

ABSTRACT

Interval Estimation for TPRs and FPRs of Two Diagnostic Tests with Unverified Negatives

Eileen M. Stock, Ph.D.

Chairpersons: Dean M. Young, Ph.D., and James D. Stamey, Ph.D.

In the clinical setting, the performance of a diagnostic or screening test is often summarized using the test's true positive rate (TPR) and false positive rate (FPR). However, estimation of the TPR and FPR for a diagnostic or screening test requires individuals to be identified as diseased or non-diseased using a gold standard, which is seldomly available. In cases for when the gold standard test is available, it may be too costly or time-consuming to implement the gold standard procedure. Furthermore, this verification procedure is frequently limited to individuals with one or more positive test results, thus yielding unverified negatives and possibly verification bias.

We present two interval estimators for the estimation of the differences in the TPRs and FPRs of two dichotomous diagnostic tests applied to members of a population stratified into two distinct groups. The two groups have varying disease prevalences with unverified negatives. We obtain maximum likelihood estimates using the EM algorithm to devise a Wald interval for the differences in the TPRs and FPRs and compare its performance to a Bayesian credible interval for a spectrum of different TPRs, FPRs, and sample sizes.

We further present a hierarchical Bayesian logit model when incorporating different locations into the model. In particular, we compare the accuracy of two low-cost

procedures for screening of high-grade cervical intra-epithelial neoplasia (CIN2+) using a cross-sectional multi-center study. We model the prevalence of CIN2+ for two different age groups assumed to have varying disease prevalences with unverified negatives. Furthermore, we include random effects to account for heterogeneity among the locations.

In the final chapter, we perform a power analysis for multivariate normality in the presence of a monotone missing data pattern. We assess the efficiency of several imputation procedures and multivariate normality tests through a simulation study where we compare the resulting powers using an alternative distribution. In particular, we consider different sample sizes and proportions of missingness for various parameter values of a multivariate skewed t -distribution. We calculate power using both the median and mean test statistics for each imputation-test combination.

Interval Estimation for TPRs and FPRs of Two Diagnostic Tests
with Unverified Negatives

by

Eileen M. Stock, B.S., M.S., M.S.

A Dissertation

Approved by the Department of Statistical Science

Jack D. Tubbs, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

Approved by the Dissertation Committee

Dean M. Young, Ph.D., Co-Chairperson

James D. Stamey, Ph.D., Co-Chairperson

Jack D. Tubbs, Ph.D.

John W. Seaman, Jr., Ph.D.

Michael E. Sherr, Ph.D.

Accepted by the Graduate School
May 2011

J. Larry Lyon, Ph.D., Dean

TABLE OF CONTENTS

LIST OF FIGURES	vii
LIST OF TABLES	viii
ACKNOWLEDGMENTS	x
DEDICATION	xi
1 Introduction	1
2 Interval Estimation for the Difference in TPRs and FPRs of Two Diagnostic Tests with Unverified Negatives	3
2.1 Introduction	3
2.2 Sampling Plan	5
2.3 Examples	9
2.3.1 Example 1	9
2.3.2 Example 2	11
2.4 Interval Estimation	12
2.4.1 A Wald Confidence Interval	12
2.4.2 A Bayesian Model	13
2.5 Simulation Studies	15
2.5.1 A Study of Four Scenarios	15
2.5.2 Adding Informative Priors	17
2.6 Discussion	20
3 A Hierarchical Approach to Modeling Multicenter Prevalences, TPRs, and FPRs for Two Populations	21

3.1	Introduction	21
3.2	Study Background	24
3.3	Model Description	26
3.4	Estimation Procedure	30
3.4.1	Individual Locations	31
3.4.2	Combined Locations	34
3.4.3	Test Accuracy for CIN2+	36
3.5	A Simulation Example	37
3.6	Conclusions	39
4	A Power Analysis for Multivariate Normality Tests and Multiple Imputation Procedures in the Presence of a Monotone Missing Data Pattern	40
4.1	Introduction	40
4.2	Multiple Imputation Schemes	42
4.2.1	EM Algorithm	43
4.2.2	Data Augmentation	44
4.2.3	Iterative Regression Imputation	46
4.2.4	Predictive Mean Matching	47
4.3	Test Statistics	47
4.3.1	The $T_{N,\beta}$ Test Statistic	48
4.3.2	The Omnibus Mardia Test Statistic	49
4.3.3	The ϵ -statistic	49
4.3.4	The Shapiro-Wilk Test Statistic	50
4.4	Simulation Design and Results	51
4.5	Conclusions	53
A	Derivations	55
A.1	EM Algorithm	55

A.2	Full Conditional Posterior Distributions	56
B	Hierarchical Model Results	58
C	Interval Estimation Code	70
C.1	EM Algorithm and Wald Interval	70
C.2	WinBUGS Code	89
C.3	R Code to Call WinBUGS	90
D	Power Analysis Plots	93
E	Power Analysis Tables	102
	BIBLIOGRAPHY	110

LIST OF FIGURES

2.1	Coverages for τ_1 and τ_2	16
2.2	Interval widths for τ_1 and τ_2	17
2.3	Coverages and interval widths for τ_1 and τ_2 (high TPR, mixed FPR scenario).	19
3.1	Density plots for Conakry's TPRs and FPRs (single site).	32
3.2	Density plots for Conakry's dependency factors and prevalences (single site).	32
3.3	Density plots for Mumbai's TPRs and FPRs (single site).	33
3.4	Density plots for Mumbai's dependency factors and prevalences (single site).	34
3.5	Density plots for Mumbai's TPRs and FPRs.	35
3.6	Density plots for Mumbai's dependency factors and prevalences.	35
3.7	Forest plot for the TPRs for VILI.	36
3.8	Forest plot for the FPRs for VILI.	37
D.1	MVSN model for $p = 2$	94
D.2	MVSN model for $p = 5$	95
D.3	MVT($df = 25$) model for $p = 2$	96
D.4	MVT($df = 25$) model for $p = 5$	97
D.5	MVT($df = 30$) model for $p = 2$	98
D.6	MVT($df = 30$) model for $p = 5$	99
D.7	MVST model for $p = 2$	100
D.8	MVST model for $p = 5$	101

LIST OF TABLES

2.1	Sampling Plan for Group 1.	6
2.2	Sampling Plan for Group 2.	6
2.3	Sampling results for 40 – 59 age group of the Hemoccult data.	10
2.4	Sampling results for 60 – 70 age group of the Hemoccult data.	10
2.5	Sampling results for < 75 age group of the Dementia data.	11
2.6	Sampling results for ≥ 75 age group of the Dementia data.	11
3.1	Sampling Plan for Group 1 at the m th center.	26
3.2	Sampling Plan for Group 2 at the m th center.	27
3.3	Resulting TPRs, FPRs, prevalences, and dependency factors.	31
B.1	Initialized chains for γ	58
B.2	Full data, individual, and pooled TPR estimates.	59
B.3	Full data, individual, and pooled FPR estimates.	60
B.4	Full, individual, and pooled prevalences and dependency factor estimates. .	61
B.5	Resulting prevalences from pooled centers.	62
B.6	Resulting TPRs and FPRs from pooled centers.	63
B.7	Resulting dependency factors from pooled centers.	64
B.8	Remaining parameter estimates resulting from pooled centers.	65
B.9	TPRs from the simulation study.	66
B.10	FPRs from the simulation study.	67
B.11	Prevalences and dependency factors from the simulation study.	68
B.12	Offset terms and other parameter results from the simulation study.	69
E.1	MVSN model for $p = 2$	102
E.2	MVSN model for $p = 5$	103

E.3	MVT model with $df = 25$ for $p = 2$	104
E.4	MVT model with $df = 25$ for $p = 5$	105
E.5	MVT model with $df = 30$ for $p = 2$	106
E.6	MVT model with $df = 30$ for $p = 5$	107
E.7	MVST model with $df = 50$, $\nu = \mathbf{100}$ for $p = 2$	108
E.8	MVST model with $df = 50$, $\nu = \mathbf{100}$ for $p = 5$	109

ACKNOWLEDGMENTS

I would first like to thank Dr. Bratcher, who opened a doorway for me into the incredible realm of statistics by providing me with a remarkable opportunity to pursue an education and career in statistics. Thank you for the inspiration and advice you have provided upon my entrance into the program that kept me striving to achieve my goals.

I would also like to thank my mentors, Dr. Young and Dr. Stamey, whose encouraging words provided me with the persistent motivation that kept me on track in my research. Thank you for your guidance and direction. Thank you for introducing several interesting research topics and providing me with the opportunity to further pursue these endeavors with you.

Most importantly, I would like to thank my parents, Paul and Marlene. I would not have accomplished everything I have thus far without the love, inspiration, and support you have provided me throughout my life and academic career. You have inspired me to set goals for my future and do my best to acquire these goals. You have always taught me the importance of a higher education, and I would not have been able to attain such a goal without your continual support. I am forever grateful for your assistance and contributions in shaping who I am as a person and for providing me with a strong foundation of values that I am able to incorporate into my personal life and work experiences.

DEDICATION

To my Parents,
Who have provided me with
Love, Guidance, and Strength
Throughout my life and academic career.
And to Josephine and Shylene,
Who will remain in my heart forever.

CHAPTER ONE

Introduction

In the clinical setting, we are often interested in comparing the performance of a new, innovative, and perhaps lower-costing diagnostic test to a conventional test that may have been in use for many years. In order to examine differences in the performances of the different screening procedures, we frequently compare the TPR and FPR of each test. However, estimation of the TPR and FPR for a diagnostic or screening test requires individuals to be verified as diseased or non-diseased using a gold standard. However, such a test may not always be readily accessible, convenient or safe to use, or even in existence. We often encounter the situation in which neither of the two tests can be considered a gold standard. When applying the gold standard, verification is often limited to individuals with one or more positive test results, thus yielding unverified negatives. When neither test is a gold standard, these unverified negatives may actually have a positive disease status. Thus, misclassification of disease status in the clinical setting may lead to incorrect inferences and misrepresent the effectiveness of each of the screening tests.

Walter (1999) has estimated the TPR and FPR in breast cancer screening for two or more tests applied to a single population in the presence of misclassification. Berry, Smith, Macaskill, and Irwig (2002) extended these results for two screening tests applied to two different samples with misclassification present. In Chapter 2, we consider the model of Berry et al. (2002) and supplement currently existing literature by considering different interval estimation procedures for the differences in the TPRs and FPRs of the two tests applied to two different groups. We introduce the sampling plan in Section 2 and discuss a Wald confidence interval and Bayesian credible interval in Section 4. We further provide a simulation study considering four realistic clinical setting scenarios involving different combinations of high and low TPRs and FPRs.

In Chapter 3, we consider a hierarchical approach to modeling multi-center prevalences, TPRs, and FPRs for the two group. We use a hierarchical Bayesian logit model to model the TPRs and FPRs of two dichotomous screening tests applied to members of two distinct groups, which is described in full detail in Section 3. In Section 4, we compare results of the hierarchical model to the non-hierarchical approach. We also further support the accuracy of our model with a simulation study, described in Section 5.

In Chapter 4, we perform a power analysis involving several multivariate normality tests and multiple imputation procedures in the presence of monotone missing data. In many research settings, we commonly encounter multivariate response data with missing values for one or more variables. Often, all variable values for an observation missing a single variable value are excluded, resulting in a substantial loss of information in the analysis. However, the introduction of imputation procedures has enabled many breakthroughs in missing data problems. In fact, we will examine several multivariate imputation procedures for cases involving multivariate normal monotone missing data and multivariate non-normal monotone missing data. The multiple imputation procedures we consider involve the EM algorithm and data augmentation using Jeffreys priors, an iterative regression approach, and a predictive mean matching approach. Each of these imputation schemes are described in further detail in Section 2.

An additional hindrance in the statistical analysis of multivariate data is that many statistical procedures implemented in the clinical setting require a multivariate normality assumption, which is seldomly tested. Thus, after imputation of the missing data, we will apply several tests for multivariate normality and compare the resulting powers. Multivariate tests we consider include Henze and Zirkler's test, an omnibus test using Mardia's tests for multivariate skewness and kurtosis, a fairly new test based on the concept of Newton's potential energy, and a generalization of the Shapiro-Wilk test for normality. These are further discussed in Section 3, and the design and results of the study are contained in Section 4.

CHAPTER TWO

Interval Estimation for the Difference in TPRs and FPRs of Two Diagnostic Tests with Unverified Negatives

2.1 Introduction

The performance of a diagnostic or screening test is often summarized using the test's sensitivity and specificity. Sensitivity, also referred to as the *True Positive Rate* (TPR), is the probability of a positive test result, given that an individual has the disease, and specificity is the probability of a negative test result, given that an individual does not have the disease (Putter, Fiocco, and Stijnen, 2010). Also commonly used in clinical settings, the term *False Positive Rate* (FPR, $1 - \text{specificity}$) is the probability of a positive test result given that an individual does not have the disease (de Leon, Soo, Bonzo, and Rudnisky, 2009). Estimation of the TPR and FPR for a diagnostic or screening test requires individuals to be identified as diseased or non-diseased using a gold standard, which is a test procedure with a 100% TPR and a 0% FPR with respect to the detection of the disease. However, in the investigation of new diagnostic tests, one commonly finds that neither of the two tests investigating the same condition or disease can be considered a gold standard. Furthermore, this verification procedure is often limited to individuals with one or more positive test results, thus yielding unverified negatives. Consequently, any parameter inference dependent on the TPR or FPR of a test will be obsolete and impractical. Thus, we assume that the probability of selecting an individual for a clinical assessment depends on only their diagnostic test results. If an individual with test results does not have a clinical assessment, then we can treat their true disease status as missing and satisfying the missing at random (MAR) assumption (Zhou, 1998).

An unverified disease status can result from several situations common in diagnostic testing. The probability of disease could be extremely small after a negative test

result so that confirmation of the disease is unwarranted, or the verification procedure may be costly to perform, unethical to administer, or risky and invasive for the patient. However, the exclusion of these unverified negatives may result in biased estimates of the TPR and FPR. This problem is known as verification bias. Recent literature has shown that one can use data collected from individuals with a confirmed disease status to estimate the proportion of diseased and non-diseased with an unconfirmed disease status to obtain estimates for TPRs and FPRs. Walter (1999) has estimated the TPR and FPR in breast cancer screening for two or more tests applied to a single population. The individuals had a verified disease status if they had one or more positive test results and an unverified disease status if they had all negative test results. He accomplished his analysis using previous methodology with the additional assumptions that the total number of individuals tested is known and that the test errors are independent.

In this paper, we consider the model proposed by Berry, Smith, Macaskill, and Irwig (2002) for two dichotomous screening tests applied to two different samples known to have varying disease prevalences in the presence of unverified negatives. We similarly relax the assumption of independence by allowing correlation between the two diagnostic tests. For example, Berry et al. (2002) have applied these methods to the detection of colorectal cancers by comparing rehydrated guaiac testing and immunochemical testing using reversed passive haemagglutination for individuals age 40-59 years old and individuals age 60-70 years old. They have considered four models: independence in both the diseased and non-diseased groups, independence in the non-diseased group only, independence in the diseased group only, and an unrestricted model in which there is no independence in either the diseased or non-diseased group. We extend these results by performing a comparison of the frequentist and Bayesian approaches to interval estimation for the differences in the TPRs and FPRs of the two tests.

Although methodology currently exists for comparing diagnostic tests in terms of their TPRs and FPRs, few papers provide comparative inference of the TPRs and

FPRs from assessing the behavior of these differences in the presence of unverified negatives. While estimates for TPR and FPR have been obtained using an EM algorithm in Berry et al. (2002), standard errors of these estimators were not given. In this paper, we contribute to the existing literature and improve existing estimation methods presented in Berry et al. (2002) by obtaining standard errors for these estimators and determining two interval estimators and further evaluating the interval estimators.

We also derive Bayesian inference for TPR and FPR differences. Hanson, Johnson, and Gastwirth (2006) have examined prevalence and diagnostic test accuracy based on dual-pooled screening for a single population involving unverified negatives using a Bayesian perspective. Joseph, Gyorkos, and Coupal (1995) have obtained simultaneous inferences about the population prevalence and the TPRs and FPRs of two diagnostic tests for a single population using a Bayesian approach. They also have extended these results by considering two populations and have applied these techniques to the estimation of the prevalence of *Strongyloides* infection in Cambodian refugees and to the investigation of the diagnostic test properties of stool examinations and serologic testing.

In the remaining sections of this article, we present the generalized sampling plan in Section 2 for modeling two screening tests applied to two populations with unverified negatives. In Section 3, we introduce some actual, real-data scenarios considered in the previous literature. In Section 4, we present the Bayesian and frequentist methodologies for obtaining point and interval estimates when we encounter a sampling scheme as described in Section 2. We compare the proposed methodologies using simulation studies in Section 5, and conclude the paper with a discussion comparing the techniques and different scenarios in Section 6.

2.2 Sampling Plan

Consider a single population stratified into two groups by some characteristic. Let x_{ijk} = the count of individuals for disease status i , Test 1 result j , and Test 2

result k for group Π_1 , where

$$\begin{aligned} i &= 0 \text{ (non-diseased), } 1 \text{ (diseased),} \\ j &= 0 \text{ (negative), } 1 \text{ (positive),} \\ k &= 0 \text{ (negative), } 1 \text{ (positive).} \end{aligned}$$

Similarly, let y_{ijk} = count for disease status i , Test 1 result j , and Test 2 result k for group Π_2 . Tables 2.1 and 2.2 show the sampling plan for Π_1 and Π_2 , respectively.

Table 2.1. Sampling Plan for Group 1.

	Diseased			Non-diseased		
	Test1+	Test1–	Total	Test1+	Test1–	Total
Test2+	x_{111}	x_{101}	$x_{111} + x_{101}$	x_{011}	x_{001}	$x_{011} + x_{001}$
Test2–	x_{110}	$[x_{100}]$	$[x_{110} + x_{100}]$	x_{010}	$[x_{000}]$	$[x_{010} + x_{000}]$
Total	$x_{111} + x_{110}$	$[x_{101} + x_{100}]$	$[n_1]$	$x_{011} + x_{010}$	$[x_{001} + x_{000}]$	$[N_1]$

Table 2.2. Sampling Plan for Group 2.

	Diseased			Non-diseased		
	Test1+	Test1–	Total	Test1+	Test1–	Total
Test2+	y_{111}	y_{101}	$y_{111} + y_{101}$	y_{011}	y_{001}	$y_{011} + y_{001}$
Test2–	y_{110}	$[y_{100}]$	$[y_{110} + y_{100}]$	y_{010}	$[y_{000}]$	$[y_{010} + y_{000}]$
Total	$y_{111} + y_{110}$	$[y_{101} + y_{100}]$	$[n_2]$	$y_{011} + y_{010}$	$[y_{001} + y_{000}]$	$[N_2]$

For Π_1 , variables x_{111} , x_{101} , x_{110} , x_{011} , x_{001} , and x_{010} have a verified disease status. Hence, these values are observed and known. Variables in brackets, such as x_{100} and x_{000} , are unknown (latent data) values. The variables x_{100} and x_{000} refer to the number of individuals sampled from Π_1 that are diseased and non-diseased, respectively, but

tested negative for both diagnostic tests. As a result, the total number of diseased and non-diseased individuals for Π_1 , n_1 and N_1 respectively, are unknown. However, the total number of double negatives for Π_1 , $M_1 = x_{100} + x_{000}$, and the total size of Π_1 , $T_1 = n_1 + N_1$, are both known. Similarly, for Π_2 , variables $y_{111}, y_{101}, y_{110}, y_{011}, y_{001}$, and y_{010} are observed and known, while variables y_{100} and y_{000} are unknown with n_2 and N_2 also unknown, but $M_2 = y_{100} + y_{000}$ and $T_2 = n_2 + N_2$ are known. Define n_1, n_2, N_1 , and N_2 by

$$\begin{aligned} n_1 &= x_{111} + x_{101} + x_{110} + x_{100}, \\ n_2 &= y_{111} + y_{101} + y_{110} + y_{100}, \\ N_1 &= x_{011} + x_{001} + x_{010} + x_{000}, \\ N_2 &= y_{011} + y_{001} + y_{010} + y_{000}. \end{aligned}$$

Also, let

$$\begin{aligned} p_g &= \text{the prevalence of disease in group } g, g = 1, 2, \\ T_g &= \text{the total sample size, } T_g = n_g + N_g \text{ for } n_g = p_g T_g, N_g = (1 - p_g) T_g, \\ M_g &= \text{the missing data, } M_1 = x_{100} + x_{000} \text{ and } M_2 = y_{100} + y_{000}, \\ s_l &= \text{TPR (sensitivity) of test } l, l = 1, 2, \\ f_l &= \text{FPR (1-specificity) of test } l, \\ \lambda &= \text{dependency factor between tests in the diseased group (reference standard positive),} \\ \delta &= \text{dependency factor between tests in the non-diseased group (reference standard negative),} \end{aligned}$$

and assume the TPRs, FPRs, and dependency factors are independent of prevalence. Our interest lies in the differences of the TPRs and FPRs for the two tests. Therefore, define a new parameter, τ_1 , to be the difference in the TPRs (or sensitivities) of the two tests (Test 1 - Test 2) and a new parameter, τ_2 , to be the difference in the FPRs of the two tests (Test 1 - Test 2) such that

$$\begin{aligned} \tau_1 &= s_1 - s_2, \\ \tau_2 &= f_1 - f_2. \end{aligned}$$

Note that the corresponding differences in the specificities of the two tests (Test 1 - Test 2) can be denoted by $-\tau_2$, or $f_2 - f_1$. We now estimate the TPR and FPR differences, represented by τ_1 and τ_2 , using maximum likelihood (ML) and the expectation-maximization (EM) algorithm.

If \mathbf{v} is the observed data and \mathbf{z} is the missing or latent data with parameter vector

$$\boldsymbol{\theta} = (p_1, p_2, s_2, f_2, \tau_1, \tau_2, \lambda, \delta)',$$

then the multinomial complete-data density can be represented by the following function,

$$\begin{aligned} f(\mathbf{v}, \mathbf{z} \mid \boldsymbol{\theta}) &= A_1 A_2 [p_1(\tau_1 + s_2)s_2\lambda]^{x_{111}} [p_1s_2(1 - [\tau_1 + s_2]\lambda)]^{x_{101}} \\ &\times [p_1(\tau_1 + s_2)(1 - s_2\lambda)]^{x_{110}} [(1 - p_1)(\tau_2 + f_2)f_2\delta]^{x_{011}} \\ &\times [(1 - p_1)f_2(1 - [\tau_2 + f_2]\delta)]^{x_{001}} [(1 - p_1)(\tau_2 + f_2)(1 - f_2\delta)]^{x_{010}} \\ &\times [1 - p_1(\tau_1 + 2s_2 - [\tau_1 + s_2]s_2\lambda) - (1 - p_1)(\tau_2 + 2f_2 - [\tau_2 + f_2]f_2\delta)]^{M_1} \\ &\times [p_2(\tau_1 + s_2)s_2\lambda]^{y_{111}} [p_2s_2(1 - [\tau_1 + s_2]\lambda)]^{y_{101}} \\ &\times [p_2(\tau_1 + s_2)(1 - s_2\lambda)]^{y_{110}} [(1 - p_2)(\tau_2 + f_2)f_2\delta]^{y_{011}} \\ &\times [(1 - p_2)f_2(1 - [\tau_2 + f_2]\delta)]^{y_{001}} [(1 - p_2)(\tau_2 + f_2)(1 - f_2\delta)]^{y_{010}} \\ &\times [1 - p_2(\tau_1 + 2s_2 - [\tau_1 + s_2]s_2\lambda) - (1 - p_2)(\tau_2 + 2f_2 - [\tau_2 + f_2]f_2\delta)]^{M_2}, \end{aligned}$$

where

$$\begin{aligned} A_1 &= \binom{T_1}{x_{111}, x_{101}, x_{110}, x_{011}, x_{001}, x_{010}, M_1}, \\ A_2 &= \binom{T_2}{y_{111}, y_{101}, y_{110}, y_{011}, y_{001}, y_{010}, M_2}. \end{aligned}$$

Furthermore, we can obtain the complete-data log-likelihood by taking the natural logarithm of the previous density to get

$$\begin{aligned} \ln[L(\boldsymbol{\theta} \mid \mathbf{v}, \mathbf{z})] &\propto x_{111} \ln[p_1(\tau_1 + s_2)s_2\lambda] + x_{101} \ln[p_1s_2(1 - [\tau_1 + s_2]\lambda)] \\ &+ x_{110} \ln[p_1(\tau_1 + s_2)(1 - s_2\lambda)] + x_{011} \ln[(1 - p_1)(\tau_2 + f_2)f_2\delta] \\ &+ x_{001} \ln[(1 - p_1)f_2(1 - [\tau_2 + f_2]\delta)] \\ &+ x_{010} \ln[(1 - p_1)(\tau_2 + f_2)(1 - f_2\delta)] \\ &+ M_1 \ln[1 - p_1(\tau_1 + 2s_2 - [\tau_1 + s_2]s_2\lambda) \\ &\quad - (1 - p_1)(\tau_2 + 2f_2 - [\tau_2 + f_2]f_2\delta)] \\ &+ y_{111} \ln[p_2(\tau_1 + s_2)s_2\lambda] + y_{101} \ln[p_2s_2(1 - [\tau_1 + s_2]\lambda)] \\ &+ y_{110} \ln[p_2(\tau_1 + s_2)(1 - s_2\lambda)] + y_{011} \ln[(1 - p_2)(\tau_2 + f_2)f_2\delta] \\ &+ y_{001} \ln[(1 - p_2)f_2(1 - [\tau_2 + f_2]\delta)] \\ &+ y_{010} \ln[(1 - p_2)(\tau_2 + f_2)(1 - f_2\delta)] \\ &+ M_2 \ln[1 - p_2(\tau_1 + 2s_2 - [\tau_1 + s_2]s_2\lambda) \\ &\quad - (1 - p_2)(\tau_2 + 2f_2 - [\tau_2 + f_2]f_2\delta)]. \end{aligned}$$

From the complete-data log-likelihood, we obtain bounds for λ and δ which will be discussed further in Section 4. However, if the estimates $\hat{\lambda}$ and $\hat{\delta}$ fall outside of this

acceptable range of values, we generally set their estimate to be this boundary value. Details of the EM algorithm are in Appendix A.1.

2.3 Examples

The performance of diagnostic tests in determining disease status is of great importance in medical and epidemiological studies. We briefly overview two real-data examples from the biostatistical literature where the procedures we propose apply. In general, there are many instances in the medical sciences and epidemiology where these procedures apply. We discuss two examples of such cases. We first examine the colorectal example in Berry et al. (2002) by comparing disease prevalences, TPRs and FPRs, and test dependency factors. Additionally, we introduce an example involving data obtained from a study on dementia (Zhou, 1998). We consider the scenario for which the disease status of individuals with double negatives is unknown (34 individuals with age < 75 years and 56 individuals with age ≥ 75 years) and known, otherwise, while excluding the 831 unverified individuals with age < 75 years and the 439 unverified individuals with age ≥ 75 years.

2.3.1 Example 1

Berry et al. (2002) have compared rehydrated guaiac testing (Hemoccult) on three consecutive bowel movements with immunochemical testing using reversed passive haemagglutination (RPHA-Hemeselect) on the first bowel movement to detect colorectal cancers or adenomas greater than 10mm in 8353 males and females aged 40 – 70 years old that were recruited between March 1992 and September 1995. This group was further divided by age into a 40 – 59 age group and a 60 – 70 age group. Tables 2.3 and 2.4 show the results of the tests, where Test 1 corresponds to the Hemoccult test and Test 2 corresponds to the RPHA test results.

We considered four models: independence in both the non-diseased and diseased groups ($\delta = \lambda = 1$), independence in the non-diseased group ($\delta = 1$), independence in the diseased group ($\lambda = 1$), and an unrestricted model in which there is no inde-

pendence in either group. The unrestricted model resulted in the parameter estimates $\hat{p}_1 = 0.005$ and $\hat{p}_2 = 0.017$ for the disease prevalences in the two populations, $\hat{\tau}_1 = 0$ for the estimated difference in the TPRs, $\hat{\tau}_2 = 0.031$ for the estimated difference in the FPRs, $\hat{s}_2 = 0.726$ for the estimated TPR of Test 2, $\hat{f}_2 = 0.024$ for the estimated FPR of Test 2, and $\hat{\lambda} = 0.858$ and $\hat{\delta} = 5.385$ are dependency factors between tests for the diseased and non-diseased groups, respectively. Our estimate for δ is indicative of a significant positive correlation. However, our estimate for λ suggested a non-significant negative correlation, which is unlikely. This compels us to further examine the standard errors of the estimators that were unreported in Berry et al. (2002). Additionally, these findings serve as motivation for the parameter values used in the simulations in this paper.

Table 2.3. Sampling results for 40 – 59 age group of the Hemoccult data.

	Diseased			Non-diseased		
	Test1+	Test1–	Total	Test1+	Test1–	Total
Age 40 – 59						
Test2+	11	9	20	28	82	110
Test2–	9	$[x_{100}]$	$[9 + x_{100}]$	228	$[x_{000}]$	$[228 + x_{000}]$
Total	20	$[9 + x_{100}]$	$[n_1]$	256	$[82 + x_{000}]$	$[N_1]$
						$[x_{100} + x_{000}] = 5360$

Table 2.4. Sampling results for 60 – 70 age group of the Hemoccult data.

	Diseased			Non-diseased		
	Test1+	Test1–	Total	Test1+	Test1–	Total
Age 60 – 70						
Test2+	22	11	33	32	59	91
Test2–	11	$[y_{100}]$	$[11 + y_{100}]$	171	$[y_{000}]$	$[171 + y_{000}]$
Total	33	$[11 + y_{100}]$	$[n_2]$	203	$[59 + y_{000}]$	$[N_2]$
						$[y_{100} + y_{000}] = 2320$

2.3.2 Example 2

Zhou (1998) considered data from an epidemiological study of dementia and compared the accuracy of a new screening test with that of a standard screening procedure. The new screening test was based on information obtained from a cognitive test given to the subject suspected of having dementia and another test that was given to a relative or someone who knew the subject, while the standard screening test considered only the cognitive test given to the patient. A poor performance on the cognitive test represented a positive test result for dementia.

Table 2.5. Sampling results for < 75 age group of the Dementia data.

	Diseased			Non-diseased		
	Test 1+	Test1 1–	Total	Test 1+	Test1 1–	Total
Age < 75						
Test 2+	7	0	7	10	6	16
Test 2–	0	$[x_{100}]$	$[0 + x_{100}]$	19	$[x_{000}]$	$[19 + x_{000}]$
Total	7	$[7 + x_{100}]$	$[n_1]$	29	$[6 + x_{000}]$	$[N_1]$
						$[x_{100} + x_{000}] = 34$
						$T_1 = 76$

Table 2.6. Sampling results for ≥ 75 age group of the Dementia data.

	Diseased			Non-diseased		
	Test 1+	Test1 1–	Total	Test 1+	Test1 1–	Total
Age ≥ 75						
Test 2+	31	3	34	25	19	44
Test 2–	5	$[y_{100}]$	$[5 + y_{100}]$	10	$[y_{000}]$	$[10 + y_{000}]$
Total	36	$[3 + y_{100}]$	$[n_2]$	35	$[19 + y_{000}]$	$[N_2]$
						$[y_{100} + y_{000}] = 56$
						$T_2 = 149$

The confirmation of dementia can be a costly and time-consuming process, requiring a neurological examination, a neuropsychological test battery, laboratory tests,

computer tomography scans, and an interview with a relative of the subject. Therefore, confirmation of a subject's dementia status was reliant on their screening test results if the subject's age was less than 75 years old. Two age groups were considered for the subjects, < 75 years old and ≥ 75 years old. Furthermore, we suspect these groups have varying prevalences for dementia. We use this data to model a situation in which we have a sample of 76 individuals age < 75 years where 34 of those 76 individuals have unverified disease statuses and a sample of 149 individuals with age ≥ 75 years where 56 of those 149 individuals have unverified disease statuses. The scenario just described is shown in Tables 2.5 and 2.6.

2.4 Interval Estimation

We determine Bayesian and Wald-based methods to obtain interval estimates for τ_1 and τ_2 . Consider a sample x_1, \dots, x_n of size n with density $f(\mathbf{x} \mid \theta, \boldsymbol{\eta})$, where θ is the parameter of interest and $\boldsymbol{\eta}$ is a vector of nuisance parameters.

2.4.1 A Wald Confidence Interval

A classical method of confidence interval construction is based on the asymptotic normality of the MLE $\hat{\boldsymbol{\varphi}}$ of a parameter vector $\boldsymbol{\varphi} = (\theta, \boldsymbol{\eta}')'$. Assume our interest lies in the single scalar parameter, θ , where $\boldsymbol{\eta}$ is a vector-valued nuisance parameter. As in Stamey, Boese, and Young (2008), we define

$$I_{\theta\theta|\boldsymbol{\eta}}(\theta, \hat{\boldsymbol{\eta}}) \equiv I_{\theta\theta}(\theta, \hat{\boldsymbol{\eta}}) - [I_{\theta\boldsymbol{\eta}}(\theta, \hat{\boldsymbol{\eta}})]' [I_{\boldsymbol{\eta}\boldsymbol{\eta}}(\theta, \hat{\boldsymbol{\eta}})]^{-1} I_{\theta\boldsymbol{\eta}}(\theta, \hat{\boldsymbol{\eta}}),$$

where

$$I_{\theta,\theta} \equiv \begin{bmatrix} I_{\theta\theta} & I_{\boldsymbol{\eta}\theta} \\ I_{\theta\boldsymbol{\eta}} & I_{\boldsymbol{\eta}\boldsymbol{\eta}} \end{bmatrix}$$

is the profile information matrix. We formulate a Wald statistic to construct confidence intervals for both τ_1 and τ_2 . For τ_1 , using the expected profile information matrix, the Wald statistic is

$$W_n \equiv (\hat{\tau}_1 - \tau_1)^2 [I_{\tau_1\tau_1|\boldsymbol{\eta}}(\hat{\tau}_1, \hat{\boldsymbol{\eta}})] \sim \chi_{1,1-\alpha}^2,$$

where we define the nuisance parameter vector as

$$\boldsymbol{\eta} = (p_1, p_2, s_2, f_2, \tau_2, \lambda, \delta)'.$$

The Wald interval will result in identical conclusions to that of a confidence interval based on the asymptotic normality of $\hat{\tau}_1$, and is

$$\hat{\tau}_1 \pm z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{\tau}_1)},$$

where $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ percentile of the standard normal distribution, where $\widehat{\text{Var}}(\hat{\tau}_1)$ can be obtained using the inverse of Fisher's information matrix. For τ_2 , the Wald statistic is

$$W_n \equiv (\hat{\tau}_2 - \tau_2)^2 [I_{\tau_2\tau_2|\boldsymbol{\eta}}(\hat{\tau}_2, \hat{\boldsymbol{\eta}})] \sim \chi_{1,1-\alpha}^2,$$

where the nuisance parameter vector is

$$\boldsymbol{\eta} = (p_1, p_2, s_2, \tau_1, f_2, \lambda, \delta)'.$$

The interval based on the asymptotic normality of $\hat{\tau}_2$ is

$$\hat{\tau}_2 \pm z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{\tau}_2)}$$

where $\widehat{\text{Var}}(\hat{\tau}_2)$ can be found using the inverse of Fisher's information matrix.

