formdlim=# nodate nonumber;
libname BP 'C:\Documents and Settings\wilkar\Desktop\Dissertation!\FINAL FROM
CLUSTER - KOTI and MERLOT\BP';
libname BPemp 'C:\Documents and Settings\wilkar\Desktop\Dissertation!\FINAL FROM
CLUSTER - KOTI and MERLOT\BP\Empirical';

/* Defines what parameter set these datasets belong to */
/*%let i=101;*/
%macro EMP;
%do i=1 %to 144;

data sasparms;
        set BP.sasparms;      if _n_=&i;
        call symput('TSS',TSS);            call symput('k',k);
```

```
        call symput('AUC1',AUC1);          call symput('AUC2',AUC2);              /*
        call symput('ph',ph);    call symput('pd',pd);
        call symput('s1d',s1d);         call symput('s2d',s2d);
run;

%let pi=4*atan(1);
%let times=1000;       %let rep=1000;        %let p=2;       %let z=1.645;
%let bratio1=s1h/s1d; %let bratio2=s2h/s2d;
%let R=1;
%let N0=floor(&TSS/(&k+1));          %let N1=&TSS - &N0;
%let m1h=0;   %let s1h=1;    %let m2h=0;   %let s2h=1;
%let converge=0.001; %let iter=1000;         %let func=1000;        %let tech=nrridg;

data parms&i;

        R=&R; ph=&ph;        pd=&pd;         AUC1 = &AUC1;        AUC2 = &AUC2;
        m1h=&m1h;   s1h=&s1h;      m2h=&m2h;   s2h=&s2h;
        s1d=&s1d;     s2d=&s2d;
        b1=s1h/s1d;    b2=s2h/s2d;
        m1d = ( s1d * Sqrt(1 + (b1)**2) * probit(AUC1) ) + m1h;
        m2d = ( s2d * Sqrt(1 + (b2)**2) * probit(AUC2) ) + m2h;

        Ca1=(m1d-m1h)/s1d; Ca2=(m2d-m2h)/s2d;
        Cauc1 = probnorm( Ca1 / sqrt(1+b1**2) );
        Cauc2 = probnorm( Ca2 / sqrt(1+b2**2) );

        b1 = s1h / s1d;          a1 = (m1d - m1h) / s1d;
        b2 = s2h / s2d;          a2 = (m2d - m2h) / s2d;
run;

proc nlp data=parms&i noprint out=NRTrueValues&i tech=&tech absgconv=&converge
            maxiter=&iter maxfunc=&func;
        max trueGYI;
        parms truetheta1 = -4 to 4 by 0.1, truetheta2 = -4 to 4 by 0.1;
        trueTPR = 1-probbnrm( ((truetheta1-m1d)/s1d), ((truetheta2-m2d)/s2d), pd);
        trueFPR = 1-probbnrm( ((truetheta1-m1h)/s1h), ((truetheta2-m2h)/s2h), ph);
        trueSens = trueTPR;
        trueSpec = 1 - trueFPR;
        trueGYI = trueTPR - &R*trueFPR;
run;

data sim;
        set NRTrueValues&i;
        call symput('t1',truetheta1);
        call symput('t2',truetheta2);
run;

data Parms_TrueValues&i;
        N_d=&N1;      N_h=&N0;       R=&R;
```

```
        merge Parms&i (keep=AUC1 AUC2 b1 b2 ph pd) NRTrueValues&i
        (keep=truetheta1 truetheta2 trueGYI trueSens trueSpec);
run;

/* Simulating Data */
%let seed1=%eval(12345+&i);        %let seed2=%eval(54321+&i);
data disease&i (keep=id test1d test2d nrun h_d_flag) healthy&i (keep=id test1h test2h
nrun h_d_flag);
        set parms&i;
        cd=sqrt(1-pd**2);        ch=sqrt(1-ph**2);
                do k=1 to &times;
                        nrun=k;
                        do l=1 to &N1;
                                id=l;
                                h_d_flag=1;
                                        test1d = rannor(&seed1);
                                        test2d = pd*test1d+cd*rannor(&seed1);
                        test1d = m1d + s1d*test1d;
                        test2d = m2d + s2d*test2d;
                                output disease&i;
                        end;
                        do j=1 to &N0;
                                id=j;
                                h_d_flag=0;
                                        test1h = rannor(&seed2);
                                test2h = ph*test1h+ch*rannor(&seed2);
                                test1h = m1h + s1h*test1h;
                                test2h = m2h + s2h*test2h;
                                output healthy&i;
                        end;
                end;
run;

/* Computing the Empirical Estimates to Compare with those chosen by the BP strategy
*/
data disease_healthy&i (drop=id);
        set disease&i (rename=(test1d=test1 test2d=test2)) healthy&i
(rename=(test1h=test1 test2h=test2));
run;
proc sort data=disease_healthy&i;    by nrun;         run;

data all_empirical&i bpemp.empirical_final&i;        set disease_healthy&i;
        by nrun;
        retain yi_mx x_mx y_mx sens_mx spec_mx;
        N_d=&N1;      N_h=&N0;      N=&TSS;
        bpsum_fn =0; bpsum_sp=0;
        thresh_x=test1; thresh_y=test2;
        do q=1 to N;
                set disease_healthy&i point=q;
```

```
                    if test2<=thresh_y  and test1<=thresh_x then do;
                    if h_d_flag=1 then bpsum_fn =bpsum_fn +1;
                            if h_d_flag=0 then bpsum_sp=bpsum_sp+1;
                    end;
            end;
            sens_bp=1-bpsum_fn/N_D;
            spec_bp=bpsum_sp/N_H;
            youden=sens_bp+spec_bp-1;

            if first.nrun=1 then do;
                    yi_mx=youden;          x_mx=thresh_x;        y_mx=thresh_y;
            sens_mx=sens_bp;   spec_mx=spec_bp;
            end;
            if youden>yi_mx then do;
                    yi_mx=youden;          x_mx=thresh_x;        y_mx=thresh_y;
            sens_mx=sens_bp;   spec_mx=spec_bp;
            end;
            if youden=yi_mx and sens_bp>sens_mx then do;
                    x_mx=thresh_x;        y_mx=thresh_y;         sens_mx=sens_bp;
            spec_mx=spec_bp;
            end;
                    output all_empirical&i;
            if last.nrun=1 then output bpemp.empirical_final&i;
run;

data empirical_final_red&i;
        if _n_ = 1 then set Parms_TrueValues&i;
        set bpemp.empirical_final&i (keep = nrun x_mx y_mx yi_mx sens_mx spec_mx);
run;

data bias_emp&i;       set empirical_final_red&i;
        biasEmpTheta1 = x_mx - truetheta1;
        biasEmpTheta2 = y_mx - truetheta2;
        biasEmpSens = sens_mx - trueSens;
        biasEmpSpec = spec_mx - trueSpec;

        if biasEmpSens > 0 then flagOverEmpSens=1;       if biasEmpSens < 0 then
        flagOverEmpSens=0;
        if biasEmpSpec > 0 then flagOverEmpSpec=1;       if biasEmpSpec < 0 then
        flagOverEmpSpec=0; \
run;

/* Calculates the mean Bias and RMSE for each empirical theta1 and theta2 value from
the true values */
proc means data = bias_emp&i n mean median std q1 q3 noprint;
        var biasEmpTheta1 biasEmpTheta2 biasEmpSens biasEmpSpec;
        output out=bias_empirical&i (drop=_type_ _freq_)
        mean=mean_biasempT1 mean_biasempT2 mean_biasempSe
        mean_biasempSp
```

```
                median=med_biasempT1 med_biasempT2 med_biasempSe
                med_biasempSp;
run;
proc freq data=bias_emp&i noprint;
        table flagOverEmpSens / list norow nocol nofreq nocum out=Overest_EmpSe&i;
        table flagOverEmpSpec / list norow nocol nofreq nocum out=Overest_EmpSp&i;
run;

data Overest_EmpSe&i;        set Overest_EmpSe&i (keep=flagOverEmpSens percent);
        where flagOverEmpSens=1;  drop flagOverEmpSens;
        rename percent=Perc_Overest_Emp_Se;
        label percent = 'Perc_Overest_Emp_Se';
run;
data Overest_EmpSp&i;        set Overest_EmpSp&i (keep=flagOverEmpSpec percent);
        where flagOverEmpSpec=1;  drop flagOverEmpSpec;
        rename percent=Perc_Overest_Emp_Sp;
        label percent = 'Perc_Overest_Emp_Sp';
run;

data allemp&i;
        set bias_emp&i (keep=yi_mx x_mx y_mx sens_mx spec_mx biasEmpTheta1
                biasEmpTheta2 biasEmpSens biasEmpSpec flagOverEmpSens
                flagOverEmpSpec);
        parameter_set_number=&i;
run;

data allemp_results&i;
        merge bias_empirical&i Overest_EmpSe&i Overest_EmpSp&i;
        parameter_set_number=&i;
run;

proc append data=allemp&i base=bpemp.BPallempicaldata;          run;
proc append data=allemp_results&i base=bpemp.BPallempiricalresults;    run;

