12-16-2020

Novel Estimation Methods for Sensitivity And AUC in Medical Diagnostic Study

Yan Hai

Recommended Citation
Hai, Yan, "Novel Estimation Methods for Sensitivity And AUC in Medical Diagnostic Study." Dissertation, Georgia State University, 2020.
doi: https://doi.org/10.57709/20321928

NOVEL ESTIMATION METHODS FOR SENSITIVITY AND AUC IN MEDICAL DIAGNOSTIC STUDY

by

YAN HAI

Under the Direction of Gengsheng Qin, PhD

ABSTRACT

In the medical diagnostic study, the accuracy of a diagnostic test is commonly evaluated based on its sensitivity and specificity. Both sensitivity and specificity are not fixed but depend on the cutoff chosen for that test. The receiver operating characteristic (ROC) curve of the test is constructed to show how sensitivity and specificity change as the cutoff varies. The area under the ROC curve (AUC) can also be used to evaluate the discriminatory ability of diagnostic tests with continuous test results. In practice, however, the cutoff of a test is

usually chosen so that the specificity is meaningfully high. Therefore, the sensitivity under a certain specificity serves as a diagnostic measure to evaluate the diagnostic tests.

In both two and three (the normal healthy stage, the early stage of the disease, and the stage of the full development of the disease) diagnostic classes studies, we propose a new influence function-based empirical likelihood method and Bayesian empirical likelihood methods. The proposed methods are shown to perform better than the existing methods in terms of both coverage probability and interval length in simulation studies. A real data set from Alzheimer's Disease Neuroimaging Initiative (ANDI) is analyzed by using the newly proposed methods.

In two-phase diagnostic studies with both screening test and gold standard test, verification of the true disease status might be partially missing based on the results of diagnostic tests and other subjects' characteristics. Because the estimators of AUC based on partially validated subjects are usually biased, it is usually necessary to estimate AUC by bias-corrected methods. We proposed direct estimators of the AUC based on hybrid imputation (FI and MSI), inverse probability weighting (IPW), and the semi-parametric efficient (SPE) approach with verification biased data when the test result is continuous under the assumption that the true disease status, if missing, is missing at random (MAR). Simulation results show that the proposed estimators are accurate for the biased sampling. We illustrate the proposed methods with a real data set of Neonatal Hearing Screening study.

KEYWORDS: Sensitivity, Specificity, ROC Curve, AUC, Verification Bias, Empirical Likelihood, Influence Function, Bayesian Inference.

NOVEL ESTIMATION METHODS FOR SENSITIVITY AND AUC IN MEDICAL DIAGNOSTIC STUDY

by

YAN HAI

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2020

NOVEL ESTIMATION METHODS FOR SENSITIVITY AND AUC IN MEDICAL DIAGNOSTIC STUDY

by

YAN HAI

Committee Chair: Gengsheng Qin

Committee: Yi Jiang

Ruiyan Luo

Jing Zhang

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2020

DEDICATION

This dissertation is dedicated to my supportive husband, our sweet little girl, always encouraging parents, advisors, and friends.

ACKNOWLEDGEMENTS

I would like to thank to my PhD advisors, Drs. Gengsheng Qin and Xiaoyi Min, for supporting and advising me during these past five years. Dr. Qin is a brilliant and kind advisor with infinite enthusiasm for creation. I am very grateful to him for his advice and valuable time offered to me, for the inspiring discussion on research, and the friendly encouragement. Dr. Min has provided insightful discussions about the research and trained me to an independent researcher. Without their help, I can not publish my papers on the top journals.

I want to thank my dissertation committee, Dr. Yi Jiang, Dr. Jing Zhang, and Dr. Ruiyan Luo, for providing invaluable advice and comments regarding the dissertation material and the presentation. My appreciation goes to Haiqi Wang and Binghuan Wang for their enthusiastic discussions with me, sharing data and unique experiences regarding the theory behind topics. I will miss the time I spent with my officemate and fellow students, Xinjie Hu, Guanhao Wei, Guangming Jin, Yuyin Shi, Shushan He, Yan Cao, and Bin Liu. I thank them for the stimulating discussion on my research and all help they provided for my dissertation. I wish to thank the fantastic group of staffs and professors in the mathematics and statistics department, Dr. Guantao Chen, Dr. Alexandra Smirnova, Dr. Zhongshan Li, Dr. Changyong Zhong, Dr. Yichuan Zhao, Dr. Xin Qi, Dr. Michael Stewart, Sandra Ahuama-Jonas, Beth Conner, Katina Akins, and Earnestine Collier, who worked so hard to take good care of us and provided generous help for our research and teaching duties.

I would like to thank the Molecular Basis of Disease (MBD) program to provide the PhD Fellowship for me. I appreciate all the help from Dr. Jenny Yang and all the members from her Lab from the Chemistry Department. I appreciate them giving me the chance to join in their study 'Early detection and staging of chronic liver diseases with a protein MRI contrast agent', which has published in Nature Communications.

Data collection and sharing for the first two parts of he dissertation was funded by the

Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimers Association; Alzheimers Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (`www.fnih.org`). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

- GSU - Georgia State University
- ROC - Receiver Operating Characteristic
- AUC - Area Under the ROC Curve
- VUS - Area Under the ROC surface
- Se - Sensitivity
- Spe - Specificity
- EL - Empirical Likelihood
- ANDI - Alzheimers Disease Neuroimaging Initiative
- NA - Normal Approximation
- BCa - Bias-corrected Acceleration
- JEL - Jackknife Empirical Likelihood
- HEL - Hybrid Empirical Likelihood
- IFEL - Influence Function-based Empirical Likelihood
- BEL - Bayesian Empirical Likelihood
- BTI - Bootstrap I
- BTII - Bootstrap II
- BHEL1 - Bayesian Hybrid Empirical Likelihood 1
- BHEL2 - Bayesian Hybrid Empirical Likelihood 2

- BEL1 - Bayesian Empirical Likelihood 1
- BEL2 - Bayesian Empirical Likelihood 2
- BIFEL1 - Bayesian Influence Function-based Empirical Likelihood 1
- BIFEL2 - Bayesian Influence Function-based Empirical Likelihood 2
- BpHEL1 - Bayesian Pseudo Hybrid Empirical Likelihood 1
- BpHEL2 - Bayesian Pseudo Hybrid Empirical Likelihood 2
- BpIFEL1 - Bayesian Pseudo Influence Function Empirical Likelihood 1
- BpIFEL2 - Bayesian Pseudo Influence Function Empirical Likelihood 2
- FI - Full Imputation
- MSI - Mean Score Imputation
- IPW - Inverse Probability Weighting
- SPE - Semi-parametric Efficient Approach
- MSE - Mean squared error
- cdf - Cumulative Distribution Function
- TSH - Thyroid-stimulating Hormone
- PSA - Prostate-specific Antigen
- MAR - Missing at Random
- MCI - Mild Cognitive Impairment
- AD - Alzheimer's Disease

PART 1

INTRODUCTION

Over the past 100 years, diagnostic testing has become a critical part of standard medical practice[1]. Every time people go to a physician with pain symptoms or an injury, they will undergo diagnostic tests. More than 6 billion diagnostic tests are performed every year in the United States[2]. Before a new diagnostic test can be introduced into medical practice, it should be evaluated for diagnostic accuracy in discriminating patients among different disease stages [3][4][5]. Generally, the diagnostic tests are utilized to classify the individuals into two groups, i.e., the healthy group (negative) and the diseased group (positive), or three ordinal groups, i.e., the healthy group (negative), the early diseased group (intermediate), and the fully diseased (positive) group. This dissertation develops estimation methods for evaluating the diagnostic accuracy of tests for two groups and three ordinal groups.

1.1 Statistical Evaluation of Diagnostic Tests

Diagnostic tests commonly measure different bio-markers of an individual in question, and the disease status of the individual is determined based on whether such measurements meet pre-specified criteria. For diagnostic tests classifying the individuals into two groups, the clinicians would identify the individual as diseased if the test result is above a specific cutoff and non-diseased if it is below the cutoff (or vice-versa). For example, the individuals will be considered hyperthyroid state if their serum thyroid-stimulating hormone (TSH) level is below 0.35 mIU per L[6]. Accuracy of a diagnostic test is a term that is frequently used to describe the evaluation of a diagnostic test versus a gold standard[7]. The evaluation of a diagnostic test is frequently based on a study in a selected population sampled according to the true disease status determined by the gold standard. For example, the evaluation is performed using a prostate-specific antigen (PSA) to detect or diagnose prostate cancer

versus a biopsy as a gold standard.

In a two diagnostic class problem, the probability that a non-diseased individual is correctly classified is defined as the specificity, and the probability that a diseased individual is correctly identified is called sensitivity. The accuracy of a diagnostic test is commonly evaluated based on its sensitivity and specificity. Both sensitivity and specificity are functions of the cutoff value when the outcome of a diagnostic test is continuous. Both sensitivity and specificity are not fixed but depend on the cutoff(s) chosen for that test. The receiver operating characteristic (ROC) curve, a plot of sensitivity vs. (1-specificity) as the cutoff value runs through the whole range of all possible outcome values, is an effective graphic tool in measuring the accuracy of a diagnostic test. Different statistics based on the ROC curve, such as the area under the ROC curve (AUC) [8] and Youden index [9], serve as the summary index measures to evaluate the discriminatory ability of the diagnostic tests.