2.4.2 A Bayesian Model

Using the sampling scheme introduced in Section 2, we next develop a Bayesian credible interval procedure using MCMC methods. A great advantage to the Bayesian approach is that parameter constraints can be easily incorporated through the specification of a prior distribution. The bounds for our dependency factors between tests in the diseased and non-diseased groups, λ and δ , are accounted for in the Bayesian model through the prior distributions chosen for these parameters. Unlike the frequentist approach, we do not encounter situations in which the estimates for λ or δ are located outside of these bounds. Another advantage in using a Bayesian approach is that large samples are not required to determine validity because Bayesian methods do

not rely on asymptotic assumptions. Furthermore, a Bayesian approach allows for the incorporation of expert knowledge through the specification of the prior distribution. We assume that uncertainty about our model’s prevalences, TPRs, and FPRs can be represented by independent beta prior distributions and uncertainty about λ and δ can be represented by independent gamma prior distributions, given by

$$\begin{aligned} p_i &\sim \text{beta}(\alpha_{p_i}, \beta_{p_i}), \\ s_i &\sim \text{beta}(\alpha_{s_i}, \beta_{s_i}), \\ f_i &\sim \text{beta}(\alpha_{f_i}, \beta_{f_i}), \\ \lambda &\sim \text{G}(\alpha_\lambda, \beta_\lambda) \text{I}_{(\lambda_l, \lambda_u)}(\lambda), \\ \delta &\sim \text{G}(\alpha_\delta, \beta_\delta) \text{I}_{(\delta_l, \delta_u)}(\delta). \end{aligned}$$

The parameters in the indicator functions, $\text{I}_{(\lambda_l, \lambda_u)}(\lambda)$ and $\text{I}_{(\delta_l, \delta_u)}(\delta)$, are obtained from the cell probabilities and defined as

$$\begin{aligned} \lambda_l &= \max\left(-\frac{1 - s_1 - s_2}{s_1 s_2}, 0\right), \\ \delta_l &= \max\left(-\frac{1 - f_1 - f_2}{f_1 f_2}, 0\right), \\ \lambda_u &= \min\left(\frac{1}{s_1}, \frac{1}{s_2}\right), \\ \delta_u &= \min\left(\frac{1}{f_1}, \frac{1}{f_2}\right). \end{aligned}$$

These parameter constraints guarantee $\hat{\lambda}$ and $\hat{\delta}$ will be contained within their bounds. Details of the full conditional posterior distributions are in Appendix A.2. Direct calculation of the posterior distributions is mathematically intractable so we implement an iterative MCMC technique for approximating posterior densities. In regards to the elicitation of prior distributions, we will first consider relatively noninformative beta and gamma priors. If we observe poor performing intervals, we incorporate informative beta priors by directly matching moments, or matching the center of the range of observed values for one of the diagnostic tests to the means of the beta distribution to obtain the prior parameters (Joseph et al., 1995).

2.5 Simulation Studies

We considered four scenarios that are encountered in diagnostic testing that consist of different combinations of high and low TPRs and FPRs. For each scenario, we find average coverages and widths of approximate 95% intervals for varying sample sizes and compare the results of a Wald-type interval with a Bayesian interval estimator using diffuse and informative priors. Note that in such diagnostic testing studies, we generally have large sample sizes and exact methods are not applicable.

2.5.1 A Study of Four Scenarios

We consider sample sizes from 500 to 2000 in increments of 50 for each population with the following parameter configurations.

1. Low TPR, Low FPR: $s_1 = s_2 = 0.7$, $f_1 = 0.06$, $f_2 = 0.03$
2. High TPR, High FPR: $s_1 = 0.88$, $s_2 = 0.98$, $f_1 = 0.15$, $f_2 = 0.18$
3. High TPR, Low FPR: $s_1 = 0.88$, $s_2 = 0.98$, $f_1 = 0.06$, $f_2 = 0.03$
4. High TPR, Mixed FPR: $s_1 = 0.95$, $s_2 = 0.80$, $f_1 = 0.15$, $f_2 = 0.04$

The first case is motivated by the parameter estimates obtained in Berry et al. (2002). We set the remaining parameters for each scenario to $p_1 = 0.30$, $p_2 = 0.05$, $\lambda = 1.00$, and $\delta = 5.00$. For each sample size of each scenario, we generated 500 data sets. In the absence of available information, we used a diffuse uniform beta prior distribution (Kostoulas, Leontides, Browne, and Gardner, 2009) defined on the interval $(0, 1)$ for the prevalences, TPRs and FPRs such that

$$p_1, p_2, s_1, s_2, f_1, f_2 \sim \text{beta}(1, 1),$$

and we use truncated diffuse gamma priors for λ and δ , given by

$$\begin{aligned} \lambda &\sim G(0.01, 0.01) I_{(\lambda_l, \lambda_u)}(\lambda), \\ \delta &\sim G(0.01, 0.01) I_{(\delta_l, \delta_u)}(\delta). \end{aligned}$$

We initialized a single chain near the truth with a 15,000 run burn-in, 50,000 sample updates, and a thinning of 5 and observed convergence of the chains. We determined

average widths and coverages of approximate 95% intervals for each sample size of the different scenarios. We then compared the Wald-type asymptotic normal approximation interval using the EM algorithm with varying proposed Bayesian intervals.

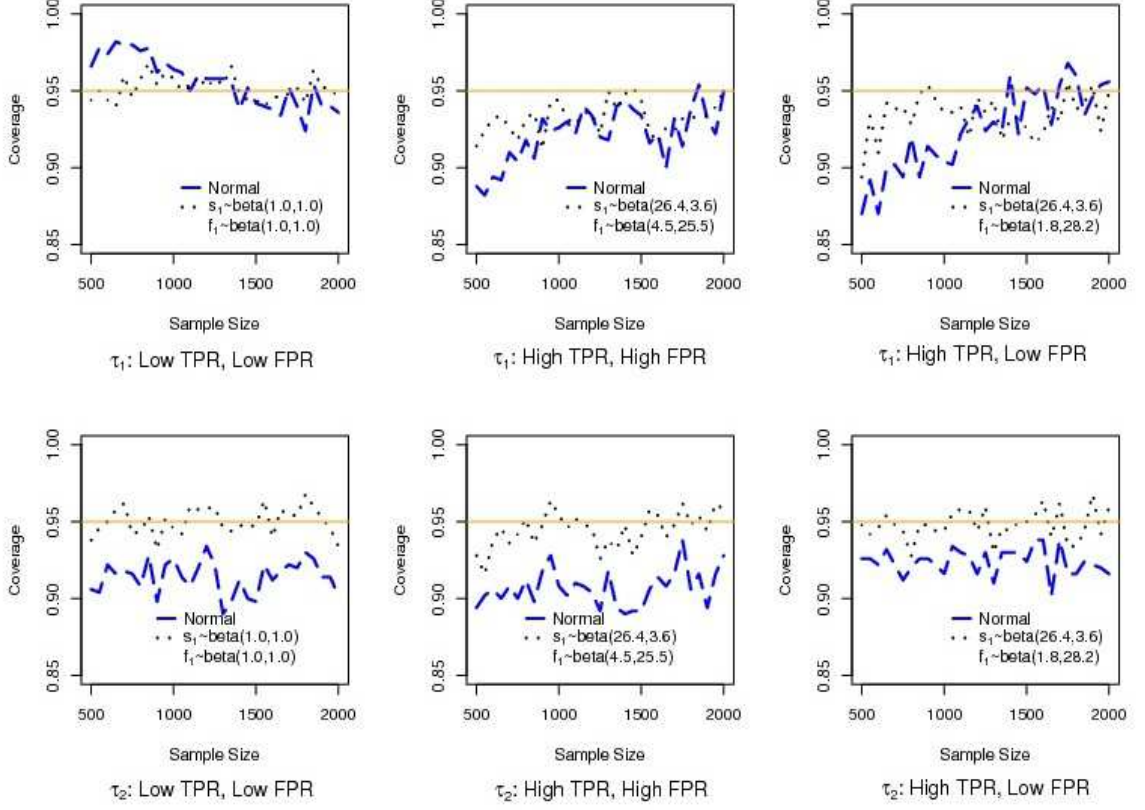


Figure 2.1. Coverages for τ_1 and τ_2 .

Coverage plots for τ_1 and τ_2 for the cases of low TPR and low FPR, high TPR and high FPR, and high TPR and low FPR are displayed in Figure 2.1. Considering the case of a low TPR and low FPR for τ_1 with relatively noninformative priors, resulting from the generation of observations using $s_1 = s_2 = 0.7$, $f_1 = 0.06$, $f_2 = 0.03$, we found that the normal approximation is overly conservative until reaching a sample size of 1100 while the Bayesian interval resulted in coverages near nominal for each sample size. For τ_2 in the low TPR and low FPR case, the Wald interval undercovered for all sample sizes while coverage for the Bayesian interval approached nominal almost immediately. Regarding the cases of high TPR, high FPR and high TPR, low FPR,

both the Wald and the Bayesian intervals with relatively noninformative priors resulted in poor coverages. Average interval widths for the low TPR and low FPR scenario for τ_1 and τ_2 , displayed in Figure 2.2, show slightly wider interval widths for the Bayesian intervals compared to the Wald intervals. The fourth scenario describes a situation where both diagnostic tests have a high TPR but only one of the tests has a low FPR while the other test has a high FPR.

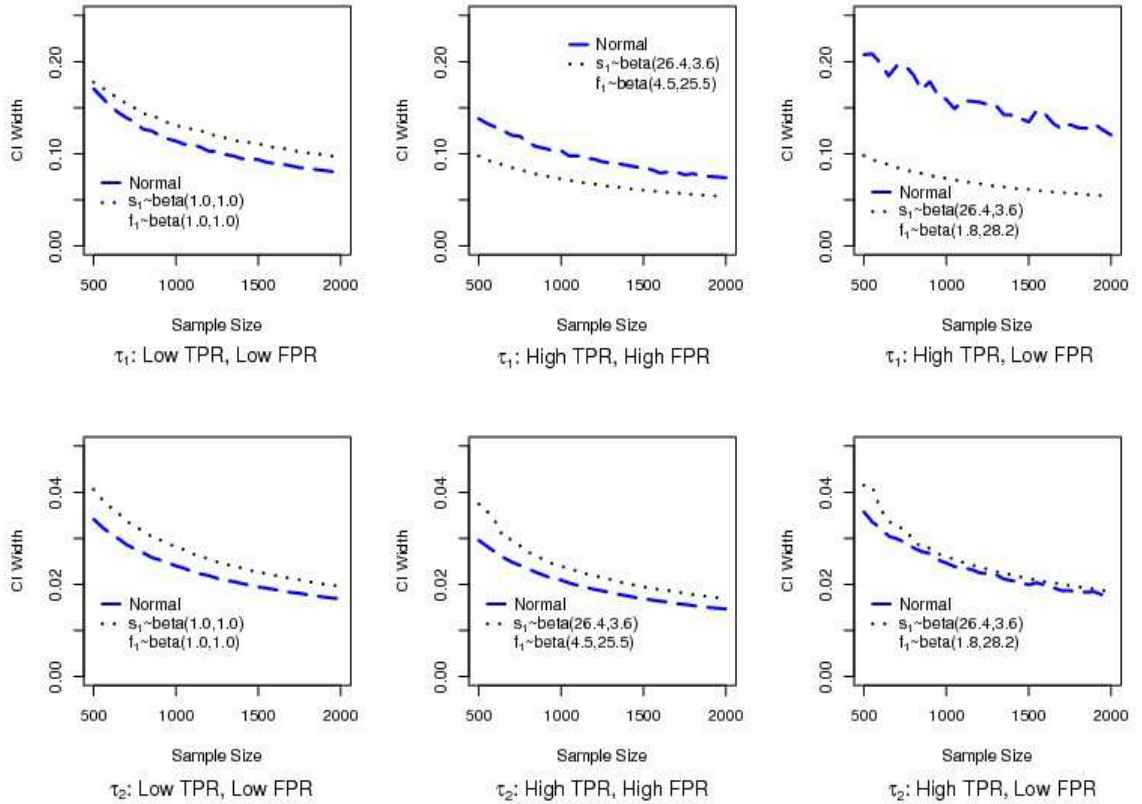


Figure 2.2. Interval widths for τ_1 and τ_2 .

2.5.2 Adding Informative Priors

Coverage for the Bayesian interval outperformed the Wald interval but still remained well below the nominal coverage for the cases of (1) high TPR and high FPR, (2) high TPR and low FPR, and (3) high TPR and mixed FPR. When comparing two diagnostic tests, we are often comparing a new, innovative, and perhaps lower-costing diagnostic test to a conventional test that may have been in use for many years. Thus,

prior information of the TPR and FPR of this test is often available. For such a conventional test, suppose we can obtain the test's TPR and FPR, which is not uncommon in a clinical setting. Thus, we can place informative priors on this conventional test's TPR and FPR. Without loss of generality, we place the informative priors given below on s_1 and f_1 for Test 1 in the three poorly performing scenarios of a high TPR and low FPR, a high TPR and high FPR, and a high TPR and mixed FPR:

- (i) High TPR, High FPR: $s_1 \sim \text{beta}(26.4, 3.6)$ and $f_1 \sim \text{beta}(4.5, 25.5)$
- (ii) High TPR, Low FPR: $s_1 \sim \text{beta}(26.4, 3.6)$ and $f_1 \sim \text{beta}(1.8, 28.2)$
- (iii) High TPR, Mixed FPR: $s_1 \sim \text{beta}(28.5, 1.5)$ and $f_1 \sim \text{beta}(4.5, 25.5)$

Therefore, the sample size equivalent of the prior information is set to 30, which is far less than the sample size equivalent of 65-85 used in studies examining *Strongyloides* prevalence in Cambodian refugees and used in estimating test sensitivity and specificity for the microscopy and serology tests (Joseph et al., 1995), (Dendukuri, Rahme, Belisle, and Joseph, 2004). Thus, for this type of study, these would be mildly informative priors.

Increasing the parameter values in the beta priors results in improved coverages for the high TPR and high FPR, high TPR and low FPR, and the high TPR and mixed FPR cases for τ_1 and τ_2 , displayed in Figures 2.1 and 2.3. Incorporating an informative prior for s_1 and f_1 into the model resulted in coverages for the Bayesian intervals reaching nominal coverage for the high TPR and high FPR case around a sample size of 1300 for τ_1 while the Wald intervals reached nominal coverage at around 1850. The Bayesian estimated intervals attained nominal coverage at about 800 for τ_2 while the normal approximated intervals undercovered for all sample sizes. For the high TPR and low FPR case, the Bayesian intervals achieve nominal coverage at about 850 for τ_1 while the Wald intervals undercovered until sample size 1400, and we observed nominal coverage almost immediately for the Bayesian intervals for τ_2 while the Wald intervals undercovered for all sample sizes. For τ_1 in the high TPR and mixed FPR

case, the average coverages reached nominal coverage immediately but then appears a little conservative by slightly overcovering. The Wald intervals did not attain nominal coverage until the sample size 1750. For τ_2 , the average coverages reached nominal coverage immediately while the Wald intervals undercovered until the sample size of 800.

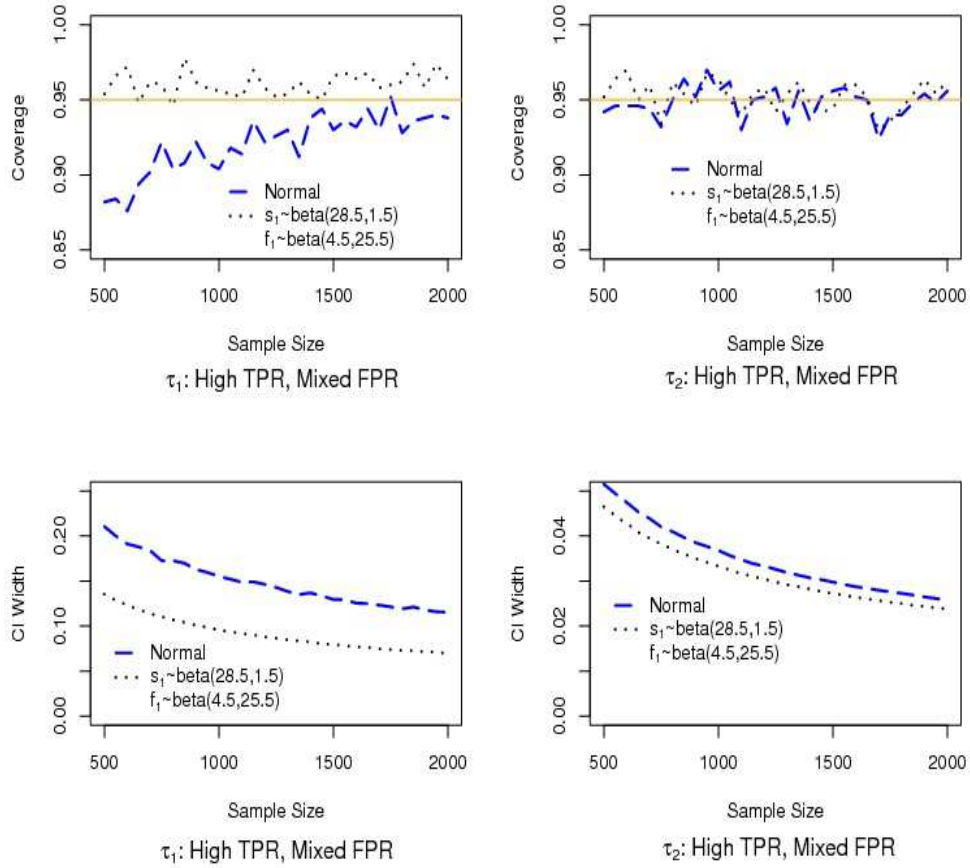


Figure 2.3. Coverages and interval widths for τ_1 and τ_2 (high TPR, mixed FPR scenario).

For the high TPR and high FPR, the average widths for the Wald intervals shown in Figure 2.2 were at least 0.02 wider than those resulting from the Bayesian intervals for τ_1 , while the average widths from the Bayesian intervals were slightly wider for τ_2 . For the high TPR and low FPR, the average widths for the Wald intervals were at least 0.06 wider than those resulting from the Bayesian intervals for τ_1 and the average widths were almost identical for τ_2 as sample size increases. For the high TPR and

mixed FPR, the average widths for the Wald intervals in Figure 2.3 were at least 0.04 wider than those resulting from the Bayesian intervals for τ_1 . For τ_2 , the average widths performed similarly with the Wald intervals slightly wider than those resulting from the Bayesian intervals.

2.6 Discussion

Although we can obtain parameter estimates using an EM algorithm that does not impose any parameter constraints, the Bayesian approach provides improved interval estimates compared to the Wald interval for all scenarios considered here. Incorporating an informative prior for one of the test's TPR and FPR results in better performing interval estimates that are near nominal coverage. Attempts to improve the Wald interval estimator failed to improve interval coverages and average widths. Agresti (2000) discussed the effect of adding very small constants to data sets resulting in cells with zero observations. However, this approach did not improve the Wald interval. The Bayesian approach is much more effective for the estimation of the difference in the TPRs and FPRs of two diagnostic tests applied to two populations when positive correlation between tests is present.

CHAPTER THREE

A Hierarchical Approach to Modeling Multicenter Prevalences, TPRs, and FPRs for Two Populations

3.1 Introduction

Cervical cancer is the third most common cancer among women worldwide, and it the second most common cancer among women in developing countries, according to Arbyn, Castellsague, de Sanjose, Bruni, Saraiya, and et al. (submitted 2010) and Ferlay, Shin, Bray, Forman, Mathers, and Parkin (2010). Although cervical cancer is preventable through effective screening programs, there were over 530,000 new cases diagnosed and 270,000 deaths from cervical cancer reported worldwide in 2008. Approximately 86% of these deaths occurred in developing countries, which lack the resources and infrastructure necessary to run organized screening programs (Arbyn, Gaffikin, Sankaranarayanan, Muwonge, Keita, Wesley, and TSU, 2006).

The Pap smear's introduction, more than 40 years ago, has remarkably transformed cervical cancer from a leading killer to a rare disease in the United States by decreasing the rate of cancer by 75 percent. However, only five percent of women in low-income countries have undergone a Pap smear in the past five years (Katz and Wright, 2006). Also, older and poorer women, which have the highest risk of developing cervical cancer, are least likely to be screened. In many developing countries, the majority of women have never undergone a pelvic exam (Gakidou, Nordhagen, and Obermeyer, 2008). Therefore new, cost-effective alternative screening methods are currently being investigated that may perform equally or possibly better than the Pap smear for such developing countries (Arbyn et al., 2006).

In this paper, we extend the model of Berry, Smith, Macaskill, and Irwig (2002) by presenting a hierarchical Bayesian logit model for modeling the TPRs and FPRs of two dichotomous screening tests applied to members of a stratified population into

two distinct groups. Estimation of the TPR and FPR for a diagnostic or screening test requires individuals to be identified as diseased or non-diseased using a gold standard. However, often in the investigation of new diagnostic tests, neither test is considered a gold standard. We compare the accuracy of two low-cost procedures for screening of high-grade cervical intra-epithelial neoplasia (CIN2+) using a five year cross-sectional multi-center study conducted by the International Agency for Research in Cancer. Similarly to Berry et al. (2002), we relax the assumption of independence by allowing correlation between the diagnostic tests. Our two groups for modeling the prevalence of CIN2+ consist of (1) women < 45 years old and (2) women ≥ 45 years old. We assume the groups have varying disease prevalences with unverified negatives. However, verification for these individuals has been performed so that we may compare the parameter estimates resulting from our model to those values obtained from the full data.

Recent developments in simulation methods and computational speed have enabled many advancements in the estimation of elaborate models. By specifying a model with a hierarchical structure, we reduce a complex model into a set of simpler submodels with more flexibility and features at each level resembling that of the original data. Furthermore, we obtain stable estimates through Bayesian methods. Martinez, Louzada-Neto, Derchain, Achcar, Gontijo, and et al. (2008) use a Bayesian method to estimate disease prevalence, sensitivity, and specificity of three cervical cancer screening tests in the presence of a covariate and in the absence of a gold standard. They further note that the method is advantageous in that the number of parameters to be estimated is not limited by the number of observations, as commonly encountered with frequentist approaches. We estimate the multiple parameters of interest through simulation methods based on MCMC using the software R, the WinBUGS program, and the R library R2WinBUGS.

Rosenberg, Andrews, and Lenk (1999) introduced a hierarchical Bayesian logistic regression model for predicting the rate of nonacceptable claims (NAC) in an attempt

to minimize unnecessary costs to health insurance companies. Six variables obtained from claim records were included into the model: gender, patient age, length of hospital stay in days, admission type, type of service, and the principal reason for admission (DIAG). It was found that the length of stay, medical type of service, and gender were influential in predicting unnecessary hospitalizations.

A hierarchical Bayesian model also has applications pertaining to models needing to account for spatial heterogeneity among different sites or locations having areal differences. MacNab (2003) quantified the magnitude of spatial variation in small-area incidence rates of intraventricular hemorrhage (IVH) among Canadian Neonatal Intensive Care Unit (NICU) neonates by proposing a Bayesian hierarchical spatial model to aid in the prediction of spatial patterns and identification of regional variations in order to make informed health service provisions and provide proper resource allocation. Furthermore, MacNab (2003) assumed a canonical logit link function for the IVH rates and presented a logistic linear mixed effects model, where the random effects represent residual spatial effects resulting from unmeasured risk factors. Several covariates, such as admission illness severity, birth weight, outborn, Apgar score, antenatal steroids usage, vasopressor infusion, and treatment of acidosis, were included into the model and their effects examined.

Farnsworth, Hoeting, Hobbs, and Miller (2006) took a hierarchical Bayesian approach to modeling chronic wasting disease (CWD) and influences on its spatial pattern in northcentral Colorado to mule deer (*Odocoileus hemionus*) movement and mixing using the standard logit transform for probability of infection for each deer. Three scales of mixing were considered to have a possible impact on spatial variation in disease prevalence: (1) individual, (2) winter subpopulation and (3) summer subpopulation. For each individual deer, the probability of being CWD-positive is modeled as a function of several individual and environmental covariates and two random effect terms to account for unobserved covariates and the spatial pattern of the disease. Farnsworth, Hoeting, Hobbs, and Miller (2006) found that local influences from the

individual mixing scale had the strongest influence on the observed spatial patterns of CWD prevalence in mule deer. Two individual-level covariates, sex and age, along with two environmental covariates, representing the percentage of private land and percentage of low-elevation grassland habitat, were found to impact disease prevalence.

Oleson and He (2004) have introduced a hierarchical Bayesian model to model a dichotomous process in the presence of nonresponse and modeled success rates with the logit function. Spatial correlations are also introduced into the model for both success and response rates. However, we further extend these results to a clinical setting for which two dichotomous processes (or screening tests) are considered and applied to two different populations at various locations. We use the logit function to model prevalence for each group at each center as a linear function consisting of an offset term with a covariate based on age. We further model the TPR and FPR of each test at each location using the logit function as a linear function of an offset term that represents an overall mean across all locations and a random effects component.

3.2 Study Background

Several countries, including China, Costa Rica, and India, have recently adopted low-cost screening techniques for cervical cancer involving visual inspection with acetic acid or iodine in an attempt to reduce the incidence rate of the disease (Katz and Wright, 2006). One alternative, low-cost screening test involves swabbing the cervix with 3 to 5% acetic acid (VIA), which will turn cervical intra-epithelial abnormalities white. Another is a visual inspection with Lugol’s iodine (VILI), where yellow non-iodine uptake areas suggest the presence of cervical lesions. A third alternative involves an inspection of the cervix with the naked eye or using a low-level magnifying device (VIAM) (Arbyn et al., 2006).

A research project conducted from 1999 to 2003 by the International Agency for Research in Cancer (IARC, Lyon) used a cross-sectional multi-center study involving 58,679 women, ages 25-65 years old, set up at 11 locations in five African countries and

India:

1. Mumbai
2. Trivandrum 1
3. Trivandrum 2
4. Calcutta 1
5. Calcutta 2
6. Bamako
7. Brazzaville
8. Conakry
9. Jaipur
10. Niamey
11. Ouagadougou

Testing with VIA occurred at all 11 locations while testing with VILI occurred at only 10 of the 11 locations. Other screening tests performed include VIAM at three locations, a Pap smear at five locations, and a Hybrid Capture II (HC2) assay used to detect the presence of high-risk HPV DNA at four locations. All screened women were verified with a gold standard, a colposcopic exploration of the cervix performed at the same time of their visit. The prevalence of cervical intra-epithelial neoplasia (CIN) was then assessed, according to four levels of disease outcome: (1) CIN I or worse (CIN1+), (2) CIN2+, (3) CIN3+, and (4) invasive cervical cancer. Different health workers, blinded to the results of the other tests, performed all of the test procedures. Since the gold standard is applied to all subjects in the study, the accuracy for CIN or cancer can be evaluated without the presence of verification bias.

We will elaborate and further extend the results of this study by comparing the accuracy of the VIA and VILI screening tests for CIN2+ in two different populations, (1) women < 45 years old and (2) women ≥ 45 years old. We propose a hierarchical Bayesian mixed logit model to examine test sensitivity and specificity along with prevalence of CIN2+ in each of the populations. We will consider only 10 of the 11 centers since VILI testing was not performed at the Calcutta 1 center, reducing the

number of women screened to 52,785 women. The random effects are modeled through Bayesian prior specifications reflecting heterogeneity among the centers.

3.3 Model Description

Consider a single population stratified into g groups by some characteristic. In our model, this stratification variable will be the age of the women screened in the study. Let $z_{gijk,m}$ represent the g^{th} group with disease status i , Test 1 result j , and Test 2 result k for group $\Pi_{g,m}$, where

$$\begin{aligned} g &= 1 \text{ (women } \geq 45), 2 \text{ (women } < 45), \\ i &= 0 \text{ (non-diseased), } 1 \text{ (diseased),} \\ j &= 0 \text{ (negative), } 1 \text{ (positive),} \\ k &= 0 \text{ (negative), } 1 \text{ (positive).} \end{aligned}$$

at location m , $m = 1$ (Mumbai), \dots , 10 (Ouagadougou). Considering a single location m , Tables 3.1 and 3.2 show the sampling plan for the two groups $\Pi_{1,m}$ and $\Pi_{2,m}$, respectively.

In Table 3.1, $z_{11\cdot 1,m}$ denotes the row sum of diseased individuals who tested positive for Test 2 (VILI) in group 1 at location m . Similarly, $z_{111\cdot,m}$ denotes the column sum of diseased individuals who tested positive for Test 1 (VIA) in group 1 at location m . Thus, we have multinomial data for each location m .

Table 3.1. Sampling Plan for Group 1 at the m th center.

	Diseased			Non-diseased		
	Test 1+	Test1 1–	Total	Test 1+	Test1 1–	Total
Test 2+	$z_{1111,m}$	$z_{1101,m}$	$z_{11\cdot 1,m}$	$z_{1011,m}$	$z_{1001,m}$	$z_{10\cdot 1,m}$
Test 2–	$z_{1110,m}$	$[z_{1100,m}]$	$[z_{11\cdot 0,m}]$	$z_{1010,m}$	$[z_{1000,m}]$	$[z_{10\cdot 0,m}]$
Total	$z_{111\cdot,m}$	$[z_{110\cdot,m}]$	$[n_{1,m}]$	$z_{101\cdot,m}$	$[z_{100\cdot,m}]$	$[N_{1,m}]$

For $\Pi_{1,m}$, the variables $z_{1111,m}$, $z_{1101,m}$, $z_{1110,m}$, $z_{1011,m}$, $z_{1001,m}$, and $z_{1010,m}$ displayed in Table 3.1 have a verified disease status. Hence, these values are observed and known. However, variables in brackets shown in Table 3.1, such as $z_{1100,m}$ and $z_{1000,m}$, are assumed unknown. The variables $z_{1100,m}$ and $z_{1000,m}$ refer to the number of individuals sampled from $\Pi_{1,m}$ that are diseased and non-diseased, respectively, but had negative test results for both screening procedures, VIA and VILI. As a result, the total number of diseased and non-diseased individuals for $\Pi_{1,m}$, represented by $n_{1,m}$ and $N_{1,m}$ respectively, are unknown. However, the total number of double negatives for $\Pi_{1,m}$, denoted by $M_{1,m} = z_{1100,m} + z_{1000,m}$, and the total size of $\Pi_{1,m}$, denoted by $T_{1,m} = n_{1,m} + N_{1,m}$, are both known.

Table 3.2. Sampling Plan for Group 2 at the m th center.

	Diseased			Non-diseased		
	Test 1+	Test1 1−	Total	Test 1+	Test1 1−	Total
Test 2+	$z_{2111,m}$	$z_{2101,m}$	$z_{21\cdot1,m}$	$z_{2011,m}$	$z_{2001,m}$	$z_{20\cdot1,m}$
Test 2−	$z_{2110,m}$	$[z_{2100,m}]$	$[z_{21\cdot0,m}]$	$z_{2010,m}$	$[z_{2000,m}]$	$[z_{20\cdot0,m}]$
Total	$z_{211\cdot,m}$	$[z_{210\cdot,m}]$	$[n_{2,m}]$	$z_{201\cdot,m}$	$[z_{200\cdot,m}]$	$[N_{2,m}]$

Similarly, for the group $\Pi_{2,m}$ at location m , the variables $z_{2111,m}$, $z_{2101,m}$, $z_{2110,m}$, $z_{2011,m}$, $z_{2001,m}$, and $z_{2010,m}$ displayed in Table 3.2 are observed and known, while the variables $z_{2100,m}$ and $z_{2000,m}$ are unknown with $n_{2,m}$ and $N_{2,m}$ also unknown, but $M_{2,m} = z_{2100,m} + z_{2000,m}$ and $T_{2,m} = n_{2,m} + N_{2,m}$ are known. We further define $n_{1,m}$, $n_{2,m}$, $N_{1,m}$ and $N_{2,m}$ by

$$\begin{aligned}
n_{1,m} &= z_{1111,m} + z_{1101,m} + z_{1110,m} + z_{1100,m}, \\
n_{2,m} &= z_{2111,m} + z_{2101,m} + z_{2110,m} + z_{2100,m}, \\
N_{1,m} &= z_{1011,m} + z_{1001,m} + z_{1010,m} + z_{1000,m}, \\
N_{2,m} &= z_{2011,m} + z_{2001,m} + z_{2010,m} + z_{2000,m}.
\end{aligned}$$

For a location m , let

- $p_{g,m}$ = the prevalence of CIN2+ in group g , $g = 1, 2$,
- $T_{g,m}$ = the total sample size, $T_{g,m} = n_{g,m} + N_{g,m}$ for $n_{g,m} = p_{g,m}T_{g,m}$ and $N_{g,m} = (1 - p_{g,m})T_{g,m}$,
- $M_{g,m}$ = the missing data, $M_{1,m} = z_{1100,m} + z_{1000,m}$ and $M_{2,m} = z_{2100,m} + z_{2000,m}$,
- $s_{l,m}$ = TPR (sensitivity) of test l , $l = 1, 2$,
- $f_{l,m}$ = FPR (1-specificity) of test l ,
- λ_m = dependency factor between tests in diseased group (reference standard positive),
- δ_m = dependency factor between tests in non-diseased group (reference standard negative),

and assume the TPRs, FPRs, and dependency factors are independent of prevalence.