%end;
%mend EMP;
%EMP;
```

8.4.3   BN Simulations and Empirical Estimates

```
options ps=65 ls=116 obs=max formdlim=# nodate nonumber;
libname BN 'C:\Documents and Settings\wilkar\Desktop\Dissertation!\FINAL FROM
CLUSTER - KOTI and MERLOT\BN';

/* Defines what parameter set these datasets belong to */
%let i=1;
```

```
data sasparms;
        set BN.sasparms3;    if _n_=&i;
        call symput('TSS',TSS);              call symput('k',k);
        call symput('AUC1',AUC1);            call symput('AUC2',AUC2);           /*
        call symput('ph',ph);   call symput('pd',pd);
        call symput('s1d',s1d);         call symput('s2d',s2d);
run;

%let pi=4*atan(1);
%let times=1000;      %let rep=1000;        %let p=2;              %let z=1.645;
%let bratio1=s1h/s1d; %let bratio2=s2h/s2d;
%let R=1;
%let N0=floor(&TSS/(&k+1));          %let N1=&TSS - &N0;
%let m1h=0;  %let s1h=1;  %let m2h=0;  %let s2h=1;
%let converge=0.001; %let iter=1000;       %let func=1000;         %let tech=nrridg;

data parms&i;

        R=&R; ph=&ph;        pd=&pd;       AUC1 = &AUC1;        AUC2 = &AUC2;
        m1h=&m1h;  s1h=&s1h;    m2h=&m2h;   s2h=&s2h;
        s1d=&s1d;      s2d=&s2d;
        b1=s1h/s1d;   b2=s2h/s2d;
        m1d = ( s1d * Sqrt(1 + (b1)**2) * probit(AUC1) ) + m1h;
        m2d = ( s2d * Sqrt(1 + (b2)**2) * probit(AUC2) ) + m2h;

        Ca1=(m1d-m1h)/s1d; Ca2=(m2d-m2h)/s2d;
        Cauc1 = probnorm( Ca1 / sqrt(1+b1**2) );
        Cauc2 = probnorm( Ca2 / sqrt(1+b2**2) );

        b1 = s1h / s1d;       a1 = (m1d - m1h) / s1d;
        b2 = s2h / s2d;       a2 = (m2d - m2h) / s2d;
run;

proc nlp data=parms&i noprint out=NRTrueValues&i tech=&tech absgconv=&converge
            maxiter=&iter maxfunc=&func;
        max trueGYI;
        parms truetheta3 = -4 to 4 by 0.1, truetheta4 = -4 to 4 by 0.1;
        trueTPR = 1-probnorm((truetheta3-m1d)/s1d)-probnorm((truetheta4-m2d)/s2d) +
            probbnrm( ((truetheta3-m1d)/s1d), ((truetheta4-m2d)/s2d), pd);
        trueFPR = 1-probnorm((truetheta3-m1h)/s1h)-probnorm((truetheta4-m2h)/s2h) +
            probbnrm( ((truetheta3-m1h)/s1h), ((truetheta4-m2h)/s2h), ph);
        trueSens = trueTPR;
        trueSpec = 1 - trueFPR;
        trueGYI = trueTPR - &R*trueFPR;
run;

data sim;
        set NRTrueValues&i;
        call symput('t3',truetheta3);
```

```
        call symput('t4',truetheta4);
run;

data Parms_TrueValues&i;
        N_d=&N1;    N_h=&N0;    R=&R;
        merge Parms&i (keep=AUC1 AUC2 b1 b2 ph pd)
        NRTrueValues&i (keep=truetheta3 truetheta4 trueGYI trueSens trueSpec);
run;

/* Simulating Data */
%let seed1=%eval(12345+&i);        %let seed2=%eval(54321+&i);
data disease&i (keep=id test3d test4d nrun h_d_flag) healthy&i (keep=id test3h test4h
nrun h_d_flag);
        set parms&i;
        cd=sqrt(1-pd**2);      ch=sqrt(1-ph**2);
            do k=1 to &times;
                    nrun=k;
                    do l=1 to &N1;
                            id=l;
                            h_d_flag=1;
                                    test3d = rannor(&seed1);
                                    test4d = pd*test3d+cd*rannor(&seed1);
                    test3d = m1d + s1d*test3d;
                    test4d = m2d + s2d*test4d;
                            output disease&i;
                    end;
                    do j=1 to &N0;
                            id=j;
                            h_d_flag=0;
                                    test3h = rannor(&seed2);
                            test4h = ph*test3h+ch*rannor(&seed2);
                            test3h = m1h + s1h*test3h;
                            test4h = m2h + s2h*test4h;
                            output healthy&i;
                    end;
            end;
run;

/* Combine the two datasets and sort by disease status */
data combined&i;      merge disease&i healthy&i;   by nrun id;     run;

/* Finding the Means, Variances, and Correlations for each simulated dataset sampled
from the specified set of parameters */
proc summary data=combined&i;
        class nrun;
        var test3d test4d test3h test4h;
        output out=simstats&i (where=(nrun>.) drop=_freq_ _type_)
        mean=m1d m2d m1h m2h std=s1d s2d s1h s2h;
run;
```

```
/*Diseased  correlation for each simulated dataset*/
proc corr data=combined&i out=Dcorr&i noprint;
    var test3d test4d;
        by nrun;
run;
data Dcorr&i; keep _TYPE_ _NAME_ test3d;
        set Dcorr&i;
        where _TYPE_='CORR' and _NAME_='test4d';
run;
data Dcorr&i (rename=(test3d=pd)); drop _TYPE_ _NAME_;       set Dcorr&i;   run;
/* Non-diseased correlation for each simulated dataset*/
proc corr data=combined&i out=Hcorr&i noprint;
    var test3h test4h;
        by nrun;
run;
data Hcorr&i; keep _TYPE_ _NAME_ test3h;
        set Hcorr&i;
        where _TYPE_='CORR' and _NAME_='test4h';
run;
data Hcorr&i (rename=(test3h=ph)); drop _TYPE_ _NAME_;       set Hcorr&i;   run;

data SimDataMSDC&i;         merge simstats&i Dcorr&i Hcorr&i;   run;

filename junk dummy;
proc printto  log=junk; run;
%macro sim;
%do m=1 %to &times;
proc nlp data=SimDataMSDC&i noprint out=AllSimDataNLP&m tech=&tech
                absgconv=&converge maxiter=&iter maxfunc=&func;
      where nrun=&m;
      max GYI;
      parms theta3 = &t3, theta4 = &t4;
      nrun=&m;
      TPR = 1-probnorm((theta3-m1d)/s1d)-probnorm((theta4-m2d)/s2d) + probbnrm(
            ((theta3-m1d)/s1d), ((theta4-m2d)/s2d), pd);
      FPR = 1-probnorm((theta3-m1h)/s1h)-probnorm((theta4-m2h)/s2h) + probbnrm(
            ((theta3-m1h)/s1h), ((theta4-m2h)/s2h), ph);
      Sens = TPR;
      Spec = 1 - FPR;
      GYI = TPR - &R*FPR;
run;
quit;
proc append data=AllSimDataNLP&m base=AllSimData&i; run;
proc datasets nolist; delete AllSimDataNLP&m; run; quit;
%end;
%mend sim;
%sim;
proc printto; run;
```