In medical practice, the cutoff value of a test is usually chosen so that a fixed meaningfully high value of specificity is achieved [10]. For example, breast MRI screening study results usually have at least 80% specificity [11]. Therefore, the sensitivity of the test at a given specificity can also be used as an important diagnostic measure. Several papers[10][12][13][14][15][16] discussed the issues of estimation of sensitivity given the fixed level of specificity for evaluating diagnostic tests. These papers proposed normal approximation-based approaches and empirical likelihood (EL) based approaches. However, these methods generally have poor performance when the specificity is high. In Part 2, we propose a new influence function-based empirical likelihood interval for the sensitivity at a given specificity to overcome such problems. Moreover, followed Lazar [17]'s study on Bayesian empirical likelihoods, we develop several Bayesian EL (BEL) intervals for the sensitivity.

1.2 Sensitivity to the Early Diseased Stage

The disease process is usually divided into two stages, as we mentioned previously. However, in practice, a disease process might be more complicated and involve three diagnostic stages: the healthy stage, the early diseased stage, and the fully diseased stage. For example,

mild cognitive impairment (MCI) is a transitional stage between the cognitive changes of normal aging and the more severe Alzheimer's Disease (AD)[18]. In this three-ordinal-group classification problem, given a pair of threshold values c_1 and c_2 ($c_1 < c_2$), the individual is identified as non-diseased if the test result is smaller than c_1, as fully diseased if the test result is greater than c_2, and as early diseased if the test result is between c_1 and c_2. The specificity P_1 is the probability that the test correctly identifies the non-diseased individuals. The sensitivity P_2 to the early diseased stage and sensitivity P_3 to the fully diseased stage is the probability that the test correctly identifies the early diseased individuals and the fully diseased individuals, respectively.

The early detection of the disease, such as AD, is crucial because this usually means that patients can receive earlier treatment. Therefore, the probability associated with the early detection of the disease is a crucial accuracy measure for the diagnostic test with three ordinal stages. Studies [19][20] provided parametric and non-parametric EL confidence intervals for the sensitivity to the early diseased stage. However, we note that the empirical likelihood ratio in Dong and Tian [20] follows a scaled chi-square distribution asymptotically. Thus an extra step is required to estimate this scale. Similarly to Part 2 for the two-group classification problem, in Part 3, we proposed an influence function empirical likelihood-based confidence interval for sensitivity at a given value of the pair (P_1, P_3), i.e., the sensitivity of the test to the early diseased stage given the specificity and the sensitivity to the fully diseased stage. The corresponding empirical log-likelihood ratio statistic converges to a standard chi-square distribution, making inference for sensitivity more convenient. In addition, we proposed the Bayesian empirical likelihood (BEL and BpEL) intervals for sensitivity to the early stage.

1.3 Verification Bias

Medical diagnostic procedure usually involves two-phrase tests, diagnostic/screening test and gold standard test that verifies the true disease status. As mentioned above, a diagnostic test needs to be evaluated by a study with the selected true non-diseased and

true diseased samples determined according to a gold standard test. However, in many situations, not all individuals with given diagnostic/screening test results ultimately have their true disease status verified through a very accurate gold standard test. That is to say, the labels referred to as true disease status of the individuals are partially missing. One reason for the missing is that the gold standard test is usually costly and invasive. So the common practice is to apply the gold standard test only on high-risk individuals based on other diagnostic/screening test results. For example, only a part of individuals with a higher risk of prostate cancer, based on their prostate-specific antigen (PSA) levels, will be referred to undergo a biopsy. Because individuals at low-risk are more likely to have their true disease status missing, simply ignoring this missingness and using only individuals with verified disease status to do statistical inference may lead to bias. Such bias is called verification bias[21]. The missing at random (MAR) assumption[22] will be adopted to deal with these missing disease status. Under the MAR assumption, the probability of an individual being verified does not depend on the true disease status.

To correct the verification bias, Alonzo and Pepe[23] proposed several methods for estimating the sensitivity and the specificity; He et al.[24] developed a closed-form expression of the AUC estimator based on one imputation method; Adimari and Chiogna[25] proposed a fully non-parametric estimation of the AUC based on K nearest-neighbor imputation. Motivated by their studies, in Part 4, we derive new closed-form estimators of AUC with verification biased data when the test result is continuous under the assumption that the true disease status, if missing, is missing at random. The proposed AUC estimators can be easily computed and directly applied in practice.

PART 2

BAYESIAN AND INFLUENCE FUNCTION BASED EMPIRICAL LIKELIHOODS FOR INFERENCE OF SENSITIVITY IN DIAGNOSTIC TESTS

2.1 Introduction

In the medical diagnostic study, the accuracy of a diagnostic test is commonly evaluated based on its sensitivity and specificity. Both sensitivity and specificity are not fixed but depend on the cutoff(s) chosen for that test. The receiver operating characteristic (ROC) curve, a plot of sensitivity vs. (1-specificity) as the cutoff varies, is an effective graphic tool in measuring the diagnostic accuracy of diagnostic tests classifying individuals into two groups (non-diseased and diseased groups). Different statistics, including the area under the ROC curve (AUC) [8] and Youden index [9] are based on the ROC curve. They are used to evaluate the discriminatory ability of diagnostic tests. In practice, however, the cutoff of a test is usually chosen so that the specificity is meaningfully high (typically 80%, 90%, or 95%) [10][20]. Hence, the sensitivity given the specificity serves as an essential measure of the accuracy of diagnostic tests.

Existing methods for interval estimation of sensitivity include normal approximation-based approaches and empirical likelihood-based approaches. Linnet [13] proposed a normal approximation (NA) based confidence interval. Platt *et al.* [10] noted that Linnet's method may be significantly affected by poor empirical density estimation and proposed to use Efron's bias-corrected acceleration (BCa) bootstrap interval [26]. Zhou and Qin [14] developed two bootstrap intervals for sensitivity based on extensions of Agresti and Coull's results on confidence intervals for binomial proportions [27] and showed that the new bootstrap intervals have better coverage accuracy than the NA and BCa intervals. Empirical likelihood (EL), introduced by Owen [28], has become a popular approach in statistical research be-

cause it does not rely on parametric assumptions on the data but still enjoys the advantages of likelihood methods. In particular, EL methods have been used widely in the evaluation of diagnostic tests. For example, Gong *et al.* [15] proposed a smoothed jackknife empirical likelihood (JEL) method for the ROC curve. Qin *et al.* [16] developed a hybrid EL (HEL) method for sensitivity.

This part provides a thorough review of these methods and comprehensive numerical studies to compare their performance in different settings. Moreover, we study two improvements of current methods. Firstly, we note that in the approach of Qin *et al.* [16], the hybrid empirical likelihood ratio follows a scaled chi-square distribution asymptotically. Thus an extra step is required to estimate this scale. We propose a new influence function-based EL (IFEL) method following Yu *et al.* [29]. The idea is to replace the estimating function in the EL with an influence function of the sensitivity. The corresponding empirical log-likelihood ratio statistic converges to a standard chi-square distribution, making inference for sensitivity more convenient. Secondly, we develop Bayesian EL (BEL) approaches. Despite the extensive use in the frequentist context, EL has only recently been used in Bayesian analysis. Lazar [17] observed that EL's properties are in many respects similar to those of parametric likelihoods, and EL could be used in Bayesian inference like parametric likelihoods. To our knowledge, Bayesian EL methods have not been used in evaluating diagnostic tests. We propose to apply Bayesian EL methods to the inference of sensitivity. A critical part of Bayesian approaches is choosing an appropriate prior, which becomes quite tricky since no parametric model is assumed for EL. We consider building EL and assigning priors on either the sensitivity parameter itself or the probability vector $(p_1, \ldots, p_n)$ in building EL. For EL on the sensitivity parameter, we follow Clarke and Yuan [30] to derive reference priors [31][32][33]. Also, we apply the idea of Rao and Wu [34] to employ Bayesian EL based on $(p_1, \ldots, p_n)$.

This part is organized as follows. In Section 2.2, we review several existing methods for interval estimation of sensitivity. In Section 2.3, we introduce a new EL ratio statistic for sensitivity based on influence function. In Section 2.4, we propose Bayesian EL methods

based on influence function and hybrid methods. In Section 2.5, we conduct simulation studies to compare the performance of the proposed methods with existing methods. In Section 2.6, we apply the new methods to a real dataset to assess three biomarkers' diagnostic accuracy in the detection of Alzheimer's disease.

2.2 Existing Methods of Constructing Confidence Intervals for Sensitivity

Let Y and X be the results from a continuous-scale test for diseased and non-diseased individuals, respectively. Suppose individuals are diagnosed as diseased if the results are greater than a cutoff η, and non-diseased if the results are below η. The sensitivity and specificity of this test are defined by

$$\text{Sensitivity} = P(Y > \eta) = 1 - G(\eta), \quad \text{Specificity} = P(X \leq \eta) = F(\eta),$$

where G and F are the distribution functions of Y and X, respectively. Therefore, when the specificity of the test is p $(0 < p < 1)$, the corresponding sensitivity is $\theta = 1 - G(F^{-1}(p))$. Let $\{Y_1, \ldots, Y_n\}$ and $\{X_1, \ldots, X_m\}$ be the test results of a random sample of diseased subjects and non-diseased subjects, respectively. The statistical problem is to construct confidence intervals for the sensitivity θ at a fixed specificity p based on these observations.