If \mathbf{v} is the observed data and \mathbf{z} is the missing or latent data with parameter vector

$$\boldsymbol{\theta} = (p_{1,m}, p_{2,m}, s_{1,m}, s_{2,m}, f_{1,m}, f_{2,m}, \lambda_m, \delta_m)',$$

then the multinomial complete-data log-likelihood for location m is

$$\begin{aligned} \ln[L(\boldsymbol{\theta} \mid \mathbf{v}, \mathbf{z})] &\propto z_{1111,m} \ln[p_{1,m} s_{1,m} s_{2,m} \lambda_m] + x_{101,m} \ln[p_{1,m} s_{2,m} (1 - s_{1,m} \lambda_m)] \\ &+ z_{1110,m} \ln[p_{1,m} s_{1,m} (1 - s_{2,m} \lambda_m)] + z_{1011,m} \ln[(1 - p_{1,m}) f_{1,m} f_{2,m} \delta_m] \\ &+ z_{1001,m} \ln[(1 - p_{1,m}) f_{2,m} (1 - f_{1,m} \delta_m)] \\ &+ z_{1010,m} \ln[(1 - p_{1,m}) f_{1,m} (1 - f_{2,m} \delta_m)] \\ &+ M_{1,m} \ln[1 - p_{1,m} (s_{1,m} + s_{2,m} - s_{1,m} s_{2,m} \lambda_m) \\ &\quad - (1 - p_{1,m}) (f_{1,m} + f_{2,m} - f_{1,m} f_{2,m} \delta_m)] \\ &+ z_{2111,m} \ln[p_{2,m} s_{1,m} s_{2,m} \lambda_m] + z_{2101,m} \ln[p_{2,m} s_{2,m} (1 - s_{1,m} \lambda_m)] \\ &+ z_{2110,m} \ln[p_{2,m} s_{1,m} (1 - s_{2,m} \lambda_m)] + z_{2011,m} \ln[(1 - p_{2,m}) f_{1,m} f_{2,m} \delta_m] \\ &+ z_{2001,m} \ln[(1 - p_{2,m}) f_{2,m} (1 - f_{1,m} \delta_m)] \\ &+ z_{2010,m} \ln[(1 - p_{2,m}) f_{1,m} (1 - f_{2,m} \delta_m)] \\ &+ M_{2,m} \ln[1 - p_{2,m} (s_{1,m} + s_{2,m} - s_{1,m} s_{2,m} \lambda_m) \\ &\quad - (1 - p_{2,m}) (f_{1,m} + f_{2,m} - f_{1,m} f_{2,m} \delta_m)], \end{aligned}$$

where

$$\begin{aligned} A_1 &= \begin{pmatrix} T_1 \\ z_{1111,m}, z_{1101,m}, z_{1110,m}, z_{1011,m}, z_{1001,m}, z_{1010,m}, M_{1,m} \end{pmatrix}, \\ A_2 &= \begin{pmatrix} T_2 \\ z_{2111,m}, z_{2101,m}, z_{2110,m}, z_{2011,m}, z_{2001,m}, z_{2010,m}, M_{2,m} \end{pmatrix}. \end{aligned}$$

Now let the log-odds for prevalence be represented by a linear function using the *logit* function,

$$\text{logit}(p_{g,m}) = \beta_{0g,m} + \beta_1 X_{g,m},$$

where prevalence depends on the observed covariates, $X_{g,m}$, defined by

$$X_{g,m} = \begin{cases} 1 & \text{if age} < 45 \\ 0 & \text{if age} \geq 45, \end{cases}$$

for an individual in group g at location m , and β_1 is an unobserved regression coefficient of the covariate, assigned a prior normal distribution, $N(0, 10)$. We treat $\beta_{0g,m}$ as an offset term representing an overall mean that varies randomly and takes on different values for each group in each of the different centers. We assign a normal prior to $\beta_{0g,m}$, such that

$$\beta_{0g,m} | \mu_{\beta_0}, \sigma_{\beta_0} \sim N(\mu_{\beta_0}, \sigma_{\beta_0}^2)$$

with hyperpriors,

$$\begin{aligned} \mu_{\beta_0} &\sim N(0, 10), \\ \sigma_{\beta_0} | W &\sim U(0, W), \end{aligned}$$

where W is a pre-specified upperbound for the variance component of $\beta_{0g,m}$. We employ a uniform prior for the standard deviation parameter of the fixed effect component in order to provide a conjugate prior that contains little information and provides little influence to the posterior distribution of the model parameters. Furthermore, we may note that $e^{\beta_{0g,m}}$ represents the overall rate odds for group g at location m (MacNab, 2003).

Similarly, let the log-odds of $s_{l,m}$ and $f_{l,m}$ be represented by the linear functions,

$$\begin{aligned} \text{logit}(s_{1,m}) &= \gamma_1 + e_{1(m)} \\ \text{logit}(s_{2,m}) &= \gamma_2 + e_{2(m)} \\ \text{logit}(f_{1,m}) &= \gamma_3 + e_{3(m)} \\ \text{logit}(f_{2,m}) &= \gamma_4 + e_{4(m)}, \end{aligned}$$

where γ_q , $q = 1, \dots, 4$, is an overall mean component with prior distribution $\gamma_q \sim N(0, 10)$ and $e_{q(m)}$ represents a random effects component for the 10 locations, normally distributed with zero mean and variance σ_q^2 , $e_{q(m)} | \sigma_q \sim N(0, \sigma_q^2)$, where $\sigma_q | W \sim U(0, W)$.

To reflect our lack of knowledge of the dependency factors between tests for the diseased and non-diseased for each location m , we assign uniform priors to λ_m and δ_m , respectively, given by

$$\begin{aligned}\lambda_m | \alpha_\lambda, \beta_\lambda &\sim U(\alpha_\lambda, \beta_\lambda) \\ \delta_m | \alpha_\delta, \beta_\delta &\sim U(\alpha_\delta, \beta_\delta),\end{aligned}$$

with restrictions placed on the lower and upper limits of the uniform priors for λ_m and δ_m similar to that of Dendukuri and Joseph (2001),

$$\begin{aligned}\alpha_\lambda &= \max\left(-\frac{1 - s_{1,m} - s_{2,m}}{s_{1,m}s_{2,m}}, 0\right) \\ \alpha_\delta &= \max\left(-\frac{1 - f_{1,m} - f_{2,m}}{f_{1,m}f_{2,m}}, 0\right) \\ \beta_\lambda &= \min\left(\frac{1}{s_{1,m}}, \frac{1}{s_{2,m}}\right) \\ \beta_\delta &= \min\left(\frac{1}{f_{1,m}}, \frac{1}{f_{2,m}}\right),\end{aligned}$$

for locations $m = 1$ (Mumbai), \dots , 10 (Ouagadougou).

3.4 Estimation Procedure

We are unable to derive closed-form expressions for the posterior density because of the large dimensionality of the hierarchical model. Hence, we use MCMC simulation to evaluate the posterior distribution. First, we obtain parameter estimates for the prevalence of CIN2+ in each of the two age groups, the TPRs and FPRs of each test, the overall mean logit components, and the dependency factors between the two tests for each location individually. Therefore, there are no random effects incorporated into the model. We compare these estimates for each of the locations to those estimates obtained from the hierarchical Bayesian mixed logit model obtained by pooling the ten locations and introducing random effects into the model. For both the individual locations and pooled scenarios, we initialize two chains with a 50,000 run burn-in, 500,000 sample updates, and keep every fifth iteration for inference.

3.4.1 Individual Locations

We first examine the hierarchical model for each location separately. Hence, random effects to reflect the heterogeneity across centers are not induced into the model. When we consider the absence of random effects, the parameters γ_q , $q = 1, \dots, 4$, reduce to the simple logit of the location's TPRs and FPRs. We further set $\mu_{\beta_0} = 0$ and $\sigma_{\beta_0} = 3.2$, providing a relatively non-informative prior for $\beta_{0g,m}$. Initials used for each location are displayed in Table B.1.

Table 3.3. Resulting TPRs, FPRs, prevalences, and dependency factors.

Parameter	Full Data	Single Site	Hierarchical
$s_{1,1}$	0.617	0.0837	0.8126
$s_{1,7}$	0.911	0.9011	0.8999
$s_{1,8}$	0.875	0.0504	0.8935
$s_{2,1}$	0.741	0.1006	0.9571
$s_{2,7}$	0.970	0.9612	0.9650
$s_{2,8}$	0.875	0.0504	0.9287
$f_{1,1}$	0.119	0.4220	0.1182
$f_{1,7}$	0.062	0.0623	0.0626
$f_{1,8}$	0.249	0.7096	0.2483
$f_{2,1}$	0.160	0.5671	0.1587
$f_{2,7}$	0.096	0.0959	0.0962
$f_{2,8}$	0.270	0.7680	0.2687
$p_{1,1}$	0.0247	0.6828	0.0194
$p_{1,7}$	0.0357	0.0359	0.0316
$p_{1,8}$	0.0125	0.6596	0.0140
$p_{2,1}$	0.0180	0.5927	0.0141
$p_{2,7}$	0.0144	0.0147	0.0158
$p_{2,8}$	0.0116	0.5934	0.0103
λ_1	—	39.7	1.011
λ_7	—	1.028	1.021
λ_8	—	57.94	1.022
δ_1	—	2.003	5.227
δ_7	—	8.92	8.894
δ_8	—	1.394	3.488

Tables B.2-B.4 in Appendix B show the resulting parameter estimates for each location when the ten centers are analyzed individually. We observe convergence for some of the locations. For the Brazzaville location, we observe the estimates 0.8192 and 0.9737 for the TPRs of Test 1 and Test 2, respectively, and the estimates 0.2338

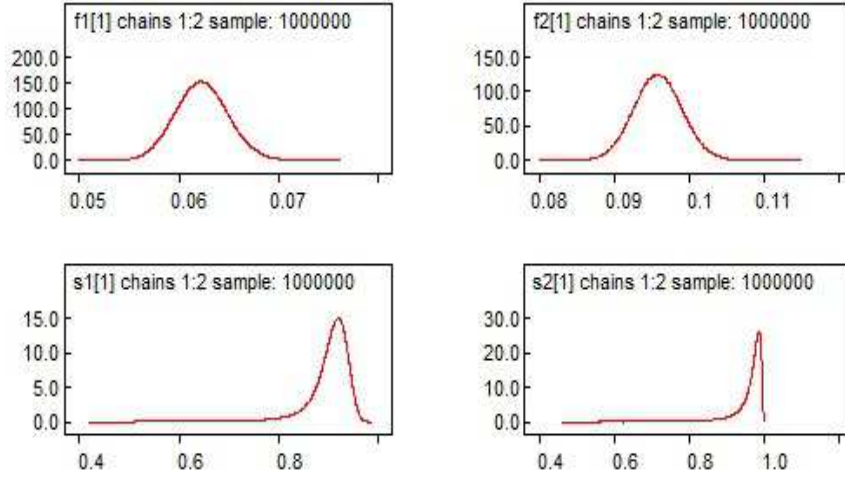


Figure 3.1. Density plots for Conakry's TPRs and FPRs (single site).

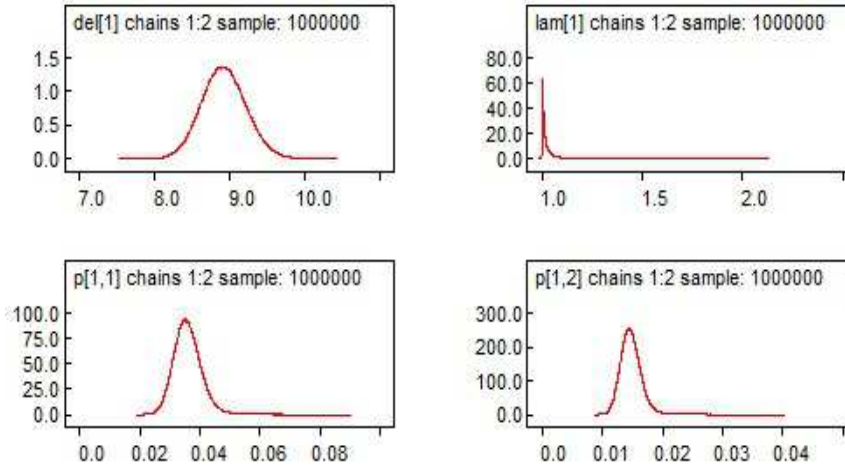


Figure 3.2. Density plots for Conakry's dependency factors and prevalences (single site).

and 0.1103 for the FPRs of Test 1 and Test 2, respectively. The full data TPRs for this location are 0.805 and 0.956, respectively, and the full data FPRs are 0.234 and 0.110, respectively. Furthermore, we obtain estimates similar to the full data values for the Conakry location. Additionally, we obtain estimates close to their full data values, but not as similar to the two previous locations, for the locations Trivandrum 2, Niamey, and Ouagadougou. However, we fail to observe convergence for the remaining centers due to small observed cell counts. Thus, for several locations we have a sparse data concern. The model results for the Mumbai location demonstrate a scenario in which

we observe a lack of convergence. Posterior densities and history plots further demonstrate the lack of convergence for several of the individual sites.

Table 3.3 shows the resulting parameter estimates for a subset of the different locations. All tables are available for the interested reader, but we include results for locations 1 (Mumbai), 7 (Conakry), and 8 (Jaipur) only for this paper. We observe convergence for some of the locations. For the Conakry site (center 7), we observe the estimates 0.9011 and 0.9612 for the TPRs of Test 1 and Test 2, respectively, and the estimates 0.0623 and 0.0959 for the FPRs of Test 1 and Test 2, respectively. The full data TPRs for this location are 0.911 and 0.970, respectively, and the true FPRs are 0.062 and 0.096, respectively. Posterior density plots for Conakry's TPRs and FPRs, prevalences, and dependency factors can be obtained from the author.

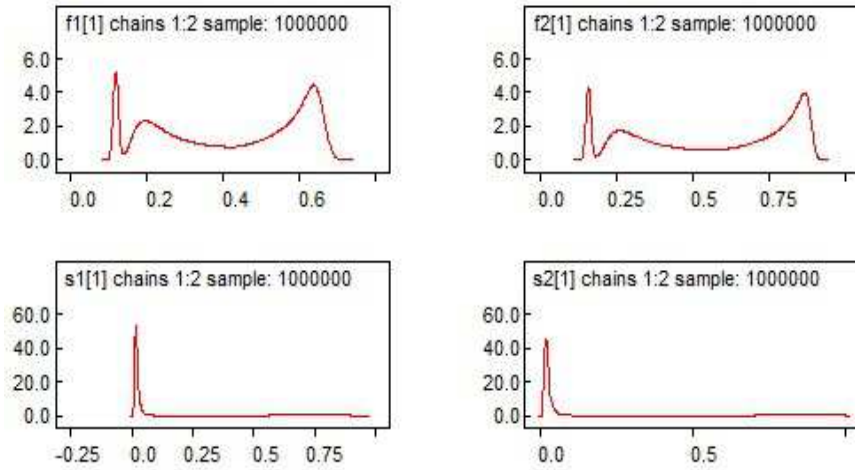


Figure 3.3. Density plots for Mumbai's TPRs and FPRs (single site).

We additionally obtain estimates similar to the full data values for the Brazzaville center. We obtain estimates close to their full data values, but not as similar to the two previous locations, for the locations of Trivandrum 2, Niamey, and Ouagadougou. However, we fail to observe convergence for the remaining centers due to small observed cell counts. Thus, for several locations we have a sparse data concern. The model results for Mumbai (center 1) demonstrate a scenario in which we observe

a lack of convergence. Posterior densities for this center’s prevalences and dependency factors are also available from the author. Posterior densities and history plots further demonstrate the lack of convergence for several of the other individual sites.

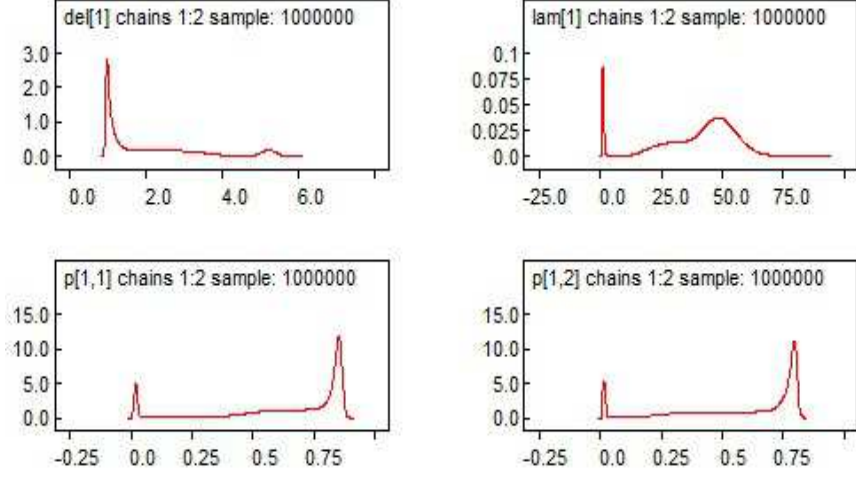


Figure 3.4. Density plots for Mumbai’s dependency factors and prevalences (single site).

3.4.2 Combined Locations

We now examine convergence for the hierarchical Bayesian logit model when analyzing data from all ten centers simultaneously and adding random effects into the model to account for heterogeneity across the centers. We now allow μ_{β_0} and σ_{β_0} to vary. Two chains are initialized at values for $\hat{\gamma}$, μ_{β_0} , σ_{β_0} , and the random effects $e_{1(m)}$, $e_{2(m)}$, $e_{3(m)}$, and $e_{4(m)}$,

1. $\hat{\gamma} = (1.5, 3.0, -1.5, -1.5)$, $\mu_{\beta_0} = -3.0$, $\sigma_{\beta_0} = 0.05$,
 $e_{1(m)}, e_{2(m)}, e_{3(m)}, e_{4(m)} = -1$ for all m
2. $\hat{\gamma} = (2.0, 3.5, -2.0, -2.0)$, $\mu_{\beta_0} = -4.0$, $\sigma_{\beta_0} = 0.20$,
 $e_{1(m)}, e_{2(m)}, e_{3(m)}, e_{4(m)} = 1$ for all m .

Resulting estimates for the full data truths for locations 1, 7, and 8 are displayed in Table 3.3. The FPR estimates $\{\hat{f}_{1,m}, \hat{f}_{2,m}\}$, $m = 1$ (Mumbai), \dots , 10 (Ouagadougou), are almost identical to their full data values for all of the locations. Although the estimates for the TPRs are not exceptionally close to the full data values, they are closer

than those estimates obtained from examining the centers individually. In addition to obtaining better estimates, we observe smaller standard deviations in the hierarchical model. To demonstrate the improvement in the estimates from the hierarchical model, we include the posterior densities for Mumbai's TPRs and FPRs in Figure 3.5 and this center's prevalences and dependency factors in Figure 3.6. Convergence is also observed for all locations in the hierarchical model.

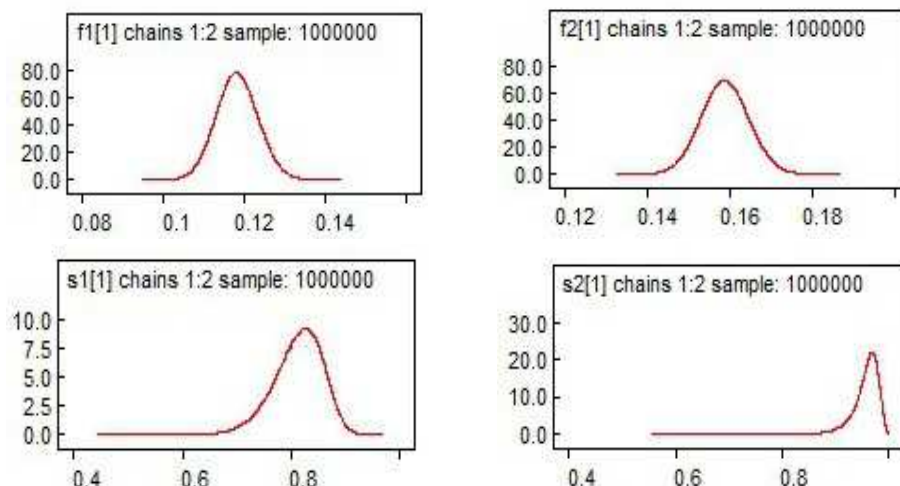


Figure 3.5. Density plots for Mumbai's TPRs and FPRs.

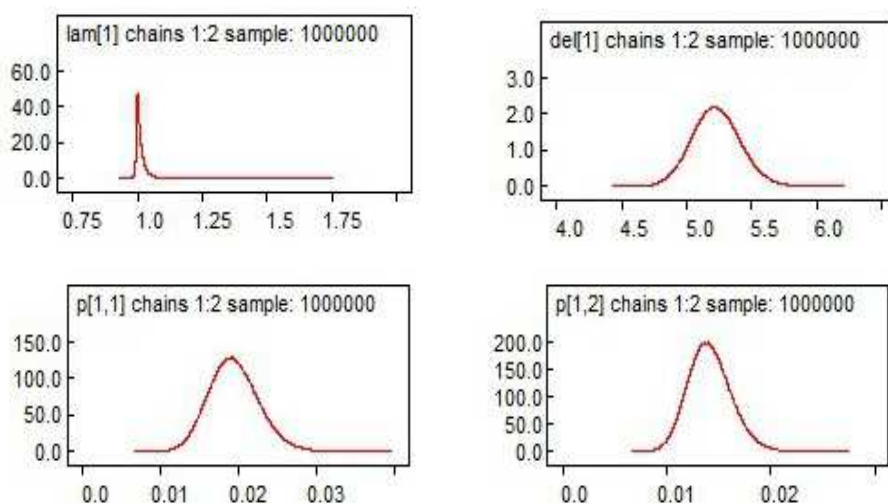


Figure 3.6. Density plots for Mumbai's dependency factors and prevalences.

3.4.3 Test Accuracy for CIN2+

The forest plots in Figures 3.7 and 3.8 shows the variation of the TPRs and FPRs for CIN2+. The TPR of VILI varied between 0.935 in Jaipur (CI: 0.840 to 0.988) to 0.977 in Bamako (CI: 0.932 to 0.999). The FPR of VILI varied between 0.085 in Niamey (CI: 0.074 to 0.096) to 0.269 in Jaipur (CI: 0.257 to 0.280). Results for VIA are also available for the interested reader from the author.

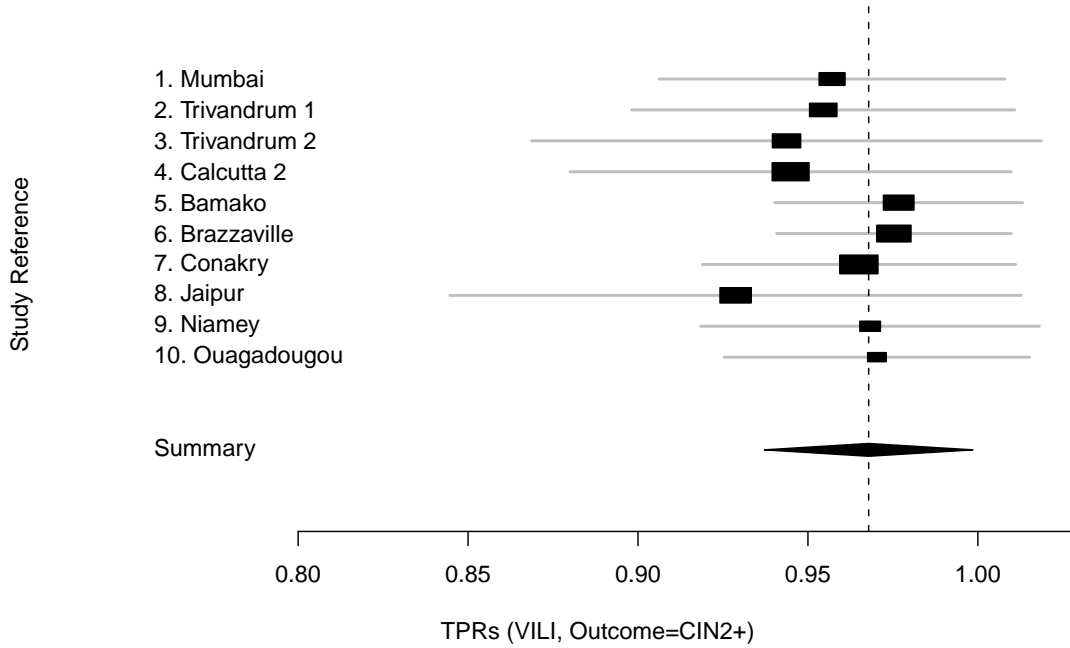


Figure 3.7. Forest plot for the TPRs for VILI.

The pooled estimates for the TPR of VIA and VILI are 0.871 and 0.968, respectively. Similarly, the pooled estimates for the FPR of VIA and VILI are 0.134 and 0.146, respectively. The pooled prevalences of CIN2+ for women age 45 and older was 0.024 and was 0.005 for women under 45 years old. We further compare the overall TPR and FPR between the tests by estimating the difference (Test 1 – Test 2) between the logits of the TPRs and FPRs. Thus, the estimate for the logarithm of the odds ratio for the difference in TPRs is -1.500 (CI: -2.655 to -0.578), demonstrating a significantly greater sensitivity for VILI than that observed for VIA. This equates to an

odds ratio of 0.223 (CI: 0.070 to 0.561). However, no difference in the tests' FPRs can be distinguished as seen by the estimate -0.088 (CI: -0.705 to 0.539) for the logarithm of the odds ratio for the difference in FPRs. This converts to an odds ratio of 0.916 (CI: 0.494 to 1.714).

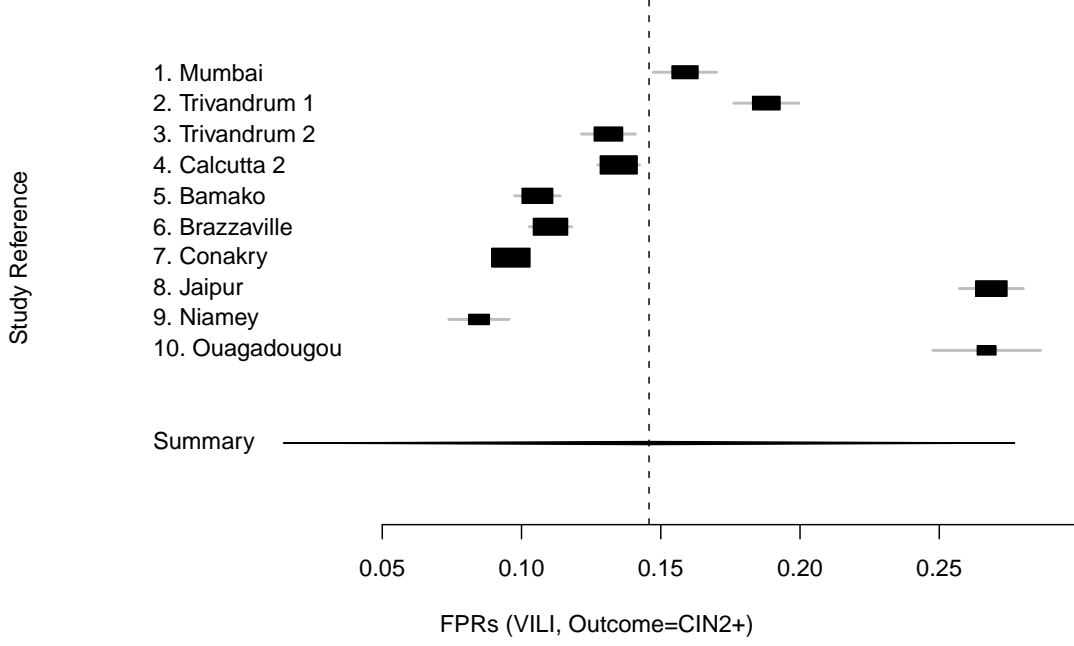


Figure 3.8. Forest plot for the FPRs for VILI.

3.5 A Simulation Example

We perform a simulation study based on the sampling plan of the cervical cancer example and estimates obtained from the model. We generate 1000 data sets having ten locations each. For each generated data set, we use a sample size of 1000 for each of the two populations. Hence, $T_{g,m} = 1000$ for each group g and location m , where $g = 1, 2$ and $m = 1, \dots, 10$. We use the following parameter configuration for the hyperpriors of $\beta_{0g,m}$,

$$\begin{aligned}\mu_{\beta_0} &= -2.5 \\ \sigma_{\beta_0} &= 0.2.\end{aligned}$$

Additionally, set $\beta_1 = 2$. We define the standard deviations of the random effects components $e_{q(m)}$ as $\sigma_q = 0.5$ for all q , $q = 1, \dots, 4$. We set the overall mean components for the log-odds of the TPRs and FPRs be

$$\begin{aligned}\gamma_1 &= 1.9 \\ \gamma_2 &= 3.4 \\ \gamma_3 &= -1.9 \\ \gamma_4 &= -1.8.\end{aligned}$$

From the previous configurations, we generate each of the random effects components from a normal distribution with mean zero and standard deviation σ_q , $e_{q(m)} \sim N(0, \sigma_q^2)$. TPRs and FPRs for each of the 10 locations are obtained from

$$\begin{aligned}s_{1,m} &= \frac{e^{\gamma_1 + e_{1(m)}}}{1 + e^{\gamma_1 + e_{1(m)}}} \\ s_{2,m} &= \frac{e^{\gamma_2 + e_{2(m)}}}{1 + e^{\gamma_2 + e_{2(m)}}} \\ f_{1,m} &= \frac{e^{\gamma_3 + e_{3(m)}}}{1 + e^{\gamma_3 + e_{3(m)}}} \\ f_{2,m} &= \frac{e^{\gamma_4 + e_{4(m)}}}{1 + e^{\gamma_4 + e_{4(m)}}}.\end{aligned}$$

From the TPR and FPR estimates, we find the minimum and maximum values for λ_m and δ_m for location m and generate these dependency factors from a uniform distribution,

$$\begin{aligned}\lambda_m &= U(\alpha_\lambda, \beta_\lambda) \\ \delta_m &= U(\alpha_\delta, \beta_\delta),\end{aligned}$$

where $\alpha_\lambda, \alpha_\delta, \beta_\lambda$, and β_δ are defined in Section 3. The parameter $\beta_{0g,m}$ is also generated from a normal distribution with mean μ_{β_0} and standard deviation σ_{β_0} such that $\beta_{0g,m} \sim N(\mu_{\beta_0}, \sigma_{\beta_0}^2)$. Then, we generate multinomial data using the log-likelihood provided in Section 3 for each group of the ten locations for every simulated data set.

For the study, we find average coverages and widths of approximate 95% credible intervals and obtain average bias for each parameter. Results are shown in Tables B.9-B.12 of Appendix B. For TPRs in Table B.9, we observe average interval widths ranging from 0.0640 to 0.0820, average coverages from 0.938 to 0.962, and average biases

from -0.0031 to 0.0004 . For FPRs in Table B.10, average interval widths ranging from 0.0331 to 0.0350 , average coverages from 0.938 to 0.967 , and average biases from -0.0010 to 0.0095 . Prevalences in Table B.11 have average interval widths ranging from 0.0295 to 0.0635 , average coverages from 0.938 to 0.964 , and average biases from 0.0006 to 0.0180 . The dependency factors, λ_m and δ_m , take on values in the ranges $(1.0163, 1.0174)$ and $(3.0415, 3.1634)$, respectively. Average interval coverages for the parameters were all well above 93% coverage.

3.6 Conclusions

We presented a hierarchical Bayesian logit framework for modeling population prevalences for two distinct groups and the TPRs and FPRs of two dichotomous screening tests, which can be used when subjects test negative for all screening tests and are not verified with a gold standard. Convergence of these estimates to their true full data value was rarely observed when examining the locations individually. However, when we consider the pooled hierarchical Bayesian logit model with random effects, estimates for the FPRs are nearly exact. Furthermore, estimates for the TPRs are more similar to their true values with smaller standard deviations, demonstrating the superiority of the hierarchical Bayesian logit model. We found a statistical difference in the TPRs between VIA and VILI with VILI having a larger overall TPR but found no statistically significant difference in the FPRs of the tests. Although, a few of the TPRs are fairly different from the single site estimates, the simulation study in Section 5 further demonstrates the accuracy and usefulness of the hierarchical model.

CHAPTER FOUR

A Power Analysis for Multivariate Normality Tests and Multiple Imputation Procedures in the Presence of a Monotone Missing Data Pattern

4.1 Introduction

For many research settings, it is not uncommon to encounter multivariate response data with missing values for one or more variables. Many statistical procedures implemented in these settings require a multivariate normality assumption, which is seldom tested despite the numerous tests that have been proposed and are available for detecting departures from multivariate normality (Farrell, Salibian-Barrera, and Naczkz, 2006). Furthermore, a missing value for any one variable may exclude all variable values for that observation, possibly causing a substantial loss of information in the analysis.

Often a missing value is replaced with some imputed value. The procedure of replacing missing values with plausible values and applying complete-data analysis methods to make inferences is referred to as imputation of missing data. There are three types of uncertainties, as noted by Zhang (2003), involved in the imputation process: (1) uncertainty due to modeling the joint distribution of the response variables and the missingness indicators, (2) uncertainty from sampling from a given imputation model assuming that the observed-data and the model parameters' values are known, and (3) uncertainty about the values of the model parameters and for selecting the imputation model.

One method we might consider to impute the missing values involves taking a random sample with replacement from the observed values. However, this method frequently underestimates the variances of the population parameter estimates, and thus, does not sufficiently reflect the extra variability resulting from the imputation. A method more appropriate is multiple imputation, which is suitable for handling nonre-

sponse data while also incorporating uncertainty into the imputation model (Srivastava and Dolatabadi, 2009).

The first step of multiple imputation involves creating $G > 1$ data sets by replacing missing values with random draws from an imputation model given the observed values. Then, each of the G data sets are analyzed using complete-data procedures, resulting in G total analyses. Imputing missing data in this way allows us to conveniently use readily available complete-data software to perform analyses on the newly imputed data sets. Finally, these results are combined into a single analysis for inference. Thus, multiple imputation may result in more valid statistical inferences when we are presented with nonresponse data.

We consider several multiple imputation procedures and apply them to monotone missing data having a pattern described in Section 2. After applying an imputation procedure and obtaining an imputed data set, we apply a test for multivariate normality. There are more than 50 multivariate normality tests available, and an ample amount of reviews on these tests. Henze (2002) provides an extensive review on multivariate normality tests that are affine invariant and consistent against all alternatives. Mecklin and Mundfrom (2004) provide a comprehensive review on multiple tests including goodness-of-fit tests to the multivariate normal distribution, skewness and kurtosis tests, and consistent tests.

Despite the large number of approaches to testing multivariate normality and developing multiple imputation procedures, there does not appear to be any studies involving the power of these tests after performing different multiple imputation schemes when encountered with monotone missing data. Therefore, we examine three multiple imputation schemes and four tests for multivariate normality for monotone incomplete data and perform a power comparison for the different combinations of tests and imputation procedures. We do this for multivariate normal data and data generated from a multivariate skewed t -distribution with different degrees of freedom and skewness parameter values.

For the remaining sections of this chapter, we discuss the multiple imputation schemes and multivariate normality tests under consideration in Sections 3 and 4, respectively. In Section 3, we discuss the use of the EM algorithm and data augmentation for incomplete multivariate normal data under the assumption of non-informative and informative priors. We further discuss an iterative regression imputation procedure and a predictive mean matching procedure. In Section 4, we introduce several multivariate normality tests, including one proposed by Henze and Zirkler (1990), an omnibus test using Mardia's tests for multivariate skewness and kurtosis, a fairly new test based on the concept of Newton's potential energy, and a generalization of the Shapiro-Wilk test for normality. In Section 5, we provide a detailed explanation of our simulation study procedure and the results, and we provide conclusions in Section 6.

4.2 Multiple Imputation Schemes

Let $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_N$ be N mutually independent random $d \times 1$ vectors from a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Now consider a case in which we observe a two-step monotone incomplete data pattern (Yamada, Romer, and Richards, 2010). Suppose we have N observations with $d = p + q$ total population characteristics but only observe $n < N$ complete observations for d of the population characteristics while observing only $q = d - p$ population characteristics on the last $N - n$ observations. Thus, our data takes the form

$$\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{Y}_1 \end{pmatrix} \quad \begin{pmatrix} \mathbf{X}_2 \\ \mathbf{Y}_2 \end{pmatrix} \quad \dots \quad \begin{pmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{pmatrix} \quad \mathbf{Y}_{n+1} \quad \mathbf{Y}_{n+2} \quad \dots \quad \mathbf{Y}_N,$$

where each vector \mathbf{Y}_i , $1 \leq i \leq N$, is $q \times 1$ and each vector \mathbf{X}_j , $1 \leq j \leq n$, is $p \times 1$. We consider several imputation procedures for the above missing data pattern.