```
data AllSimData&i;
      set AllSimData&i (drop=_obs_ _type_);
      lcutoff1 = m1h - (&z*s1h);      lcutoff2 = m2h - (&z*s2h);
      hcutoff1 = m1d + (&z*s1d);    hcutoff2 = m2d + (&z*s2d);
run;

data extreme_sim&i;
      set AllSimData&i;
            if theta3<lcutoff1 or theta4<lcutoff2 or theta3>hcutoff1 or theta4>hcutoff2;
run;

proc sort data=AllSimData&i;          by nrun;          run;
data finalsimdata&i (keep = nrun m1h s1h m2h s2h m1d s1d m2d s2d ph pd lcutoff1
            hcutoff1 lcutoff2 hcutoff2 finalsimtheta3 finalsimtheta4 finalsimGYI
            finalsimSens finalsimSpec good_sim_thetas prob_sim3 prob_sim4
            prob_sim34);
      set AllSimData&i;
      by nrun;

      if lcutoff1 <= theta3 <= hcutoff1 and lcutoff2 <= theta4 <= hcutoff2
      then good_sim_thetas=1;
            else good_sim_thetas=0;

      if good_sim_thetas=1 then do;
            finalsimtheta3 = theta3;
            finalsimtheta4 = theta4;
            finalsimGYI = GYI;
            finalsimSens = Sens;
            finalsimSpec = Spec;
      end;
      if good_sim_thetas=0 then do;
            finalsimtheta3 = .;
            finalsimtheta4 = .;
            finalsimGYI = .;
            finalsimSens = .;
            finalsimSpec = .;
      end;

      /* Determine if theta1, theta2, or both were bad */
      if good_sim_thetas = 0 then do;
            if theta3 < lcutoff1 or theta3 > hcutoff1 then prob_sim3=1;
            if theta4 < lcutoff2 or theta4 > hcutoff2 then prob_sim4=1;
      end;

      if good_sim_thetas = 1 then do;
            prob_sim3=0;  prob_sim4=0;
      end;

      if prob_sim3=. then prob_sim3=0;    if prob_sim4=. then prob_sim4=0;
```

```
        if prob_sim3=1 and prob_sim4=1 then prob_sim34=1;       else
prob_sim34=0;
run;
```

/* Bootstrapping the Variance of the Theta3 and Theta4 Estimates for each Test for each Simulated Dataset (j of them)*/

/* Merge the disease and healthy datasets with the final simulated datasets to make sure that we're only bootstrapping to calculate variances for acceptable simulated theta BN pairs */

```
data cd_finalsim&i;
        merge disease&i finalsimdata&i (keep=nrun good_sim_thetas);
        by nrun;
run;
data ch_finalsim&i;
        merge healthy&i finalsimdata&i (keep=nrun good_sim_thetas);
        by nrun;
run;
```

/* Re-sample with Replacement 1,000 times from each of the j simulated datasets, */
/* with each Re-sample (Replicate) having 100 observations, N1 from diseased and */
/* N0 from healthy */
```
%let seedval1=%eval(123456789+&i);       %let seedval0=%eval(987654321+&i);

%macro boot;
        %do q=1 %to &times;
                proc surveyselect data=cd_finalsim&i noprint out=bootsim_d&q
                        seed=&seedval1 method=urs samprate=1 outhits rep=&rep;
                        where nrun=&q;
                run;
                proc surveyselect data=ch_finalsim&i noprint out=bootsim_h&q
                        seed=&seedval0 method=urs samprate=1 outhits rep=&rep;
                        where nrun=&q;
                run;

                /* Calculate means and sd's for disease and healthy */
                proc summary data=bootsim_d&q;
                        class nrun replicate;
                        var test3d test4d;
                        output out=bootstats_d&q (where=(replicate ne . and nrun ne .)
                        drop=_freq_ _type_) mean=m1d m2d std=s1d s2d;
                run;
                proc summary data=bootsim_h&q;
                        class nrun replicate;
                        var test3h test4h;
                        output out=bootstats_h&q (where=(replicate ne . and nrun ne .)
                        drop=_freq_ _type_) mean=m1h m2h std=s1h s2h;
                run;
```

```
        data bootsim_means&q; merge bootstats_d&q bootstats_h&q;
run;

        proc sort data=bootsim_d&q; by nrun replicate;        run;
        proc sort data=bootsim_h&q; by nrun replicate;        run;

        /*Diseased  correlation for each bootstrap dataset*/
        proc corr data=bootsim_d&q out=Dcorrboot_&q noprint;
        var test3d test4d;
                by nrun replicate;
        run;
        data Dcorrboot_&q; keep _TYPE_ _NAME_ test3d;
                set Dcorrboot_&q;
                where _TYPE_='CORR' and _NAME_='test4d';
        run;
        data Dcorrboot_&q (rename=(test3d=pd)); drop _TYPE_ _NAME_;
                set Dcorrboot_&q;        run;
        /* Non-diseased correlation for each bootstrap dataset*/
        proc corr data=bootsim_h&q out=Hcorrboot_&q noprint;
        var test3h test4h;
                by nrun replicate;
        run;
        data Hcorrboot_&q; keep _TYPE_ _NAME_ test3h;
                set Hcorrboot_&q;
                where _TYPE_='CORR' and _NAME_='test4h';
        run;
        data Hcorrboot_&q (rename=(test3h=ph)); drop _TYPE_ _NAME_;
                set Hcorrboot_&q;        run;

        data BootDataMSDC_&q;
                merge bootsim_means&q Dcorrboot_&q Hcorrboot_&q;        run;

        proc datasets nolist;   delete bootsim_d&q bootsim_h&q
                        bootsim_comb&q bootsim_means&q Dcorrboot_&q
                        Hcorrboot_&q; run; quit;
        data first&q;   set BootDataMSDC_&q;
                if 1<=replicate<=333; run;
        data middle&q;        set BootDataMSDC_&q;
                if 334<=replicate<=667;        run;
        data last&q;   set BootDataMSDC_&q;
                if 668<=replicate<=1000;        run;

filename junk dummy;
proc printto  log=junk; run;
proc nlp data=first&q noprint out=AllBootDataNLP_1 tech=&tech
                absgconv=&converge maxiter=&iter maxfunc=&func;
        by replicate;
                max GYI;
                parms boottheta3 = &t3, boottheta4 = &t4;
```

```
                    nrun=&q;
                    TPR = 1-probnorm((boottheta3-m1d)/s1d)-probnorm((boottheta4-
                            m2d)/s2d) + probbnrm( ((boottheta3-m1d)/s1d),
                            ((boottheta4-m2d)/s2d), pd);
                    FPR = 1-probnorm((boottheta3-m1h)/s1h)-probnorm((boottheta4-
                            m2h)/s2h) + probbnrm( ((boottheta3-m1h)/s1h),
                            ((boottheta4-m2h)/s2h), ph);
                    Sens = TPR;
                    Spec = 1 - FPR;
                    GYI = TPR - &R*FPR;
run;    quit;

            proc nlp data=middle&q noprint out=AllBootDataNLP_2 tech=&tech
                        absgconv=&converge maxiter=&iter maxfunc=&func;
            by replicate;
                    max GYI;
                    parms boottheta3 = &t3, boottheta4 = &t4;
                    nrun=&q;
                    TPR = 1-probnorm((boottheta3-m1d)/s1d)-probnorm((boottheta4-
                            m2d)/s2d) + probbnrm( ((boottheta3-m1d)/s1d),
                            ((boottheta4-m2d)/s2d), pd);
                    FPR = 1-probnorm((boottheta3-m1h)/s1h)-probnorm((boottheta4-
                            m2h)/s2h) + probbnrm( ((boottheta3-m1h)/s1h),
                            ((boottheta4-m2h)/s2h), ph);
                    Sens = TPR;
                    Spec = 1 - FPR;
                    GYI = TPR - &R*FPR;
run;    quit;

            proc nlp data=last&q noprint out=AllBootDataNLP_3 tech=&tech
                        absgconv=&converge maxiter=&iter maxfunc=&func;
            by replicate;
                    max GYI;
                    parms boottheta3 = &t3, boottheta4 = &t4;
                    nrun=&q;
                    TPR = 1-probnorm((boottheta3-m1d)/s1d)-probnorm((boottheta4-
                            m2d)/s2d) + probbnrm( ((boottheta3-m1d)/s1d),
                            ((boottheta4-m2d)/s2d), pd);
                    FPR = 1-probnorm((boottheta3-m1h)/s1h)-probnorm((boottheta4-
                            m2h)/s2h) + probbnrm( ((boottheta3-m1h)/s1h),
                            ((boottheta4-m2h)/s2h), ph);
                    Sens = TPR;
                    Spec = 1 - FPR;
                    GYI = TPR - &R*FPR;
run;    quit;

data AllBootDataNLP;
        set AllBootDataNLP_1 AllBootDataNLP_2 AllBootDataNLP_3;
run;
```