2.2.1 Normal approximation-based bootstrap methods

As mentioned in the introduction section, Zhou and Qin [14] developed two bootstrap intervals for the sensitivity, which have better coverage accuracy than other normal-approximation based confidence intervals. Their methods used Agresti and Coull's idea for the construction of the confidence interval of a binomial proportion [27]. The first confidence interval called bootstrap I (BTI) interval, for θ is defined by

$$\left(\tilde{\theta} - z_{1-\alpha/2}\sqrt{V^*(\tilde{\theta})},\ \tilde{\theta} + z_{1-\alpha/2}\sqrt{V^*(\tilde{\theta})}\right),$$

where

$$\tilde{\theta} = \frac{\sum_{j=1}^{n} I[Y_j \geq \hat{F}^{-1}(p)] + \frac{1}{2}z_{1-\alpha/2}^2}{n + z_{1-\alpha/2}^2},$$

$\hat{F}^{-1}(p)$ is the p-th quantile of $\hat{F}$, $\hat{F}(x) = \frac{1}{m}\sum_{j=1}^{m} I(X_j \leq x)$ is the empirical distribution function of F, $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-th quantile of the standard normal distribution, and the variance $V^*(\tilde{\theta})$ of $\tilde{\theta}$ is estimated by the following bootstrap procedure:

(1) Draw a resample Y_j^*'s of size n and a separate resample X_i^*'s of size m with replacement, from the diseased sample Y_j's and the non-diseased sample X_i's, respectively.

(2) Calculate the bootstrap version of $\tilde{\theta}$:

$$\tilde{\theta}^* = \frac{\sum_{j=1}^{n} I[Y_j^* \geq \hat{F}^{*-1}(p)] + \frac{1}{2}z_{1-\alpha/2}^2}{n + z_{1-\alpha/2}^2},$$

where $\hat{F}^{*-1}(p)$ is the p-th sample quantile based on the bootstrap resample X_i^*'s.

(3) Repeat the first two steps B times ($B \geq 200$ is recommended) to obtain a set of bootstrap replications $\tilde{\theta}^{*(b)} (b = 1, \ldots, B)$. Then, the bootstrap variance estimator $V^*(\tilde{\theta})$ is defined by

$$V^*(\tilde{\theta}) = \frac{1}{B-1}\sum_{b=1}^{B}(\tilde{\theta}^{*(b)} - \bar{\theta}^*)^2,$$

where $\bar{\theta}^* = \frac{1}{B}\sum_{b=1}^{B}\tilde{\theta}^{*(b)}$.

The second confidence interval, called bootstrap II (BTII) interval, for θ is defined by

$$\left(\bar{\theta}^* - z_{1-\alpha/2}\sqrt{V^*(\tilde{\theta})},\ \bar{\theta}^* + z_{1-\alpha/2}\sqrt{V^*(\tilde{\theta})}\right).$$

BTI and BTII intervals are shown to have better coverage accuracy and shorter interval lengths than the normal approximation-based interval and the BCa interval regardless of the sample sizes and for both normal and non-normal data when specificity p is high. However, the coverage accuracy usually deteriorates for small sample sizes.

2.2.2 Jackknife empirical likelihood (JEL) method

Motivated by the JEL method for a U-statistic [35], Gong *et al.* [15] proposed a smoothed jackknife method by applying the standard EL method to the jackknife sample mean. A simple empirical estimator for θ is defined as

$$\tilde{\theta}_{m,n} = 1 - \hat{G}(\hat{F}^{-1}(p)), \tag{2.1}$$

where $\hat{G}(y) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \leq y)$ is the empirical distribution functions of G.

Gong *et al.* [15] defined a smooth version of $\tilde{\theta}_{m,n}$ as

$$\hat{\theta}_{m,n} = 1 - \frac{1}{n}\sum_{i=1}^{n} K(\frac{p - \hat{F}(Y_i)}{h}),$$

where $K(x) = \int_{-\infty}^{x} w(y)dy$, w is a symmetric density function with support $[-1, 1]$, and $h = h(n) > 0$ is a bandwidth.

Let

$$\hat{\theta}_{m,n,j} = 1 - \frac{1}{n-1}\sum_{1\leq i\leq n, i\neq j} K\left(\frac{p - \hat{F}(Y_i)}{h}\right), \text{ for } 1 \leq j \leq n,$$

and

$$\hat{\theta}_{m,n,j} = 1 - \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{p - \hat{F}_{m,j-n}(Y_i)}{h}\right), \text{ for } n < j \leq n+m,$$

where

$$\hat{F}_{m,k}(x) = \frac{1}{m-1}\sum_{1\leq j\leq m, j\neq k} I(X_j \leq x),\ k = 1, \ldots, m.$$

The jackknife pseudo-sample is then defined as

$$\hat{V}_j(p) = (n+m)\hat{\theta}_{m,n} - (n+m-1)\hat{\theta}_{m,n,j},\ j = 1, \ldots, n+m,$$

and the JEL for the sensitivity θ is defined as

$$L_{n,m}(\theta) = \sup_{\boldsymbol{p}} \left\{ \prod_{j=1}^{n+m} p_j : \sum_{j=1}^{n+m} p_j = 1, \sum_{j=1}^{n+m} p_j \hat{V}_j(p) = \theta \right\}.$$

By standard Lagrange multiplier arguments, the empirical log-likelihood ratio can be derived as

$$l_{n,m}(\theta) = -\sum_{j=1}^{n+m} \log \left\{ 1 + \lambda \big[\hat{V}_j(p) - \theta \big] \right\},$$

where λ satisfies

$$\frac{1}{n+m} \sum_{j=1}^{n+m} \frac{\hat{V}_j(p) - \theta}{1 + \lambda[\hat{V}_j(p) - \theta]} = 0.$$

Gong *et al.* [15] showed that $-2l_{n,m}(\theta)$ converges in distribution to the standard chi-square distribution with one degree of freedom, and a $100(1-\alpha)\%$ level JEL-based confidence interval on θ is given by

$$CI_{JEL}(\theta) = \{\theta : -2l_{n,m}(\theta) \leq \chi_1^2(1-\alpha)\},$$

where $\chi_1^2(1-\alpha)$ is the $(1-\alpha)$-th quantile of χ_1^2. As previously mentioned, JEL method needs to choose a bandwidth for kernel estimation.

2.2.3 Hybrid empirical likelihood (HEL) method

A challenge in constructing confidence intervals for sensitivity is to estimate the cut-off point that yields the desired specificity. Qin *et al.* [16] proposed a hybrid EL-based procedure which does not estimate an explicit cut-off point.

For a test value Y from a diseased subject, let $U = 1 - F(Y)$. U is called the placement value of Y and can be interpreted as the proportion of the non-diseased population with test value greater than Y. It essentially marks the placement of Y within the non-diseased distribution[36]. We have

$$E\left[I(U \leq 1-p)\right] = P(F(Y) \geq p) = P(Y \geq F^{-1}(p)) = \theta.$$

Based on this relationship between θ and U,

$$W_{Hj}(\theta, p) = I(U_j \leq 1 - p) - \theta, \tag{2.2}$$

can be used in EL as the estimating function. Qin *et al.* [16] proposed a profile EL which replaces F with the corresponding empirical distribution function as follow:

$$L_H(\theta) = \sup_{\boldsymbol{p}} \left\{ \prod_{j=1}^{n} p_j : \sum_{j=1}^{n} p_j = 1, \sum_{j=1}^{n} p_j \hat{W}_{Hj}(\theta, p) = 0 \right\}, \tag{2.3}$$

where $\hat{W}_{Hj}(\theta, p) = I(\hat{U}_j \leq 1 - p) - \theta$ with $\hat{U}_j = 1 - \hat{F}(Y_j)$, for $j = 1, \ldots, n$. The corresponding empirical log-likelihood ratio is

$$l_H(\theta) = -\sum_{j=1}^{n} \log\{1 + \lambda \hat{W}_{Hj}(\theta, p)\}, \tag{2.4}$$

where λ is the solution of

$$\frac{1}{n} \sum_{j=1}^{n} \frac{\hat{W}_{Hj}(\theta, p)}{1 + \lambda \hat{W}_{Hj}(\theta, p)} = 0.$$

The asymptotic distribution of $-2l_H(\theta)$ is a scaled chi-squared distribution with one degree of freedom. Thus, a $100(1 - \alpha)\%$ level hybrid EL and bootstrap confidence interval for θ can be constructed as follows:

$$CI_H(\theta) = \{\theta : -2c^* l_H(\theta) \leq \chi_1^2(1 - \alpha)\},$$

where c^* can be estimated from the following bootstrap procedure:

(1) Draw a resample Y_j^*'s of size n and a separate resample X_i^*'s of size m with replacement, from the diseased sample Y_j's and the non-diseased sample X_i's, respectively.

(2) Calculate the bootstrap estimator of θ:

$$\theta^* = \frac{\sum_{j=1}^{n} I[Y_j^* \geq \hat{F}^{*-1}(p)]}{n},$$

where $\hat{F}^{*-1}(p)$ is the p-th sample quantile based on the bootstrap resample X_i^*'s.

(3) Repeat the first two steps B times to obtain a set of bootstrap replications $\theta^{*b}(b = 1, \ldots, B)$. Then c^* is defined by

$$c^* = \frac{\bar{\theta}^{*b}(1-\bar{\theta}^{*b})}{\frac{n}{B-1}\sum_{b=1}^{B}(\theta^{*b}-\bar{\theta}^{*b})^2}, \tag{2.5}$$

where $\bar{\theta}^{*b} = \frac{1}{B}\sum_{b=1}^{B}\theta^{*b}$.

Qin *et al.* [16] showed that this HEL interval has good coverage probability when sample size (m, n) is greater than (50, 50). However, estimating c^* could be computationally expensive and not desirable.