A single imputation scheme replaces a missing value in a data set with a single randomly imputed value. However, a multiple imputation procedure replaces each missing value in a data set with several imputed values, incorporating uncertainty into

our imputation model. Multiple imputation typically creates several imputed values for each missing value, where each is predicted from a slightly different model. Then, each of the imputed data sets are analyzed using standard complete-data procedures, and the multiple results are pooled into a single analysis to make inferences (Ganjali and Ranji, 2008). The differences between these imputed data sets reflect our uncertainty of the missing values. We discuss several multiple imputation schemes that will be implemented in our simulation study.

4.2.1 EM Algorithm

We introduce the EM algorithm as a simple iterative algorithm that will converge to a mode in the likelihood function of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Let \mathbf{Z} be $N \times d$ and partition \mathbf{Z} such that $\mathbf{Z} = (\mathbf{Z}_{mis}, \mathbf{Z}_{obs})$, where \mathbf{Z}_{obs} represents all observed values and \mathbf{Z}_{mis} represents all missing values. Note that \mathbf{Z}_{obs} can be further represented by observed values such that $\mathbf{Z}_{obs} = (\mathbf{z}_{obs,1}, \dots, \mathbf{z}_{obs,N})$. Further define $\boldsymbol{\mu}^{(t)}$ and $\boldsymbol{\Sigma}^{(t)}$ to be the parameter estimates for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively, at iteration t . We use the sample moments as starting values for $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\Sigma}^{(1)}$ and run the EM algorithm until the change in the log-likelihood is less than some specified tolerance (Reiter, Kohnen, Karr, Lin, and Sanil, 2004).

The EM algorithm involves cycling between an E-step and an M-step. For the E-step, we calculate

$$E \left(\sum_{i=1}^N z_{ij} | \mathbf{Z}_{obs}, \boldsymbol{\mu}^{(t)}, \boldsymbol{\Sigma}^{(t)} \right) = \sum_{i=1}^N z_{ij}^{(t)}, \quad j = 1, \dots, d,$$

where

$$z_{ij} = \begin{cases} z_{ij} & \text{if } z_{ij} \text{ is observed} \\ \hat{z}_{ij} & \text{if } z_{ij} \text{ is missing,} \end{cases}$$

and $\hat{z}_{ij} = E(z_{ij} | \mathbf{Z}_{obs}, \boldsymbol{\mu}^{(t)}, \boldsymbol{\Sigma}^{(t)})$. Reiter et al. (2004) show that \hat{z}_{ij} can be calculated from the predicted value in the regression of z_{ij} on \mathbf{z}_{Bi} , given by

$$\hat{z}_{ij} = \mu_j^{(t)} + \boldsymbol{\Sigma}_{j,Bi}^{(t)} \left(\boldsymbol{\Sigma}_{Bi,Bi}^{(t)} \right)^{-1} \left(\mathbf{z}_{obs,i} - \boldsymbol{\mu}_{Bi}^{(t)} \right),$$

where

- $\mu_j^{(t)}$ = the element of $\boldsymbol{\mu}^{(t)}$ for j ,
- $\boldsymbol{\mu}_{Bi}^{(t)}$ = the vector of elements of $\boldsymbol{\mu}$ for all variables in $\mathbf{z}_{obs,i}$,
- $\Sigma_{j,Bi}^{(t)}$ = the elements of $\Sigma^{(t)}$ corresponding to the covariances between variable j and all the variables in $\mathbf{z}_{obs,i}$,
- $\Sigma_{Bi,Bi}^{(t)}$ = the elements of $\Sigma^{(t)}$ corresponding to the covariance matrix of all variables in $\mathbf{z}_{obs,i}$.

For the E-step, we also want to calculate

$$E \left(\sum_{i=1}^N z_{ij} z_{ik} | \mathbf{Z}_{obs}, \boldsymbol{\mu}^{(t)}, \Sigma^{(t)} \right) = \sum_{i=1}^N z_{ij}^{(t)} = \sum_{i=1}^N z_{ij}^{(t)} z_{ik}^{(t)} + c_{jki}^{(t)}, \quad j, k = 1, \dots, d,$$

where $c_{jki}^{(t)} = 0$ when z_{ij} or z_{ik} is observed and $c_{jki}^{(t)} = \text{Cov}(z_{ij}, z_{ik} | \mathbf{Z}_{obs}, \boldsymbol{\mu}^{(t)}, \Sigma^{(t)})$ when z_{ij} and z_{ik} are missing. Reiter et al. (2004) have shown that the covariances can be obtained from the off-diagonal element in the 2×2 matrix

$$\Sigma_{jk}^{(t)} - \Sigma_{jk,Bi}^{(t)} \left(\Sigma_{Bi,Bi}^{(t)} \right)^{-1} \Sigma_{jk,Bi}^{(t)},$$

where

- $\Sigma_{jk}^{(t)}$ = the covariance submatrix of $\Sigma^{(t)}$ for variables j and k
- $\Sigma_{jk,Bi}^{(t)}$ = the submatrix of $\Sigma^{(t)}$ representing the covariances between (z_{ij}, z_{ik}) and the variables in $\mathbf{z}_{obs,i}$.

For the M-step, we maximize the likelihood after completing the data with the expectations from the E-step to obtain new estimates, $\boldsymbol{\mu}^{(t+1)}$ and $\Sigma^{(t+1)}$. Thus, for each j and k ,

$$\begin{aligned} \mu_j^{(t+1)} &= \frac{1}{N} \sum_{i=1}^N z_{ij}^{(t)} \\ \sigma_{jk}^{(t+1)} &= \frac{1}{N} \sum_{i=1}^N \left(z_{ij} - \mu_j^{(t+1)} \right) \left(z_{ik} - \mu_k^{(t+1)} \right) + \frac{1}{N} \sum_{i=1}^N c_{jki}^{(t)}. \end{aligned}$$

4.2.2 Data Augmentation

Data augmentation for incomplete multivariate normal data is an iterative process consisting of an imputation (I-step) stage and a posterior (P-step) stage (Ganjali

and Ranji, 2008). In the I-step for iteration t , we perform independent simulations of random normal vectors for each row of \mathbf{Z} with missing data given the observed data and new parameter estimates,

$$\mathbf{Z}_{mis}^{(t+1)} \sim P(\mathbf{Z}_{mis} | \mathbf{Z}_{obs}, \boldsymbol{\mu}^{(t)}, \boldsymbol{\Sigma}^{(t)}).$$

We use the estimates from the EM algorithm for our initial estimate for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in the I-step.

Suppose we can partition $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ accordingly,

$$\begin{aligned} \boldsymbol{\mu} &= (\boldsymbol{\mu}'_1, \boldsymbol{\mu}'_2)' \\ \boldsymbol{\Sigma} &= \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}'_{12} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \end{aligned}$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ represent the mean vectors of \mathbf{Z}_{mis} and \mathbf{Z}_{obs} , respectively, and their corresponding covariance submatrices are given by the above partitioning of $\boldsymbol{\Sigma}$. The covariance matrix for \mathbf{Z}_{mis} is $\boldsymbol{\Sigma}_{11}$, \mathbf{Z}_{obs} has the covariance matrix $\boldsymbol{\Sigma}_{22}$, and the covariance matrix between \mathbf{Z}_{mis} and \mathbf{Z}_{obs} is $\boldsymbol{\Sigma}_{12}$. Then, the conditional distribution of \mathbf{Z}_{mis} given $\mathbf{Z}_{obs} = \mathbf{c}_i$ is multivariate normal with mean vector $\boldsymbol{\mu}_{1.2}$ and covariance $\boldsymbol{\Sigma}_{11.2}$,

$$\mathbf{Z}_{mis} | \mathbf{Z}_{obs} = \mathbf{c}_i \sim N_d(\boldsymbol{\mu}_{1.2}, \boldsymbol{\Sigma}_{11.2}),$$

where

$$\begin{aligned} \boldsymbol{\mu}_{1.2} &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{c}_i - \boldsymbol{\mu}_2) \\ \boldsymbol{\Sigma}_{11.2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}'_{12}. \end{aligned}$$

For the P-step, we assume some prior structure, such as an informative conjugate prior distribution for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ or Jeffreys prior, and obtain the resulting posterior distributions. Let us consider both cases. Let $\tilde{\mathbf{Z}}$ denote \mathbf{Z} containing the imputed values from the I-step. Then, by Schafer (1999), using Jeffreys prior structure yields the posterior distributions

$$\begin{aligned} \boldsymbol{\Sigma}^{(t+1)} | \tilde{\mathbf{Z}} &\sim W^{-1} \left(N-1, \frac{1}{N} \mathbf{S}_{\tilde{\mathbf{Z}}}^{-1} \right) \\ \boldsymbol{\mu}^{(t+1)} | \left(\boldsymbol{\Sigma}^{(t+1)}, \tilde{\mathbf{Z}} \right) &\sim N_d \left(\bar{\mathbf{Z}}, \frac{1}{N} \boldsymbol{\Sigma}^{(t+1)} \right). \end{aligned}$$

Using the informative priors for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and specifying degrees of freedom m and precision matrix $\boldsymbol{\Lambda}$ in the inverse Wishart prior for $\boldsymbol{\Sigma}$ and a mean vector $\boldsymbol{\mu}_0$ and constant n_0 in the multivariate normal prior for $\boldsymbol{\mu}$,

$$\begin{aligned}\boldsymbol{\Sigma} &\sim W^{-1}(m, \boldsymbol{\Lambda}) \\ \boldsymbol{\mu}|\boldsymbol{\Sigma} &\sim N_d\left(\boldsymbol{\mu}_0, \frac{1}{n_0}\boldsymbol{\Sigma}\right),\end{aligned}$$

we obtain the posterior distributions,

$$\begin{aligned}\boldsymbol{\Sigma}^{(t+1)}|\tilde{\mathbf{Z}} &\sim W^{-1}\left(N+m, [N\mathbf{S}_Z + \boldsymbol{\Lambda}^{-1} + \mathbf{S}_{n'}]^{-1}\right) \\ \boldsymbol{\mu}^{(t+1)}|\left(\boldsymbol{\Sigma}^{(t+1)}, \tilde{\mathbf{Z}}\right) &\sim N_d\left(\frac{1}{N+n_0}(N\bar{\mathbf{Z}} + n_0\boldsymbol{\mu}_0), \frac{1}{N+n_0}\boldsymbol{\Sigma}^{(t+1)}\right),\end{aligned}$$

where

$$\mathbf{S}_{n'} = \frac{Nn_0}{N+n_0}(\bar{\mathbf{Z}} - \boldsymbol{\mu}_0)(\bar{\mathbf{Z}} - \boldsymbol{\mu}_0)'.$$

Note, using the parameterization of $m = -1$, $n_0 = 0$, and $\boldsymbol{\Lambda} = \mathbf{0}_{d \times d}$ results in the posterior distributions obtained using Jeffreys prior. For our simulation study, we will only consider the case of using Jeffreys prior.

4.2.3 Iterative Regression Imputation

To use an iterative regression approach to obtain estimates for the missing data, we need to first obtain initial values for the missing data. Our initial guess will consist of a convenient approach in which we impute values for the missing data of each variable by randomly selecting from observed values for that variable. After obtaining these values as a starting point, we will fit a regression to the observed values and impute the resulting predictions into the missing values. Denote the columns of \mathbf{Z} containing missingness by

$$(\mathbf{Z}_{mis}, \mathbf{Z}_{obs}) = (\mathbf{z}_{mis,1}, \mathbf{z}_{mis,2}, \dots, \mathbf{z}_{mis,p}, \mathbf{Z}_{obs}).$$

Then, the iterative regression approach involves imputing each variable vector with missingness given the values of the others. That is, we impute $\mathbf{z}_{mis,1}$ given the observed and missing estimates $(\mathbf{z}_{mis,2}, \mathbf{z}_{mis,3}, \dots, \mathbf{z}_{mis,p}, \mathbf{Z}_{obs})$. Similarly, we impute $\mathbf{z}_{mis,2}$ given

the values $(z_{mis,1}, z_{mis,3}, \dots, z_{mis,p}, Z_{obs})$ using the imputed estimates obtained from the previous imputation for $z_{mis,1}$. Once we have imputed the p^{th} vector, we continue the iterative scheme. This imputation involving a regression model also allows us to incorporate our uncertainty about the regression coefficients as they are changing with each imputation. However, some data sets may contain hundreds of variables. Thus, for imputation procedures, it would be tedious and unnecessary to include all variables into the model. Hence, an adequate subset of the variables should be considered.

4.2.4 Predictive Mean Matching

The predictive mean matching method is an imputation method similar to the regression method. However, for each missing value, it imputes a value randomly from a set of observed values whose predicted values are closest to the predicted value for the missing value from the simulated regression model. The predictive mean matching method ensures that imputed values are plausible and may be more appropriate than the regression method if the normality assumption is violated. We consider the predictive mean matching procedure along with data augmentation using Jeffreys priors and the iterative regression imputation for our power study.

4.3 Test Statistics

Many tests have been proposed for testing multivariate normality. Some tests are extremely popular and have been implemented for years, such as the Shapiro-Wilk test of normality and Mardia's tests of multivariate skewness and kurtosis. However, numerous tests for multivariate normality since the development of these have been proposed, such as the test of multivariate normality proposed by Henze and Zirkler (1990) and generalizations of the Shapiro-Wilk test for multivariate normal data. Today, new tests for multivariate normality are continuing to surface, such as the ϵ -test proposed by Szekely and Rizzo (2005). We review several of the tests for multivariate normality that will be considered in our simulation study.

4.3.1 The $T_{N,\beta}$ Test Statistic

We first consider a consistent and affine invariant test proposed by Henze and Zirkler (1990) for assessing multivariate normality. Let the sample mean vector and sample covariance matrix for the observations $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_N$ be denoted by $\bar{\mathbf{Z}}_N$ and \mathbf{S}_Z , respectively, and define $D_{N,\beta}$ as

$$D_{N,\beta} = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \exp \left[-\frac{\beta^2}{2} \|\mathbf{Z}_i - \mathbf{Z}_j\|^2 \right] - \frac{2}{N} (1 + \beta^2)^{-d/2} \sum_{i=1}^N \exp \left[-\frac{\beta^2}{2(1 + \beta^2)} \|\mathbf{Z}_i\|^2 \right] + (1 + 2\beta^2)^{-d/2},$$

where $\|\mathbf{Z}_i - \mathbf{Z}_j\|^2 = (\mathbf{Z}_i - \mathbf{Z}_j)' \mathbf{S}_Z^{-1} (\mathbf{Z}_i - \mathbf{Z}_j)$ and $\|\mathbf{Z}_i\|^2 = (\mathbf{Z}_i - \bar{\mathbf{Z}}_N)' \mathbf{S}_Z^{-1} (\mathbf{Z}_i - \bar{\mathbf{Z}}_N)$.

Also, the parameter β is a function of both d and N ,

$$\beta = \beta_d(N) = \frac{1}{\sqrt{2}} \left(\frac{2d+1}{4} \right)^{\frac{1}{d+4}} N^{\frac{1}{d+4}}.$$

Then, the test statistic proposed by Henze and Zirkler (1990), $T_{N,\beta}$, is given by

$$T_{N,\beta} = N * (4I_{\{\mathbf{S}_Z \text{ singular}\}} + D_{N,\beta} I_{\{\mathbf{S}_Z \text{ nonsingular}\}}),$$

where $I\{\cdot\}$ is the indicator function. Note that $D_{N,\beta}$ is undefined if \mathbf{S}_Z is singular and is replaced by its maximum possible value in the formula for $T_{N,\beta}$.

Under the assumption of multivariate normality, the distribution of $T_{N,\beta}$ can be approximated by a lognormal distribution if we equate the first two moments, $\mu_{\beta,d} = E[T_{N,\beta}(d)]$ and $\sigma_{\beta,d}^2 = Var[T_{N,\beta}(d)]$, where

$$\begin{aligned} E[T_{N,\beta}(d)] &= 1 - (1 + 2\beta^2)^{-d/2} \left[1 + \frac{d\beta^2}{1 + 2\beta^2} + \frac{d(d+2)\beta^4}{2(1 + 2\beta^2)^2} \right], \\ Var[T_{N,\beta}(d)] &= 2(1 + 4\beta^2)^{-d/2} + 2(1 + 2\beta^2)^{-d} \left[1 + \frac{2d\beta^4}{(1 + 2\beta^2)^2} + \frac{3d(d+2)\beta^8}{4(1 + 2\beta^2)^4} \right] \\ &\quad - 4w(\beta)^{-d/2} \left[1 + \frac{3d\beta^4}{2w(\beta)} + \frac{d(d+2)\beta^8}{2w(\beta)^2} \right], \end{aligned}$$

and $w(\beta) = (1 + \beta^2)(1 + 3\beta^2)$. This results in

$$q_{\beta,d}(a) = \mu_{\beta,d} \left(1 + \frac{\sigma_{\beta,d}^2}{\mu_{\beta,d}^2} \right)^{-1/2} \exp \left[\Phi^{-1}(a) \left(\log \left[1 + \frac{\sigma_{\beta,d}^2}{\mu_{\beta,d}^2} \right] \right)^{1/2} \right]$$

as an approximation to the α -quantile of $T_{N,\beta}(d)$, where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal distribution function. Henze and Zirkler (1990) have shown that there exists a strong agreement between the empirical quantiles of $T_{N,\beta}(d)$ and $q_{\beta,d}(1-\alpha)$ as an approximate critical value for a level α test when the sample size N is at least 20.

4.3.2 The Omnibus Mardia Test Statistic

An omnibus test for normality using measures of skewness and kurtosis is discussed in Doornik and Hansen (2008) and Mecklin and Mundfrom (2004). We may represent the measures of multivariate skewness and kurtosis proposed by Mardia (1975) as

$$\begin{aligned} b_{1,d} &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N g_{ij}^3 \\ b_{2,d} &= \frac{1}{N} \sum_{i=1}^N g_{ii}^4, \end{aligned}$$

respectively, where

$$g_{ij} = (\mathbf{Z}_i - \bar{\mathbf{Z}}_N)' \mathbf{S}_Z^{-1} (\mathbf{Z}_j - \bar{\mathbf{Z}}_N).$$

Then, an omnibus test based on these measures has the statistic

$$M_d = \frac{Nb_{1d}}{6} + \frac{N[b_{2d} - d(d+2)]^2}{8d(d+2)},$$

which can be approximated by a χ^2 distribution with $1 + \frac{1}{6}d(d+1)(d+2)$ degrees of freedom.

4.3.3 The ϵ -statistic

The ϵ -test (energy) for multivariate normality proposed by Szekely and Rizzo (2005) is affine invariant and consistent against all nonnormal fixed alternatives. It was based on the concept of Newton's potential energy and uses Euclidean distances between sampled elements. The test applies to any multivariate distribution with finite second moments. Suppose $\mathbf{Z}_1, \dots, \mathbf{Z}_N$ is a random sample from a d -variate population with distribution F , and that $\mathbf{z}_1, \dots, \mathbf{z}_N$ are the observed values of the random sample.

Then, the statistic to test $H_0 : F = F_0$ vs. $H_1 : F \neq F_0$ is given by

$$\epsilon_{N,d} = N \left(\frac{2}{N} \sum_{i=1}^N E \|\mathbf{z}_i - \mathbf{Z}\| - E \|\mathbf{Z} - \mathbf{Z}'\| - \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \|\mathbf{z}_i - \mathbf{z}_j\| \right),$$

where \mathbf{Z} and \mathbf{Z}' are independently and identically distributed with distribution F_0 .

If the hypothesized distribution is $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the statistic is

$$\epsilon_{N,d} = N \left(\frac{2}{N} \sum_{i=1}^N E \|\mathbf{v}_i - \mathbf{U}\| - E \|\mathbf{U} - \mathbf{U}'\| - \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \|\mathbf{v}_i - \mathbf{v}_j\| \right),$$

where \mathbf{v}_i is a transformation of the sample,

$$\mathbf{v}_i = \boldsymbol{\Sigma}^{-1/2}(\mathbf{z}_i - \boldsymbol{\mu}), i = 1, \dots, N,$$

and \mathbf{U} and \mathbf{U}' are independently and identically distributed $N_d(\mathbf{0}, \mathbf{I})$. The test statistic can be equivalently expressed as

$$\epsilon_{N,d} = N \left(\frac{2}{N} \sum_{i=1}^N E \|\mathbf{v}_i - \mathbf{U}\| - 2 \frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} - \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \|\mathbf{v}_i - \mathbf{v}_j\| \right),$$

where

$$E \|\mathbf{a} - \mathbf{U}\| = \frac{\sqrt{2} \Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} + \sqrt{\frac{2}{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k \|\mathbf{a}\|^{2(k+1)} \Gamma(\frac{d+1}{2}) \Gamma(k + \frac{3}{2})}{2k! 2^k (2k+1)(k+1) \Gamma(k + \frac{d}{2} + 1)}.$$

However, since the parameters are unknown, we estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ with the sample mean vector $\bar{\mathbf{Z}}$ and sample covariance matrix \mathbf{S}_Z so that the test is based on $\hat{\epsilon}_{N,d}$. Then the standardized sample vectors are represented by

$$\mathbf{v}_i = \mathbf{S}^{-1/2}(\mathbf{Z}_i - \bar{\mathbf{Z}}), i = 1, \dots, N.$$

The test is performed by parametric bootstrapping with $R = 200$ replicates. We reject multivariate normality for large values of $\epsilon_{N,d}$.

4.3.4 The Shapiro-Wilk Test Statistic

The fourth test for multivariate normality we consider is based on the Shapiro-Wilk univariate normality test and on an empirical standardization of the observations

(Villasenor-Alva and Gonzalez-Estrada, 2009). The Shapiro and Wilk statistic for testing univariate normality is

$$W_U = \frac{\sum_{i=1}^N a_i \mathbf{u}_{(i)}^2}{\sum_{i=1}^N (\mathbf{u}_i - \bar{\mathbf{u}})^2},$$

where $\mathbf{u}_{(i)}$ are univariate order statistics from a standard normal random sample of size N , and a_i are coefficients defined by Shapiro and Wilk (1965) as

$$\mathbf{a} = (a_1, \dots, a_N)' = \frac{\mathbf{m}' \mathbf{V}^{-1}}{(\mathbf{m}' \mathbf{V}^{-1} \mathbf{V}^{-1} \mathbf{m})^{1/2}},$$

where $\mathbf{m} = E(\mathbf{U})$ and $\mathbf{V} = \text{Cov}(\mathbf{U})$.

To test the hypothesis of $\mathbf{Z}_1, \dots, \mathbf{Z}_N$ sampled from $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, Villasenor-Alva and Gonzalez-Estrada (2009) propose the test statistic

$$W^* = \frac{1}{d} \sum_{k=1}^d W_{V_k},$$

where W_{V_k} is the Shapiro-Wilk statistic evaluated on the k^{th} row vector of the transformation,

$$\mathbf{V}_k = \mathbf{S}_Z^{-1/2}(\mathbf{Z}_k - \bar{\mathbf{Z}}), \quad k = 1, \dots, d.$$

The critical values are approximated by a transformation of the univariate standard normal distribution, and the test rejects multivariate normality at a test size α if $W^* < c_{\alpha;N,d}$ such that

$$\alpha = P(W^* < c_{\alpha;N,d} \mid \text{multivariate normality holds}).$$

4.4 Simulation Design and Results

We perform a Monte Carlo power study for multivariate normality comparing the four multivariate tests listed in Section 3 after performing the three imputation procedures described in Section 2. We first generate 10,000 multivariate normal data sets having a monotone missing data pattern with mean vector $\boldsymbol{\mu} = \mathbf{0}$ and assuming an

intra-class covariance matrix for Σ , and then impute 50 times each. Once values have been imputed for a single data set, we calculate the mean and median test statistics for each imputation-test combination. We repeat this procedure for each of the 10,000 generated data sets in order to determine the critical value of the 12 imputation-test combinations for multivariate normality that results in a $\alpha = 0.10$ test. We obtain two sets of 12 critical values, with one set resulting from using median test statistics and the other from using mean test statistics. We do this for several different sample sizes and parameter configurations.

We consider the case for when we have $N = 100$ total observations with $n = 60$ observed observations having all $d = 10$ population characteristics. We further consider the values $p = 2$ (8% missing) and $p = 5$ (20% missing). We similarly consider the case of having the same proportion of missing data for a larger sample size. We also consider the cases for when we have $N = 200$ and $N = 300$ total observations. When considering $N = 200$ observations, we let $n = 120$ represent the number of observed values having all $d = 10$ population characteristics and then consider the values $p = 2$ (8% missing) and $p = 5$ (20% missing). Similarly, when considering $N = 300$ observations, we let $n = 180$ represent the number of observed values having all $d = 10$ population characteristics using the values $p = 2$ (8% missing) and $p = 5$ (20% missing).

We then calculate the empirical Type I error rates for the imputation-test combinations using an alternative distribution, the multivariate skewed t -distribution with skewness parameter vector $\boldsymbol{\nu}$ and df degrees of freedom. We compare the performance of the imputation procedures and tests for different degrees of freedom and skewness. In particular, we consider three cases: (1) no skewness with $df = 25$ (and $df = 30$ is also included), (2) a large number of degrees of freedom ($df = 200$) with skewness $\boldsymbol{\nu} = \mathbf{1000}$ of dimension 10, and (3) a combination of skewness and degrees of freedom with $df = 50$ and $\boldsymbol{\nu} = \mathbf{100}$ having dimension 10. Thus, we examine the performance of the imputation procedures and tests based on skewness and degrees of freedom both separately and combined.

Power was compared for a simulation of $M = 50,000$ multivariate t data sets with monotone missing data patterns for each of the different parameterizations under consideration. For each simulation, missing values were imputed 50 times. Then, the mean and median test statistics for each imputation-test combination is saved, resulting in 50,000 mean (and median) test statistics, where the 10% quantile is used to estimate power. Power (P) and its standard error (se_P) were both estimated, where the standard error can be approximated by

$$se_P = \left[\frac{P(1-P)}{M} \right]^{1/2}$$

and will have a maximum of $se_{\max} = 0.0022$.

4.5 Conclusions

The figures in Appendix D display the results of our power analysis study for different multivariate normality tests and multiple imputation procedures in the presence of a monotone missing data pattern. Exact values are shown in the tables found in Appendix E. The energy test outperformed all other tests. The second overall best-performing test appears to be that using the $T_{N,\beta}$ statistic. However, there are a few instances in which the multivariate Shapiro-Wilk test outperforms the test using the $T_{N,\beta}$ statistic for when the iterative regression and predictive mean matching methods were used. We observed this mainly for smaller sample sizes. Power was generally smallest for the multivariate skewed normal distribution and greatest for the multivariate skewed t -distribution. Also, when less missingness is encountered ($p = 2$), power for the tests were typically greater. Mardia's omnibus test performed poorly in all situations. There were also a few instances in which power differed by more than 0.01 when comparing results using the means and medians.

APPENDICES

APPENDIX A

Derivations

A.1 EM Algorithm

We use the EM algorithm to determine the MLEs for the full likelihood for τ_1 and τ_2 . Given the density $f(\mathbf{v}, \mathbf{z} | \boldsymbol{\theta})$ with complete (observed and missing) data (\mathbf{v}, \mathbf{z}) , where $\mathbf{v} = (\{x_{ijk}\}, \{y_{ijk}\}, M_1, M_2)$ for $(j, k) = \{(0, 1), (1, 0), (1, 1)\}$ denotes the observed data and $\mathbf{z} = (x_{100}, x_{000}, y_{100}, y_{000})$ denotes the missing data with parameter vector, $\boldsymbol{\theta} = (p_1, p_2, s_2, f_2, \tau_1, \tau_2, \lambda, \delta)'$. For the *E*-Step, we have

$$\begin{aligned} E(x_{100} | \mathbf{v}, \boldsymbol{\theta}^{(r)}) &= \frac{M_1 p_1 [1 - \tau_1 - 2s_2 + (\tau_1 + s_2)s_2\lambda]}{1 - p_1[\tau_1 + 2s_2 - (\tau_1 + s_2)s_2\lambda] - (1 - p_1)[\tau_2 + 2f_2 - (\tau_2 + f_2)f_2\delta]} \\ &\equiv x_{100}^{(r)}, \\ E(y_{100} | \mathbf{v}, \boldsymbol{\theta}^{(r)}) &= \frac{M_2 p_2 [1 - \tau_1 - 2s_2 + (\tau_1 + s_2)s_2\lambda]}{1 - p_2[\tau_1 + 2s_2 - (\tau_1 + s_2)s_2\lambda] - (1 - p_1)[\tau_2 + 2f_2 - (\tau_2 + f_2)f_2\delta]} \\ &\equiv y_{100}^{(r)}, \\ E(x_{000} | \mathbf{v}, \boldsymbol{\theta}^{(r)}) &= M_1 - x_{100}^{(r)} \equiv x_{000}^{(r)}, \\ E(y_{000} | \mathbf{v}, \boldsymbol{\theta}^{(r)}) &= M_2 - y_{100}^{(r)} \equiv y_{000}^{(r)}. \end{aligned}$$

After taking the partial derivatives, we obtain

$$\begin{aligned} p_1^{(r+1)} &= \frac{x_{111} + x_{101} + x_{110} + x_{100}^{(r)}}{T_1}, \\ p_2^{(r+1)} &= \frac{y_{111} + y_{101} + y_{110} + y_{100}^{(r)}}{T_2}, \\ s_1^{(r+1)} &= \frac{x_{111} + y_{111} + x_{110} + y_{110}}{x_{111} + y_{111} + x_{101} + y_{101} + x_{110} + y_{110} + x_{100}^{(r)} + y_{100}^{(r)}}, \\ s_2^{(r+1)} &= \frac{x_{111} + y_{111} + x_{101} + y_{101}}{x_{111} + y_{111} + x_{101} + y_{101} + x_{110} + y_{110} + x_{100}^{(r)} + y_{100}^{(r)}}, \\ f_1^{(r+1)} &= \frac{x_{011} + y_{011} + x_{010} + y_{010}}{x_{011} + y_{011} + x_{001} + y_{001} + x_{010} + y_{010} + x_{000}^{(r)} + y_{000}^{(r)}}, \\ f_2^{(r+1)} &= \frac{x_{011} + y_{011} + x_{001} + y_{001}}{x_{011} + y_{011} + x_{001} + y_{001} + x_{010} + y_{010} + x_{000}^{(r)} + y_{000}^{(r)}}. \end{aligned}$$

as new estimates for the prevalences, TPRs, and FPRs. Similarly, we obtain new estimates for the dependency factors,

$$\begin{aligned}\lambda^{(r+1)} &= \frac{(x_{111} + y_{111})(x_{111} + y_{111} + x_{101} + y_{101} + x_{110} + y_{110} + x_{100}^{(r)} + y_{100}^{(r)})}{(x_{111} + y_{111} + x_{101} + y_{101})(x_{111} + y_{111} + x_{110} + y_{110})}, \\ \delta^{(r+1)} &= \frac{(x_{011} + y_{011})(x_{011} + y_{011} + x_{001} + y_{001} + x_{010} + y_{010} + x_{000}^{(r)} + y_{000}^{(r)})}{(x_{011} + y_{011} + x_{001} + y_{001})(x_{011} + y_{011} + x_{010} + y_{010})}.\end{aligned}$$

We use each iteration of parameter estimates to obtain improved estimates for the differences in sensitivities and specificities,

$$\begin{aligned}\tau_1^{(r+1)} &= s_1^{(r+1)} - s_2^{(r+1)}, \\ \tau_2^{(r+1)} &= f_1^{(r+1)} - f_2^{(r+1)}.\end{aligned}$$

A.2 Full Conditional Posterior Distributions

The likelihood in Section 3 can be rewritten as

$$\begin{aligned}f(\mathbf{v}, \mathbf{z} \mid \boldsymbol{\theta}) &= A_1 A_2 [p_1 s_1 s_2 \lambda]^{x_{111}} [p_1 s_2 (1 - s_1 \lambda)]^{x_{101}} \\ &\quad \times [p_1 s_1 (1 - s_2 \lambda)]^{x_{110}} [(1 - p_1) f_1 f_2 \delta]^{x_{011}} \\ &\quad \times [(1 - p_1) f_2 (1 - f_1 \delta)]^{x_{001}} [(1 - p_1) f_1 (1 - f_2 \delta)]^{x_{010}} \\ &\quad \times [p_1 (1 - s_1 - s_2 + s_1 s_2 \lambda)]^{x_{100}} [(1 - p_1) (1 - f_1 - f_2 + f_1 f_2 \delta)]^{x_{000}} \\ &\quad \times [p_2 s_1 s_2 \lambda]^{y_{111}} [p_2 s_2 (1 - s_1 \lambda)]^{y_{101}} \\ &\quad \times [p_2 s_1 (1 - s_2 \lambda)]^{y_{110}} [(1 - p_2) f_1 f_2 \delta]^{y_{011}} \\ &\quad \times [(1 - p_2) f_2 (1 - f_1 \delta)]^{y_{001}} [(1 - p_2) f_1 (1 - f_2 \delta)]^{y_{010}} \\ &\quad \times [p_2 (1 - s_1 - s_2 + s_1 s_2 \lambda)]^{y_{100}} [(1 - p_2) (1 - f_1 - f_2 + f_1 f_2 \delta)]^{y_{000}}.\end{aligned}$$

Using the results from Appendix A.1, the full conditional posterior distributions for the latent data are

$$\begin{aligned}\pi(x_{100} \mid \cdot) &\sim \text{binom}(M_1, p_{x_{100}}), \\ \pi(y_{100} \mid \cdot) &\sim \text{binom}(M_2, p_{y_{100}}), \\ \pi(x_{000} \mid \cdot) &\sim \text{binom}(M_1, p_{x_{000}}), \\ \pi(y_{000} \mid \cdot) &\sim \text{binom}(M_2, p_{y_{000}}),\end{aligned}$$

where

$$\begin{aligned}
p_{x_{100}} &= \frac{p_1[1 - \tau_1 - 2s_2 + (\tau_1 + s_2)s_2\lambda]}{1 - p_1[\tau_1 + 2s_2 - (\tau_1 + s_2)s_2\lambda] - (1 - p_1)[\tau_2 + 2f_2 - (\tau_2 + f_2)f_2\delta]}, \\
p_{y_{100}} &= \frac{p_2[1 - \tau_1 - 2s_2 + (\tau_1 + s_2)s_2\lambda]}{1 - p_2[\tau_1 + 2s_2 - (\tau_1 + s_2)s_2\lambda] - (1 - p_2)[\tau_2 + 2f_2 - (\tau_2 + f_2)f_2\delta]}, \\
p_{x_{000}} &= \frac{(1 - p_1)[1 - \tau_2 - 2f_2 + (\tau_2 + f_2)f_2\delta]}{1 - p_1[\tau_1 + 2s_2 - (\tau_1 + s_2)s_2\lambda] - (1 - p_1)[\tau_2 + 2f_2 - (\tau_2 + f_2)f_2\delta]}, \\
p_{y_{000}} &= \frac{(1 - p_2)[1 - \tau_2 - 2f_2 + (\tau_2 + f_2)f_2\delta]}{1 - p_2[\tau_1 + 2s_2 - (\tau_1 + s_2)s_2\lambda] - (1 - p_2)[\tau_2 + 2f_2 - (\tau_2 + f_2)f_2\delta]}.
\end{aligned}$$

We obtain closed-form full conditional posterior distributions for the prevalences,

$$\begin{aligned}
\pi(p_1|\cdot) &\sim \text{beta}(\alpha_{p_1} + x_{111} + x_{101} + x_{110} + x_{100}, \beta_{p_1} + x_{011} + x_{001} + x_{010} + x_{000}), \\
\pi(p_2|\cdot) &\sim \text{beta}(\alpha_{p_2} + y_{111} + y_{101} + y_{110} + y_{100}, \beta_{p_2} + y_{011} + y_{001} + y_{010} + y_{000}),
\end{aligned}$$

where the notation $\theta|\cdot$ denotes θ given all other variables. The full conditional posterior distribution of the TPRs, FPRs and dependency factors are proportional to the joint distribution such that

$$\begin{aligned}
\pi(s_1|\cdot) &\propto s_1^{\alpha_{s_1} + x_{111} + x_{110} + y_{111} + y_{110} - 1} (1 - s_1\lambda)^{x_{101} + y_{101}} (1 - s_1 - s_2 + s_1s_2\lambda)^{x_{100} + y_{100}} \\
&\quad \times (1 - s_1)^{\beta_{s_1} - 1}, \\
\pi(s_2|\cdot) &\propto s_2^{\alpha_{s_2} + x_{111} + x_{101} + y_{111} + y_{101} - 1} (1 - s_2\lambda)^{x_{110} + y_{110}} (1 - s_1 - s_2 + s_1s_2\lambda)^{x_{100} + y_{100}} \\
&\quad \times (1 - s_2)^{\beta_{s_2} - 1}, \\
\pi(f_1|\cdot) &\propto f_1^{\alpha_{f_1} + x_{011} + x_{010} + y_{011} + y_{010} - 1} (1 - f_1\delta)^{x_{001} + y_{001}} (1 - f_1 - f_2 + f_1f_2\delta)^{x_{000} + y_{000}} \\
&\quad \times (1 - f_1)^{\beta_{f_1} - 1}, \\
\pi(f_2|\cdot) &\propto f_2^{\alpha_{f_2} + x_{011} + x_{001} + y_{011} + y_{001} - 1} (1 - f_2\delta)^{x_{010} + y_{010}} (1 - f_1 - f_2 + f_1f_2\delta)^{x_{000} + y_{000}} \\
&\quad \times (1 - f_2)^{\beta_{f_2} - 1}, \\
\pi(\lambda|\cdot) &\propto \lambda^{\alpha_\lambda + x_{111} + y_{111} - 1} (1 - s_1\lambda)^{x_{101} + y_{101}} (1 - s_2\lambda)^{x_{110} + y_{110}} \\
&\quad \times (1 - s_1 - s_2 + s_1s_2\lambda)^{x_{100} + y_{100}} e^{-\beta_\lambda \lambda} \mathbf{I}_{(\lambda_l, \lambda_u)}(\lambda), \\
\pi(\delta|\cdot) &\propto \delta^{\alpha_\delta + x_{011} + y_{011} - 1} (1 - f_1\delta)^{x_{001} + y_{001}} (1 - f_2\delta)^{x_{010} + y_{010}} \\
&\quad \times (1 - f_1 - f_2 + f_1f_2\delta)^{x_{000} + y_{000}} e^{-\beta_\delta \delta} \mathbf{I}_{(\delta_l, \delta_u)}(\delta).
\end{aligned}$$

APPENDIX B

Hierarchical Model Results

Table B.1. Initialized chains for γ .