```
        proc datasets nolist;
                append data=AllBootDataNLP base=AllBootData&i;
                delete BootDataMSDC_&q first&q middle&q last&q AllBootDataNLP_1
                        AllBootDataNLP_2 AllBootDataNLP_3 AllBootDataNLP;
        run; quit;
        proc printto;    run;
%end;
%mend;
%boot;


data AllBootData&i;
        set AllBootData&i (keep = nrun replicate m1h s1h m2h s2h m1d s1d m2d s2d ph
        pd boottheta3 boottheta4 GYI Sens Spec);
                lcutoff1 = m1h - (&z*s1h);      lcutoff2 = m2h - (&z*s2h);
                hcutoff1 = m1d + (&z*s1d);    hcutoff2 = m2d + (&z*s2d);
run;


data extreme&i;
        set allbootdata&i;
                if boottheta3<lcutoff1 or boottheta4<lcutoff2 or boottheta3>hcutoff1 or
                boottheta4>hcutoff2;
run;


data finalbootdata&i (keep = nrun replicate lcutoff1 lcutoff2 hcutoff1 hcutoff2
finalboottheta3 finalboottheta4 finalbootSens finalbootSpec good_boot_thetas);
        set AllBootData&i;
        by nrun replicate;

        if lcutoff1 <= boottheta3 <= hcutoff1 and lcutoff2 <= boottheta4 <= hcutoff2 then
                good_boot_thetas=1;
                else good_boot_thetas=0;

        if good_boot_thetas=1 then do;
                finalboottheta3 = boottheta3;
                finalboottheta4 = boottheta4;
                finalbootSens = Sens;
                finalbootSpec = Spec;
        end;
        if good_boot_thetas=0 then do;
                finalboottheta3 = .;
                finalboottheta4 = .;
                finalbootSens = .;
                finalbootSpec = .;
        end;
run;

/* Calculate the Variances of  Theta1 and Theta2 for each of the Bootstrapped Datasets'
1,000 Replicates */
proc sort data=finalbootdata&i;        by nrun;        run;
```

```
proc corr data = finalbootdata&i cov out=OutCov&i (type=cov) nocorr noprint;
      var finalboottheta3 finalboottheta4;
      by nrun;
      where good_boot_thetas=1;
run;
data OutCov1_&i;
      set OutCov&i;
      where _TYPE_='COV';
      by nrun;
run;
data S1_&i S2_&i;
      set OutCov1_&i;
      by nrun;
      first=first.nrun;
      last=last.nrun;
      if first.nrun = 1 then output S1_&i;
      if last.nrun = 1 then output S2_&i;
run;
data S1_&i;
      set S1_&i (keep = nrun finalboottheta3 finalboottheta4);
      rename finalboottheta3=VarT3;
      rename finalboottheta4=CovT3T4;
run;
data S2_&i;
      set S2_&i (keep = nrun finalboottheta4);
      rename finalboottheta4=VarT4;
run;
data VarCov&i;
      merge S1_&i S2_&i;
      by nrun;
run;
/* Calculating the Mean Bootstrapped Theta's */
proc means data = finalbootdata&i noprint;
      by nrun;
      var finalboottheta3;
      where good_boot_thetas=1;
      output out = Means1_&i mean = MeanTheta3;
run;
proc means data = finalbootdata&i noprint;
      by nrun;
      var finalboottheta4;
      where good_boot_thetas=1;
      output out = Means2_&i mean = MeanTheta4;
run;
data Means&i;
      merge Means1_&i Means2_&i;
      by nrun;
      drop _type_ _freq_;
run;
```

```
proc summary data=finalbootdata&i;
        class nrun;
        var finalbootSens finalbootSpec;
        output out=bootvarSS&i (where=(nrun>.) drop=_freq_ _type_) var=VarSe VarSp;
        where good_boot_thetas=1;
run;

data Means_VarCov&i;        merge Means&i VarCov&i;    by nrun;        run;
data SimThetaVarCov&i;       merge finalsimdata&i Means_VarCov&i bootvarSS&i;
        by nrun;                run;

/* Calculation of the Simultaneous (1-alpha)100% CI's for the Estimated Thresholds */
data SimConfInt&i;
        set SimThetaVarCov&i;
        f = finv(0.95, &p, &TSS-&p);
        cv = probit(1-0.05/(2*&p));
        lobon1 = finalsimtheta3 - cv*sqrt(varT3);
        upbon1 = finalsimtheta3 + cv*sqrt(varT3);
        lobon2 = finalsimtheta4 - cv*sqrt(varT4);
        upbon2 = finalsimtheta4 + cv*sqrt(varT4);
run;

/* Combine Original "True" Threshold Values with the Estimated Simultaneous CI's to ID
if the True Values lie within the SCIs/CRs.. Done with an Indicator Variable 0 or 1. */
/* This will be used to estimate the Coverage. Additionally the area of each SCI/CR is
calculated. */
data SCI_CR&i;
        if _n_ = 1 then set Parms_TrueValues&i;
        set SimConfInt&i (keep=nrun good_sim_thetas finalsimtheta3 finalsimtheta4
        finalsimSens finalsimSpec finalsimGYI MeanTheta3 MeanTheta4 VarT3 VarT4
        CovT3T4 VarSe VarSp f lobon1 upbon1 lobon2 upbon2 prob_sim3 prob_sim4
        prob_sim34);

        if good_sim_thetas=0 then do;
                MeanTheta3=.;        MeanTheta4=.;
                VarT3=.;        VarT4=.;        CovT3T4=.;    VarSe=.;        VarSp=.;
        end;

        /* 95% Simultaneous Confidence Regions for each of the Simulated Dataset
        Theta1 and Theta2 Optimal Threshold Values */
        if lobon1 <= truetheta3 <= upbon1 AND lobon2 <= truetheta4 <= upbon2 then
        TrueInSCI = 1;
                else TrueInSCI=0;
        if lobon1 = . or upbon1 = . or lobon2 = . or upbon2 = . then TrueInSCI=.;

        /* Calculates the width of each CI for each test */
        SCI_wTest1 = upbon1-lobon1;        SCI_wTest2 = upbon2-lobon2;

        /* Calculates the area of the rectangular Simultaneous CI region */
```

```
        SCI_area = SCI_wTest1 * SCI_wTest2;

        /* 95% Confidence Regions for each of the Simulated Dataset Theta1 and
        Theta2 Optimal Threshold Values */
        /* The Variances and Covariances were estimated using the Bootstrapped
        Datasets from each of the Simulated Datasets */
        if good_sim_thetas=1 then do;
                det = (VarT3*VarT4) - ((CovT3T4)**2);
                matmult = (VarT4*(finalsimtheta3 - truetheta3)**2)-
                        (2*CovT3T4*(finalsimtheta3 - truetheta3)*(finalsimtheta4 -
                        truetheta4)) + (VarT3*(finalsimtheta4 - truetheta4)**2);
                        LHS = (1/det) * matmult;
                RHS = ( (&p*(&TSS-1)) / (&TSS-&p) ) * f;
        end;
        if good_sim_thetas=0 then do;
                det=.;  matmult=.;    LHS=.;        RHS=.;
        end;

        if LHS <= RHS then TrueInConfReg = 1;
        if LHS > RHS then TrueInConfReg = 0;
        if LHS = . and RHS = . then TrueInConfReg = .;

/* Calculate the Area of the CRs */
        trace = VarT3 + VarT4;
        lambda1 = ( trace + Sqrt( trace**2 - 4*det ) ) / 2;
        lambda2 = ( trace - Sqrt( trace**2 - 4*det ) ) / 2;
        HLength_lambda1 = Sqrt(lambda1) * Sqrt( RHS );
        HLength_lambda2 = Sqrt(lambda2) * Sqrt( RHS );
        Area_CR = &pi * HLength_lambda1 * Hlength_lambda2;

        correlation = ( CovT3T4 / (sqrt(VarT3*VarT4)) );

run;
data all_SCI_CR&i;
        set SCI_CR&i (drop = f det matmult LHS RHS trace lambda1 lambda2
        HLength_lambda1 HLength_lambda2);
run;

data bias_MSE_SCI_CR&i;   set all_SCI_CR&i;
        biasTheta3 = finalsimtheta3 - truetheta3;
        biasTheta4 = finalsimtheta4 - truetheta4;
        biasSens = finalsimSens - trueSens;
        biasSpec = finalsimSpec - trueSpec;

        rmse_theta3 = sqrt( VarT3 + (biasTheta3)**2 );
        rmse_theta4 = sqrt( VarT4 + (biasTheta4)**2 );
        rmse_Sens = sqrt( VarSe + (biasSens)**2 );
        rmse_Spec = sqrt( VarSp + (biasSpec)**2 );
```