2.3 Influence Function-Based Empirical Likelihood (IFEL) Method

Yu *et al.* [29] proposed an EL function based on influence functions of parameters of interest. Motivated by their study, we propose a new influence function-based EL method to construct confidence intervals for sensitivity. Recall that $\theta = 1 - G(F^{-1}(p))$, and $\tilde{\theta}_{m,n} = 1 - \hat{G}(\hat{F}^{-1}(p))$. Denote $\eta = F^{-1}(p) = G^{-1}(1-\theta)$, $\hat{\eta} = \hat{F}^{-1}(p)$ (i.e., the p-th sample quantile of X_i's), and the combined samples as

$$Z_k = \begin{cases} Y_k, & k = 1, \ldots, n, \\ X_{k-n}, & k = n+1, \ldots, n+m. \end{cases}$$

We have the following decomposition:

$$\tilde{\theta}_{m,n} - \theta \;\; = \;\; [G(\eta) - \hat{G}(\eta)] + [\hat{G}(\eta) - \hat{G}(\hat{\eta})]$$

with

$$G(\eta) - \hat{G}(\eta) = \frac{1}{n}\sum_{i=1}^{n}[I(Y_i > \eta) - \theta] = \frac{1}{m+n}\sum_{k=1}^{n}\frac{m+n}{n}[I(Z_k > \eta) - \theta].$$

From the Bahadur representation for the sample quantile $\hat{\eta}$ [37],

$$\hat{\eta} - \eta = \frac{p - \frac{1}{m}\sum_{i=1}^{m} I(X_i \le \eta)}{f(\eta)} + o_p(m^{-\frac{1}{2}}),$$

it follows that

$$\begin{aligned}
\hat{G}(\eta) - \hat{G}(\hat{\eta}) &= \int [I(y \le \eta) - I(y \le \hat{\eta})] d\hat{G}(y) \\
&= \int [I(y \le \eta) - I(y \le \hat{\eta})] dG(y) + o_p(n^{-1/2}) \\
&= -g(\eta)(\hat{\eta} - \eta) + o_p(m^{-1/2} + n^{-1/2}) \\
&= \frac{1}{m}\frac{g(\eta)}{f(\eta)} \sum_{i=1}^{m} [I(X_i \le \eta) - p] + o_p(m^{-1/2} + n^{-1/2}) \\
&= \frac{1}{m+n} \sum_{k=n+1}^{n+m} \frac{m+n}{m}\frac{g(\eta)}{f(\eta)} [I(Z_k \le \eta) - p] + o_p((m+n)^{-1/2}),
\end{aligned}$$

where f and g are the densities for X and Y, respectively.

Therefore,

$$\tilde{\theta}_{m,n} - \theta = \frac{1}{m+n} \sum_{k=1}^{n+m} W_k(\theta, p) + o_p((m+n)^{-1/2}) \tag{2.6}$$

where

$$W_k(\theta, p) = \begin{cases} \frac{m+n}{n}[I(Z_k > \eta) - \theta], & k = 1, \ldots, n, \\ \frac{m+n}{m}\frac{g(\eta)}{f(\eta)}[I(Z_k \le \eta) - p], & k = n+1, \ldots, n+m, \end{cases} \tag{2.7}$$

is called the influence function of θ.

From (2.6), we can easily get the following asymptotic distribution of the empirical estimator for θ.

Proposition 2.1: Assume that F and G are continuous distribution functions with density functions f and g, respectively, $f(\eta)$ is strictly positive, $g'(x)$ and $\frac{g(x)}{f(x)}$ are bounded in a neighborhood of $\eta = F^{-1}(p)$. If $\lim \frac{m}{n} = \rho$ $(0 < \rho < \infty)$, then

$$\sqrt{m+n}(\tilde{\theta}_{m,n} - \theta) \xrightarrow{d} N(0, \sigma^2), \tag{2.8}$$

where $\sigma^2 = (1+\rho)\theta(1-\theta) + (1+\rho^{-1})p(1-p)\frac{g^2(\eta)}{f^2(\eta)}$.

Linnet [13] heuristically derived the conclusion in Proposition 1, but he didn't explicitly give the formula for the asymptotic variance (see Zhou and Qin [14]).

Based on the influence function, an EL for the sensitivity θ can be defined as follows:

$$L_{IF}(\theta) = \sup_{\boldsymbol{p}} \left\{ \prod_{k=1}^{m+n} p_k : \sum_{k=1}^{m+n} p_k = 1, \sum_{k=1}^{m+n} p_k \hat{W}_k(\theta, p) = 0 \right\}, \tag{2.9}$$

where $\hat{W}_k(\theta, p)$ is the estimated influence function of θ given as follows

$$\hat{W}_k(\theta, p) = \begin{cases} \frac{m+n}{n}[I(Z_k > \hat{\eta}) - \theta], & k = 1, \ldots, n, \\ \frac{m+n}{m}\frac{\hat{g}(\hat{\eta})}{\hat{f}(\hat{\eta})}[I(Z_k \leq \hat{\eta}) - p], & k = n+1, \ldots, n+m, \end{cases}$$

where $\hat{g}$ and $\hat{f}$ are the density estimators for g and f, respectively.

Here we use the following kernel estimator for f:

$$\hat{f}(x) = \frac{1}{mh_X} \sum_{i=1}^{m} K(\frac{x - X_i}{h_X}),$$

where $K(\cdot)$ is a Gaussian kernel function, and the bandwidth is the "rule-of-thumb" bandwidth [38] defined by:

$$h_X = 0.9min(s_X, \frac{iqr_X}{1.34})m^{-1/5},$$

where s_X and iqr_X are, respectively, the standard deviation and the inter-quartile range, of the sample X_i's. The above bandwidth is also adopted by Zou *et al.* [39].

The density function g can be estimated similarly. When f and g are uniformly continuous, the kernel estimators $\hat{f}$ and $\hat{g}$ defined above are almost surely and uniformly consistent [40].

By the Lagrange multiplier, the maximization of (2.9) is achieved at

$$\tilde{p}_k = \frac{1}{m+n}[1 + \lambda \hat{W}_k(\theta, p)]^{-1}, k = 1, \ldots, m+n,$$

where λ is the solution of

$$\frac{1}{m+n}\sum_{k=1}^{m+n}\frac{\hat{W}_k(\theta,p)}{1+\lambda\hat{W}_k(\theta,p)}=0. \tag{2.10}$$

The corresponding empirical log-likelihood ratio statistic is

$$l_{IF}(\theta)=-\sum_{k=1}^{m+n}\log\{1+\lambda\hat{W}_k(\theta,p)\}. \tag{2.11}$$

When test results Y_k's are not all greater/smaller than $\hat{\eta}$, the empirical log-likelihood ratio $l_{IF}(\theta)$ is well defined on $(0,1)$. The following theorem establishes the asymptotic distribution of $l_{IF}(\theta)$ and the proof is given in the Appendix A.

Theorem 2.1: Assume that F and G are distribution functions with uniformly continuous density functions f and g, respectively, $f(\eta)$ is strictly positive, $f'(x)$ and $g'(x)$ and $\frac{g(x)}{f(x)}$ are bounded in a neighborhood of $\eta=F^{-1}(p)$. If θ_0 is the true value of the sensitivity at a fixed level p of specificity, and $\lim\frac{m}{n}=\rho$ $(0<\rho<\infty)$, then the limiting distribution of $-2l_{IF}(\theta_0)$ is a standard chi-squared distribution with one degree of freedom as $m,n\to\infty$.

From Theorem 1, a $100(1-\alpha)\%$ level influence function-based EL confidence interval for θ can be constructed as

$$CI_{IF}(\theta)=\{\theta:-2l_{IF}(\theta)\leq\chi_1^2(1-\alpha)\}.$$

2.4 Bayesian Empirical Likelihood (BEL) Method

Lazar[17] noted that EL has many of the same asymptotic properties as those derived from parametric models. In this sense, EL could be used as the basis for Bayesian inference. Conceptually, Bayesian EL enjoys the advantages of both EL and Bayesian methods: (i) no parametric assumption is needed by incorporating EL; (ii) Bayesian framework quantifies uncertainty more naturally, and with proper choices of priors, the Bayesian EL methods can outperform the classical EL methods[41]. We propose two types of Bayesian EL methods to construct credible intervals for sensitivity.

2.4.1 Bayesian empirical likelihood based on sensitivity

We noticed that the classical EL intervals (e.g, HEL and JEL) for sensitivity at a high specificity level (e.g, $p = 0.95$) sometimes have under-coverage problems with small sample sizes (e.g., $(n; m) = (20; 20), (50; 50)$. See Section 2.5). With prior knowledge on the diagnostic accuracy (i.e., sensitivity/specificity) of a test, Bayesian EL methods could improve small sample performances of the classical EL methods, which motivated us propose Bayesian EL methods for inference in medical diagnostics. To our knowledge, this article is the first application of Bayesian EL in medical diagnostics. We follow Lazar[17] to combine empirical likelihood $L(\theta)$ with a specified prior $\pi(\theta)$ on θ via the Bayes theorem to obtain a posterior

$$\pi(\theta|data) \propto L(\theta)\pi(\theta).$$

Instead of using parametric likelihood in traditional Bayesian framework, we use empirical likelihood here. An important step is to choose an appropriate prior on sensitivity θ. We consider reference priors in this study. Reference priors, originally introduced by Bernardo[31], and further developed by Berger, Bernardo, and Sun [32][33], are a popular choice for objective priors. They are an important type of objective priors which only depend on the assumed model and the available data. In our problem, since we do not have a parametric model, we follow Clarke and Yuan[30] to derive reference priors for EL. The following proposition gives the reference priors for the Bayesian hybrid EL method where $L_H(\theta)^{c^*}$ is used as the likelihood, and $L_H(\theta)$ and c^* are defined in Equations (2.3) and (2.5) from Section 2.3.