	Center	Chain 1	Chain 2
1.	Mumbai	$\hat{\gamma} = (0.55, 1.10, -2.05, -1.75)'$	$\hat{\gamma} = (0.40, 1.00, -1.95, -1.60)'$
2.	Trivandrum 1	$\hat{\gamma} = (2.10, 3.40, -1.30, -1.50)'$	$\hat{\gamma} = (2.05, 3.30, -1.25, -1.40)'$
3.	Trivandrum 2	$\hat{\gamma} = (1.45, 1.45, -2.15, -1.95)'$	$\hat{\gamma} = (1.35, 1.35, -2.05, -1.85)'$
4.	Calcutta 2	$\hat{\gamma} = (1.05, 1.55, -2.20, -1.90)'$	$\hat{\gamma} = (0.95, 1.40, -2.05, -1.80)'$
5.	Bamako	$\hat{\gamma} = (1.40, 3.55, -2.35, -2.20)'$	$\hat{\gamma} = (1.30, 3.40, -2.25, -2.10)'$
6.	Brazzaville	$\hat{\gamma} = (1.50, 3.15, -1.25, -2.15)'$	$\hat{\gamma} = (1.35, 3.00, -1.10, -2.05)'$
7.	Conakry	$\hat{\gamma} = (2.40, 3.55, -2.80, -2.30)'$	$\hat{\gamma} = (2.25, 3.40, -2.65, -2.20)'$
8.	Jaipur	$\hat{\gamma} = (2.00, 2.00, -1.15, -1.05)'$	$\hat{\gamma} = (1.90, 1.90, -1.05, -0.95)'$
9.	Niamey	$\hat{\gamma} = (0.70, 2.25, -2.90, -2.45)'$	$\hat{\gamma} = (0.55, 2.15, -2.80, -2.35)'$
10.	Ouagadougou	$\hat{\gamma} = (2.25, 3.95, -1.10, -1.05)'$	$\hat{\gamma} = (2.15, 3.85, -1.00, -0.95)'$

Table B.2. Full data, individual, and pooled TPR estimates.

Parameter	Full Data	Single Site	Hierarchical
$s_{1,1}$	0.617	0.0837	0.8126
$s_{1,2}$	0.886	0.4764	0.9139
$s_{1,3}$	0.802	0.8839	0.9170
$s_{1,4}$	0.743	0.5074	0.8520
$s_{1,5}$	0.793	0.4212	0.8124
$s_{1,6}$	0.804	0.8192	0.8252
$s_{1,7}$	0.911	0.9011	0.8999
$s_{1,8}$	0.875	0.0504	0.8935
$s_{1,9}$	0.650	0.5898	0.7923
$s_{1,10}$	0.900	0.8161	0.8860
$s_{2,1}$	0.741	0.1006	0.9571
$s_{2,2}$	0.913	0.4919	0.9545
$s_{2,3}$	0.802	0.8838	0.9437
$s_{2,4}$	0.814	0.5589	0.9449
$s_{2,5}$	0.970	0.5141	0.9767
$s_{2,6}$	0.956	0.9737	0.9753
$s_{2,7}$	0.970	0.9612	0.9650
$s_{2,8}$	0.875	0.0504	0.9287
$s_{2,9}$	0.900	0.8065	0.9683
$s_{2,10}$	0.980	0.8878	0.9703

Table B.3. Full data, individual, and pooled FPR estimates.

Parameter	Full Data	Single Site	Hierarchical
$f_{1,1}$	0.119	0.4220	0.1182
$f_{1,2}$	0.219	0.4873	0.2185
$f_{1,3}$	0.110	0.1203	0.1096
$f_{1,4}$	0.107	0.1930	0.1070
$f_{1,5}$	0.092	0.2401	0.0919
$f_{1,6}$	0.234	0.2338	0.2336
$f_{1,7}$	0.062	0.0623	0.0626
$f_{1,8}$	0.249	0.7096	0.2483
$f_{1,9}$	0.055	0.0937	0.0563
$f_{1,10}$	0.258	0.2970	0.2556
$f_{2,1}$	0.160	0.5671	0.1587
$f_{2,2}$	0.189	0.4190	0.1879
$f_{2,3}$	0.131	0.1440	0.1312
$f_{2,4}$	0.135	0.2436	0.1349
$f_{2,5}$	0.105	0.2763	0.1057
$f_{2,6}$	0.110	0.1103	0.1104
$f_{2,7}$	0.096	0.0959	0.0962
$f_{2,8}$	0.270	0.7680	0.2687
$f_{2,9}$	0.084	0.0142	0.0847
$f_{2,10}$	0.269	0.3103	0.2670

Table B.4. Full, individual, and pooled prevalences and dependency factor estimates.

Parameter	Full Data	Single Site	Hierarchical
$p_{1,1}$	0.0247	0.6828	0.0194
$p_{2,1}$	0.0180	0.5927	0.0141
$p_{1,2}$	0.0392	0.3548	0.0378
$p_{2,2}$	0.0296	0.3847	0.0277
$p_{1,3}$	0.0201	0.0402	0.0169
$p_{2,3}$	0.0152	0.0368	0.0134
$p_{1,4}$	0.0096	0.1950	0.0100
$p_{2,4}$	0.0085	0.1856	0.0074
$p_{1,5}$	0.0344	0.3988	0.0355
$p_{2,5}$	0.0279	0.3505	0.0268
$p_{1,6}$	0.0773	0.0773	0.0746
$p_{2,6}$	0.0491	0.0478	0.0478
$p_{1,7}$	0.0357	0.0359	0.0316
$p_{2,7}$	0.0144	0.0147	0.0158
$p_{1,8}$	0.0125	0.6596	0.0140
$p_{2,8}$	0.0116	0.5934	0.0103
$p_{1,9}$	0.0127	0.0887	0.0112
$p_{2,9}$	0.0068	0.0846	0.0073
$p_{1,10}$	0.0314	0.0902	0.0322
$p_{2,10}$	0.0341	0.0732	0.0225
λ_1	—	39.7	1.011
λ_2	—	11.96	1.025
λ_3	—	2.69	1.031
λ_4	—	24.93	1.015
λ_5	—	12.69	1.018
λ_6	—	1.021	1.017
λ_7	—	1.028	1.021
λ_8	—	57.94	1.022
λ_9	—	11.72	1.013
λ_{10}	—	3.18	1.020
δ_1	—	2.003	5.227
δ_2	—	2.365	3.657
δ_3	—	6.155	6.300
δ_4	—	4.988	6.087
δ_5	—	4.248	6.457
δ_6	—	2.862	2.863
δ_7	—	8.92	8.894
δ_8	—	1.394	3.488
δ_9	—	10.15	10.85
δ_{10}	—	3.136	3.334

Table B.5. Resulting prevalences from pooled centers.

Parameter	Mean	SD	MC error	2.5%	median	97.5%
$p_{1,1}$	0.01937	0.003187	6.615E-6	0.01358	0.01921	0.02607
$p_{2,1}$	0.01412	0.002056	3.158E-6	0.01045	0.014	0.01849
$p_{1,2}$	0.03779	0.004278	1.045E-5	0.02986	0.03763	0.04664
$p_{2,2}$	0.0277	0.003065	7.421E-6	0.02217	0.02754	0.03416
$p_{1,3}$	0.01686	0.00277	7.919E-6	0.0118	0.01674	0.02266
$p_{2,3}$	0.01338	0.001992	4.979E-6	0.009876	0.01324	0.01767
$p_{1,4}$	0.01004	0.002027	5.061E-6	0.006363	0.009942	0.01433
$p_{2,4}$	0.007381	0.001001	1.597E-6	0.005575	0.007328	0.009489
$p_{1,5}$	0.03553	0.004492	1.311E-5	0.02711	0.03541	0.04469
$p_{2,5}$	0.0268	0.002453	5.771E-6	0.02226	0.0267	0.03187
$p_{1,6}$	0.07464	0.006182	1.536E-5	0.0632	0.07439	0.08745
$p_{2,6}$	0.04783	0.003019	8.673E-6	0.04215	0.04775	0.05399
$p_{1,7}$	0.03158	0.004108	1.637E-5	0.02424	0.03136	0.04021
$p_{2,7}$	0.01577	0.001635	5.752E-6	0.01273	0.01571	0.01912
$p_{1,8}$	0.014	0.002913	7.731E-6	0.008734	0.01385	0.02019
$p_{2,8}$	0.0103	0.001405	2.459E-6	0.007776	0.01022	0.01326
$p_{1,9}$	0.01116	0.003125	7.277E-6	0.006048	0.01082	0.01825
$p_{2,9}$	0.007253	0.001652	2.575E-6	0.004397	0.007124	0.01084
$p_{1,10}$	0.03219	0.006854	1.543E-5	0.02011	0.0317	0.0471
$p_{2,10}$	0.02254	0.003341	4.829E-6	0.01651	0.02236	0.02958

Table B.6. Resulting TPRs and FPRs from pooled centers.

Parameter	Mean	SD	MC error	2.5%	median	97.5%
$s_{1,1}$	0.8126	0.04521	8.96E-5	0.7124	0.8171	0.8881
$s_{1,2}$	0.9139	0.02981	1.648E-4	0.843	0.9182	0.9597
$s_{1,3}$	0.917	0.03643	1.812E-4	0.8333	0.922	0.9735
$s_{1,4}$	0.852	0.04087	1.018E-4	0.761	0.8558	0.92
$s_{1,5}$	0.8124	0.03183	1.401E-4	0.7449	0.8144	0.8687
$s_{1,6}$	0.8252	0.02322	1.258E-4	0.7751	0.8268	0.8657
$s_{1,7}$	0.8999	0.0273	1.436E-4	0.8383	0.9029	0.9441
$s_{1,8}$	0.8935	0.03832	1.392E-4	0.8093	0.8973	0.956
$s_{1,9}$	0.7923	0.07269	1.595E-4	0.6201	0.8044	0.9013
$s_{1,10}$	0.886	0.03677	1.134E-4	0.8062	0.889	0.9486
$s_{2,1}$	0.9571	0.02596	9.271E-5	0.8914	0.9625	0.9898
$s_{2,2}$	0.9545	0.02875	1.845E-4	0.8763	0.9619	0.9886
$s_{2,3}$	0.9437	0.03826	2.201E-4	0.8399	0.9548	0.9879
$s_{2,4}$	0.9449	0.0331	1.627E-4	0.8594	0.953	0.9841
$s_{2,5}$	0.9767	0.01861	1.642E-4	0.9319	0.9805	0.9992
$s_{2,6}$	0.9753	0.01762	1.506E-4	0.932	0.9793	0.9964
$s_{2,7}$	0.965	0.02352	1.601E-4	0.9036	0.9701	0.9934
$s_{2,8}$	0.9287	0.04289	2.201E-4	0.8177	0.9402	0.9789
$s_{2,9}$	0.9683	0.02542	1.449E-4	0.9054	0.9729	0.9986
$s_{2,10}$	0.9703	0.02295	1.33E-4	0.913	0.9745	0.9987
$f_{1,1}$	0.1182	0.005125	1.062E-5	0.1083	0.1181	0.1284
$f_{1,2}$	0.2185	0.006288	1.337E-5	0.2063	0.2184	0.2309
$f_{1,3}$	0.1096	0.004548	1.042E-5	0.1009	0.1096	0.1187
$f_{1,4}$	0.107	0.003441	7.318E-6	0.1003	0.1069	0.1138
$f_{1,5}$	0.09185	0.003923	7.067E-6	0.08431	0.0918	0.09968
$f_{1,6}$	0.2336	0.005231	7.327E-6	0.2234	0.2335	0.2439
$f_{1,7}$	0.06256	0.002631	5.541E-6	0.0575	0.06253	0.06783
$f_{1,8}$	0.2483	0.005708	2.037E-5	0.2372	0.2483	0.2596
$f_{1,9}$	0.05633	0.004577	1.313E-5	0.04769	0.05622	0.06563
$f_{1,10}$	0.2556	0.009714	3.206E-5	0.2368	0.2556	0.2749
$f_{2,1}$	0.1587	0.005786	1.322E-5	0.1475	0.1587	0.1702
$f_{2,2}$	0.1879	0.005941	1.222E-5	0.1764	0.1878	0.1997
$f_{2,3}$	0.1312	0.004913	1.184E-5	0.1217	0.1311	0.1409
$f_{2,4}$	0.1349	0.003807	8.658E-6	0.1275	0.1349	0.1424
$f_{2,5}$	0.1057	0.004171	7.703E-6	0.09764	0.1056	0.114
$f_{2,6}$	0.1104	0.003861	5.35E-6	0.1029	0.1104	0.1181
$f_{2,7}$	0.09618	0.003199	7.359E-6	0.09	0.09616	0.1025
$f_{2,8}$	0.2687	0.005861	2.19E-5	0.2573	0.2687	0.2803
$f_{2,9}$	0.08473	0.005525	1.804E-5	0.0742	0.08462	0.09585
$f_{2,10}$	0.267	0.009852	3.34E-5	0.2479	0.2669	0.2866

Table B.7. Resulting dependency factors from pooled centers.

Parameter	Mean	SD	MC error	2.5%	median	97.5%
λ_1	1.011	0.02165	6.525E-5	0.9903	1.006	1.065
λ_2	1.025	0.02965	1.966E-4	0.9991	1.016	1.109
λ_3	1.031	0.03772	2.205E-4	0.999	1.019	1.139
λ_4	1.015	0.02737	1.106E-4	0.9913	1.007	1.084
λ_5	1.018	0.01913	1.681E-4	0.9993	1.013	1.066
λ_6	1.017	0.01834	1.576E-4	0.999	1.012	1.063
λ_7	1.021	0.02462	1.733E-4	0.9992	1.014	1.087
λ_8	1.022	0.03485	1.626E-4	0.9943	1.011	1.117
λ_9	1.013	0.02087	7.785E-5	0.9908	1.008	1.063
λ_{10}	1.02	0.0222	1.15E-4	0.9992	1.013	1.077
δ_1	5.227	0.1838	3.928E-4	4.879	5.223	5.6
δ_2	3.657	0.09769	2.004E-4	3.471	3.655	3.854
δ_3	6.3	0.2229	5.127E-4	5.879	6.295	6.752
δ_4	6.087	0.1637	3.5E-4	5.775	6.084	6.416
δ_5	6.457	0.2411	3.988E-4	6.0	6.451	6.945
δ_6	2.863	0.07841	9.324E-5	2.711	2.863	3.019
δ_7	8.894	0.2915	6.306E-4	8.341	8.887	9.484
δ_8	3.488	0.07334	2.72E-4	3.348	3.487	3.635
δ_9	10.85	0.6993	0.002204	9.563	10.82	12.3
δ_{10}	3.334	0.1154	3.85E-4	3.117	3.33	3.569

Table B.8. Remaining parameter estimates resulting from pooled centers.

Parameter	Mean	SD	MC error	2.5%	median	97.5%
$\beta_{01,1}$	-3.757	0.2842	0.003208	-4.33	-3.752	-3.203
$\beta_{02,1}$	-3.704	0.2954	0.00337	-4.298	-3.701	-3.121
$\beta_{01,2}$	-3.725	0.2762	0.003261	-4.28	-3.723	-3.181
$\beta_{02,2}$	-3.674	0.2947	0.003479	-4.263	-3.674	-3.087
$\beta_{01,3}$	-3.804	0.2874	0.003234	-4.39	-3.798	-3.251
$\beta_{02,3}$	-3.665	0.2956	0.00338	-4.255	-3.665	-3.076
$\beta_{01,4}$	-3.796	0.2974	0.003237	-4.408	-3.787	-3.229
$\beta_{02,4}$	-3.724	0.2993	0.003345	-4.325	-3.72	-3.136
$\beta_{01,5}$	-3.743	0.2771	0.003254	-4.301	-3.741	-3.199
$\beta_{02,5}$	-3.659	0.2941	0.003471	-4.246	-3.659	-3.074
$\beta_{01,6}$	-3.621	0.2787	0.003409	-4.169	-3.623	-3.059
$\beta_{02,6}$	-3.721	0.2967	0.003671	-4.313	-3.72	-3.135
$\beta_{01,7}$	-3.547	0.2808	0.003273	-4.094	-3.551	-2.98
$\beta_{02,7}$	-3.883	0.3022	0.003543	-4.503	-3.875	-3.309
$\beta_{01,8}$	-3.782	0.2954	0.00326	-4.387	-3.774	-3.215
$\beta_{02,8}$	-3.708	0.2966	0.003344	-4.305	-3.707	-3.122
$\beta_{01,9}$	-3.732	0.2993	0.003254	-4.339	-3.727	-3.148
$\beta_{02,9}$	-3.783	0.3072	0.003316	-4.418	-3.774	-3.197
$\beta_{01,10}$	-3.717	0.2907	0.003257	-4.301	-3.714	-3.144
$\beta_{02,10}$	-3.7	0.2952	0.003348	-4.292	-3.698	-3.117
β_1	-0.3713	0.1183	6.209E-4	-0.5936	-0.3765	-0.124
γ_1	1.907	0.2508	0.001279	1.43	1.899	2.426
γ_2	3.405	0.5066	0.005312	2.501	3.379	4.479
γ_3	-1.864	0.2667	0.006488	-2.435	-1.864	-1.34
γ_4	-1.768	0.1894	0.00453	-2.15	-1.77	-1.393
μ_{β_0}	-3.722	0.2475	0.003312	-4.218	-3.72	-3.235
σ_{β_0}	0.1742	0.09195	5.73E-4	0.02744	0.1622	0.3911

Table B.9. TPRs from the simulation study.

Parameter	Average Posterior Mean	Average Interval Width	Average Interval Coverage	Average Bias
$s_{1,1}$	0.8601	0.0807	0.954	-0.0012
$s_{1,2}$	0.8606	0.0808	0.946	-0.0031
$s_{1,3}$	0.8590	0.0815	0.954	-0.0001
$s_{1,4}$	0.8555	0.0820	0.95	-0.0007
$s_{1,5}$	0.8584	0.0797	0.95	-0.0011
$s_{1,6}$	0.8580	0.0807	0.943	-0.0011
$s_{1,7}$	0.8581	0.0806	0.948	-0.0010
$s_{1,8}$	0.8563	0.0814	0.946	-0.0001
$s_{1,9}$	0.8583	0.0805	0.945	-0.0007
$s_{1,10}$	0.8567	0.0811	0.948	-0.0018
$s_{2,1}$	0.9621	0.0657	0.938	-0.0010
$s_{2,2}$	0.9636	0.0654	0.944	-0.0012
$s_{2,3}$	0.9631	0.0658	0.95	-0.0017
$s_{2,4}$	0.9626	0.0662	0.949	-0.0017
$s_{2,5}$	0.9628	0.0640	0.962	-0.0010
$s_{2,6}$	0.9632	0.0651	0.938	-0.0007
$s_{2,7}$	0.9629	0.0649	0.939	-0.0010
$s_{2,8}$	0.9629	0.0656	0.938	0.0004
$s_{2,9}$	0.9632	0.0650	0.951	-0.0013
$s_{2,10}$	0.9638	0.0651	0.959	-0.0016

Table B.10. FPRs from the simulation study.

Parameter	Average Posterior Mean	Average Interval Width	Average Interval Coverage	Average Bias
$f_{1,1}$	0.1394	0.0333	0.943	0.0052
$f_{1,2}$	0.1395	0.0334	0.938	0.0026
$f_{1,3}$	0.1377	0.0332	0.945	0.0041
$f_{1,4}$	0.1397	0.0333	0.949	0.0061
$f_{1,5}$	0.1406	0.0334	0.944	0.0031
$f_{1,6}$	0.1376	0.0331	0.949	0.0095
$f_{1,7}$	0.1402	0.0334	0.95	0.0028
$f_{1,8}$	0.1407	0.0334	0.951	0.0049
$f_{1,9}$	0.1393	0.0333	0.947	0.0002
$f_{1,10}$	0.1398	0.0333	0.967	0.0040
$f_{2,1}$	0.1505	0.0345	0.947	0.0045
$f_{2,2}$	0.1497	0.0344	0.965	0.0040
$f_{2,3}$	0.1476	0.0342	0.95	0.0047
$f_{2,4}$	0.1562	0.0350	0.938	-0.0010
$f_{2,5}$	0.1537	0.0347	0.96	0.0020
$f_{2,6}$	0.1562	0.0350	0.947	0.0043
$f_{2,7}$	0.1521	0.0346	0.939	0.0020
$f_{2,8}$	0.1537	0.0347	0.952	0.0062
$f_{2,9}$	0.1556	0.0349	0.96	0.0045
$f_{2,10}$	0.1506	0.0344	0.941	0.0057

Table B.11. Prevalences and dependency factors from the simulation study.

Parameter	Average Posterior Mean	Average Interval Width	Average Interval Coverage	Average Bias
$p_{1,1}$	0.0770	0.0295	0.95	0.0093
$p_{1,2}$	0.0771	0.0296	0.947	0.0120
$p_{1,3}$	0.0772	0.0296	0.946	0.0127
$p_{1,4}$	0.0770	0.0295	0.938	0.0080
$p_{1,5}$	0.0779	0.0297	0.942	0.0068
$p_{1,6}$	0.0773	0.0296	0.948	0.0125
$p_{1,7}$	0.0770	0.0295	0.945	0.0159
$p_{1,8}$	0.0772	0.0296	0.951	0.0070
$p_{1,9}$	0.0775	0.0297	0.95	0.0087
$p_{1,10}$	0.0769	0.0295	0.94	0.0180
$p_{2,1}$	0.3798	0.0634	0.951	0.0040
$p_{2,2}$	0.3799	0.0635	0.941	0.0027
$p_{2,3}$	0.3782	0.0635	0.955	0.0007
$p_{2,4}$	0.3779	0.0635	0.964	0.0038
$p_{2,5}$	0.3792	0.0631	0.945	0.0043
$p_{2,6}$	0.3784	0.0633	0.944	0.0025
$p_{2,7}$	0.3805	0.0634	0.961	0.0041
$p_{2,8}$	0.3759	0.0632	0.949	0.0006
$p_{2,9}$	0.3780	0.0631	0.956	0.0016
$p_{2,10}$	0.3787	0.0633	0.944	0.0048
λ_1	1.0170	0.0660	0.949	0.0019
λ_2	1.0174	0.0662	0.953	0.0018
λ_3	1.0173	0.0666	0.946	0.0017
λ_4	1.0172	0.0669	0.951	0.0022
λ_5	1.0163	0.0640	0.949	0.0016
λ_6	1.0170	0.0656	0.932	0.0013
λ_7	1.0168	0.0652	0.935	0.0018
λ_8	1.0168	0.0660	0.941	0.0005
λ_9	1.0166	0.0652	0.944	0.0017
λ_{10}	1.0171	0.0657	0.955	0.0024
δ_1	3.1059	0.9546	0.955	0.0423
δ_2	3.1318	0.9532	0.95	0.0340
δ_3	3.1634	0.9838	0.958	0.0727
δ_4	3.0415	0.9234	0.947	0.0324
δ_5	3.0539	0.9258	0.952	0.0252
δ_6	3.0923	0.9346	0.951	0.0358
δ_7	3.1012	0.9517	0.951	0.0368
δ_8	3.0702	0.9363	0.941	0.1035
δ_9	3.1122	0.9330	0.941	0.0568
δ_{10}	3.0933	0.9388	0.949	0.0259

Table B.12. Offset terms and other parameter results from the simulation study.

Parameter	Average Posterior Mean	Average Interval Width	Average Interval Coverage	Average Bias
$\beta_{01,1}$	-2.5033	0.4225	0.95	0.0016
$\beta_{01,2}$	-2.5013	0.4223	0.947	-0.0032
$\beta_{01,3}$	-2.5002	0.4213	0.946	0.0001
$\beta_{01,4}$	-2.5026	0.4221	0.938	0.0021
$\beta_{01,5}$	-2.4901	0.4194	0.942	0.0030
$\beta_{01,6}$	-2.4980	0.4211	0.948	0.0002
$\beta_{01,7}$	-2.5015	0.4215	0.945	-0.0009
$\beta_{01,8}$	-2.5008	0.4216	0.953	0.0028
$\beta_{01,9}$	-2.4951	0.4210	0.95	0.0022
$\beta_{01,10}$	-2.5039	0.4222	0.939	-0.0017
$\beta_{02,1}$	-2.4920	0.4888	0.957	-0.0018
$\beta_{02,2}$	-2.4916	0.4890	0.956	-0.0006
$\beta_{02,3}$	-2.4990	0.4897	0.967	0.0003
$\beta_{02,4}$	-2.4999	0.4893	0.952	-0.0016
$\beta_{02,5}$	-2.4942	0.4880	0.954	-0.0019
$\beta_{02,6}$	-2.4977	0.4888	0.959	-0.0004
$\beta_{02,7}$	-2.4887	0.4889	0.964	-0.0017
$\beta_{02,8}$	-2.5085	0.4893	0.95	0.0006
$\beta_{02,9}$	-2.4996	0.4886	0.952	-0.0004
$\beta_{02,10}$	-2.4964	0.4890	0.954	-0.0023
μ_{β_0}	-2.4976	0.3149	0.95	-0.0010
σ_{β_0}	0.2128	0.1866	0.945	0.0638
γ_1	1.8862	0.8400	0.961	-0.0072
γ_2	3.4625	1.4490	0.96	0.0184
γ_3	-1.9027	0.7667	0.955	0.0014
γ_4	-1.7914	0.7696	0.957	-0.0048
β_1	1.9965	0.4246	0.951	-0.0018

APPENDIX C

Interval Estimation Code

C.1 EM Algorithm and Wald Interval

```
#####
#####
#-----#
#----- EM ALGORITHM -----#
#-----#
#####
#####

EM<-function(Obsvec,Paramvec,tol)
{
  #####
  ###----- FITTING BY MAXIMUM -----###
  #####
  ###---- Method by Berry et al. ----###
  #####

  #####
  ###-- Obsvec = vector of observed values (disease status
  ###--          verified for patients with at least one
  ###--          positive test result)
  ### Paramvec = vector of parameter estimates (initial
  ###--          values read into function first)
  ###-- tol = specified tolerance
  #####

  #####
  ###-- T = vector of total sample sizes for groups 1/2
  ###-- a = vector of total diseased with Test1+ and
  ###--      Test2+ for group 1, group 2
  ###-- b = vector of total diseased with Test1- and
  ###--      Test2+ for group 1, group 2
  ###-- c = vector of total diseased with Test1+ and
  ###--      Test2- for group 1, group 2
  ###-- A = vector of total non-diseased with Test1+ and
  ###--      Test2+ for group 1, group 2
```

```

###--      B = vector of total non-diseased with Test1+ and
###--      Test2- for group 1, group 2
###--      C = vector of total non-diseased with Test1- and
###--      Test2+ for group 1, group 2
#####
T<-c(Obsvec[1],Obsvec[2])
a<-c(Obsvec[3],Obsvec[4])
b<-c(Obsvec[5],Obsvec[6])
c<-c(Obsvec[7],Obsvec[8])
A<-c(Obsvec[9],Obsvec[10])
B<-c(Obsvec[11],Obsvec[12])
C<-c(Obsvec[13],Obsvec[14])

#####
###--- Initial Values/Estimates ---###
#####
#####
###--      p = vector of prevalences for groups 1/2
###--      tau = vector of s1-s2 (difference in sensitivities)
###--      and f1-f2 (difference in FPRs)
###--      s = or s2, sensitivity/TPR of first test
###--      f = or f2, FPR (1-specificity) of first test
###--      lambda = dependency factor between tests in the
###--      diseased group
###--      delta = dependency factor between tests in the non-
###--      diseased group
#####
p<-c(Paramvec[1],Paramvec[2])
tau<-c(Paramvec[3],Paramvec[5])
s<-Paramvec[4]
f<-Paramvec[6]
lambda<-Paramvec[7]
delta<-Paramvec[8]

#####
### Diseased/Non-diseased Estimates ###
#####
n<-p*T
N<-(c(1,1)-p)*T

#####
###----- STEP 1 -----###
#####

#####
###--- Estimate Double Negatives ---###

```

```
#####
#####
###--      d = vector of total diseased with Test1- and
###--      Test2- for group 1, group 2
###--      D = vector of total non-diseased with Test1-
###--      and Test2- for group 1, group 2
###--      M = vector of total double negatives for groups
###--      1 and 2
#####
d<-n-(a+b+c)
D<-N-(A+B+C)
M<-d+D

#####
###---- Summations of Equivalent Cells in Each Group ----###
#####
ddot<-sum(d)
adot<-sum(a)
bdot<-sum(b)
cdot<-sum(c)
Adot<-sum(A)
Bdot<-sum(B)
Cdot<-sum(C)

#####
###---- New Parameter Estimates ----###
#####
pnew<-p
taunew<-tau
snew<-c(tau[1]+s,s)
fnew<-c(tau[2]+f,f)

#####
###----- STEP 2 -----###
#####

#####
###--- Initials for d1,d2,D1,D2 ---###
#####
dnew<-(a+b+c)*ddot/(adot+bdot+cdot)
Dnew<-M-dnew
ddot<-sum(dnew)
Ddot<-sum(Dnew)

#####
###----- STEP 3 -----###
```

```
#####

#####
###----- Conditional MLEs -----###
#####
pdiff<-c(1,1)
count<-0
diffLog<-1
maxdiff<-1
while(maxdiff>tol)
{ #####
  ###--- Old PARAMETER ESTIMATES ---###
  ParaOld<-c(pnew[1],pnew[2],taunew[1],snew[2],taunew[2],fnew[2],
             lambda,delta)
  n<-pnew*T
  N<-(c(1,1)-pnew)*T

  #####
  ###--- NEW PARAMETER ESTIMATES ---###
  #####
  ### Obtained from LogLik=sum of cell #
  ### probabilities (Berry, pg 856) & #
  ### set first derivative=0 #
  ### FirstDerivatives.wCellProbs.nb #
  #####
  pnew<-c((a[1]+b[1]+c[1]+dnew[1])/T[1],(a[2]+b[2]+c[2]+
             dnew[2])/T[2])
  snew<-c((adot+cdot)/(adot+bdot+cdot+sum(dnew)),(adot+bdot)/
             (adot+bdot+cdot+sum(dnew)))
  fnew<-c((Adot+Cdot)/(Adot+Bdot+Cdot+sum(Dnew)),(Adot+Bdot)/
             (Adot+Bdot+Cdot+sum(Dnew)))
  lambda<-adot*sum(n)/((adot+bdot)*(adot+cdot))
  delta<-Adot*sum(N)/((Adot+Bdot)*(Adot+Cdot))

  taunew<-c(snew[1]-snew[2],fnew[1]-fnew[2])

#####
###----- STEP 4 -----###
#####

#####
###----- NEW d/D/M ESTIMATES -----###
dnew<-M*pnew*(1-(taunew[1]+snew[2])-snew[2]+((taunew[1]+
             snew[2])*snew[2]*lambda))/(1-pnew*((taunew[1]+snew[2])+
             snew[2]-((taunew[1]+snew[2])*snew[2]*lambda))-(1-pnew)*
             ((taunew[2]+fnew[2])+fnew[2]-((taunew[2]+fnew[2])*
```

```

        fnew[2]*delta)))
Dnew<-M-dnew
M<-dnew+Dnew

#####
###----- STEP 5 -----###
#####

#####
###----- d, D SUMS -----###
ddot<-sum(dnew)
Ddot<-sum(Dnew)

#####
###----- NUMBER OF ITERATIONS ----###
count<-count+1

#####
###----- NEW ESTIMATES -----###
ParaNew<-c(pnew[1],pnew[2],taunew[1],snew[2],taunew[2],fnew[2],
           lambda,delta)

#####
## DISEASED/NON-DISEASED ESTIMATES ##
Num<-c(T[1],T[2],n[1],n[2],N[1],N[2])

#####
###-- MAXIMUM PARAMETER CHANGE --###
maxdiff<-max(abs(ParaNew-ParaOld))
}

#####
###- Check Parameter Constraints -###
#####

#####
###----- lambda, delta > 0 -----###
###----- lambda < 1/s1 -----###
###----- lambda < 1/s2 -----###
###----- delta < 1/f1 -----###
###----- delta < 1/f2 -----###
###- lambda > -(1-s1-s2)/(s1s2), -###
###- delta > -(1-f1-f2)/(f1f2), -###
#####
if((pnew[1]*(taunew[1]+snew[2])*snew[2]*lambda<=0)||((1-pnew[1])*