```
              if biasSens > 0 then flagOverSens=1;          if biasSens < 0 then
       flagOverSens=0;          if biasSens = . then flagOverSens = .;
              if biasSpec > 0 then flagOverSpec=1;          if biasSpec < 0 then
       flagOverSpec=0;          if biasSpec = . then flagOverSpec = .;
run;


/* Estimate the Coverage of the Simultaneous CI's and Confidence Regions */
/* Calculates the number of times the sensitivity and specificity from the simulated data
is over/underestimated from the true values */
proc freq data = bias_MSE_SCI_CR&i noprint;
       table good_sim_thetas / list norow nocol nofreq nocum out=num_missing&i;
       table prob_sim3 / list norow nocol nofreq nocum out=prob_T3_&i;
       table prob_sim4 / list norow nocol nofreq nocum out=prob_T4_&i;
       table prob_sim34 / list norow nocol nofreq nocum out=prob_T3T4_&i;
       table TrueInSCI / list norow nocol nofreq nocum out=Coverage_SCI&i;
       table TrueInConfReg / list norow nocol nofreq nocum out=Coverage_CR&i;
       table flagOverSens / list norow nocol nofreq nocum out=Overest_SimSe&i;
       table flagOverSpec / list norow nocol nofreq nocum out=Overest_SimSp&i;
run;


/* Calculates the mean Bias and RMSE for each simulated theta1 and theta2 value from
the true values */
proc means data = bias_MSE_SCI_CR&i n mean median std q1 q3 noprint;
       var biasTheta3 biasTheta4 biasSens biasSpec rmse_theta3 rmse_theta4
       rmse_Sens rmse_Spec;
       output out=bias_sim&i (drop=_type_ _freq_)
       mean=mean_biasT3 mean_biasT4 mean_biasSe mean_biasSp mean_RMSET3
       mean_RMSET4 mean_RMSESe mean_RMSESp
       median=med_biasT3 med_biasT4 med_biasSe med_biasSp med_RMSET3
       med_RMSET4 med_RMSESe med_RMSESp;
run;
/* Calculates the mean/median of the Variances, Covariances, and computed correlation
*/
/* Compute the mean/median of the width of the SCIs, and the area of the CR */
proc summary data=bias_MSE_SCI_CR&i;
       var VarT3 VarT4 CovT3T4 correlation SCI_wTest1 SCI_wTest2    Area_CR;
       output out=VarCovCorrWidthArea_Boot&i (drop=_freq_ _type_)
       mean=mean_VarT3 mean_VarT4 mean_CovT3T4 mean_comp_corr
       meanSCI_width1 meanSCI_width2 meanCR_area
       median=med_VarT3 med_VarT4 med_CovT3T4 med_comp_corr
       medSCI_width1 medSCI_width2 medCR_area;
run;


/* Calculates the correlation between the simulated theta1 and theta2 values */
proc corr data=finalsimdata&i outp=auto_corr_simthetas&i noprint;
       var finalsimtheta3 finalsimtheta4;
       where good_sim_thetas=1;
run;
data auto_corr_simthetas&i; keep _TYPE_ _NAME_ finalsimtheta3;
```

```
        set auto_corr_simthetas&i;
        where _TYPE_='CORR' and _NAME_='finalsimtheta4';
run;
data auto_corr_simthetas&i (rename=(finalsimtheta3=auto_corr));

drop _TYPE_ _NAME_;        set auto_corr_simthetas&i;    run;


/* Create Parameter Set Data File */
proc sort data=bias_MSE_SCI_CR&i;        by nrun;        run;
proc sort data=finalsimdata&i;        by nrun;        run;
data final_parm_set_nosim&i;
        merge bias_MSE_SCI_CR&I;
        by nrun;
run;
data final_parm_set_test&i;
        merge final_parm_set_nosim&i finalsimdata&i (keep = nrun m1h m2h m1d m2d
                s1h s2h s1d s2d ph pd);
        by nrun;
        rename m1h=sim_m1h;        rename m2h=sim_m2h;
        rename m1d=sim_m1d;        rename m2d=sim_m2d;
        rename s1h=sim_s1h;        rename s2h=sim_s2h;
        rename s1d=sim_s1d;        rename s2d=sim_s2d;
        rename ph=sim_ph;    rename pd=sim_pd;
run;

data bn.final_parm_set&i;
        merge final_parm_set_nosim&i (keep=n_d n_h R ph pd AUC1 AUC2 b1 b2)
                final_parm_set_test&i (drop=n_d n_h R AUC1 AUC2 b1 b2);
        TSS = &TSS;  k=&k;
run;

/* Create Parameterization File - Results of Simulation Studies */
data num_missing&i; set num_missing&i (keep=good_sim_thetas count);
        where good_sim_thetas=0;    drop good_sim_thetas;
        rename count=Num_Sim_Missing;    label count = 'Num_Sim_Missing';
run;
data prob_T3_&i;        set prob_T3_&i (keep= prob_sim3 count);
        where prob_sim3=1;  drop prob_sim3;
        rename count=Num_Prob_T3;        label count = 'Num_Prob_T3';
run;
data prob_T4_&i;        set prob_T4_&i (keep= prob_sim4 count);
        where prob_sim4=1;  drop prob_sim4;
        rename count=Num_Prob_T4;        label count = 'Num_Prob_T4';
run;
data prob_T3T4_&i;   set prob_T3T4_&i (keep= prob_sim34 count);
        where prob_sim34=1; drop prob_sim34;
        rename count=Num_Prob_T3T4;     label count = 'Num_Prob_T3T4';
run;
data Coverage_SCI&i;        set Coverage_SCI&i (keep=TrueInSCI percent);
```

```
            where TrueInSCI=1;   drop TrueInSCI;
            rename percent=SCI_Coverage;      label percent = 'SCI_Coverage';
run;
data Coverage_CR&i;        set Coverage_CR&i (keep=TrueInConfReg percent);
            where TrueInConfReg=1;     drop TrueInConfReg;
            rename percent=CR_Coverage;       label percent = 'CR_Coverage';
run;
data Overest_SimSe&i;        set Overest_SimSe&i (keep=flagOverSens percent);
            where flagOverSens=1;       drop flagOverSens;
            rename percent=Perc_Overest_Se;  label percent = 'Perc_Overest_Se';
run;
data Overest_SimSp&i;        set Overest_SimSp&i (keep=flagOverSpec percent);
            where flagOverSpec=1;       drop flagOverSpec;
            rename percent=Perc_Overest_Sp;  label percent = 'Perc_Overest_Sp';
run;

/* Creates Final Parameterization Results - 1 observation for each Parameter set (Wide)
*/
data results&i;
        merge sasparms NRTrueValues&i VarCovCorrWidthArea_Boot&i
                auto_corr_simthetas&i bias_sim&I Num_missing&i Coverage_SCI&i
        Coverage_CR&i Overest_SimSe&i;
        N_H = &N0;    N_D=&N1;
        if Num_Sim_Missing = . then Num_Sim_Missing=0;
run;
proc append data=results&i base=bn.BNallresults; run;

/* Creates "Problem" file */
data problems&i;
        merge sasparms Num_missing&i prob_T3_&i prob_T4_&i prob_T3T4_&i;
        N_H = &N0;    N_D=&N1;
        /* If the numbers below are missing, there were no missing values so no dataset
        was created. For completeness, force these as 0's */
        if Num_Sim_Missing = . then Num_Sim_Missing=0;
        if Num_Prob_T3 = . then Num_Prob_T3=0;
        if Num_Prob_T4 = . then Num_Prob_T4=0;
        if Num_Prob_T3T4 = . then Num_Prob_T3T4=0;
run;
proc append data=problems&i base=bn.BNallproblems;    run;

/* Empirical */
options ps=65 ls=116 obs=max formdlim=# nodate nonumber;
libname BN 'C:\Documents and Settings\wilkar\Desktop\Dissertation!\FINAL FROM
CLUSTER - KOTI and MERLOT\BN';
libname BNemp 'C:\Documents and Settings\wilkar\Desktop\Dissertation!\FINAL FROM
CLUSTER - KOTI and MERLOT\BN\Empirical';