Proposition 2.2: The reference prior based on the relative entropy for HEL using $W_{Hj}(\theta, p)$ from (2.2) is

$$\pi_{H,1}(\theta) = \beta(\frac{3}{2}, \frac{3}{2}),$$

and the reference prior based on Hellinger distance is

$$\pi_{H,2}(\theta) = \beta(\frac{1}{2}, \frac{1}{2}),$$

where $\beta(a, b)$ is the beta distribution with parameters a and b. The corresponding posterior is

$$\pi_H(\theta|Y) \propto \prod_{j=1}^{n}[1 + \tilde{\lambda}\hat{W}_{Hj}(\theta, p)]^{-c^*}\pi_H(\theta),$$

where $\pi_H(\theta) = \pi_{H,1}(\theta),$ or $\pi_{H,2}(\theta)$, and c^* is from Equation (2.5).

Based on these posteriors, we can calculate two equal-tail credible intervals for θ. We call them as the Bayesian Hybrid Empirical Likelihood 1 (BHEL1) interval and the Bayesian Hybrid Empirical Likelihood 2 (BHEL2) interval using priors $\pi_{H,1}(\theta)$ and $\pi_{H,2}(\theta)$, respectively.

Similarly, to construct Bayesian credible intervals for θ based on the IFEL using $W_k(\theta, p)$ in (2.7), we propose the following reference priors:

$$\pi_{IF,1}(\theta) \propto \left[(1 + \frac{m}{n})\theta(1-\theta) + (1 + \frac{n}{m})p(1-p)\frac{g^2(\eta)}{f^2(\eta)}\right]^{\frac{1}{2}},$$

and

$$\pi_{IF,2}(\theta) \propto \left[(1 + \frac{m}{n})\theta(1-\theta) + (1 + \frac{n}{m})p(1-p)\frac{g^2(\eta)}{f^2(\eta)}\right]^{-\frac{1}{2}}.$$

These two priors are both proper since $\pi_{IF,1}(\theta)$ is bounded by a constant and $\pi_{IF,2}(\theta)$ is bounded by a beta distribution. In practice, we use $\hat{W}_k(\theta, p)$ to estimate the influence function $W_k(\theta, p)$, and replace f, g, and η with their estimates since they are generally unknown. The posterior based on this approach is then

$$\pi_{IF}(\theta|Z) \propto \prod_{k=1}^{m+n}[1 + \tilde{\lambda}\hat{W}_k(\theta, p)]^{-1}\pi_{IF}(\theta).$$

where $\pi_{IF}(\theta) = \pi_{IF,1}(\theta),$ or $\pi_{IF,2}(\theta)$.

Based on these posteriors, we also can calculate two equal-tail credible intervals for θ. We call them as Bayesian Influence Function-based Empirical Likelihood 1 (BIFEL1) and Bayesian Influence Function-based Empirical Likelihood 2 (BIFEL2) intervals using priors $\pi_{IF,1}(\theta)$ and $\pi_{If,2}(\theta)$, respectively.

2.4.2 Bayesian pseudo empirical likelihood (BpEL) based on probability vector

The methods presented in Section 4.1 are based on the posterior distributions of θ. In this section, instead of applying priors on θ, we apply Rao and Wu's method [34] to obtain an alternative approach for Bayesian EL inference on θ based on probability vector $(p_1, \ldots, p_l)$. We treat $(p_1, \ldots, p_l)$ as unknown parameters and the EL function is:

$$L_{EL}(p_1, \ldots, p_l) = \prod_{i=1}^{l} p_i,$$

where $l = n$ for hybrid EL, and $l = m + n$ for influence function EL. Consider the Dirichlet prior $D(\alpha_1, \ldots, \alpha_l)$ on $(p_1, \ldots, p_l)$:

$$\pi(p_1, \ldots, p_l) = c(\alpha_1, \ldots, \alpha_l) \prod_{i=1}^{l} p_i^{\alpha_i - 1},$$

where $c(\alpha_1, \ldots, \alpha_l) = \Gamma(\sum_{i=1}^{l} \alpha_i) / \prod_{i=1}^{l} \Gamma(\alpha_i)$. The posterior distribution of $(p_1, \ldots, p_l)$ given the data is Dirichlet $D(1 + \alpha_1, \ldots, 1 + \alpha_l)$ and is given by:

$$\pi(p_1, \ldots, p_l | data) = c(1 + \alpha_1, \ldots, 1 + \alpha_n) \prod_{i=1}^{l} p_i^{\alpha_i}.$$

The posterior of sensitivity θ satisfies the following equation:

$$\sum_{i=1}^{l} p_i \hat{Q}_i(\theta) = 0, \tag{2.12}$$

where $\hat{Q}_i(\theta)$ is an estimating/influence function and $(p_1, \ldots, p_l)$ follows the Dirichlet distribution $D(1 + \alpha_1, \ldots, 1 + \alpha_l)$. In practice, we can generate samples of $(p_1, \ldots, p_l)$ from $D(1 + \alpha_1, \ldots, 1 + \alpha_l)$, and by solving (2.12), we get the posterior samples of θ. Based on these posterior samples, we can calculate the equal-tail credible intervals for sensitivity θ.

Similar to Section 2.4.1, we consider two types of EL: hybrid EL (2.3) and influence function EL (2.9). We call them Bayesian pseudo hybrid EL (BpHEL) and Bayesian pseudo

influence function EL (BpIFEL), respectively. For BpHEL, we use $\hat{W}_{Hj}(\theta, p)$ to replace $\hat{Q}_i(\theta)$ in (2.12), and consider $D(c^*, \ldots, c^*)$ and $D(c^* + \frac{1}{n}, \ldots, c^* + \frac{1}{n})$ as the priors (labeled BpHEL1 and BpHEL2, respectively), where c^* is the scale estimate defined in Section 2.2. For BpIFEL, similarly, we use $\hat{W}_k(\theta, p)$ to replace $\hat{Q}_i(\theta)$ in (2.12), and consider $D(1, \ldots, 1)$ and $D(1 + \frac{1}{n+m}, \ldots, 1 + \frac{1}{n+m})$ as the priors (labeled BpIFEL1 and BpIFEL2, respectively).

2.5 Simulation Study

Simulation studies are conducted to examine the finite sample performance of the proposed approaches: influence function-based empirical likelihood (IFEL), Bayesian influence function empirical likelihood methods (BIFEL1 and BIFEL2) with reference priors $\pi_{IF,1}(\theta)$ and $\pi_{IF,2}(\theta)$, Bayesian hybrid empirical likelihood methods (BHEL1 and BHEL2) with reference priors $\pi_{H,1}(\theta)$ and $\pi_{H,2}(\theta)$, Bayesian pseudo influence function empirical likelihood method (BpIFEL), and Bayesian pseudo hybrid empirical likelihood method (BpHEL). We compare them with existing approaches, including BTI, BTII, smoothed JEL method, hybrid empirical likelihood method (HEL), and the modified normal approximation (NA) method proposed by Linnet [13].

2.5.1 Simulation settings

We consider seven simulation settings for the underlying non-diseased distribution F and diseased distribution G: (i) Normal distributions with $F = N(0, 1)$ and $G = N(1, 1)$, (ii) Normal distributions with $F = N(0, 1)$ and $G = N(2, 1)$, (iii) Normal distributions with $F = N(0, 1)$ and $G = N(2.5, 1)$, (iv) Exponential distributions with $F = Exponential(1)$ and $G = Exponential(0.25)$, (v) Exponential distributions with $F = Exponential(4.25)$ and $G = Exponential(0.25)$, (vi) Mixed distributions with $F = N(1, 1)$ and $G = Exponential(0.5)$, and (vii) Mixed distributions with $F = N(0, 1)$ and $G = Exponential(0.1)$. Under settings (i), (ii) and (iii), both the diseased and non-diseased distributions of test results are normal distributions but with different degrees of separation and corresponding to low, medium, and high sensitivity, respectively. Similarly, settings (iv), (vi) and settings (v), (vii) are