```

```

    (taunew[2]+fnew[2])*fnew[2]*delta<=0)){
      return(c(0,ParaNew,dnew[1],dnew[2],Dnew[1],Dnew[2],Num,maxdiff))}
if((pnew[1]*snew[2]*(1-(taunew[1]+snew[2])*lambda)<=0)||((1-pnew[1])*
  fnew[2]*(1-(taunew[2]+fnew[2])*delta)<=0)){
  return(c(0,ParaNew,dnew[1],dnew[2],Dnew[1],Dnew[2],Num,maxdiff))}
if((pnew[1]*(taunew[1]+snew[2])*(1-snew[2]*lambda)<=0)||((1-pnew[1])*
  (taunew[2]+fnew[2])*(1-fnew[2]*delta)<=0)){
  return(c(0,ParaNew,dnew[1],dnew[2],Dnew[1],Dnew[2],Num,maxdiff))}
if(1-pnew[1]*((taunew[1]+snew[2])+snew[2]-(taunew[1]+snew[2])*
  snew[2]*lambda)-(1-pnew[1])*((taunew[2]+fnew[2])+fnew[2]-(taunew[2]+
  fnew[2])*fnew[2]*delta)<=0 ){
  return(c(0,ParaNew,dnew[1],dnew[2],Dnew[1],Dnew[2],Num,maxdiff))}

if((pnew[2]*(taunew[1]+snew[2])*snew[2]*lambda<=0)||((1-pnew[2])*
  (taunew[2]+fnew[2])*fnew[2]*delta<=0)){
  return(c(0,ParaNew,dnew[1],dnew[2],Dnew[1],Dnew[2],Num,maxdiff))}
if((pnew[2]*snew[2]*(1-(taunew[1]+snew[2])*lambda)<=0)||((1-pnew[2])*
  fnew[2]*(1-(taunew[2]+fnew[2])*delta)<=0)){
  return(c(0,ParaNew,dnew[1],dnew[2],Dnew[1],Dnew[2],Num,maxdiff))}
if((pnew[2]*(taunew[1]+snew[2])*(1-snew[2]*lambda)<=0)||((1-pnew[2])*
  (taunew[2]+fnew[2])*(1-fnew[2]*delta)<=0)){
  return(c(0,ParaNew,dnew[1],dnew[2],Dnew[1],Dnew[2],Num,maxdiff))}
if(1-pnew[2]*((taunew[1]+snew[2])+snew[2]-(taunew[1]+snew[2])*
  snew[2]*lambda)-(1-pnew[2])*((taunew[2]+fnew[2])+fnew[2]-(taunew[2]+
  fnew[2])*fnew[2]*delta)<=0 ){
  return(c(0,ParaNew,dnew[1],dnew[2],Dnew[1],Dnew[2],Num,maxdiff))}

#####
###----- NEW LOG-LIKELIHOOD -----###
LogL<-sum(a*log(pnew*(taunew[1]+snew[2])*snew[2]*lambda)+b*log(
  pnew*snew[2]*(1-(taunew[1]+snew[2])*lambda))+c*log(pnew*(
  taunew[1]+snew[2])*(1-snew[2]*lambda))+A*log((1-pnew)*(taunew[2]+
  fnew[2])*fnew[2]*delta)+B*log((1-pnew)*fnew[2]*(1-(taunew[2]+
  fnew[2])*delta))+C*log((1-pnew)*(taunew[2]+fnew[2])*(1-fnew[2]*
  delta))+M*log(1-pnew*((taunew[1]+snew[2])+snew[2]-(taunew[1]+
  snew[2])*snew[2]*lambda)-(1-pnew)*((taunew[2]+fnew[2])+fnew[2]-
  (taunew[2]+fnew[2])*fnew[2]*delta)))

#####
###-- Return LL and New Estimates --###
#####
return(c(LogL,ParaNew,dnew[1],dnew[2],Dnew[1],Dnew[2],Num,maxdiff))
}

```

```
#####
#####
#-----#
#----- CALCULATE WALD/NORMAL INTERVAL -----#
#-----#
#####
#####
WaldInt<-function(Obsvec,ParaNew,dnew,Dnew){
  #####
  ###-- Obsvec = vector of observed values
  ###- ParaNew = vector of parameter estimates from EM alg.
  ###-- tol = specified tolerance
  ###---- dnew = unverified positives
  ###---- Dnew = unverified negatives
  #####

  #####
  ###--- Initial Values/Estimates ---###
  #####
  #####
  ###--      p = vector of prevalences for group 1, group 2
  ###--      tau = vector of s1-s2 (difference in sensitivities)
  ###--           and f1-f2 (difference in FPRs)
  ###--      s = or s2, sensitivity/TPR of first test
  ###--      f = or f2, FPR (1-specificity) of first test
  ###-- lambda = dependency factor between tests in the
  ###--           diseased group
  ###-- delta = dependency factor between tests in the
  ###--           non-diseased group
  #####

p1<-ParaNew[1]
p2<-ParaNew[2]
tau1<-ParaNew[3]
s2<-ParaNew[4]
tau2<-ParaNew[5]
f2<-ParaNew[6]
lambda<-ParaNew[7]
delta<-ParaNew[8]
s1<-tau1+s2
f1<-tau2+f2
d1<-dnew[1]
d2<-dnew[2]
D1<-Dnew[1]
D2<-Dnew[2]

  #####

```



```

###--      T = vector of total sample sizes for groups 1/2
###--      a = vector of total diseased with Test1+ and
###--          Test2+ for group 1, group 2
###--      b = vector of total diseased with Test1- and
###--          Test2+ for group 1, group 2
###--      c = vector of total diseased with Test1+ and
###--          Test2- for group 1, group 2
###--      A = vector of total non-diseased with Test1+
###--          and Test2+ for group 1, group 2
###--      B = vector of total non-diseased with Test1+
###--          and Test2- for group 1, group 2
###--      C = vector of total non-diseased with Test1-
###--          and Test2+ for group 1, group 2
#####
T<-c(Obsvec[1],Obsvec[2])
a<-c(Obsvec[3],Obsvec[4])
b<-c(Obsvec[5],Obsvec[6])
c<-c(Obsvec[7],Obsvec[8])
A<-c(Obsvec[9],Obsvec[10])
B<-c(Obsvec[11],Obsvec[12])
C<-c(Obsvec[13],Obsvec[14])

a1<-a[1]
a2<-a[2]
b1<-b[1]
b2<-b[2]
c1<-c[1]
c2<-c[2]
A1<-A[1]
A2<-A[2]
B1<-B[1]
B2<-B[2]
C1<-C[1]
C2<-C[2]

#####
##--- Fishers Information Matrix ---##
##-- Matrix of Second Derivatives --##
##----- 8x8 -----##
#####
#####
##  J={J_ij}, i,j=p1,p2,s2,tau2,f2,lambda,delta
##
##          dp1          dp2          ....
##
##  dp1 :    dp1(dp1)    dp1(dp2)    ....

```

```

##      dp2 :      dp2(dp1)      dp2(dp2)      ....
##      ds1 :      ds1(dp1)      ds1(dp2)      ....
##      ds2 :      ds2(dp1)      ds2(dp2)      ....
##      df1 :      df1(dp1)      df1(dp2)      ....
##      df2 :      df2(dp1)      df2(dp2)      ....
##      dlambd:dlambd(dp1) dlambd(dp2)      ....
##      ddelta: ddelta(dp1) ddelta(dp2)      ....
#####

J11=-(A1/(1-p1)^2)-B1/(1-p1)^2-C1/(1-p1)^2-a1/p1^2-b1/p1^2-
      c1/p1^2-(M1*(f1+f2-delta*f1*f2-s1-s2+lambda*s1*s2)^2)/
      (1-(f1+f2-delta*f1*f2)*(1-p1)-p1*(s1+s2-lambda*s1*s2))^2
J12=0
J13=(M1*p1*(1-lambda*s2)*(f1+f2-delta*f1*f2-s1-s2+lambda*s1*
      s2))/(1-(f1+f2-delta*f1*f2)*(1-p1)-p1*(s1+s2-lambda*s1*
      s2))^2+(M1*(-1+lambda*s2))/(1-(f1+f2-delta*f1*f2)*(1-p1)-
      p1*(s1+s2-lambda*s1*s2))
J14=(M1*p1*(1-lambda*s1)*(f1+f2-delta*f1*f2-s1-s2+lambda*s1*
      s2))/(1-(f1+f2-delta*f1*f2)*(1-p1)-p1*(s1+s2-lambda*s1*
      s2))^2+(M1*(-1+lambda*s1))/(1-(f1+f2-delta*f1*f2)*(1-p1)-
      p1*(s1+s2-lambda*s1*s2))
J15=((1-delta*f2)*M1*(1-p1)*(f1+f2-delta*f1*f2-s1-s2+lambda*
      s1*s2))/(1-(f1+f2-delta*f1*f2)*(1-p1)-p1*(s1+s2-lambda*s1*
      s2))^2+((1-delta*f2)*M1)/(1-(f1+f2-delta*f1*f2)*(1-p1)-p1*
      (s1+s2-lambda*s1*s2))
J16=((1-delta*f1)*M1*(1-p1)*(f1+f2-delta*f1*f2-s1-s2+lambda*
      s1*s2))/(1-(f1+f2-delta*f1*f2)*(1-p1)-p1*(s1+s2-lambda*s1*
      s2))^2+((1-delta*f1)*M1)/(1-(f1+f2-delta*f1*f2)*(1-p1)-p1*
      (s1+s2-lambda*s1*s2))
J17=-((M1*p1*s1*s2*(f1+f2-delta*f1*f2-s1-s2+lambda*s1*s2))/(1-
      (f1+f2-delta*f1*f2)*(1-p1)-p1*(s1+s2-lambda*s1*s2))^2)+(M1*
      s1*s2)/(1-(f1+f2-delta*f1*f2)*(1-p1)-p1*(s1+s2-lambda*s1*s2))
J18=-((f1*f2*M1*(1-p1)*(f1+f2-delta*f1*f2-s1-s2+lambda*s1*s2))/
      (1-(f1+f2-delta*f1*f2)*(1-p1)-p1*(s1+s2-lambda*s1*s2))^2)-
      (f1*f2*M1)/(1-(f1+f2-delta*f1*f2)*(1-p1)-p1*(s1+s2-lambda*
      s1*s2))

J21=0
J22=-(A2/(1-p2)^2)-B2/(1-p2)^2-C2/(1-p2)^2-a2/p2^2-b2/p2^2-c2/
      p2^2-(M2*(f1+f2-delta*f1*f2-s1-s2+lambda*s1*s2)^2)/(1-(f1+
      f2-delta*f1*f2)*(1-p2)-p2*(s1+s2-lambda*s1*s2))^2
J23=(M2*p2*(1-lambda*s2)*(f1+f2-delta*f1*f2-s1-s2+lambda*s1*s2))/
      (1-(f1+f2-delta*f1*f2)*(1-p2)-p2*(s1+s2-lambda*s1*s2))^2+
      (M2*(-1+lambda*s2))/(1-(f1+f2-delta*f1*f2)*(1-p2)-p2*(s1+s2-
      lambda*s1*s2))
J24=(M2*p2*(1-lambda*s1)*(f1+f2-delta*f1*f2-s1-s2+lambda*s1*s2))/

```

$$\begin{aligned}
& (1-(f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2-\text{lambda}*s1*s2))^2+(M2* \\
& (-1+\text{lambda}*s1))/(1-(f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2- \\
& \text{lambda}*s1*s2)) \\
J25 = & ((1-\text{delta}*f2)*M2*(1-p2)*(f1+f2-\text{delta}*f1*f2-s1-s2+\text{lambda}*s1* \\
& s2))/(1-(f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2-\text{lambda}*s1* \\
& s2))^2+((1-\text{delta}*f2)*M2)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(\\
& s1+s2-\text{lambda}*s1*s2)) \\
J26 = & ((1-\text{delta}*f1)*M2*(1-p2)*(f1+f2-\text{delta}*f1*f2-s1-s2+\text{lambda}*s1* \\
& s2))/(1-(f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2-\text{lambda}*s1* \\
& s2))^2+((1-\text{delta}*f1)*M2)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(\\
& s1+s2-\text{lambda}*s1*s2)) \\
J27 = & -((M2*p2*s1*s2*(f1+f2-\text{delta}*f1*f2-s1-s2+\text{lambda}*s1*s2))/(1-(f1+ \\
& f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2-\text{lambda}*s1*s2))^2)+(M2*s1* \\
& s2)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2-\text{lambda}*s1*s2)) \\
J28 = & -((f1*f2*M2*(1-p2)*(f1+f2-\text{delta}*f1*f2-s1-s2+\text{lambda}*s1*s2))/(1- \\
& (f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2-\text{lambda}*s1*s2))^2)-(f1*f2* \\
& M2)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2-\text{lambda}*s1*s2)) \\
J31 = & (M1*p1*(1-\text{lambda}*s2)*(f1+f2-\text{delta}*f1*f2-s1-s2+\text{lambda}*s1*s2))/ \\
& (1-(f1+f2-\text{delta}*f1*f2)*(1-p1)-p1*(s1+s2-\text{lambda}*s1*s2))^2-(M1* \\
& (1-\text{lambda}*s2))/(1-(f1+f2-\text{delta}*f1*f2)*(1-p1)-p1*(s1+s2-\text{lambda}* \\
& s1*s2)) \\
J32 = & (M2*p2*(1-\text{lambda}*s2)*(f1+f2-\text{delta}*f1*f2-s1-s2+\text{lambda}*s1*s2))/ \\
& (1-(f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2-\text{lambda}*s1*s2))^2-(M2* \\
& (1-\text{lambda}*s2))/(1-(f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2-\text{lambda}* \\
& s1*s2)) \\
J33 = & -(a1/s1^2)-a2/s1^2-c1/s1^2-c2/s1^2-(b1*\text{lambda}^2)/(1-\text{lambda}* \\
& s1)^2-(b2*\text{lambda}^2)/(1-\text{lambda}*s1)^2-(M1*p1^2*(1-\text{lambda}* \\
& s2)^2)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p1)-p1*(s1+s2-\text{lambda}*s1* \\
& s2))^2-(M2*p2^2*(1-\text{lambda}*s2)^2)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p2)- \\
& p2*(s1+s2-\text{lambda}*s1*s2))^2 \\
J34 = & -((M1*p1^2*(1-\text{lambda}*s1)*(1-\text{lambda}*s2))/(1-(f1+f2-\text{delta}*f1*f2)* \\
& (1-p1)-p1*(s1+s2-\text{lambda}*s1*s2))^2+(\text{lambda}*M1*p1)/(1-(f1+f2- \\
& \text{delta}*f1*f2)*(1-p1)-p1*(s1+s2-\text{lambda}*s1*s2))-(M2*p2^2*(1-\text{lambda}* \\
& s1)*(1-\text{lambda}*s2))/(1-(f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2- \\
& \text{lambda}*s1*s2))^2+(\text{lambda}*M2*p2)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p2)- \\
& p2*(s1+s2-\text{lambda}*s1*s2)) \\
J35 = & -(((1-\text{delta}*f2)*M1*(1-p1)*p1*(1-\text{lambda}*s2))/(1-(f1+f2-\text{delta}*f1* \\
& f2)*(1-p1)-p1*(s1+s2-\text{lambda}*s1*s2))^2)-((1-\text{delta}*f2)*M2*(1-p2)* \\
& p2*(1-\text{lambda}*s2))/(1-(f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2- \\
& \text{lambda}*s1*s2))^2 \\
J36 = & -(((1-\text{delta}*f1)*M1*(1-p1)*p1*(1-\text{lambda}*s2))/(1-(f1+f2-\text{delta}*f1* \\
& f2)*(1-p1)-p1*(s1+s2-\text{lambda}*s1*s2))^2)-((1-\text{delta}*f1)*M2*(1-p2)* \\
& p2*(1-\text{lambda}*s2))/(1-(f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2- \\
& \text{lambda}*s1*s2))^2 \\
J37 = & -((b1*\text{lambda}*s1)/(1-\text{lambda}*s1)^2)-(b2*\text{lambda}*s1)/(1-\text{lambda}*s1)^2-
\end{aligned}$$

$$\begin{aligned}
& b1/(1-\lambda*s1)-b2/(1-\lambda*s1)+(M1*p1^2*s1*s2*(1-\lambda*s2))/((1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2+(M1*p1*s2)/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2)))+(M2*p2^2*s1*s2*(1-\lambda*s2))/((1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2+(M2*p2*s2)/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))) \\
J38 &= (f1*f2*M1*(1-p1)*p1*(1-\lambda*s2))/((1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2+(f1*f2*M2*(1-p2)*p2*(1-\lambda*s2))/((1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2 \\
J41 &= (M1*p1*(1-\lambda*s1)*(f1+f2-\delta*f1*f2-s1-s2+\lambda*s1*s2))/((1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2-(M1*(1-\lambda*s1))/((1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))) \\
J42 &= (M2*p2*(1-\lambda*s1)*(f1+f2-\delta*f1*f2-s1-s2+\lambda*s1*s2))/((1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2-(M2*(1-\lambda*s1))/((1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))) \\
J43 &= -((M1*p1^2*(1-\lambda*s1)*(1-\lambda*s2))/((1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2+(\lambda*M1*p1)/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))-(M2*p2^2*(1-\lambda*s1)*(1-\lambda*s2))/((1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2+(\lambda*M2*p2)/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))) \\
J44 &= -(a1/s^2)-a2/s^2-b1/s^2-b2/s^2-(c1*\lambda^2)/(1-\lambda*s2)^2-(c2*\lambda^2)/(1-\lambda*s2)^2-(M1*p1^2*(1-\lambda*s1)^2)/((1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2-(M2*p2^2*(1-\lambda*s1)^2)/((1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2 \\
J45 &= -(((1-\delta*f2)*M1*(1-p1)*p1*(1-\lambda*s1))/((1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2)-((1-\delta*f2)*M2*(1-p2)*p2*(1-\lambda*s1))/((1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2 \\
J46 &= -(((1-\delta*f1)*M1*(1-p1)*p1*(1-\lambda*s1))/((1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2)-((1-\delta*f1)*M2*(1-p2)*p2*(1-\lambda*s1))/((1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2 \\
J47 &= -((c1*\lambda*s2)/(1-\lambda*s2)^2)-(c2*\lambda*s2)/(1-\lambda*s2)^2-c1/(1-\lambda*s2)-c2/(1-\lambda*s2)+(M1*p1^2*s1*(1-\lambda*s1)*s2)/((1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2+(M1*p1*s1)/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2)))+(M2*p2^2*s1*(1-\lambda*s1)*s2)/((1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2+(M2*p2*s1)/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))) \\
J48 &= (f1*f2*M1*(1-p1)*p1*(1-\lambda*s1))/((1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2+(f1*f2*M2*(1-p2)*p2*(1-\lambda*s1))/((1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2
\end{aligned}$$

$$\begin{aligned}
J51 &= ((1-\delta*f2)*M1*(1-p1)*(f1+f2-\delta*f1*f2-s1-s2+\lambda*s1*s2))/ \\
&\quad (1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2+((1-\delta*f2)*M1)/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2)) \\
J52 &= ((1-\delta*f2)*M2*(1-p2)*(f1+f2-\delta*f1*f2-s1-s2+\lambda*s1*s2))/ \\
&\quad (1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2+((1-\delta*f2)*M2)/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2)) \\
J53 &= -(((1-\delta*f2)*M1*(1-p1)*p1*(1-\lambda*s2))/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2)-((1-\delta*f2)*M2*(1-p2)* \\
&\quad p2*(1-\lambda*s2))/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2 \\
J54 &= -(((1-\delta*f2)*M1*(1-p1)*p1*(1-\lambda*s1))/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2)-((1-\delta*f2)*M2*(1-p2)* \\
&\quad p2*(1-\lambda*s1))/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2 \\
J55 &= -(A1/f1^2)-A2/f1^2-C1/f1^2-C2/f1^2-(B1*\delta^2)/(1-\delta*f1)^2- \\
&\quad (B2*\delta^2)/(1-\delta*f1)^2-((1-\delta*f2)^2*M1*(1-p1)^2)/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2-((1-\delta*f2)^2*M2*(1-p2)^2)/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2 \\
J56 &= -(((1-\delta*f1)*(1-\delta*f2)*M1*(1-p1)^2)/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2)+(delta*M1*(1-p1))/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))-((1-\delta*f1)*(1-\delta*f2)*M2*(1-p2)^2)/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2+(delta*M2*(1-p2))/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2)) \\
J57 &= ((1-\delta*f2)*M1*(1-p1)*p1*s1*s2)/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2+((1-\delta*f2)*M2*(1-p2)*p2*s1*s2)/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2 \\
J58 &= -((B1*\delta*f1)/(1-\delta*f1)^2)-(B2*\delta*f1)/(1-\delta*f1)^2- \\
&\quad B1/(1-\delta*f1)-B2/(1-\delta*f1)+(f1*f2*(1-\delta*f2)*M1*(1-p1)^2)/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2+(f2*M1*(1-p1))/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))+(f1*f2*(1-\delta*f2)*M2*(1-p2)^2)/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2+(f2*M2*(1-p2))/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2)) \\
J61 &= ((1-\delta*f1)*M1*(1-p1)*(f1+f2-\delta*f1*f2-s1-s2+\lambda*s1*s2))/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2+ \\
&\quad ((1-\delta*f1)*M1)/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2)) \\
J62 &= ((1-\delta*f1)*M2*(1-p2)*(f1+f2-\delta*f1*f2-s1-s2+\lambda*s1*s2))/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2+((1-\delta*f1)*M2)/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))
\end{aligned}$$

$$\begin{aligned}
J63 &= -(((1-\delta*f1)*M1*(1-p1)*p1*(1-\lambda*s2))/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2 - ((1-\delta*f1)*M2*(1-p2)*p2*(1-\lambda*s2))/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2 \\
J64 &= -(((1-\delta*f1)*M1*(1-p1)*p1*(1-\lambda*s1))/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2 - ((1-\delta*f1)*M2*(1-p2)*p2*(1-\lambda*s1))/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2 \\
J65 &= -(((1-\delta*f1)*(1-\delta*f2)*M1*(1-p1)^2)/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2 + (\delta*M1*(1-p1))/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2)) - ((1-\delta*f1)*(1-\delta*f2)*M2*(1-p2)^2)/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2 + (\delta*M2*(1-p2))/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2)) \\
J66 &= -(A1/f2^2) - A2/f2^2 - B1/f2^2 - B2/f2^2 - (C1*\delta^2)/(1-\delta*f2)^2 - (C2*\delta^2)/(1-\delta*f2)^2 - ((1-\delta*f1)^2*M1*(1-p1)^2)/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2 - ((1-\delta*f1)^2*M2*(1-p2)^2)/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2 \\
J67 &= ((1-\delta*f1)*M1*(1-p1)*p1*s1*s2)/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2 + ((1-\delta*f1)*M2*(1-p2)*p2*s1*s2)/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2 \\
J68 &= -((C1*\delta*f2)/(1-\delta*f2)^2) - (C2*\delta*f2)/(1-\delta*f2)^2 - C1/(1-\delta*f2) - C2/(1-\delta*f2) + (f1*(1-\delta*f1)*f2*M1*(1-p1)^2)/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2 + (f1*M1*(1-p1))/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2)) + (f1*(1-\delta*f1)*f2*M2*(1-p2)^2)/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2 + (f1*M2*(1-p2))/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2)) \\
J71 &= -((M1*p1*s1*s2*(f1+f2-\delta*f1*f2-s1-s2+\lambda*s1*s2))/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2) + (M1*s1*s2)/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2)) \\
J72 &= -((M2*p2*s1*s2*(f1+f2-\delta*f1*f2-s1-s2+\lambda*s1*s2))/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2) + (M2*s1*s2)/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2)) \\
J73 &= -((b1*\lambda*s1)/(1-\lambda*s1)^2) - (b2*\lambda*s1)/(1-\lambda*s1)^2 - b1/(1-\lambda*s1) - b2/(1-\lambda*s1) + (M1*p1^2*s1*s2*(1-\lambda*s2))/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2 + (M1*p1*s2)/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2)) + (M2*p2^2*s1*s2*(1-\lambda*s2))/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2))^2 + (M2*p2*s2)/(1-(f1+f2-\delta*f1*f2)*(1-p2)-p2*(s1+s2-\lambda*s1*s2)) \\
J74 &= -((c1*\lambda*s2)/(1-\lambda*s2)^2) - (c2*\lambda*s2)/(1-\lambda*s2)^2 - c1/(1-\lambda*s2) - c2/(1-\lambda*s2) + (M1*p1^2*s1*(1-\lambda*s1)*s2)/(1-(f1+f2-\delta*f1*f2)*(1-p1)-p1*(s1+s2-\lambda*s1*s2))^2 + (M1*p1*
\end{aligned}$$

$$\begin{aligned}
& s1)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p1)-p1*(s1+s2-\text{lambda}*s1*s2))+(M2* \\
& p2^2*s1*(1-\text{lambda}*s1)*s2)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+ \\
& s2-\text{lambda}*s1*s2))^2+(M2*p2*s1)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p2)-p2* \\
& (s1+s2-\text{lambda}*s1*s2)) \\
J75=& ((1-\text{delta}*f2)*M1*(1-p1)*p1*s1*s2)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p1)- \\
& p1*(s1+s2-\text{lambda}*s1*s2))^2+((1-\text{delta}*f2)*M2*(1-p2)*p2*s1*s2)/(\\
& 1-(f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2-\text{lambda}*s1*s2))^2 \\
J76=& ((1-\text{delta}*f1)*M1*(1-p1)*p1*s1*s2)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p1)- \\
& p1*(s1+s2-\text{lambda}*s1*s2))^2+((1-\text{delta}*f1)*M2*(1-p2)*p2*s1*s2)/(1- \\
& (f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2-\text{lambda}*s1*s2))^2 \\
J77=& -(a1/\text{lambda}^2)-a2/\text{lambda}^2-(b1*s1^2)/(1-\text{lambda}*s1)^2-(b2*s1^2)/(\\
& 1-\text{lambda}*s1)^2-(c1*s2^2)/(1-\text{lambda}*s2)^2-(c2*s2^2)/(1-\text{lambda}* \\
& s2)^2-(M1*p1^2*s1^2*s2^2)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p1)-p1*(s1+ \\
& s2-\text{lambda}*s1*s2))^2-(M2*p2^2*s1^2*s2^2)/(1-(f1+f2-\text{delta}*f1*f2)*(\\
& 1-p2)-p2*(s1+s2-\text{lambda}*s1*s2))^2 \\
J78=& -((f1*f2*M1*(1-p1)*p1*s1*s2)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p1)-p1*(\\
& s1+s2-\text{lambda}*s1*s2))^2)-(f1*f2*M2*(1-p2)*p2*s1*s2)/(1-(f1+f2- \\
& \text{delta}*f1*f2)*(1-p2)-p2*(s1+s2-\text{lambda}*s1*s2))^2 \\
J81=& -((f1*f2*M1*(1-p1)*(f1+f2-\text{delta}*f1*f2-s1-s2+\text{lambda}*s1*s2))/(1-(\\
& f1+f2-\text{delta}*f1*f2)*(1-p1)-p1*(s1+s2-\text{lambda}*s1*s2))^2)-(f1*f2* \\
& M1)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p1)-p1*(s1+s2-\text{lambda}*s1*s2)) \\
J82=& -((f1*f2*M2*(1-p2)*(f1+f2-\text{delta}*f1*f2-s1-s2+\text{lambda}*s1*s2))/(1-(\\
& f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2-\text{lambda}*s1*s2))^2)-(f1*f2* \\
& M2)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2-\text{lambda}*s1*s2)) \\
J83=& (f1*f2*M1*(1-p1)*p1*(1-\text{lambda}*s2))/(1-(f1+f2-\text{delta}*f1*f2)*(1-p1)- \\
& p1*(s1+s2-\text{lambda}*s1*s2))^2+(f1*f2*M2*(1-p2)*p2*(1-\text{lambda}*s2))/(1- \\
& (f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2-\text{lambda}*s1*s2))^2 \\
J84=& (f1*f2*M1*(1-p1)*p1*(1-\text{lambda}*s1))/(1-(f1+f2-\text{delta}*f1*f2)*(1-p1)- \\
& p1*(s1+s2-\text{lambda}*s1*s2))^2+(f1*f2*M2*(1-p2)*p2*(1-\text{lambda}*s1))/(1- \\
& (f1+f2-\text{delta}*f1*f2)*(1-p2)-p2*(s1+s2-\text{lambda}*s1*s2))^2 \\
J85=& -(B1*\text{delta}*f1)/(1-\text{delta}*f1)^2-(B2*\text{delta}*f1)/(1-\text{delta}*f1)^2- \\
& B1/(1-\text{delta}*f1)-B2/(1-\text{delta}*f1)+(f1*f2*(1-\text{delta}*f2)*M1*(1- \\
& p1)^2)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p1)-p1*(s1+s2-\text{lambda}*s1*s2))^2+ \\
& (f2*M1*(1-p1))/(1-(f1+f2-\text{delta}*f1*f2)*(1-p1)-p1*(s1+s2-\text{lambda}*s1* \\
& s2))+(f1*f2*(1-\text{delta}*f2)*M2*(1-p2)^2)/(1-(f1+f2-\text{delta}*f1*f2)* \\
& (1-p2)-p2*(s1+s2-\text{lambda}*s1*s2))^2+(f2*M2*(1-p2))/(1-(f1+f2-\text{delta}* \\
& f1*f2)*(1-p2)-p2*(s1+s2-\text{lambda}*s1*s2)) \\
J86=& -((C1*\text{delta}*f2)/(1-\text{delta}*f2)^2)-(C2*\text{delta}*f2)/(1-\text{delta}*f2)^2-C1/(\\
& 1-\text{delta}*f2)-C2/(1-\text{delta}*f2)+(f1*(1-\text{delta}*f1)*f2*M1*(1-p1)^2)/(1- \\
& (f1+f2-\text{delta}*f1*f2)*(1-p1)-p1*(s1+s2-\text{lambda}*s1*s2))^2+(f1*M1* \\
& (1-p1))/(1-(f1+f2-\text{delta}*f1*f2)*(1-p1)-p1*(s1+s2-\text{lambda}*s1*s2))+ \\
& (f1*(1-\text{delta}*f1)*f2*M2*(1-p2)^2)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p2)- \\
& p2*(s1+s2-\text{lambda}*s1*s2))^2+(f1*M2*(1-p2))/(1-(f1+f2-\text{delta}*f1* \\
& f2)*(1-p2)-p2*(s1+s2-\text{lambda}*s1*s2)) \\
J87=& -((f1*f2*M1*(1-p1)*p1*s1*s2)/(1-(f1+f2-\text{delta}*f1*f2)*(1-p1)-p1*(s1+
\end{aligned}$$

```

      s2-lambda*s1*s2))^2)-(f1*f2*M2*(1-p2)*p2*s1*s2)/(1-(f1+f2-delta*
      f1*f2)*(1-p2)-p2*(s1+s2-lambda*s1*s2))^2
J88=-(A1/delta^2)-A2/delta^2-(B1*f1^2)/(1-delta*f1)^2-(B2*f1^2)/(1-
      delta*f1)^2-(C1*f2^2)/(1-delta*f2)^2-(C2*f2^2)/(1-delta*f2)^2-
      (f1^2*f2^2*M1*(1-p1)^2)/(1-(f1+f2-delta*f1*f2)*(1-p1)-p1*(s1+s2-
      lambda*s1*s2))^2-(f1^2*f2^2*M2*(1-p2)^2)/(1-(f1+f2-delta*f1*f2)*
      (1-p2)-p2*(s1+s2-lambda*s1*s2))^2

#####
###----- Inverse Fishers -----###
###----- Information Matrix -----###
#####
F<--matrix(c(J11,J21,J31,J41,J51,J61,J71,J81,J12,J22,J32,J42,J52,
      J62,J72,J82,J13,J23,J33,J43,J53,J63,J73,J83,J14,J24,
      J34,J44,J54,J64,J74,J84,J15,J25,J35,J45,J55,J65,J75,
      J85,J16,J26,J36,J46,J56,J66,J76,J86,J17,J27,J37,J47,
      J57,J67,J77,J87,J18,J28,J38,J48,J58,J68,J78,J88),
      nrow=8,ncol=8)
I<-diag(x=1,nrow=8,ncol=8)
InvF<-solve(F,I)

#####
###----- Check Assumptions -----###
#####
if( (InvF[3,3]<0)|| (InvF[4,4]<0) ){return(rbind(0,0,0,0))}
if( (InvF[5,5]<0)|| (InvF[6,6]<0) ){return(rbind(0,0,0,0))}

#####
###----- Normal Approximation ----###
###----- for Tau1 -----###
#####

#####
###----- vars1 = var of s1 -----###
###----- vars2 = var of s2 -----###
###----- covs1s2 = cov of s1,s2 ----###
###----- SDtau1 = sd of tau1 -----###
###- Norm_LBtau1 = Wald LB of tau1 -###
###- Norm_UBtau1 = Wald UB of tau1 -###
###----- 95% Wald CI -----###
#####
vars1<-InvF[3,3]
vars2<-InvF[4,4]
covs1s2<-InvF[3,4]
if(vars1+vars2-2*covs1s2<0){return(rbind(0,0,0,0))}
SDtau1<-sqrt(vars1+vars2-2*covs1s2)