%macro EMP;
%do i=1 %to 144;
```

```
data sasparms;
      set BN.sasparms3;   if _n_=&i;
      call symput('TSS',TSS);              call symput('k',k);
      call symput('AUC1',AUC1);            call symput('AUC2',AUC2);          /*
      call symput('ph',ph);   call symput('pd',pd);
      call symput('s1d',s1d);         call symput('s2d',s2d);
run;

%let pi=4*atan(1);
%let times=1000;      %let rep=1000;        %let p=2;              %let z=1.645;
%let bratio1=s1h/s1d; %let bratio2=s2h/s2d;
%let R=1;
%let N0=floor(&TSS/(&k+1));          %let N1=&TSS - &N0;
%let m1h=0;   %let s1h=1;   %let m2h=0;   %let s2h=1;
%let converge=0.001; %let iter=1000;        %let func=1000;        %let tech=nrridg;

data parms&i;
      R=&R; ph=&ph;        pd=&pd;        AUC1 = &AUC1;       AUC2 = &AUC2;
      m1h=&m1h;   s1h=&s1h;     m2h=&m2h;   s2h=&s2h;
      s1d=&s1d;     s2d=&s2d;
      b1=s1h/s1d;   b2=s2h/s2d;
      m1d = ( s1d * Sqrt(1 + (b1)**2) * probit(AUC1) ) + m1h;
      m2d = ( s2d * Sqrt(1 + (b2)**2) * probit(AUC2) ) + m2h;
      Ca1=(m1d-m1h)/s1d; Ca2=(m2d-m2h)/s2d;
      Cauc1 = probnorm( Ca1 / sqrt(1+b1**2) );
      Cauc2 = probnorm( Ca2 / sqrt(1+b2**2) );

      b1 = s1h / s1d;       a1 = (m1d - m1h) / s1d;
      b2 = s2h / s2d;       a2 = (m2d - m2h) / s2d;
run;

proc nlp data=parms&i noprint out=NRTrueValues&i tech=&tech absgconv=&converge
            maxiter=&iter maxfunc=&func;
      max trueGYI;
      parms truetheta3 = -4 to 4 by 0.1, truetheta4 = -4 to 4 by 0.1;
      trueTPR = 1-probnorm((truetheta3-m1d)/s1d)-probnorm((truetheta4-m2d)/s2d) +
            probbnrm( ((truetheta3-m1d)/s1d), ((truetheta4-m2d)/s2d), pd);
      trueFPR = 1-probnorm((truetheta3-m1h)/s1h)-probnorm((truetheta4-m2h)/s2h) +
            probbnrm( ((truetheta3-m1h)/s1h), ((truetheta4-m2h)/s2h), ph);
      trueSens = trueTPR;
      trueSpec = 1 - trueFPR;
      trueGYI = trueTPR - &R*trueFPR;
run;

data sim;
      set NRTrueValues&i;
      call symput('t1',truetheta1);
      call symput('t2',truetheta2);
run;
```

```
data Parms_TrueValues&i;
        N_d=&N1;      N_h=&N0;      R=&R;
        merge Parms&i (keep=AUC1 AUC2 b1 b2 ph pd)
        NRTrueValues&i (keep=truetheta3 truetheta4 trueGYI trueSens trueSpec);
run;

/* Simulating Data */
%let seed1=%eval(12345+&i);       %let seed2=%eval(54321+&i);
data disease&i (keep=id test1d test2d nrun h_d_flag) healthy&i (keep=id test1h test2h
nrun h_d_flag);
        set parms&i;
        cd=sqrt(1-pd**2);      ch=sqrt(1-ph**2);
               do k=1 to &times;
                     nrun=k;
                     do l=1 to &N1;
                            id=l;
                            h_d_flag=1;
                                   test1d = rannor(&seed1);
                                   test2d = pd*test1d+cd*rannor(&seed1);
                     test1d = m1d + s1d*test1d;
                     test2d = m2d + s2d*test2d;
                            output disease&i;
                     end;
                     do j=1 to &N0;
                            id=j;
                            h_d_flag=0;
                                   test1h = rannor(&seed2);
                            test2h = ph*test1h+ch*rannor(&seed2);
                            test1h = m1h + s1h*test1h;
                            test2h = m2h + s2h*test2h;
                            output healthy&i;
                     end;
               end;
run;

/* Computing the Empirical Estimates to Compare with those chosen by the BN strategy
*/
data disease_healthy&i (drop=id);
        set disease&i (rename=(test1d=test1 test2d=test2))
        healthy&i (rename=(test1h=test1 test2h=test2));
run;
proc sort data=disease_healthy&i;    by nrun;        run;

data all_empirical&i bnemp.empirical_final&i;        set disease_healthy&i;
        by nrun;
        retain yi_mx x_mx y_mx sens_mx spec_mx;
        set disease_healthy&i;
        N_d=&N1;     N_h=&N0;     N=&TSS;
        bnsum_sn=0; bnsum_fp=0;
```

```
            thresh_x=test1; thresh_y=test2;
            do q=1 to N;
                    set disease_healthy&i point=q;
                    if test2>thresh_y  and test1>thresh_x then do;
                    if h_d_flag=1 then bnsum_sn=bnsum_sn+1;
                            if h_d_flag=0 then bnsum_fp =bnsum_fp +1;
                    end;
            end;
            sens_bn=bnsum_sn/N_d;
            fp_bn =bnsum_fp/N_h;
            spec_bn=1-fp_bn;
            youden=sens_bn+spec_bn-1;

            if first.nrun=1 then do;
                    yi_mx=youden;        x_mx=thresh_x;        y_mx=thresh_y;
            sens_mx=sens_bn;    spec_mx=spec_bn;
            end;
            if youden>yi_mx then do;
                    yi_mx=youden;        x_mx=thresh_x;        y_mx=thresh_y;
            sens_mx=sens_bn;    spec_mx=spec_bn;
            end;
            if youden=yi_mx and sens_bn>sens_mx then do;
                    x_mx=thresh_x;        y_mx=thresh_y;        sens_mx=sens_bn;
            spec_mx=spec_bn;
            end;
                    output all_empirical&i;
            if last.nrun=1 then output bnemp.empirical_final&i;
    run;

    data empirical_final_red&i;
            if _n_ = 1 then set Parms_TrueValues&i;
            set bnemp.empirical_final&i (keep = nrun x_mx y_mx yi_mx sens_mx spec_mx);
    run;

    data bias_emp&i;      set empirical_final_red&i;
            biasEmpTheta3 = x_mx - truetheta3;
            biasEmpTheta4 = y_mx - truetheta4;
            biasEmpSens = sens_mx - trueSens;
            biasEmpSpec = (spec_mx) - trueSpec;

            if biasEmpSens > 0 then flagOverEmpSens=1;      if biasEmpSens < 0 then
                    flagOverEmpSens=0; if biasEmpSens = . then flagOverEmpSens = .;
            if biasEmpSpec > 0 then flagOverEmpSpec=1;      if biasEmpSpec < 0 then
                    flagOverEmpSpec=0; if biasEmpSpec = . then flagOverEmpSpec = .;

    run;
    /* Calculates the mean Bias and RMSE for each empirical theta1 and theta2 value from
    the true values */
    proc means data = bias_emp&i n mean median std q1 q3 noprint;
```

```
            var biasEmpTheta3 biasEmpTheta4 biasEmpSens biasEmpSpec;
            output out=bias_empirical&i (drop=_type_ _freq_)
            mean=mean_biasempT3 mean_biasempT4 mean_biasempSe
            mean_biasempSp
            median=med_biasempT3 med_biasempT4 med_biasempSe med_biasempSp;
run;
proc freq data=bias_emp&i noprint;
            table flagOverEmpSens / list norow nocol nofreq nocum out=Overest_EmpSe&i;
            table flagOverEmpSpec / list norow nocol nofreq nocum out=Overest_EmpSp&i;
run;
data Overest_EmpSe&i; set Overest_EmpSe&i (keep=flagOverEmpSens percent);
            where flagOverEmpSens=1;  drop flagOverEmpSens;
            rename percent=Perc_Overest_Emp_Se;
            label percent = 'Perc_Overest_Emp_Se';
run;
data Overest_EmpSp&i;        set Overest_EmpSp&i (keep=flagOverEmpSpec percent);
            where flagOverEmpSpec=1;  drop flagOverEmpSpec;
            rename percent=Perc_Overest_Emp_Sp;
            label percent = 'Perc_Overest_Emp_Sp';
run;

data allemp&i;
            set bias_emp&i (keep=yi_mx x_mx y_mx sens_mx spec_mx biasEmpTheta3
                        biasEmpTheta4 biasEmpSens biasEmpSpec flagOverEmpSens
                        flagOverEmpSpec);
            parameter_set_number=&i;
run;

data allemp_results&i;
            merge bias_empirical&i Overest_EmpSe&i Overest_EmpSp&i;
            parameter_set_number=&i;
run;

proc append data=allemp&i base=bnemp.BNallempicaldata;          run;
proc append data=allemp_results&i base=bnemp.BNallempiricalresults;    run;