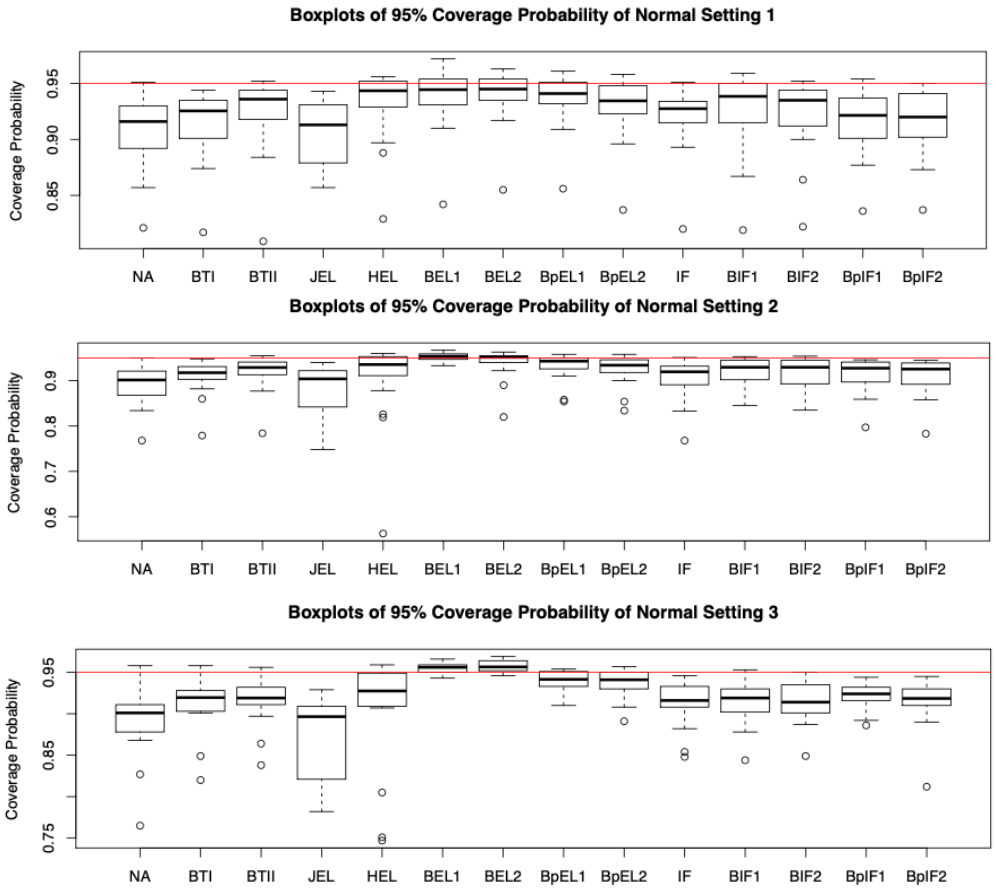
corresponding to medium and high sensitivity, respectively. Random samples of size m and n are generated from F and G respectively. Sample sizes (m, n)=(20, 20), (50, 50), (100, 100), (50, 100), (100, 50), (500, 500) are considered. We construct 95% level confidence (credible) intervals for sensitivity θ at specificity levels $p = 80\%$, 90% and 95%, respectively. The simulation procedure is repeated $5,000$ times to find the frequentist coverage probabilities and average lengths of the intervals. For JEL method, we use the kernel $w(x) = \frac{15}{16}(1 - x^2)^2 I(|x| \leq 1)$ and $h = n^{-1/3}$ as suggested by Gong *et al.* [15].

2.5.2 Simulation results

The simulation results under the normal distribution settings are reported in Tables 2.1 - 2.3 and Figure 2.1. From Table 2.1, where the sensitivity is at a low level, we observe that HEL and Bayesian approaches based on HEL have the best overall performance. Bayesian approaches generally have similar or improved performance over HEL. We note that BpHEL2 intervals generally have lower coverage probabilities than the three other Bayesian HEL intervals when the sample size is (20, 20). The possible reason is that the modified prior is not suitable for the small sample size. IFEL and corresponding Bayesian approaches have slightly worse performance comparing with HEL related approaches, especially when specificity is high.

BTI and BTII are generally not as good, especially when the sample size is small. JEL performs the best under one setting (when the sample size is $(20, 20)$ and specificity $p = 95\%$) but has poor performance comparing with our methods for other settings. Under the unbalanced sample setting, our new methods are similar to or better than others, and BHEL and BpHEL methods are much better than HEL. We also notice that the performance for sample size $(100, 50)$ is much better than that of sample size $(50, 100)$. It indicates that non-diseased samples are more critical in the inference of sensitivity with high specificity. Comparing the results from the normal distribution setting (i) with those from the normal distribution settings (ii) and (iii), which have a higher degree of separation and a higher sensitivity, we can see from Table 2.2 and 2.3 that the performance of methods does not

Figure (2.1) Boxplots of 95% coverage probabilities under Normal distribution settings.

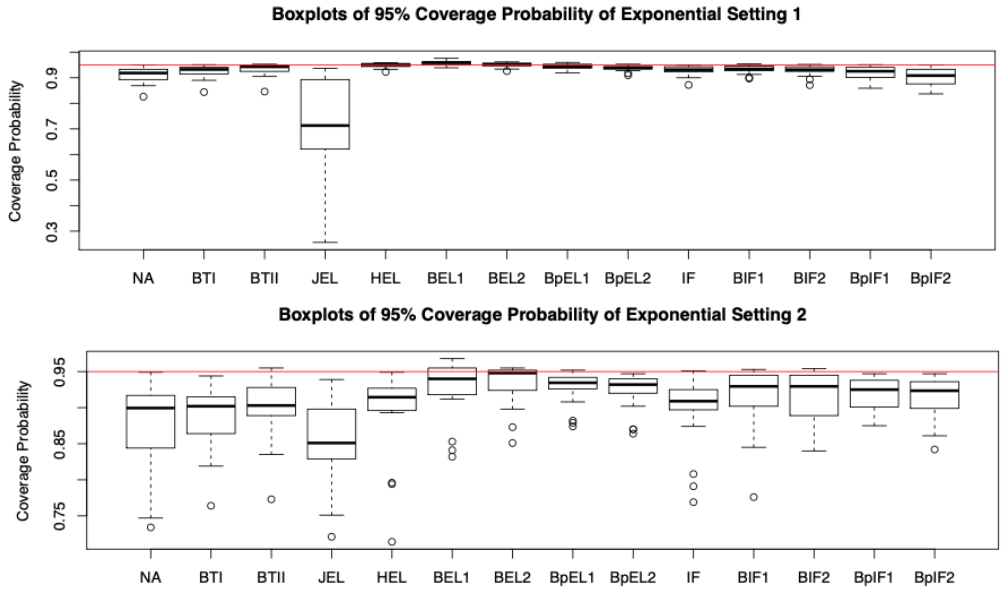


obviously depend on the degree of separation of test outcomes in the diseased and non-diseased groups. The methods have similar or slightly worse finite sample performance with the higher degree of separation of test results in terms of coverage probability. For example, when the sample size is (20, 20) and (50, 50) with specificity $p = 0.8$ and sensitivity $\theta = 0.95$, the performance of IF related methods is worse than that with the lower sensitivity settings.

Place Tables 2.1 to 2.3 here

The simulation results under the exponential distribution settings are reported in Table 2.4, 2.5, and Figure 2.2. HEL and BHEL intervals still perform well in both balanced and unbalanced settings. The performance of BHEL1 and BHEL2 intervals is improved compared with that of HEL intervals, and the coverage probabilities of BHEL1 and BHEL2

Figure (2.2) Boxplots of 95% coverage probabilities under Exponential distribution settings.

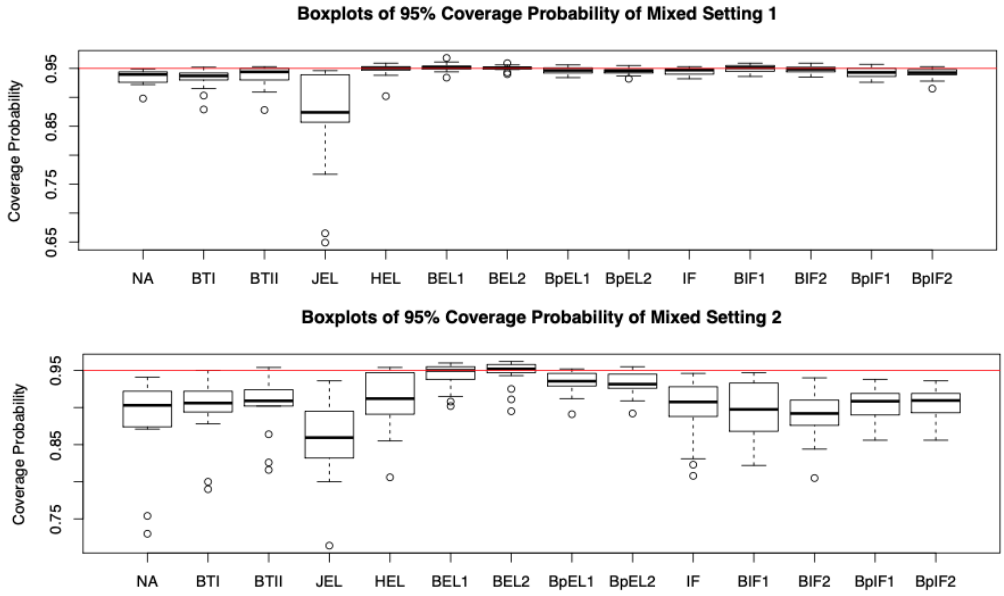


intervals are very close to 95% when sensitivity is at a medium level. Especially when the sample size is $(20, 20)$ and specificity $p = 0.95$, BHEL1 interval has the highest coverage probability 94.3%, which is very close to the nominal confidence level 95%. Influence function related IFEL, BIFEL, and BpIFEL intervals do not work well here, possibly because of the poor density estimation. NA, BTI, and BTII intervals have poor performance with the small sample size and high specificity. JEL interval has very poor performance, especially when $p = 0.95$. Compared with Table 2.5 where the sensitivity is at high level, we note that the overall performance of the methods is worse than that of Table 2.4, especially when the sample sizes are (20,20) and (50,50).