```



```

Norm_LBtau1<-(s1-s2)-qnorm(.95)*SDtau1
Norm_UBtau1<-(s1-s2)+qnorm(.95)*SDtau1

#####
###----- Normal Approximation ----###
###----- for Tau1 -----###
#####

#####
###----- varf1 = var of f1 -----###
###----- varf2 = var of f2 -----###
###----- covf1f2 = cov of f1,f2 ----###
###----- SDtau2 = sd of tau2 -----###
###- Norm_LBtau2 = Wald LB of tau2 -###
###- Norm_UBtau2 = Wald UB of tau2 -###
###----- 95% Wald CI -----###
#####
varf1<-InvF[5,5]
varf2<-InvF[6,6]
covf1f2<-InvF[5,6]
if(varf1+varf2-2*covf1f2<0){return(rbind(0,0,0,0))}
SDtau2<-sqrt(varf1+varf2-2*covf1f2)
Norm_LBtau2<-(f1-f2)-qnorm(.95)*SDtau2
Norm_UBtau2<-(f1-f2)+qnorm(.95)*SDtau2

#####
###-- Return Normal Approximation -###
###----- LB,UB for Tau1,Tau2 -----###
#####
return(rbind(Norm_LBtau1,Norm_UBtau1,Norm_LBtau2,Norm_UBtau2))

}

```

sectionTPR and FPR Generation

```

#####
#####
#-----#
#----- HIGH TPR, HIGH FPR SIMULATION CODE -----#
#-----#
#####
#####

#####
###---- q = number of sample size to generate --###
###---- j = total combined sample size, T1+T2 --###
### j.len = length of j, length(j)=20          --###

```

```

###--- qj = length(j) x q = 20 x 500 = 10,000 ---###
#####
q=500 #generate 500 for each sample size, j
j=c(1000,1100,1200,1300,1400,1500,1600,1700,1800,1900,
    2000,2100,2200,2300,2400,2500,2600,2700,2800,2900,
    3000,3100,3200,3300,3400,3500,3600,3700,3800,3900,
    4000)
j.len<-length(j)
qj<-j.len*q

#####
###-- Prob.HH = matrix of cell probs -----###
###--- Gen.HH = generated data for HH scenario -----###
###----- Prob = temporary matrix of cell probs -----###
###-- CInorm1.HH = norm intervals for tau1 -----###
### CInorm1ss.HH = norm interval widths for tau1 ---###
###- Norm1cov.HH = norm coverages for tau1 -----###
###-- CInorm2.HH = norm intervals for tau2 -----###
### CInorm2ss.HH = norm interval widths for tau2 ---###
###- Norm2cov.HH = norm coverages for tau2 -----###
###--- LogLik.HH = Keep track of logL -----###
###----- MLEs.HH = keep MLEs -----###
###----- Dd.HH = keep record of d,D -----###
###--- LogErr.HH = times lambda,delta out-of-bounds ###
###--- NegErr.HH = times run into neg var -----###
#####
# Initial Fitted Frequencies generated by Multinomial
  Prob.HH<-matrix(0,nrow=qj,ncol=14)
  Gen.HH<-matrix(0,nrow=qj,ncol=15)
  Prob<-matrix(0,nrow=2,ncol=7)

  CInorm1.HH<-matrix(0,nrow=qj,ncol=2)
  CInorm1ss.HH<-matrix(0,nrow=j.len,ncol=1)
  Norm1cov.HH<-matrix(0,nrow=j.len,ncol=1)
  CInorm2.HH<-matrix(0,nrow=qj,ncol=2)
  CInorm2ss.HH<-matrix(0,nrow=j.len,ncol=1)
  Norm2cov.HH<-matrix(0,nrow=j.len,ncol=1)

  LogLik.HH<-matrix(0,nrow=qj,ncol=1)
  MLEs.HH<-matrix(0,nrow=qj,ncol=8)
  Dd.HH<-matrix(0,nrow=qj,ncol=4)
  LogErr.HH=rep(0,j.len)
  NegErr.HH=rep(0,j.len)

#####
###----- Generate 500 for -----###

```

```

    ###----- sample size, n -----###
    #####

for(z in 1:j.len){

  ntemp<-j[z] #temp n; going through each value of j
  print(ntemp)

  T[1]<-(ntemp/2) #assume equal sample sizes for both pops
  T[2]<-(ntemp/2)
  N1tot<-0
  N2tot<-0
  ptm <- proc.time()

  v=1
  while(v<=q){
    print(v)

    p<-c(0.3,0.05) #assume high TPR, high FPR
    s<-c(0.88,0.98)
    f<-c(0.15,0.18)
    tau<-c(s[1]-s[2],f[1]-f[2])
    lambda<-1.0
    delta<-5.0

    ParamV=c(p[1],p[2],tau[1],s[2],tau[2],f[2],lambda,delta)
    Tol=1e-04

    for(i in 1:2)
    {
      Prob[i,<math>] <- c(p[i] * (tau[1] + s[2]) * s[2] * lambda,
        p[i] * s[2] * (1 - (tau[1] + s[2]) * lambda),
        p[i] * (tau[1] + s[2]) * (1 - s[2] * lambda),
        (1 - p[i]) * (tau[2] + f[2]) * f[2] * delta,
        (1 - p[i]) * f[2] * (1 - (tau[2] + f[2]) * delta),
        (1 - p[i]) * (tau[2] + f[2]) * (1 - f[2] * delta),
        1 - p[i] * ((tau[1] + s[2]) + s[2] - ((tau[1] + s[2]) * s[2] * lambda)) -
        (1 - p[i]) * ((tau[2] + f[2]) + f[2] - ((tau[2] + f[2]) * f[2] * delta)))
    }

    x<-rmultinom(1,round(T[1]),Prob[1,<math>])
    y<-rmultinom(1,round(T[2]),Prob[2,<math>])

    x111<-x[1]
    x101<-x[2]
    x110<-x[3]

```

```

x011<-x[4]    #small number
x001<-x[5]
x010<-x[6]

y111<-y[1]
y101<-y[2]
y110<-y[3]

y011<-y[4]    #small number
y001<-y[5]
y010<-y[6]

M1<-x[7]
M2<-y[7]

Gen.HH[(q*(z-1)+v),]<-c(v,T[1],T[2],x111,y111,x101,y101,x110,y110,
                        x011,y011,x001,y001,x010,y010)

#####
###----- Call EM Algorithm -----###
#####

Etemp<-EM(Gen.HH[(q*(z-1)+v),2:15],ParamV,Tol)

print(Etemp[1])
print(Etemp[2:9])
print(Etemp[10:13])

if(Etemp[1]==0){
  LogErr.HH[z]=LogErr.HH[z]+1
}else{

Prob.HH[(q*(z-1)+v),]=c(Prob[1,],Prob[2,])
LogLik.HH[(q*(z-1)+v),]=Etemp[1]
MLEs.HH[(q*(z-1)+v),]=Etemp[2:9]
Dd.HH[(q*(z-1)+v),]<-Etemp[10:13]

#####
###----- Obtain Normal CIs -----###
#####

LWNtemp<-WaldInt(Gen.HH[(q*(z-1)+v),2:15],Etemp[2:9],Etemp[10:11],
                Etemp[12:13])

if( (LWNtemp[1]==0)&&(LWNtemp[2]==0)&&(LWNtemp[3]==0)&&(

```

```

        LWNtemp[4]==0) ){
            NegErr.HH[z]=NegErr.HH[z]+1
        }else{
            CInorm1.HH[(q*(z-1)+v),]<-LWNtemp[1:2]
            CInorm2.HH[(q*(z-1)+v),]<-LWNtemp[3:4]

            N1parm<-tau[1]
            N1low<-CInorm1.HH[(q*(z-1)+v),1]
            N1upp<-CInorm1.HH[(q*(z-1)+v),2]
            N2parm<-tau[2]
            N2low<-CInorm2.HH[(q*(z-1)+v),1]
            N2upp<-CInorm2.HH[(q*(z-1)+v),2]

            if((N1low<=N1parm)&&(N1parm<=N1upp)){N1tot<-N1tot+1}
            if(v==q){Norm1cov.HH[z]<-(N1tot/q)}
            if((N2low<=N2parm)&&(N2parm<=N2upp)){N2tot<-N2tot+1}
            if(v==q){Norm2cov.HH[z]<-(N2tot/q)}
            v=v+1
        }
    }
}

proc.time() - ptm

```

C.2 WinBUGS Code

```

#####
#####
#-----#
#----- Informative prior (HH) -----#
#-----#
#####
#####

model fixed{

p1~dbeta(1,1);
p2~dbeta(1,1);
s1~dbeta(26.4,3.6);
s2~dbeta(1,1);
f1~dbeta(4.5,25.5);
f2~dbeta(1,1);
lam~dgamma(0.01,0.01);
del~dgamma(0.01,0.01);

pop1[1]<-p1*s1*s2*lam

```

```

pop1[2]<-p1*(1-s1*lam)*s2
pop1[3]<-p1*s1*(1-s2*lam)
pop1[4]<-(1-p1)*f1*f2*del
pop1[5]<-(1-p1)*(1-f1*del)*f2
pop1[6]<-(1-p1)*f1*(1-f2*del)
pop1[7]<-(p1*(1-s1-s2+s1*s2*lam)+(1-p1)*(1-f1-f2+f1*f2*del))

t1[1:7]~dmulti(pop1[1:7],N1)
t1[1]<-a1
t1[2]<-b1
t1[3]<-c1
t1[4]<-A1
t1[5]<-B1
t1[6]<-C1
t1[7]<-M1

pop2[1]<-p2*s1*s2*lam
pop2[2]<-p2*(1-s1*lam)*s2
pop2[3]<-p2*s1*(1-s2*lam)
pop2[4]<-(1-p2)*f1*f2*del
pop2[5]<-(1-p2)*(1-f1*del)*f2
pop2[6]<-(1-p2)*f1*(1-f2*del)
pop2[7]<-(p2*(1-s1-s2+s1*s2*lam)+(1-p2)*(1-f1-f2+f1*f2*del))

t2[1:7]~dmulti(pop2[1:7],N2)
t2[1]<-a2
t2[2]<-b2
t2[3]<-c2
t2[4]<-A2
t2[5]<-B2
t2[6]<-C2
t2[7]<-M2

tau1<-(s1-s2)
tau2<-(f1-f2)
}

```

C.3 R Code to Call WinBUGS

```

#####
#####
#-----#
#----- Generate Bayesian Intervals -----#
#-----#
#####
#####
library(R2WinBUGS)

```

```
#####
###-- Using a sample size equivalent of 30 --###
###----- Informative Prior -----###
#####
BayesWB.HH.x3<-function(Gen,T,s){

  m<- dim(Gen)[1] #500

  set.seed(s)
  ave.vec1<-rep(NA,m)
  coverage1<-rep(NA,m)
  length1<-rep(NA,m)
  power1<-rep(NA,m)
  ave.vec2<-rep(NA,m)
  coverage2<-rep(NA,m)
  length2<-rep(NA,m)
  power2<-rep(NA,m)

  for(i in 1:m){

    N1<-T/2
    N2<-T/2

    a1<-Gen[i,4]
    b1<-Gen[i,6]
    c1<-Gen[i,8]

    A1<-Gen[i,10]    #small number
    B1<-Gen[i,12]
    C1<-Gen[i,14]

    a2<-Gen[i,5]
    b2<-Gen[i,7]
    c2<-Gen[i,9]

    A2<-Gen[i,11]    #small number
    B2<-Gen[i,13]
    C2<-Gen[i,15]

    M1<-N1-(a1+b1+c1+A1+B1+C1)
    M2<-N2-(a2+b2+c2+A2+B2+C2)

    data<-list("a1","b1","c1","A1","B1","C1","M1","N1","a2","b2",
               "c2","A2","B2","C2","M2","N2") ## save data
    parameters<-list("p1","p2","s1","s2","f1","f2","lam","del",
                     "tau1","tau2")
  }
}
```

```

inits<-list(p1=0.3,p2=0.05,s1=0.88,s2=0.98,f1=0.15,f2=0.18,
            lam=1.0,del=5.0) ### HH Case ###

tau=c(0.88-0.98,0.15-0.18)

inits<-list(inits)

ss.sim<-bugs(data,inits,parameters,
             "DependPriors.Beta_HHx3_20100620.txt",
             n.chains=1,n.burnin=15000,n.iter=50000,n.thin=5)

ave.vec1[i]<-ss.sim$summary[9,1]
coverage1[i]<-(tau[1]>=ss.sim$summary[9,3])&(
            tau[1]<=ss.sim$summary[9,7])
length1[i]<-ss.sim$summary[9,7]-ss.sim$summary[9,3]

ave.vec2[i]<-ss.sim$summary[10,1]
coverage2[i]<-(tau[2]>=ss.sim$summary[10,3])&(
            tau[2]<=ss.sim$summary[10,7])
length2[i]<-ss.sim$summary[10,7]-ss.sim$summary[10,3]

}

#####
###-- Return Bayesian CI coverages/widths --###
#####

#####
###-- ave1_c = avg posterior coverage for tau1 --###
###-- ave1_l = avg posterior CI width for tau1 --###
###-- ave2_c = avg posterior coverage for tau2 --###
###-- ave2_l = avg posterior CI width for tau2 --###
#####
ave1_c<-mean(coverage1)
ave1_l<-mean(length1)
ave2_c<-mean(coverage2)
ave2_l<-mean(length2)

return(c(ave1_c,ave1_l,ave2_c,ave2_l))
}

WB.HH.x3.500<-BayesWB.HH.x3(Gen.HH[1:500,],sum(Gen.HH[1,2:3]),10)

```

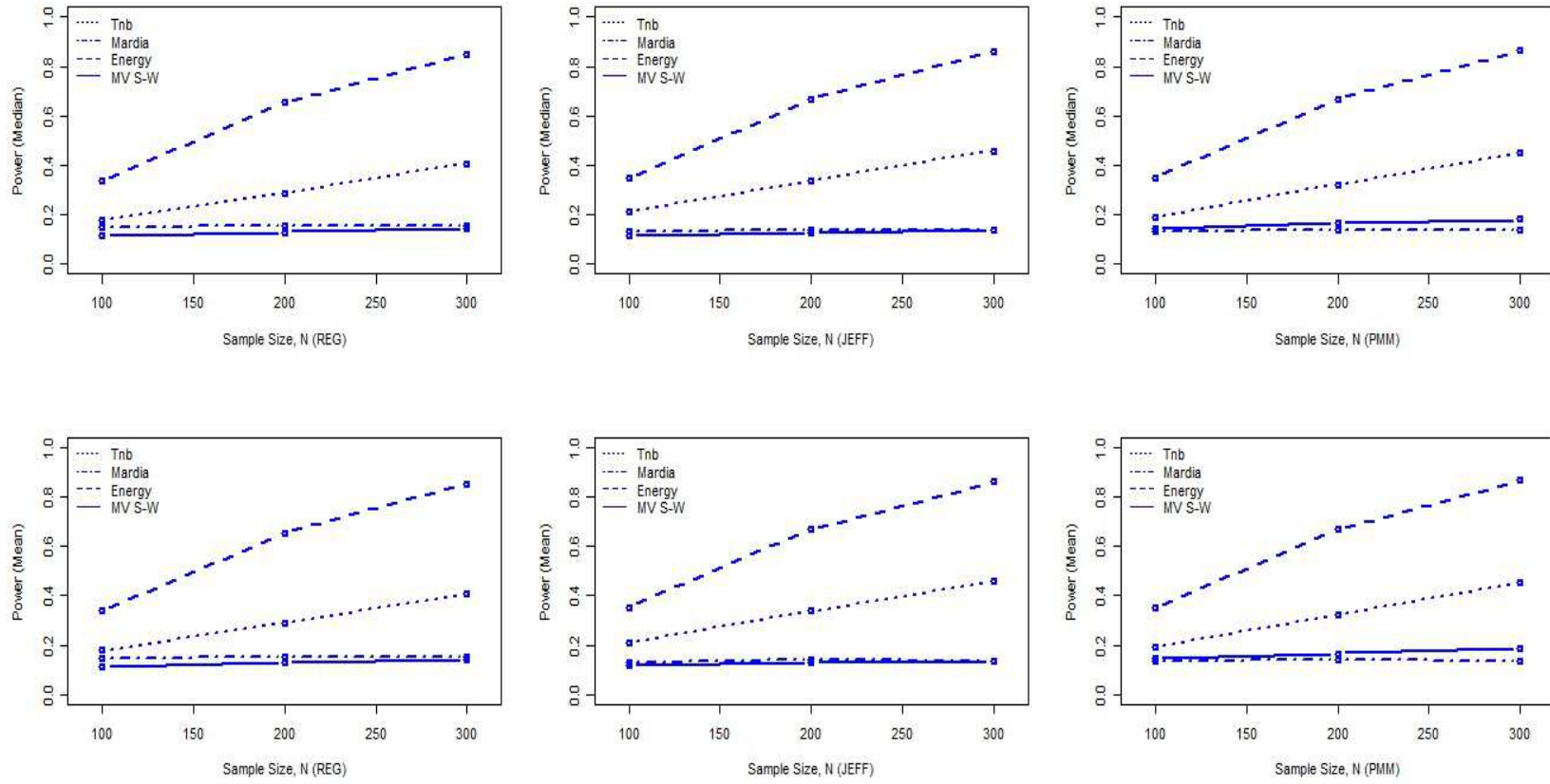

APPENDIX D

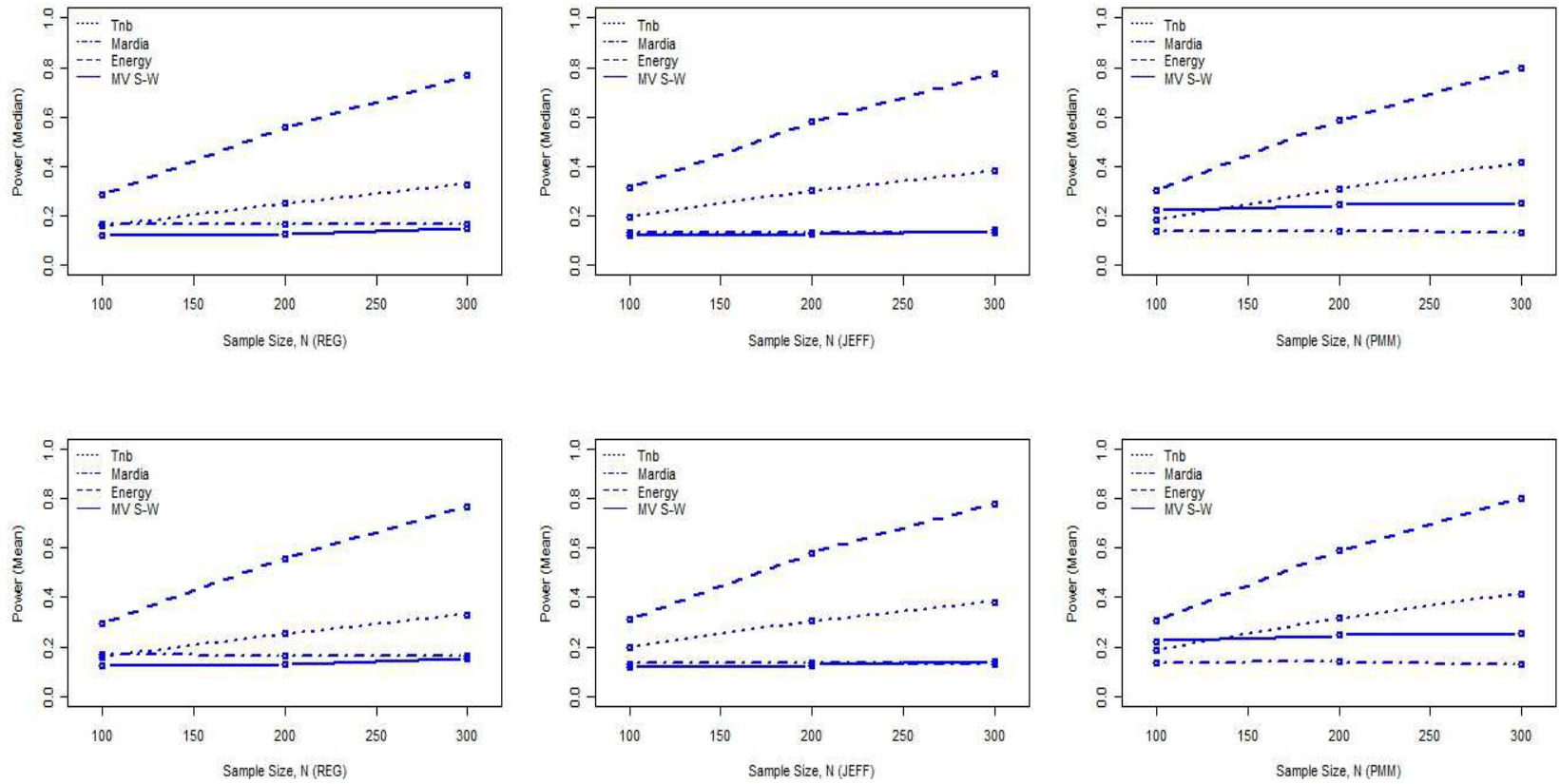
Power Analysis Plots

The following set of figures display the results of our power analysis study for different multivariate normality tests and multiple imputation procedures in the presence of a monotone missing data pattern. For each of the following figures, the top row represents the power of the four tests for the three different imputation procedures when we use the 50,000 generated median test statistics. The bottom row represents the power of the imputation-test combinations when using the 50,000 generated mean values.

Figure D.1 was obtained using a multivariate skewed-normal distribution with skewness parameter $\nu = \mathbf{1000}$ having a dimension of 10 with $p = 2$. Similarly, Figure D.2 consider the case of when $p = 5$. Figures D.3 and D.4 were created using a multivariate t distribution with $df = 25$ and with $\nu = \mathbf{0}$ of dimension 10 when $p = 2$ and $p = 5$, respectively. Figures D.5 and D.6 consider the case for when $df = 30$. The remaining two figures, Figures D.7 and D.8, consider both a skewness parameter, $\nu = \mathbf{100}$ with dimension 10, and a degrees of freedom, $df = 50$, for when $p = 2$ and $p = 5$.

Exact values of the values in these figures are displayed in tables found in Appendix E. For each of the imputation methods (listed in the first column) and each sample size (listed in the second column) there are two rows of values. The top of these two rows represents the powers using the generated means and the bottom of the two rows represents the powers using the generated medians for the tests specified by the column names. Values in parentheses are the respective standard errors.

Figure D.1. MVSN model for $p = 2$.

Figure D.2. MVSN model for $p = 5$.

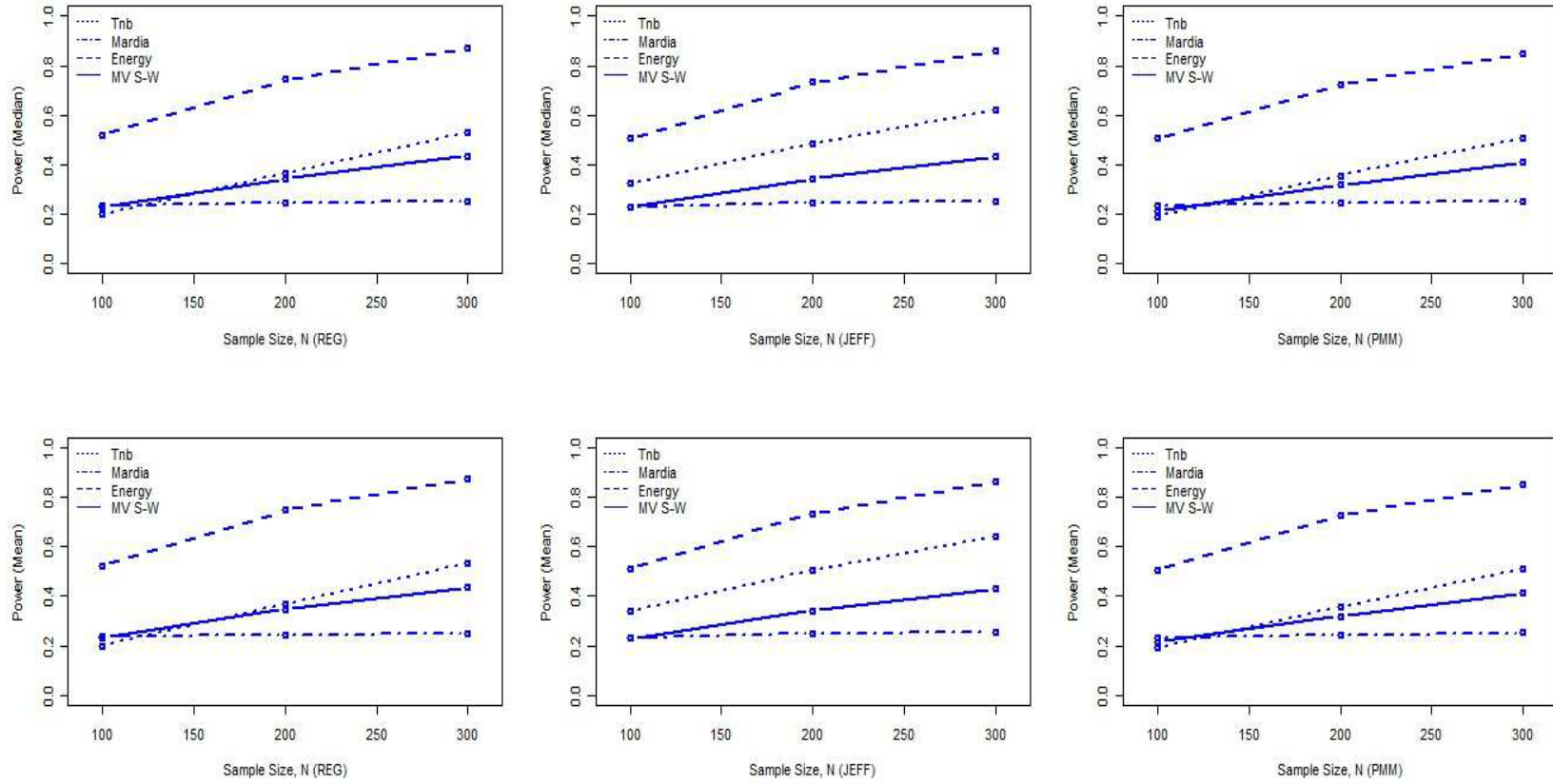
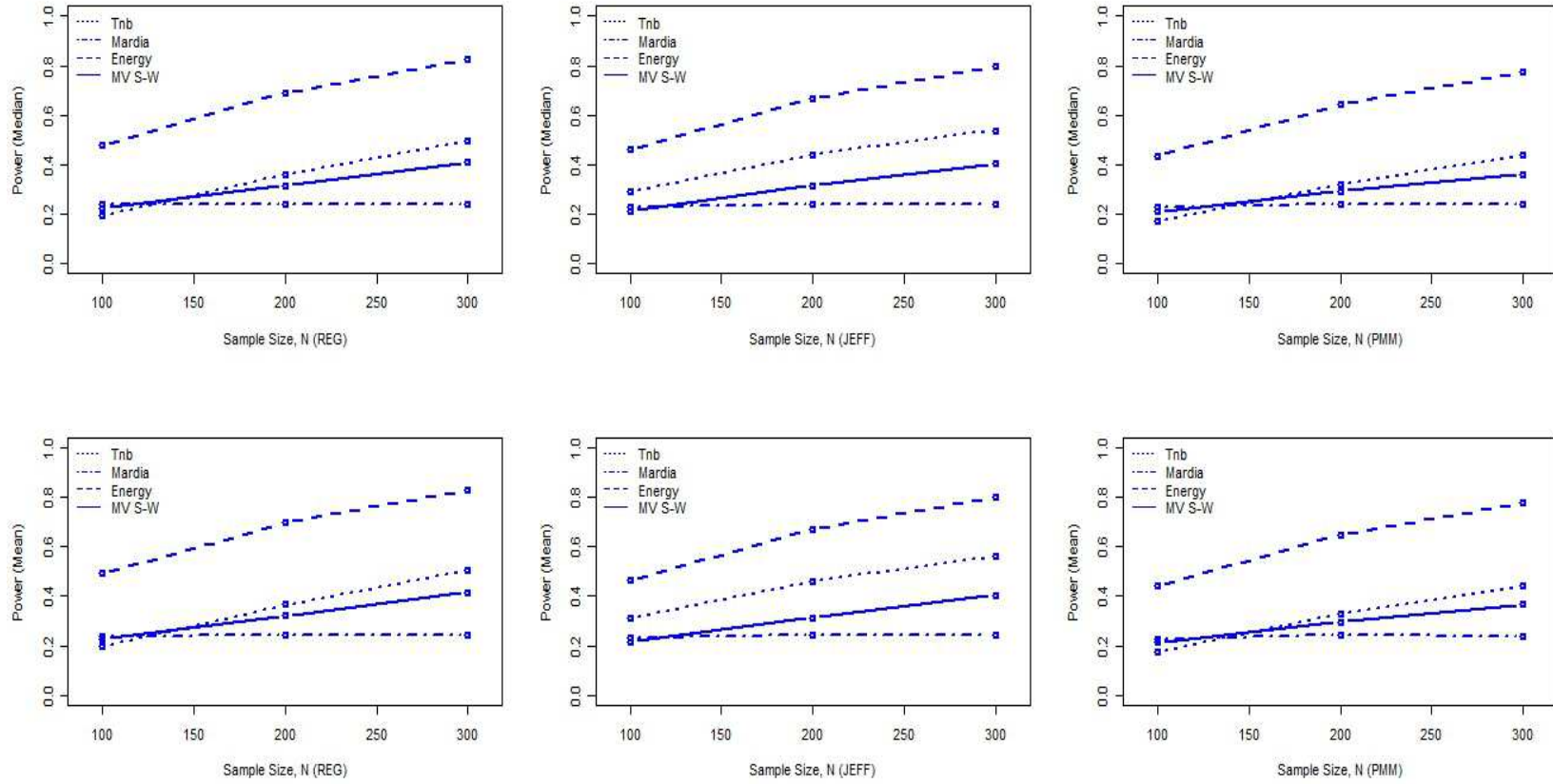
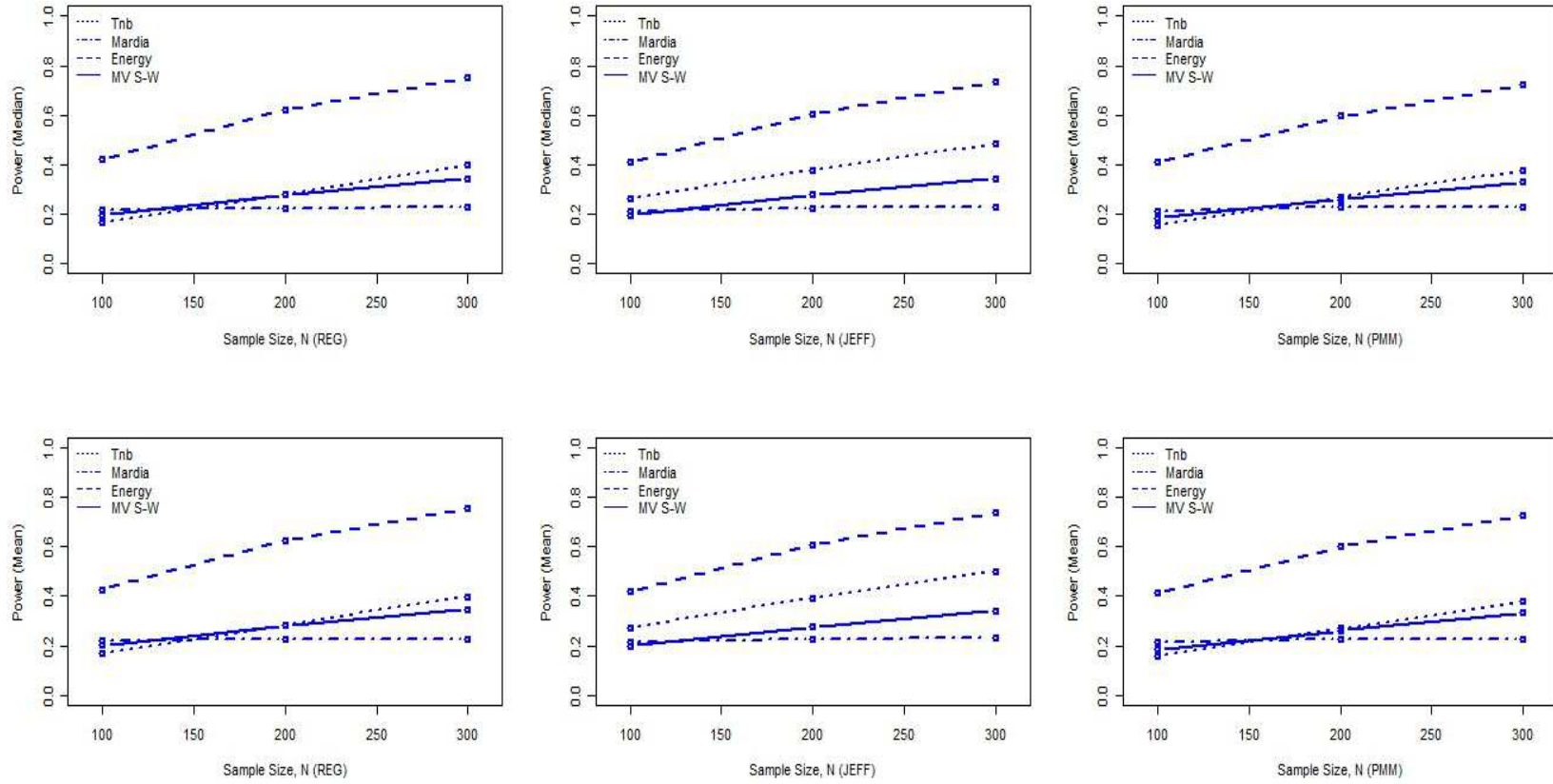
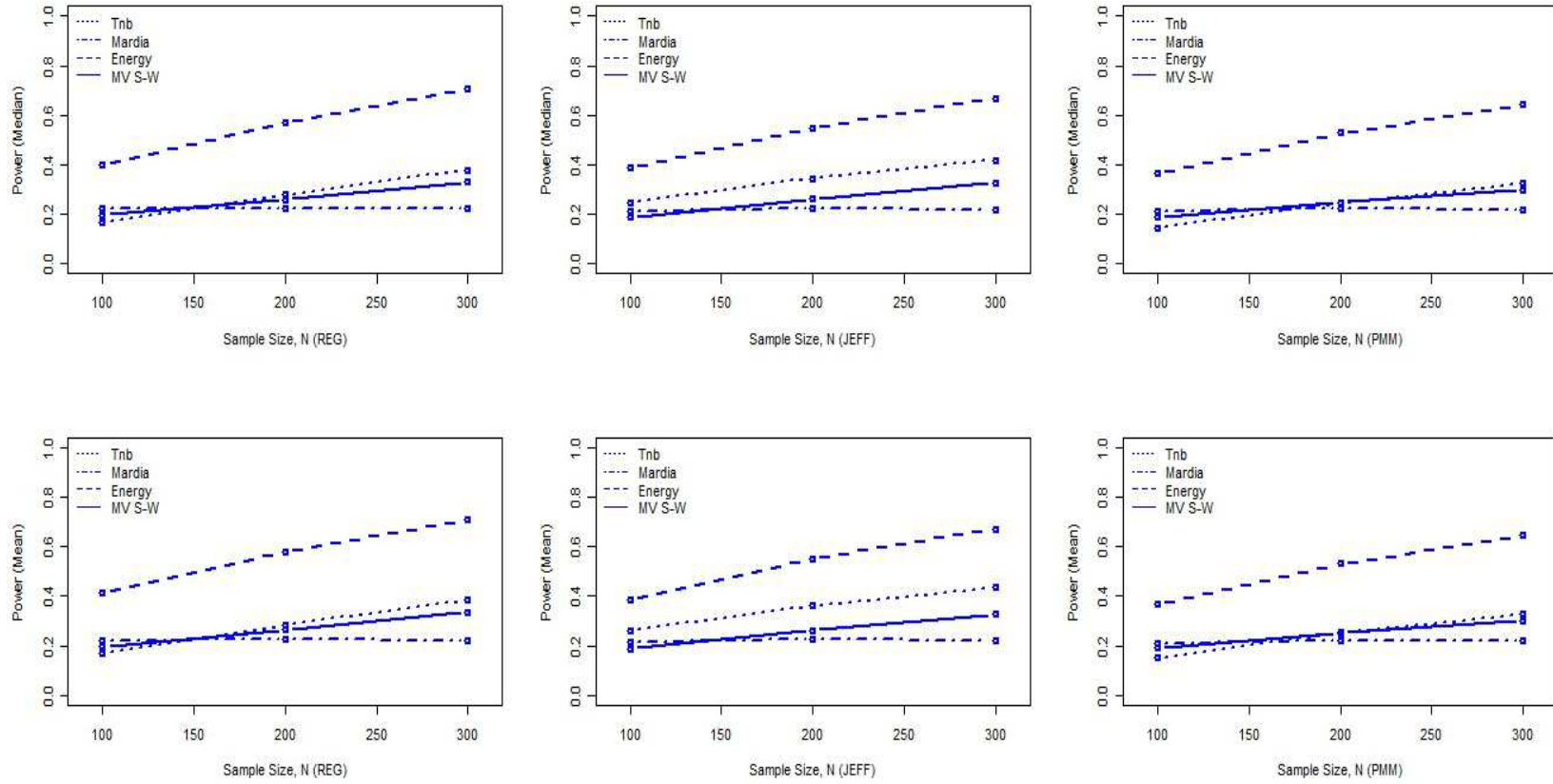
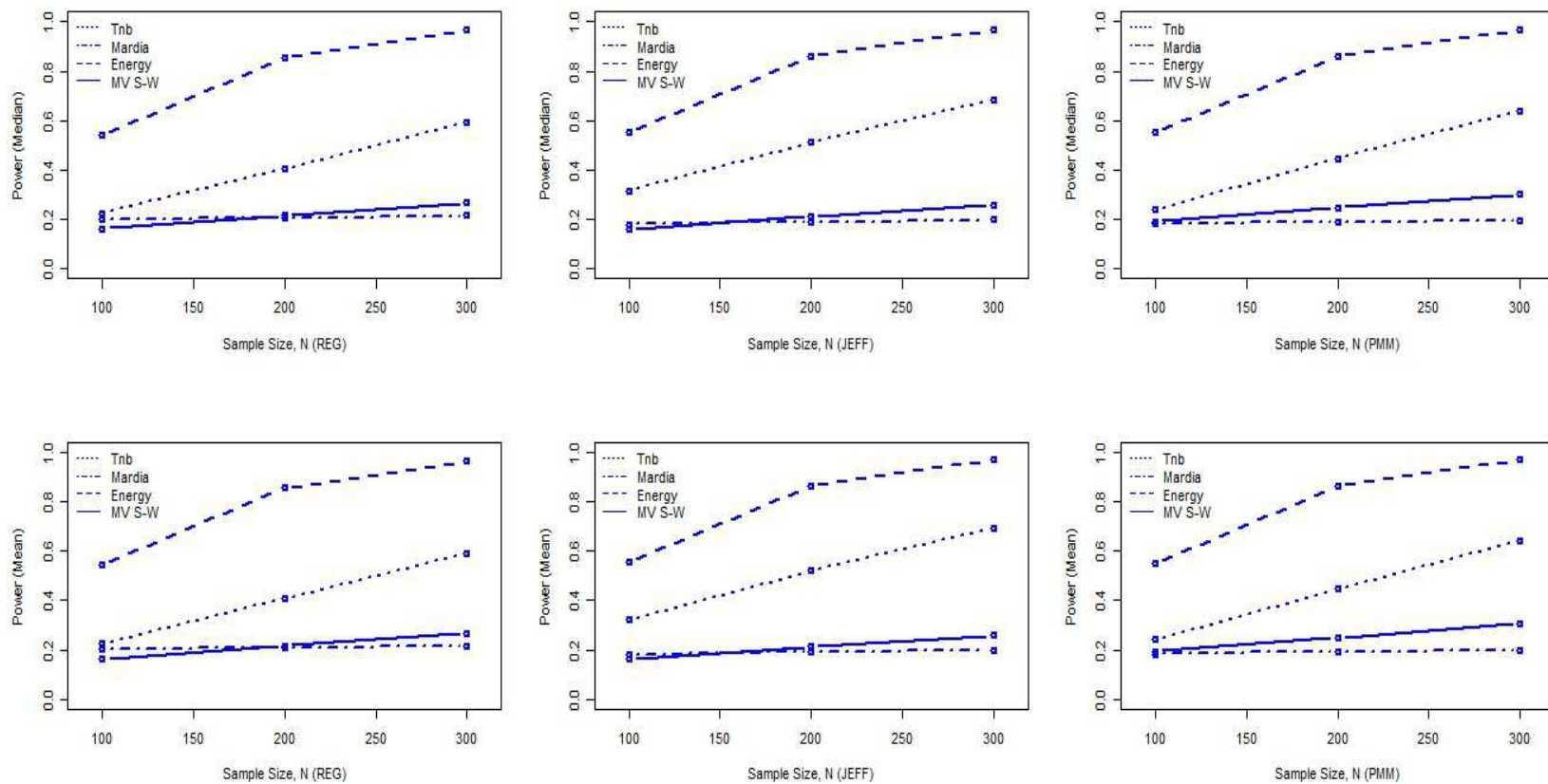