%end;

%mend EMP;

%EMP;
```

8.4.4 Assessing Normality

libname normalbp 'C:\Documents and Settings\wilkar\Desktop\Dissertation!\FINAL FROM CLUSTER - KOTI and MERLOT\BP';
libname normalbn 'C:\Documents and Settings\wilkar\Desktop\Dissertation!\FINAL FROM CLUSTER - KOTI and MERLOT\BN';
libname bp 'C:\Documents and Settings\wilkar\Desktop\Dissertation!\FINAL FROM CLUSTER - KOTI and MERLOT\BP\New_Individual';
libname bn 'C:\Documents and Settings\wilkar\Desktop\Dissertation!\FINAL FROM CLUSTER - KOTI and MERLOT\BN\Individual';

```
/* Used a random number generator (http://www.random.org/) to select the seeds.
Bounds were 1 to 1,000,000 */
/* Select 14 datasets for each sequential testing strategy (~10%) to look at the normality
of the OOP components */
/* BELIEVE THE POSITIVE */
proc surveyselect data = normalbp.sasparms
        method = SRS rep = 1 sampsize = 14 seed = 319595 out = normalbp;
run;
proc print data = normalbp noobs;   run;
data bp3;
        set bp.final_parm_set3;
        parmset=3;
run;
data bp15;
        set bp.final_parm_set15;
        parmset=15;
run;
data bp17;
        set bp.final_parm_set17;
        parmset=17;
run;
data bp27;
        set bp.final_parm_set27;
        parmset=27;
run;
data bp48;
        set bp.final_parm_set48;
        parmset=48;
run;
data bp57;
        set bp.final_parm_set57;
        parmset=57;
run;
data bp59;
        set bp.final_parm_set59;
        parmset=59;
run;
```

```
data bp67;
        set bp.final_parm_set67;
        parmset=67;
run;
data bp69;
        set bp.final_parm_set69;
        parmset=69;
run;
data bp78;
        set bp.final_parm_set78;
        parmset=78;
run;
data bp80;
        set bp.final_parm_set80;
        parmset=80;
run;
data bp113;
        set bp.final_parm_set113;
        parmset=113;
run;
data bp124;
        set bp.final_parm_set124;
        parmset=124;
run;
data bp129;
        set bp.final_parm_set129;
        parmset=129;
run;


data allbp;
        set bp3 bp15 bp17 bp27 bp48 bp57 bp59 bp67 bp69 bp78 bp80 bp113 bp124
        bp129;
run;


proc greplay igout=work.gseg nofs;
        delete _all_;
run;   quit;

filename bp 'C:\Documents and Settings\wilkar\Desktop\Dissertation!\FINAL FROM
CLUSTER - KOTI and MERLOT\Normality';
goptions gsfname=bp dev=gif373;
proc univariate data=allbp plot normal;
        var finalsimtheta1 finalsimtheta2;
        where good_sim_thetas=1;
        histogram / normal;
        qqplot;
        by parmset;
run;
```

```
/* BELIEVE THE NEGATIVE */
proc surveyselect data = normalbn.sasparms3
     method = SRS rep = 1 sampsize = 14 seed = 834213 out = normalbn;
run;
proc print data = normalbn noobs;   run;

data bn21;
     set bn.final_parm_set21;
     parmset=21;
run;
data bn24;
     set bn.final_parm_set24;
     parmset=24;
run;
data bn29;
     set bn.final_parm_set29;
     parmset=29;
run;
data bn36;
     set bn.final_parm_set36;
     parmset=36;
run;
data bn40;
     set bn.final_parm_set40;
     parmset=40;
run;
data bn82;
     set bn.final_parm_set82;
     parmset=82;
run;
data bn85;
     set bn.final_parm_set85;
     parmset=85;
run;
data bn86;
     set bn.final_parm_set86;
     parmset=86;
run;
data bn90;
     set bn.final_parm_set90;
     parmset=90;
run;
data bn96;
     set bn.final_parm_set96;
     parmset=96;
run;
```

```
data bn97;
        set bn.final_parm_set97;
        parmset=97;
run;
data bn105;
        set bn.final_parm_set105;
        parmset=105;
run;
data bn115;
        set bn.final_parm_set115;
        parmset=115;
run;
data bn139;
        set bn.final_parm_set139;
        parmset=139;
run;

data allbn;
        set bn21 bn24 bn29 bn36 bn40 bn82 bn85 bn86 bn90 bn96 bn97 bn105 bn115
        bn139;
run;

proc greplay igout=work.gseg nofs;
        delete _all_;
run;    quit;

filename bn 'C:\Documents and Settings\wilkar\Desktop\Dissertation!\FINAL FROM
CLUSTER - KOTI and MERLOT\Normality\bn';
goptions gsfname=bn dev=gif373;
proc univariate data=allbn plot normal;
        var finalsimtheta3 finalsimtheta4;
        where good_sim_thetas=1;
        histogram / normal;
        qqplot;
        by parmset;
run;
```

8.4.5 Prediction Models for SCI Widths and TSS

```
libname all 'C:\Documents and Settings\wilkar\Desktop\Dissertation!\FINAL FROM
CLUSTER - KOTI and MERLOT\All';
libname bpall 'C:\Documents and Settings\wilkar\Desktop\Dissertation!\FINAL FROM
CLUSTER - KOTI and MERLOT\BP\New_Individual';
libname bnall 'C:\Documents and Settings\wilkar\Desktop\Dissertation!\FINAL FROM
CLUSTER - KOTI and MERLOT\BN\Individual';
```

```sas
data all.all_bp_estimation all.all_bp_test;
        set bpall.final_parm_set1 - bpall.final_parm_set144 (keep=nrun N_d N_h TSS k
                ph pd AUC1 AUC2 b1 b2 sim_ph sim_pd sim_m1h sim_m2h sim_m1d
                sim_m2d sim_s1h sim_s1d sim_s2h sim_s2d TrueTheta1 TrueTheta2
                finalsimtheta1 finalsimtheta2 good_sim_thetas prob_sim1 prob_sim2
                VarT1 VarT2 CovT1T2 SCI_wTest1 SCI_wTest2);
        sim_a1=(sim_m1d-sim_m1h) / sim_s1d;        sim_b1=sim_s1h/sim_s1d;
        sim_a2=(sim_m2d-sim_m2h) / sim_s2d;        sim_b2=sim_s2h/sim_s2d;
        AUCTest1=probnorm( sim_a1 / sqrt(1+sim_b1**2) );
        AUCTest2=probnorm( sim_a2 / sqrt(1+sim_b2**2) );

        if nrun le 500 then output all.all_bp_estimation;
                else output all.all_bp_test;
run;
data all.all_bn_estimation all.all_bn_test;
        set bnall.final_parm_set1 - bnall.final_parm_set144 (keep=nrun N_d N_h
        TSS k  ph pd AUC1 AUC2 b1 b2 sim_ph sim_pd sim_m1h sim_m2h
        sim_m1d sim_m2d sim_s1h sim_s1d sim_s2h sim_s2d TrueTheta3
        TrueTheta4 finalsimtheta3 finalsimtheta4 good_sim_thetas prob_sim3 prob_sim4
        VarT3 VarT4 CovT3T4 SCI_wTest1 SCI_wTest2);
        sim_a1=(sim_m1d-sim_m1h) / sim_s1d;        sim_b1=sim_s1h/sim_s1d;
        sim_a2=(sim_m2d-sim_m2h) / sim_s2d;        sim_b2=sim_s2h/sim_s2d;
        AUCTest1=probnorm( sim_a1 / sqrt(1+sim_b1**2) );
        AUCTest2=probnorm( sim_a2 / sqrt(1+sim_b2**2) );

        if nrun le 500 then output all.all_bn_estimation;
                else output all.all_bn_test;
run;