Place Table 2.4 and 2.5 here

The simulation results under the mixed distribution settings are reported in Table 2.6, 2.7 and Figure 2.3. Similar to the exponential distribution settings, the performances of BHEL1 and BHEL2 intervals are much improved comparing with that of HEL intervals when the sample size is $(20, 20)$ and specificity $p = 0.95$. The performance of influence

Figure (2.3) Boxplots of 95% coverage probabilities under Mixed distribution settings.



function related IFEL, BIFEL, and BpIFEL intervals are even better than that of the HEL related intervals in Table 2.6 with the medium level of sensitivity. BIFEL2 interval has the best coverage probability 94.3%, close to the nominal confidence level 95% when the sample size is $(20, 20)$ and specificity $p = 95\%$. However, when sensitivity is at a higher level (Table 2.7), the overall performance is poor when the sample size is small. The performance of NA method, which also needs density estimation, is acceptable in most of the settings. However, when the sample size is small, and specificity is high, the new methods always perform better. BTI and BTII intervals have poor performance with the small sample size and high specificity. JEL interval also has very poor performance. The poor performance of JEL might be due to the bandwidth problem in the smoothed jackknife method. The IF-related EL intervals have acceptable coverage probabilities with a large sample size $(500, 500)$, although some of them have slightly under-coverage problems due to the possible bandwidth selection problem for the kernel estimators of the density function g and f.

Place Table 2.6 and 2.7 here

In summary, HEL interval and new Bayesian intervals, especially BHEL1 and BIFEL2 intervals, have coverage probabilities closer to the nominal confidence level and shorter average

interval lengths than other intervals.

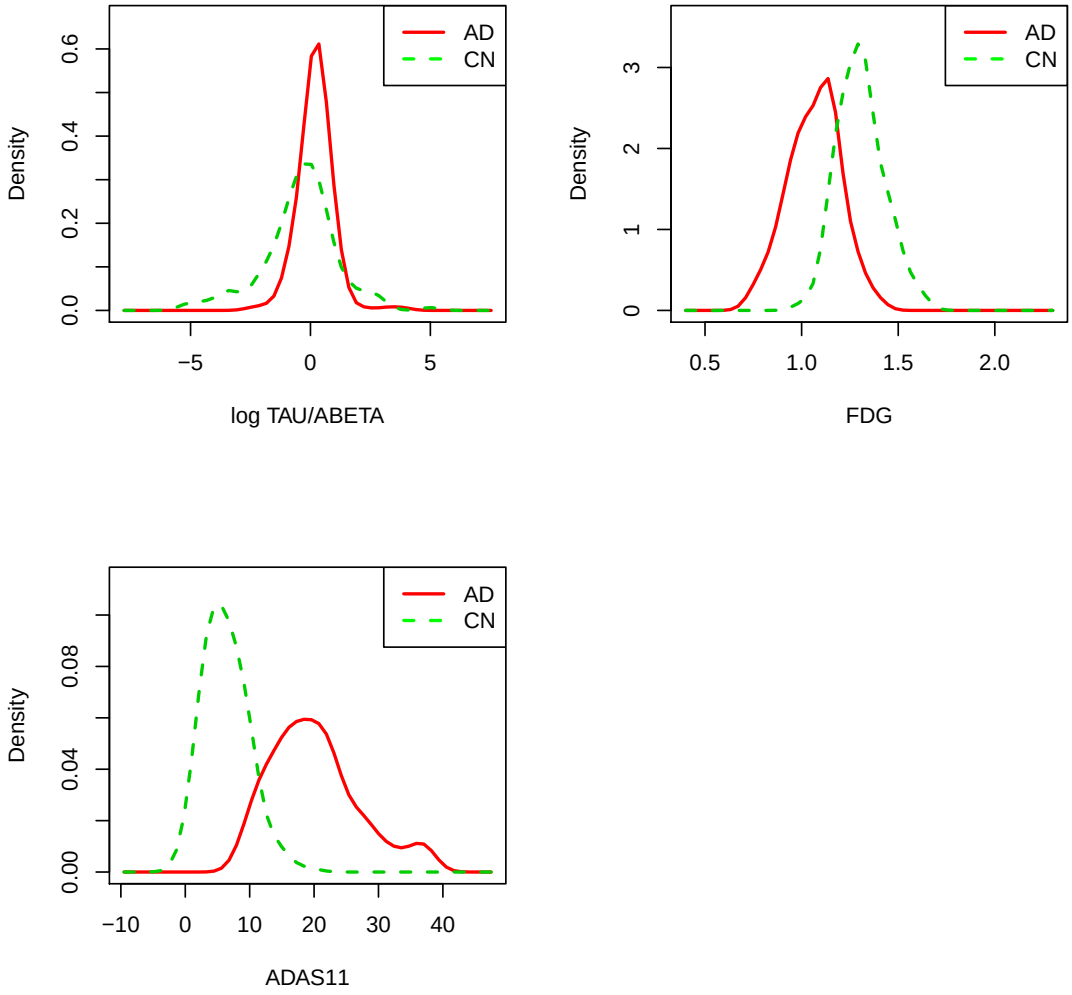
2.6 A Real Example in the Detection of Alzheimer's Disease

In this section, we illustrate the application of the proposed methods to assess the diagnostic accuracy of biomarkers in detecting Alzheimer's disease (AD). Alzheimer's disease is the most common cause of dementia. There are an estimated 5.8 million Americans of all ages living with Alzheimer's dementia in 2019, and the total Medicaid spending of the United States for people with Alzheimer's or other dementia is projected to be $49 billion in 2019 [42]. The data used in this section were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`adni.loni.usc.edu`). The ADNI study's goal is to track the progression of the diseases, mild cognitive impairment (MCI) and AD, using biomarkers and clinical measures.

We apply the proposed methods to a small subset of a data-freeze named "QT-PAD Project Data" downloaded on June 29th 2017. It is available in the "Test Data/Data for Challenges" section of the LONI website (ADNI database). Here we only consider non-missing records based on three commonly used biomarkers [43][44][20]: the ratio of levels of total protein Tau and protein $A\beta_{42}$ (TAU/ABETA), fluorodeoxyglucose (FDG), and Alzheimer's Disease Assessment Scale 11 (ADAS11). The dataset we used consists of 170 AD patients and 152 control subjects (CN). The distribution of tau-related biomarker was skewed to the right, and TAU/ABETA was log-transformed before analysis to reduce skewness. Figure 2.4 presents the estimated density curves of log TAU/ABETA, FDG and ADAS11 for the two groups.

The point estimates and confidence intervals for sensitivity of these three biomarkers when specificity $p = 0.95, 0.90, 0.80$ are reported in Tables 2.8–2.10, respectively. TAU/ABETA has very low (0.01 to 0.25) sensitivity when the specificity is above 0.80 and achieves 0.5 when specificity $p = 0.7$ (results are not reported here). It suggests that TAU/ABETA is not a good biomarker for the diagnosis of AD. FDG has moderate (0.69) to high (0.87)

Figure (2.4) Estimated densities for log TAU/ABETA, FDG, and ADAS11 in the ADNI data.



sensitivity when the specificity is fixed at $p = 0.95, 0.90, 0.80$. The sensitivity for FDG drops by around 13 percentage points if specificity is increased from 0.8 to 0.9, and drops by around 7 to 10 percentage points when specificity is further increased to 0.95. ADAS11 achieves very high (0.85 to 1) sensitivity when the specificity is above 0.80, suggesting it has high diagnostic accuracy in detecting the Alzheimer's Disease. Comparing these 95% level confidence/credible intervals for sensitivity, influence function-based approaches, especially BIFEL1, always have shorter interval lengths.

Place Table 2.8 , 2.9, and 2.10 here

Table (2.1) Coverage probabilities and average lengths of 95% confidence intervals for sensitivity θ at a fixed level of specificity p when $F = N(0,1)$ and $G = N(1,1)$

(m,n)	Methods	$p = 0.95$, $\theta = 0.26$		$p = 0.90$, $\theta = 0.39$		$p = 0.80$, $\theta = 0.56$	
		Coverage Probability	Average Length	Coverage Probability	Average Length	Coverage Probability	Average Length
(20,20)	NA	0.821	0.539	0.865	0.597	0.898	0.603
	BTI	0.817	0.596	0.874	0.609	0.898	0.573
	BTII	0.809	0.596	0.884	0.609	0.918	0.573
	JEL	0.901	0.314	0.892	0.378	0.906	0.456
	HEL	0.829	0.540	0.932	0.574	0.953	0.547
	BHEL1	0.842	0.521	0.939	0.529	0.972	0.509
	BHEL2	0.855	0.552	0.919	0.566	0.955	0.539
	BpHEL1	0.856	0.553	0.909	0.582	0.932	0.553
	BpHEL2	0.837	0.525	0.896	0.553	0.916	0.528
	IFEL	0.820	0.485	0.902	0.594	0.924	0.614
	BIFEL1	0.819	0.457	0.909	0.560	0.951	0.594
	BIFEL2	0.822	0.465	0.900	0.581	0.932	0.620
	BpIFEL1	0.836	0.571	0.877	0.625	0.919	0.625
	BpIFEL2	0.837	0.564	0.873	0.616	0.914	0.618
(50,50)	NA	0.874	0.379	0.902	0.405	0.921	0.394
	BTI	0.901	0.425	0.918	0.435	0.927	0.401
	BTII	0.901	0.425	0.931	0.435	0.940	0.401
	JEL	0.931	0.236	0.879	0.288	0.922	0.332
	HEL	0.888	0.384	0.944	0.423	0.951	0.395
	BHEL1	0.926	0.386	0.945	0.400	0.964	0.375
	BHEL2	0.943	0.392	0.945	0.414	0.963	0.393
	BpHEL1	0.943	0.390	0.942	0.422	0.960	0.401
	BpHEL2	0.923	0.378	0.936	0.411	0.956	0.391
	IFEL	0.915	0.388	0.919	0.422	0.933	0.405
	BIFEL1	0.915	0.364	0.934	0.414	0.951	0.408
	BIFEL2	0.912	0.370	0.931	0.426	0.944	0.417
	BpIFEL1	0.893	0.401	0.901	0.405	0.924	0.391
	BpIFEL2	0.902	0.407	0.913	0.414	0.940	0.402
(100,100)	NA	0.899	0.277	0.921	0.294	0.934	0.284
	BTI	0.929	0.318	0.935	0.317	0.941	0.293
	BTII	0.944	0.318	0.944	0.317	0.952	0.293
	JEL	0.907	0.183	0.865	0.224	0.931	0.249
	HEL	0.934	0.309	0.943	0.314	0.953	0.290
	BHEL1	0.927	0.304	0.947	0.304	0.954	0.283
	BHEL2	0.935	0.310	0.945	0.312	0.954	0.287
	BpHEL1	0.934	0.312	0.942	0.316	0.955	0.291
	BpHEL2	0.933	0.306	0.943	0.311	0.948	0.287
		0.915	0.291	0.933	0.304	0.940	0.288
	BIFEL1	0.907	0.284	0.940	0.306	0.950	0.287
	BIFEL2	0.907	0.289	0.937	0.311	0.944	0.290
	BpIFEL1	0.902	0.281	0.925	0.296	0.934	0.284
	BpIFEL2	0.898	0.285	0.932	0.299	0.943	0.285