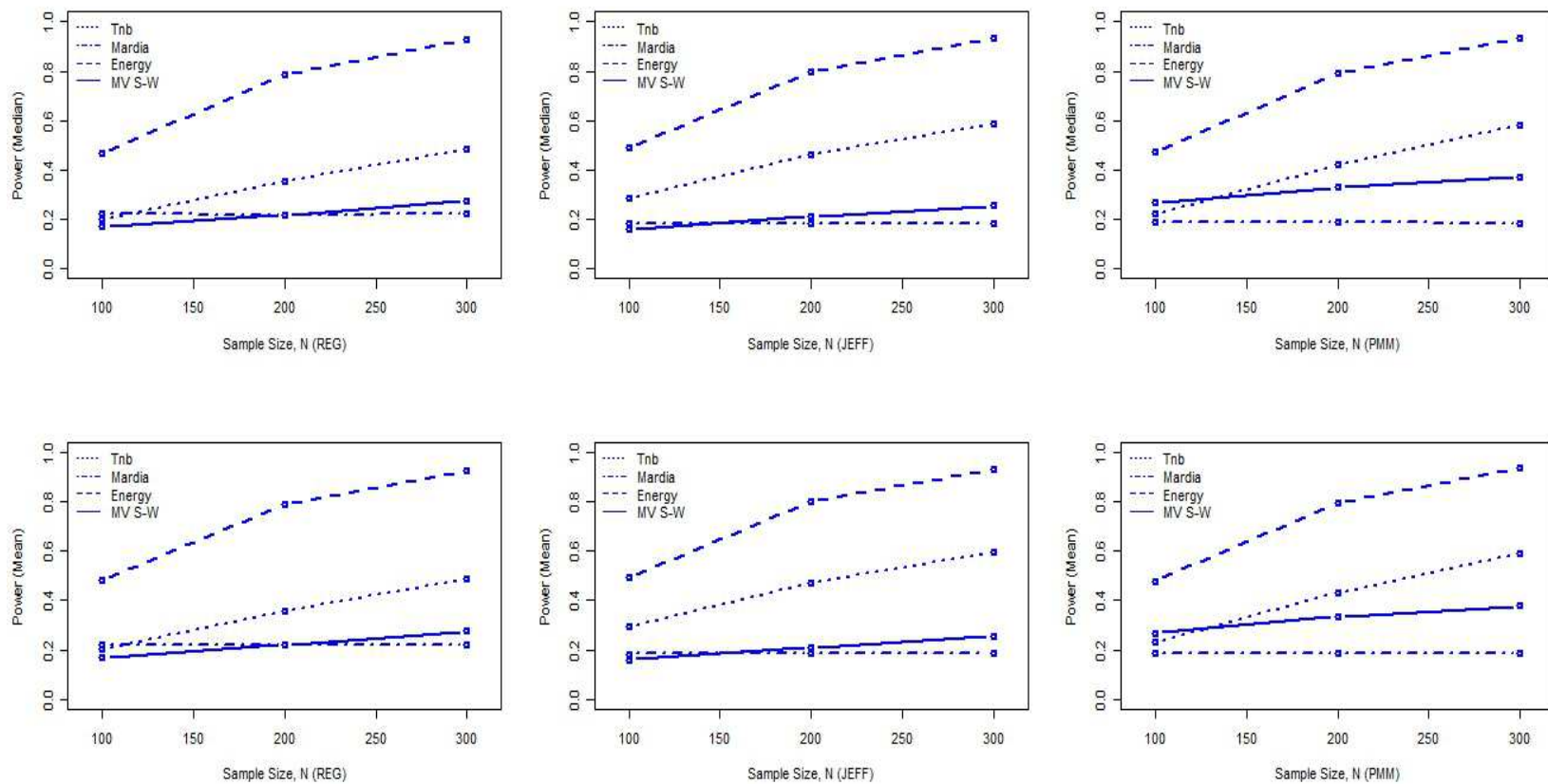
Figure D.3. $MVT(df = 25)$ model for $p = 2$.

Figure D.4. MVT($df = 25$) model for $p = 5$.

Figure D.5. MVT($df = 30$) model for $p = 2$.

Figure D.6. $MVT(df = 30)$ model for $p = 5$.

Figure D.7. MVST model for $p = 2$.

Figure D.8. MVST model for $p = 5$.

APPENDIX E

Power Analysis Tables

Table E.1. MVSN model for $p = 2$.

		$T_{N,\beta}$	Mardia	Energy	Shapiro-Wilks
REG	100	0.1782(0.0017)	0.1486(0.0016)	0.3388(0.0021)	0.1143(0.0014)
		0.1784(0.0017)	0.1480(0.0016)	0.3375(0.0021)	0.1140(0.0014)
	200	0.2914(0.0020)	0.1553(0.0016)	0.6541(0.0021)	0.1308(0.0015)
		0.2890(0.0020)	0.1542(0.0016)	0.6545(0.0021)	0.1298(0.0015)
	300	0.4078(0.0022)	0.1542(0.0016)	0.8515(0.0016)	0.1415(0.0016)
		0.4078(0.0022)	0.1543(0.0016)	0.8508(0.0016)	0.1418(0.0016)
JEFF	100	0.2094(0.0018)	0.1315(0.0015)	0.3545(0.0021)	0.1158(0.0014)
		0.2110(0.0018)	0.1308(0.0015)	0.3499(0.0021)	0.1146(0.0014)
	200	0.3377(0.0021)	0.1391(0.0015)	0.6687(0.0021)	0.1282(0.0015)
		0.3372(0.0021)	0.1386(0.0015)	0.6689(0.0021)	0.1285(0.0015)
	300	0.4607(0.0022)	0.1380(0.0015)	0.8613(0.0015)	0.1360(0.0015)
		0.4596(0.0022)	0.1371(0.0015)	0.8599(0.0016)	0.1372(0.0015)
PMM	100	0.1938(0.0018)	0.1328(0.0015)	0.3500(0.0021)	0.1470(0.0016)
		0.1918(0.0018)	0.1331(0.0015)	0.3508(0.0021)	0.1456(0.0016)
	200	0.3231(0.0021)	0.1394(0.0015)	0.6703(0.0021)	0.1662(0.0017)
		0.3231(0.0021)	0.1394(0.0015)	0.6686(0.0021)	0.1650(0.0017)
	300	0.4559(0.0022)	0.1376(0.0015)	0.8665(0.0015)	0.1849(0.0017)
		0.4529(0.0022)	0.1372(0.0015)	0.8643(0.0015)	0.1811(0.0017)

Table E.2. MVSN model for $p = 5$.

		$T_{N,\beta}$	Mardia	Energy	Shapiro-Wilks
REG	100	0.1594(0.0016)	0.1673(0.0017)	0.2971(0.0020)	0.1226(0.0015)
		0.1603(0.0016)	0.1677(0.0017)	0.2882(0.0020)	0.1234(0.0015)
	200	0.2526(0.0019)	0.1656(0.0017)	0.5588(0.0022)	0.1305(0.0015)
		0.2505(0.0019)	0.1658(0.0017)	0.5580(0.0022)	0.1288(0.0015)
	300	0.3303(0.0021)	0.1634(0.0017)	0.7690(0.0019)	0.1518(0.0016)
		0.3266(0.0021)	0.1650(0.0017)	0.7674(0.0019)	0.1507(0.0016)
JEFF	100	0.2004(0.0018)	0.1327(0.0015)	0.3135(0.0021)	0.1188(0.0014)
		0.1954(0.0018)	0.1328(0.0015)	0.3144(0.0021)	0.1197(0.0015)
	200	0.3045(0.0021)	0.1366(0.0015)	0.5807(0.0022)	0.1265(0.0015)
		0.3020(0.0021)	0.1350(0.0015)	0.5792(0.0022)	0.1274(0.0015)
	300	0.3826(0.0022)	0.1314(0.0015)	0.7784(0.0019)	0.1424(0.0016)
		0.3804(0.0022)	0.1309(0.0015)	0.7774(0.0019)	0.1413(0.0016)
PMM	100	0.1877(0.0017)	0.1374(0.0015)	0.3041(0.0021)	0.2228(0.0019)
		0.1834(0.0017)	0.1383(0.0015)	0.3011(0.0021)	0.2220(0.0019)
	200	0.3162(0.0021)	0.1387(0.0015)	0.5883(0.0022)	0.2473(0.0019)
		0.3097(0.0021)	0.1384(0.0015)	0.5871(0.0022)	0.2440(0.0019)
	300	0.4164(0.0022)	0.1307(0.0015)	0.8010(0.0018)	0.2567(0.0020)
		0.4147(0.0022)	0.1309(0.0015)	0.7987(0.0018)	0.2516(0.0019)

Table E.3. MVT model with $df = 25$ for $p = 2$.

		$T_{N,\beta}$	Mardia	Energy	Shapiro-Wilks
REG	100	0.2007(0.0018)	0.2386(0.0019)	0.5235(0.0022)	0.2348(0.0019)
		0.1994(0.0018)	0.2369(0.0019)	0.5200(0.0022)	0.2313(0.0019)
	200	0.3697(0.0022)	0.2460(0.0019)	0.7475(0.0019)	0.3485(0.0021)
		0.3655(0.0022)	0.2446(0.0019)	0.7444(0.0020)	0.3436(0.0021)
	300	0.5337(0.0022)	0.2497(0.0019)	0.8731(0.0015)	0.4376(0.0022)
		0.5323(0.0022)	0.2499(0.0019)	0.8709(0.0015)	0.4358(0.0022)
JEFF	100	0.3396(0.0021)	0.2344(0.0019)	0.5127(0.0022)	0.2303(0.0019)
		0.3255(0.0021)	0.2316(0.0019)	0.5067(0.0022)	0.2297(0.0019)
	200	0.5062(0.0022)	0.2486(0.0019)	0.7337(0.0020)	0.3431(0.0021)
		0.4879(0.0022)	0.2461(0.0019)	0.7330(0.0020)	0.3445(0.0021)
	300	0.6425(0.0021)	0.2555(0.0020)	0.8612(0.0015)	0.4316(0.0022)
		0.6240(0.0022)	0.2530(0.0019)	0.8586(0.0016)	0.4342(0.0022)
PMM	100	0.1942(0.0018)	0.2329(0.0019)	0.5071(0.0022)	0.2164(0.0018)
		0.1924(0.0018)	0.2327(0.0019)	0.5061(0.0022)	0.2142(0.0018)
	200	0.3587(0.0021)	0.2453(0.0019)	0.7269(0.0020)	0.3202(0.0021)
		0.3571(0.0021)	0.2442(0.0019)	0.7237(0.0020)	0.3187(0.0021)
	300	0.5122(0.0022)	0.2522(0.0019)	0.8529(0.0016)	0.4148(0.0022)
		0.5098(0.0022)	0.2517(0.0019)	0.8500(0.0016)	0.4119(0.0022)

Table E.4. MVT model with $df = 25$ for $p = 5$.

		$T_{N,\beta}$	Mardia	Energy	Shapiro-Wilks
REG	100	0.1981(0.0018)	0.2386(0.0019)	0.4912(0.0022)	0.2258(0.0019)
		0.1951(0.0018)	0.2385(0.0019)	0.4784(0.0022)	0.2259(0.0019)
	200	0.3659(0.0022)	0.2434(0.0019)	0.6962(0.0021)	0.3217(0.0021)
		0.3587(0.0021)	0.2432(0.0019)	0.6913(0.0021)	0.3171(0.0021)
	300	0.5024(0.0022)	0.2412(0.0019)	0.8306(0.0017)	0.4156(0.0022)
		0.4941(0.0022)	0.2414(0.0019)	0.8261(0.0017)	0.4106(0.0022)
JEFF	100	0.3127(0.0021)	0.2334(0.0019)	0.4631(0.0022)	0.2149(0.0018)
		0.2920(0.0020)	0.2288(0.0019)	0.4613(0.0022)	0.2148(0.0018)
	200	0.4587(0.0022)	0.2460(0.0019)	0.6720(0.0021)	0.3143(0.0021)
		0.4387(0.0022)	0.2416(0.0019)	0.6694(0.0021)	0.3164(0.0021)
	300	0.5599(0.0022)	0.2434(0.0019)	0.7994(0.0018)	0.4046(0.0022)
		0.5386(0.0022)	0.2407(0.0019)	0.7967(0.0018)	0.4036(0.0022)
PMM	100	0.1758(0.0017)	0.2280(0.0019)	0.4403(0.0022)	0.2130(0.0018)
		0.1698(0.0017)	0.2278(0.0019)	0.4349(0.0022)	0.2109(0.0018)
	200	0.3275(0.0021)	0.2428(0.0019)	0.6484(0.0021)	0.2956(0.0020)
		0.3184(0.0021)	0.2418(0.0019)	0.6423(0.0021)	0.2927(0.0020)
	300	0.4445(0.0022)	0.2399(0.0019)	0.7777(0.0019)	0.3662(0.0022)
		0.4384(0.0022)	0.2394(0.0019)	0.7731(0.0019)	0.3629(0.0022)

Table E.5. MVT model with $df = 30$ for $p = 2$.

		$T_{N,\beta}$	Mardia	Energy	Shapiro-Wilks
REG	100	0.1679(0.0017)	0.2180(0.0018)	0.4268(0.0022)	0.2013(0.0018)
		0.1677(0.0017)	0.2172(0.0018)	0.4231(0.0022)	0.1980(0.0018)
	200	0.2823(0.0020)	0.2259(0.0019)	0.6253(0.0022)	0.2814(0.0020)
		0.2791(0.0020)	0.2243(0.0019)	0.6227(0.0022)	0.2779(0.0020)
	300	0.3997(0.0022)	0.2271(0.0019)	0.7560(0.0019)	0.3461(0.0021)
		0.3984(0.0022)	0.2273(0.0019)	0.7527(0.0019)	0.3443(0.0021)
JEFF	100	0.2726(0.0020)	0.2160(0.0018)	0.4171(0.0022)	0.1982(0.0018)
		0.2642(0.0020)	0.2123(0.0018)	0.4108(0.0022)	0.1975(0.0018)
	200	0.3943(0.0022)	0.2291(0.0019)	0.6078(0.0022)	0.2777(0.0020)
		0.3795(0.0022)	0.2265(0.0019)	0.6060(0.0022)	0.2796(0.0020)
	300	0.5012(0.0022)	0.2312(0.0019)	0.7381(0.0020)	0.3427(0.0021)
		0.4845(0.0022)	0.2293(0.0019)	0.7342(0.0020)	0.3444(0.0021)
PMM	100	0.1609(0.0016)	0.2140(0.0018)	0.4133(0.0022)	0.1859(0.0017)
		0.1581(0.0016)	0.2136(0.0018)	0.4114(0.0022)	0.1840(0.0017)
	200	0.2695(0.0020)	0.2275(0.0019)	0.5998(0.0022)	0.2608(0.0020)
		0.2694(0.0020)	0.2271(0.0019)	0.5965(0.0022)	0.2599(0.0020)
	300	0.3787(0.0022)	0.2292(0.0019)	0.7284(0.0020)	0.3328(0.0021)
		0.3763(0.0022)	0.2285(0.0019)	0.7249(0.0020)	0.3306(0.0021)

Table E.6. MVT model with $df = 30$ for $p = 5$.

		$T_{N,\beta}$	Mardia	Energy	Shapiro-Wilks
REG	100	0.1695(0.0017)	0.2208(0.0019)	0.4119(0.0022)	0.1982(0.0018)
		0.1686(0.0017)	0.2212(0.0019)	0.4001(0.0022)	0.1981(0.0018)
	200	0.2841(0.0020)	0.2255(0.0019)	0.5774(0.0022)	0.2645(0.0020)
		0.2791(0.0020)	0.2252(0.0019)	0.5718(0.0022)	0.2587(0.0020)
	300	0.3860(0.0022)	0.2227(0.0019)	0.7122(0.0020)	0.3354(0.0021)
		0.3789(0.0022)	0.2228(0.0019)	0.7057(0.0020)	0.3314(0.0021)
JEFF	100	0.2609(0.0020)	0.2161(0.0018)	0.3872(0.0022)	0.1882(0.0017)
		0.2463(0.0019)	0.2135(0.0018)	0.3858(0.0022)	0.1891(0.0018)
	200	0.3617(0.0021)	0.2266(0.0019)	0.5518(0.0022)	0.2609(0.0020)
		0.3458(0.0021)	0.2222(0.0019)	0.5478(0.0022)	0.2632(0.0020)
	300	0.4358(0.0022)	0.2220(0.0019)	0.6728(0.0021)	0.3272(0.0021)
		0.4176(0.0022)	0.2201(0.0019)	0.6694(0.0021)	0.3269(0.0021)
PMM	100	0.1508(0.0016)	0.2111(0.0018)	0.3680(0.0022)	0.1910(0.0018)
		0.1458(0.0016)	0.2104(0.0018)	0.3639(0.0022)	0.1894(0.0018)
	200	0.2528(0.0019)	0.2237(0.0019)	0.5317(0.0022)	0.2502(0.0019)
		0.2455(0.0019)	0.2222(0.0019)	0.5280(0.0022)	0.2477(0.0019)
	300	0.3289(0.0021)	0.2205(0.0019)	0.6491(0.0021)	0.3013(0.0021)
		0.3252(0.0021)	0.2204(0.0019)	0.6444(0.0021)	0.2985(0.0020)

Table E.7. MVST model with $df = 50$, $\nu = \mathbf{100}$ for $p = 2$.

		$T_{N,\beta}$	Mardia	Energy	Shapiro-Wilks
REG	100	0.2254(0.0019)	0.2033(0.0018)	0.5440(0.0022)	0.1630(0.0017)
		0.2253(0.0019)	0.2018(0.0018)	0.5428(0.0022)	0.1619(0.0016)
	200	0.4085(0.0022)	0.2086(0.0018)	0.8557(0.0016)	0.2175(0.0018)
		0.4050(0.0022)	0.2069(0.0018)	0.8545(0.0016)	0.2164(0.0018)
	300	0.5927(0.0022)	0.2154(0.0018)	0.9668(0.0008)	0.2674(0.0020)
		0.5930(0.0022)	0.2154(0.0018)	0.9664(0.0008)	0.2673(0.0020)
JEFF	100	0.3208(0.0021)	0.1823(0.0017)	0.5567(0.0022)	0.1619(0.0016)
		0.3172(0.0021)	0.1809(0.0017)	0.5514(0.0022)	0.1602(0.0016)
	200	0.5216(0.0022)	0.1906(0.0018)	0.8630(0.0015)	0.2125(0.0018)
		0.5128(0.0022)	0.1901(0.0018)	0.8629(0.0015)	0.2129(0.0018)
	300	0.6927(0.0021)	0.1994(0.0018)	0.9698(0.0008)	0.2580(0.0020)
		0.6829(0.0021)	0.1984(0.0018)	0.9691(0.0008)	0.2588(0.0020)
PMM	100	0.2427(0.0019)	0.1839(0.0017)	0.5525(0.0022)	0.1931(0.0018)
		0.2396(0.0019)	0.1839(0.0017)	0.5525(0.0022)	0.1914(0.0018)
	200	0.4467(0.0022)	0.1912(0.0018)	0.8622(0.0015)	0.2496(0.0019)
		0.4465(0.0022)	0.1912(0.0018)	0.8607(0.0015)	0.2485(0.0019)
	300	0.6431(0.0021)	0.1979(0.0018)	0.9707(0.0008)	0.3063(0.0021)
		0.6403(0.0021)	0.1971(0.0018)	0.9702(0.0008)	0.3010(0.0021)

Table E.8. MVST model with $df = 50$, $\nu = \mathbf{100}$ for $p = 5$.

		$T_{N,\beta}$	Mardia	Energy	Shapiro-Wilks
REG	100	0.2011(0.0018)	0.2222(0.0019)	0.4802(0.0022)	0.1697(0.0017)
		0.2017(0.0018)	0.2228(0.0019)	0.4697(0.0022)	0.1705(0.0017)
	200	0.3590(0.0021)	0.2208(0.0019)	0.7874(0.0018)	0.2206(0.0019)
		0.3549(0.0021)	0.2206(0.0019)	0.7861(0.0018)	0.2178(0.0018)
	300	0.4886(0.0022)	0.2225(0.0019)	0.9282(0.0012)	0.2766(0.0020)
		0.4838(0.0022)	0.2231(0.0019)	0.9266(0.0012)	0.2754(0.0020)
JEFF	100	0.2961(0.0020)	0.1838(0.0017)	0.4937(0.0022)	0.1598(0.0016)
		0.2856(0.0020)	0.1821(0.0017)	0.4911(0.0022)	0.1618(0.0016)
	200	0.4721(0.0022)	0.1892(0.0018)	0.8012(0.0018)	0.2102(0.0018)
		0.4621(0.0022)	0.1857(0.0017)	0.7990(0.0018)	0.2116(0.0018)
	300	0.5967(0.0022)	0.1874(0.0017)	0.9328(0.0011)	0.2568(0.0020)
		0.5865(0.0022)	0.1865(0.0017)	0.9327(0.0011)	0.2553(0.0019)
PMM	100	0.2303(0.0019)	0.1877(0.0017)	0.4789(0.0022)	0.2679(0.0020)
		0.2238(0.0019)	0.1882(0.0017)	0.4753(0.0022)	0.2666(0.0020)
	200	0.4333(0.0022)	0.1893(0.0018)	0.7971(0.0018)	0.3348(0.0021)
		0.4219(0.0022)	0.1879(0.0017)	0.7948(0.0018)	0.3302(0.0021)
	300	0.5890(0.0022)	0.1865(0.0017)	0.9374(0.0011)	0.3773(0.0022)
		0.5844(0.0022)	0.1856(0.0017)	0.9361(0.0011)	0.3725(0.0022)

BIBLIOGRAPHY

- Agresti, A. (2000), “Strategies for comparing treatments on a binary response with multi-centre data.” *Statistics in Medicine*, 1115–1139.
- Arbyn, M., Castellsague, X., de Sanjose, S., Bruni, L., Saraiya, M., and et al., F. B. (submitted 2010), “Worldwide burden of cervical cancer in 2008.” *Worldwide burden of cervical cancer in 2008*.
- Arbyn, M., Gaffikin, L., Sankaranarayanan, R., Muwonge, R., Keita, N., Wesley, R., and TSU, V. (2006), “Assessment of Innovative Approaches to Cervical Cancer Screening, Follow-Up and Treatment of Screen Detected Cervical Lesions in Developing Countries: A Pooled Analysis of ACCP Study Results.” *IPH/EPI REPORTS*.
- Berry, G., Smith, C. L., Macaskill, P., and Irwig, L. (2002), “Analytic methods for comparing two dichotomous screening or diagnostic tests applied to two populations of differing disease prevalence when individuals negative on both tests are unverified.” *Statistics in Medicine*., 21.
- Dendukuri, N. and Joseph, L. (2001), “Bayesian Approaches to Modeling the Conditional Dependence between Multiple Diagnostic Tests.” *Biometrics*., 57, 158–167.
- Dendukuri, N., Rahme, E., Belisle, P., and Joseph, L. (2004), “Bayesian Sample Size Determination for Prevalence and Diagnostic Test Studies in the Absence of a Gold Standard Test.” *Biometrics*, 60, 388–397.
- Doornik, J. A. and Hansen, H. (2008), “An Omnibus Test for Univariate and Multivariate Normality.” *Oxford Bulletin of Economics and Statistics*., 70.
- Farnsworth, M. L., Hoeting, J. A., Hobbs, N. T., and Miller, M. W. (2006), “Linking Chronic Wasting Disease to Mule Deer Movement Scales: A Hierarchical Bayesian Approach.” *Ecological Applications*., 16, 1026–1036.
- Farrell, P. J., Salibian-Barrera, M., and Naczkz, K. (2006), “On tests for multivariate normality and associated simulation studies.” *Journal of Statistical Computation & Simulation*.
- Ferlay, J., Shin, H. R., Bray, F., Forman, D., Mathers, C., and Parkin, D. M. (2010), “Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10.” *Worldwide burden of cervical cancer in 2008*.
- Gakidou, E., Nordhagen, S., and Obermeyer, Z. (2008), “Coverage of cervical cancer screening in 57 countries: low average levels and large inequalities.” *Public Library of Science Medicine*., 5, 863–868.

- Ganjali, M. and Ranji, H. (2008), "A Comparison of Several Algorithms and Models for Analyzing Multivariate Normal Data with Missing Responses." *Applications and Applied Mathematics: An International Journal.*, 3, 55–68.
- Hanson, T. E., Johnson, W. O., and Gastwirth, J. L. (2006), "Bayesian inference for prevalence and diagnostic test accuracy based on dual-pooled screening." *Biostatistics*, 7, 41–57.
- Henze, N. (2002), "Invariant tests for multivariate normality: a critical review." *Statistical Papers.*, 43.
- Henze, N. and Zirkler, B. (1990), "A class of invariant consistent tests for multivariate normality." *Communications in Statistics - Theory and Methods.*, 19, 3595–3617.
- Joseph, L., Gyorkos, T. W., and Coupal, L. (1995), "Bayesian Estimation of Disease Prevalence and the Parameters of Diagnostic Tests in the Absence of a Gold Standard." *American Journal of Epidemiology*, 141, 263–272.
- Katz, I. T. and Wright, A. A. (2006), "Preventing cervical cancer in the developing world." *The New England Journal of Medicine.*, 354, 1110.
- Kostoulas, P., Leontides, L., Browne, W. J., and Gardner, I. A. (2009), "Bayesian estimation of variance partition coefficients adjusted for imperfect test sensitivity and specificity." *Preventive Veterinary Medicine*, 89, 155–162.
- MacNab, Y. C. (2003), "Hierarchical Bayesian Modeling of Spatially Correlated Health Service Outcome and Utilization Rates." *Biometrics.*, 59, 305–316.
- Mardia, K. V. (1975), "Assessment of Multinormality and the Robustness of Hotelling's T^2 Test." *Journal of the Royal Statistical Society. Series C (Applied Statistics).*, 24, 163–171.
- Martinez, E. Z., Louzada-Neto, F., Derchain, S. F., Achcar, J. A., Gontijo, R. C., and et al., L. O. S. (2008), "Bayesian estimation of performance measures of cervical cancer screening tests in the presence of covariates and absence of a gold standard." *Cancer Informatics*.
- Mecklin, C. J. and Mundfrom, D. J. (2004), "An Appraisal and Bibliography of Tests for Multivariate Normality." *International Statistical Review.*, 72, 123–138.
- Oleson, J. J. and He, C. Z. (2004), "Hierarchical Bayesian Modeling in Dichotomous Processes in the Presence of Nonresponse." *Biometrics.*, 60, 50–59.
- Putter, H., Fiocco, M., and Stijnen, T. (2010), "Meta-Analysis of Diagnostic Test Accuracy Studies with Multiple Thresholds using Survival Methods." *Biometrical Journal*, 52, 95–110.
- Reiter, J. P., Kohnen, C. N., Karr, A. F., Lin, X., and Sanil, A. P. (2004), "Secure Regression for Vertically Partitioned, Partially Overlapping Data." *ASA Proceedings of the Joint Statistical Meetings*.

- Rosenberg, M. A., Andrews, R. W., and Lenk, P. J. (1999), “A Hierarchical Bayesian Model for Predicting the Rate of Nonacceptable In-Patient Hospital Utilization.” *Journal of Business & Economic Statistics.*, 17, 1–8.
- Schafer, J. L. (1999), *Analysis of Incomplete Multivariate Data.*, Boca Raton, FL: Chapman & Hall/CRC.
- Shapiro, S. S. and Wilk, M. B. (1965), “An Analysis of Variance Test for Normality (Complete Samples).” *Biometrika.*, 52, 591–611.
- Srivastava, M. S. and Dolatabadi, M. (2009), “Multiple imputation and other resampling schemes for imputing missing observations.” *Journal of Multivariate Analysis.*, 100.
- Stamey, J. D., Boese, D. H., and Young, D. M. (2008), “Confidence intervals for parameters of two diagnostic tests in the absence of a gold standard.” *Computational Statistics and Data Analysis*, 52, 1335–1346.
- Szekely, G. J. and Rizzo, M. L. (2005), “A New Test for Multivariate Normality.” *Journal of Multivariate Analysis.*, 93, 58–80.
- de Leon, A. R., Soo, A., Bonzo, D. C., and Rudnisky, C. J. (2009), “Joint Estimation of Diagnostic Accuracy Measures for Paired Organs Application in Ophthalmology.” *Biometrical Journal*, 51, 837–850.
- Villasenor-Alva, J. A. and Gonzalez-Estrada, E. (2009), “A generalization of Shapiro-Wilk’s test for multivariate normality.” *Communications in Statistics: Theory and Methods.*, 38, 1870–1883.
- Walter, S. D. (1999), “Estimation of test sensitivity and specificity when disease confirmation is limited to positive results.” *Epidemiology*, 10, 67–72.
- Yamada, T., Romer, M. M., and Richards, D. P. S. (2010), “Kurtosis Tests for Multivariate Normality with Monotone Incomplete Data.” *Submitted for publication.*
- Zhang, P. (2003), “Multiple Imputation: Theory and Method.” *International Statistical Review.*, 71, 581–592.
- Zhou, X. H. (1998), “Comparing Accuracies of Two Screening Tests in a Two-Phase Study for Dementia.” *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 47, 135–147.