/*proc freq data=all.all_bp_estimation;        table good_sim_thetas / list missing; run;*/
/*proc freq data=all.all_bn_estimation;        table good_sim_thetas / list missing; run;*/
/*proc freq data=all.all_bp_test;        table good_sim_thetas / list missing; run;*/
/*proc freq data=all.all_bn_test;        table good_sim_thetas / list missing; run;*/

proc reg data=all.all_bn_estimation outest=estimates;
        *model SCI_wTest1 = TSS sim_ph sim_pd sim_m1h sim_m2h sim_m1d
                sim_m2d sim_s1h sim_s1d sim_s2h sim_s2d k / AIC BIC; /* R2=0.65 */
        *model SCI_wTest1 = TSS AUCTest1 AUCTest2 sim_ph sim_pd sim_b1 sim_b2
                k / AIC BIC; /* R2=0.58 */
        *model SCI_wTest1 = TSS sim_a1 sim_a2 sim_b1 sim_b2 sim_ph sim_pd k /
                AIC BIC; /* R2=0.61 */
        model SCI_wTest2 = TSS sim_ph sim_pd sim_m1h sim_m2h sim_m1d sim_m2d
                sim_s1h sim_s1d sim_s2h sim_s2d k / AIC BIC; /* R2=0.61 */
        *model SCI_wTest2 = TSS AUCTest1 AUCTest2 sim_ph sim_pd sim_b1 sim_b2
                k / AIC BIC; /* R2=0.58 */
        *model SCI_wTest2 = TSS sim_a1 sim_a2 sim_b1 sim_b2 sim_ph sim_pd k /
                AIC BIC; /* R2=0.62 */
run;
quit;
```

```
proc reg data=all.all_bn_estimation outest=estimates;
        model TSS = SCI_wTest1 sim_ph sim_pd sim_m1h sim_m2h sim_m1d sim_m2d
                sim_s1h sim_s1d sim_s2h sim_s2d k / AIC BIC;      /* R2=0.42 */
        *model TSS = SCI_wTest2 sim_ph sim_pd sim_m1h sim_m2h sim_m1d
                sim_m2d sim_s1h sim_s1d sim_s2h sim_s2d k / AIC BIC; /* R2=0.39 */
        *model TSS = SCI_wTest1 AUCTest1 AUCTest2 sim_ph sim_pd sim_b1 sim_b2
                k / AIC BIC;    /* R2=0.36 */
        *model TSS = SCI_wTest2 AUCTest1 AUCTest2 sim_ph sim_pd sim_b1 sim_b2
                k / AIC BIC;    /* R2=0.36 */
        *model TSS = SCI_wTest1 sim_a1 sim_a2 sim_b1 sim_b2 sim_ph sim_pd k /
                AIC BIC;        /* R2=0.38 */
        *model TSS = SCI_wTest2 sim_a1 sim_a2 sim_b1 sim_b2 sim_ph sim_pd k /
                AIC BIC;        /* R2=0.39 */
run;
quit;


/* INTERACTIONS */
ods output parameterestimates=estimates;
proc glm data=all.all_bp_estimation;
        *model SCI_wTest1 = TSS sim_ph sim_pd sim_m1h sim_m2h sim_m1d
                sim_m2d sim_s1h sim_s1d sim_s2h sim_s2d k TSS*k;
        *model SCI_wTest1 = TSS AUCTest1 AUCTest2 sim_ph sim_pd sim_b1 sim_b2
                    k /*TSS*k*/ AUCTest1*sim_b1 AUCTest2*sim_b2;
        *model SCI_wTest1 = TSS sim_a1 sim_a2 sim_b1 sim_b2 sim_ph sim_pd k
                TSS*k;
        *model SCI_wTest2 = TSS sim_ph sim_pd sim_m1h sim_m2h sim_m1d
                sim_m2d sim_s1h sim_s1d sim_s2h sim_s2d k TSS*k;
        *model SCI_wTest2 = TSS sim_a1 sim_a2 sim_b1 sim_b2 sim_ph sim_pd k
                TSS*k;
        *model SCI_wTest2 = TSS AUCTest1 AUCTest2 sim_ph sim_pd sim_b1 sim_b2
                k /*TSS*k*/ AUCTest1*sim_b1 AUCTest2*sim_b2;
        *model TSS = SCI_wTest1 AUCTest1 AUCTest2 sim_ph sim_pd sim_b1 sim_b2
                k AUCTest1*sim_b1 AUCTest2*sim_b2;
        model TSS = SCI_wTest2 AUCTest1 AUCTest2 sim_ph sim_pd sim_b1 sim_b2
                k AUCTest1*sim_b1 AUCTest2*sim_b2;
run; quit;


%macro R2_widths(data=, var=);
data &data;
        set &data (drop=_model_ _type_ _depvar_ _rmse_ &var);
run;
/* Check Robustreg with calculations */
data predval;
        if _n_ = 1 then set &data (rename=(SCI_wTest1=SCI_wTest1e
        sim_m1h=sim_m1he sim_m1d=sim_m1de sim_ph=sim_phe sim_pd=sim_pde
        sim_m2h=sim_m2he sim_m2d=sim_m2de k=ke sim_s1h=sim_s1he
        sim_s2h=sim_s2he sim_s1d=sim_s1de sim_s2d=sim_s2de));
        set all.all_bn_test;
```

```
          pred_&var=
          Intercept+SCI_wTest1e*SCI_wTest1+sim_phe*sim_ph+sim_pde*sim_pd+
          sim_m 1he*sim_m2h+sim_m2he*sim_m2h+sim_m1de*sim_m1d+
          sim_m2de*sim_m2d+sim_s1he*sim_s1h+sim_s1de*sim_s1d+sim_s2he*sim_s2h
          +sim_s2de*sim_s2d+ke*k;

/*        pred_&var=
          Intercept+SCI_wTest2e*SCI_wTest2+sim_phe*sim_ph+sim_pde*sim_pd+
          sim_b1e*sim_b1+sim_b2e*sim_b2+AUCTest1e*AUCTest1+*/
/*        AUCTest2e*AUCTest2+ke*k;*/

/*        pred_&var=
          Intercept+SCI_wTest2e*SCI_wTest2+sim_phe*sim_ph+sim_pde*sim_pd+
          sim_b1e*sim_b1+sim_b2e*sim_b2+sim_a1e*sim_a1+sim_a2e*sim_a2+ke*k;*/
run;

proc summary data=all.all_bn_test;
          var &var;
          output out=mean&var mean=avg;
run;
proc sort data=all.all_bn_test;          by nrun;          run;
proc sort data=predval;          by nrun;          run;
data sums_est;
          if _n_ = 1 then set mean&var(keep=avg);
          merge all.all_bn_test predval (keep=nrun pred_&var);
          by nrun;
          tot=(&var - avg)**2;
          err=(&var - pred_&var)**2;
run;
proc summary data=sums_est;
          var tot err;
          output out=sums sum=SStot SSerr;
run;
data r;
          set sums;
          R2=1-(SSerr/SStot);
run;

proc print data=r;
          var R2;
run;
%mend R2_widths;

%R2_widths(data=estimates, var=TSS);
```

VITA

Amber Rose Wilk was born October 19, 1985 in Fayetteville, North Carolina to Edward and Annalisa Wilk. She graduated from the Copper Hills High School in West Jordan, Utah in 2004. She received her Bachelor's of Science in Mathematics and Statistics Composite Secondary Education from Utah State University in December of 2007. While at USU she was highly involved in the implementation of the Utah Mathematics Endorsement Project throughout the state of Utah as well as peer advising in the mathematics and statistics department. She graduated graduating Magna Cum Laude and was the Valedictorian for the College of Science. In August of 2008, she moved to Richmond, Virginia to begin her graduate studies towards a Ph.D. in Biostatistics at Virginia Commonwealth University. She spent her graduate career employed for two years as a statistical consultant, completed a one year internship with Merck & Co., Inc., and worked for a year with the Department of Healthcare Policy and Research. She has received the Charles C. Clayton, Biopharmaceutical Applied Statistics Scholarship, and Forbes Day Presentation award as well as several Biostatistics student presentation awards while at VCU. In addition, she has been a teaching assistant for a variety of graduate level Biostatistics courses.