(m,n)	Methods	$p = 0.95$, $\theta = 0.26$		$p = 0.90$, $\theta = 0.39$		$p = 0.80$, $\theta = 0.56$	
		Coverage Probability	Average Length	Coverage Probability	Average Length	Coverage Probability	Average Length
(500, 500)	NA	0.933	0.132	0.933	0.137	0.951	0.129
	BTI	0.944	0.145	0.944	0.145	0.944	0.145
	BTII	0.952	0.145	0.952	0.145	0.952	0.145
	JEL	0.937	0.139	0.938	0.113	0.942	0.139
	HEL	0.947	0.145	0.944	0.144	0.956	0.133
	BHEL1	0.944	0.144	0.943	0.143	0.957	0.132
	BHEL2	0.947	0.144	0.944	0.144	0.956	0.133
	BpHEL1	0.937	0.144	0.934	0.145	0.961	0.133
	BpHEL2	0.936	0.144	0.931	0.144	0.958	0.132
	IFEL	0.934	0.134	0.935	0.138	0.951	0.129
	BIFEL1	0.938	0.136	0.939	0.139	0.953	0.130
	BIFEL2	0.941	0.137	0.941	0.139	0.952	0.130
	BpIFEL1	0.928	0.135	0.939	0.138	0.954	0.130
	BpIFEL2	0.926	0.134	0.941	0.138	0.950	0.130
(50,100)	NA	0.857	0.335	0.892	0.359	0.924	0.349
	BTI	0.896	0.389	0.919	0.394	0.932	0.359
	BTII	0.903	0.389	0.928	0.394	0.941	0.359
	JEL	0.872	0.216	0.857	0.284	0.919	0.313
	HEL	0.897	0.357	0.929	0.384	0.938	0.350
	BHEL1	0.910	0.351	0.931	0.363	0.952	0.337
	BHEL2	0.917	0.358	0.934	0.379	0.954	0.346
	BpHEL1	0.917	0.357	0.928	0.386	0.948	0.351
	BpHEL2	0.906	0.348	0.925	0.377	0.944	0.345
	IFEL	0.893	0.337	0.912	0.378	0.931	0.360
	BIFEL1	0.867	0.310	0.917	0.366	0.948	0.362
	BIFEL2	0.864	0.310	0.914	0.369	0.943	0.366
	BpIFEL1	0.882	0.370	0.909	0.370	0.937	0.352
	BpIFEL2	0.883	0.369	0.907	0.368	0.936	0.351
(100,50)	NA	0.911	0.326	0.922	0.349	0.930	0.341
	BTI	0.924	0.358	0.919	0.366	0.935	0.345
	BTII	0.932	0.358	0.932	0.366	0.944	0.345
	JEL	0.943	0.196	0.875	0.237	0.924	0.283
	HEL	0.928	0.350	0.952	0.366	0.955	0.345
	BHEL1	0.937	0.344	0.953	0.349	0.960	0.333
	BHEL2	0.942	0.351	0.949	0.361	0.952	0.342
	BpHEL1	0.940	0.352	0.951	0.367	0.952	0.347
	BpHEL2	0.933	0.344	0.949	0.359	0.951	0.342
	IFEL	0.924	0.336	0.933	0.357	0.941	0.343
	BIFEL1	0.927	0.328	0.950	0.355	0.959	0.341
	BIFEL2	0.933	0.337	0.948	0.365	0.952	0.348
	BpIFEL1	0.916	0.334	0.940	0.356	0.947	0.344
	BpIFEL2	0.912	0.333	0.941	0.354	0.946	0.343

Table (2.2) Coverage probabilities and average lengths of 95% confidence intervals for sensitivity θ at a fixed level of specificity p when $F = Normal(0, 1)$ and $G = Normal(2, 1)$

(m,n)	Methods	$p = 0.95$, $\theta = 0.64$		$p = 0.90$, $\theta = 0.76$		$p = 0.80$, $\theta = 0.88$	
		Coverage Probability	Average Length	Coverage Probability	Average Length	Coverage Probability	Average Length
(20, 20)	NA	0.768	0.534	0.853	0.480	0.868	0.363
	BTI	0.779	0.558	0.860	0.496	0.903	0.359
	BTII	0.784	0.558	0.877	0.496	0.913	0.359
	JEL	0.783	0.236	0.800	0.260	0.748	0.270
	HEL	0.826	0.500	0.819	0.389	0.563	0.225
	BHEL1	0.933	0.475	0.935	0.411	0.947	0.224
	BHEL2	0.890	0.523	0.928	0.482	0.820	0.388
	BpHEL1	0.858	0.527	0.926	0.466	0.854	0.305
	BpHEL2	0.834	0.498	0.918	0.441	0.854	0.292
	IFEL	0.768	0.428	0.876	0.443	0.833	0.313
	BIFEL1	0.845	0.398	0.858	0.433	0.899	0.257
	BIFEL2	0.835	0.456	0.857	0.450	0.893	0.121
	BpIFEL1	0.797	0.546	0.859	0.497	0.876	0.363
	BpIFEL2	0.783	0.539	0.858	0.491	0.871	0.358
(50, 50)	NA	0.857	0.402	0.887	0.335	0.910	0.240
	BTI	0.888	0.456	0.910	0.363	0.918	0.245
	BTII	0.902	0.456	0.923	0.363	0.929	0.245
	JEL	0.820	0.205	0.888	0.230	0.884	0.217
	HEL	0.955	0.445	0.943	0.345	0.878	0.203
	BHEL1	0.951	0.418	0.967	0.348	0.947	0.248
	BHEL2	0.940	0.437	0.955	0.351	0.960	0.245
	BpHEL1	0.932	0.446	0.942	0.350	0.944	0.235
	BpHEL2	0.920	0.431	0.932	0.340	0.936	0.231
	IFEL	0.891	0.433	0.916	0.343	0.923	0.232
	BIFEL1	0.902	0.405	0.945	0.353	0.923	0.240
	BIFEL2	0.889	0.410	0.935	0.355	0.930	0.196
	BpIFEL1	0.897	0.441	0.928	0.348	0.941	0.245
	BpIFEL2	0.892	0.439	0.925	0.346	0.938	0.244
(100, 100)	NA	0.883	0.305	0.914	0.246	0.934	0.173
	BTI	0.917	0.352	0.930	0.264	0.937	0.176
	BTII	0.929	0.352	0.940	0.264	0.946	0.176
	JEL	0.905	0.150	0.842	0.191	0.903	0.158
	HEL	0.935	0.344	0.953	0.262	0.936	0.169
	BHEL1	0.956	0.330	0.959	0.259	0.959	0.181
	BHEL2	0.944	0.340	0.950	0.260	0.960	0.176
	BpHEL1	0.926	0.341	0.946	0.258	0.958	0.173
	BpHEL2	0.918	0.335	0.936	0.255	0.954	0.171
	IFEL	0.904	0.314	0.927	0.250	0.947	0.174
	BIFEL1	0.907	0.314	0.937	0.256	0.953	0.174
	BIFEL2	0.902	0.316	0.929	0.257	0.954	0.174
	BpIFEL1	0.897	0.314	0.927	0.251	0.943	0.174
	BpIFEL2	0.897	0.313	0.926	0.250	0.944	0.173

(m,n)	Methods	$p = 0.95$, $\theta = 0.64$		$p = 0.90$, $\theta = 0.76$		$p = 0.80$, $\theta = 0.88$	
		Coverage Probability	Average Length	Coverage Probability	Average Length	Coverage Probability	Average Length
(500, 500)	NA	0.921	0.149	0.938	0.115	0.950	0.078
	BTI	0.948	0.165	0.944	0.120	0.947	0.078
	BTII	0.953	0.165	0.950	0.120	0.955	0.078
	JEL	0.940	0.236	0.930	0.233	0.932	0.212
	HEL	0.953	0.163	0.950	0.120	0.949	0.078
	BHEL1	0.955	0.162	0.952	0.120	0.954	0.079
	BHEL2	0.954	0.163	0.950	0.120	0.955	0.078
	BpHEL1	0.958	0.163	0.944	0.120	0.936	0.078
	BpHEL2	0.958	0.162	0.948	0.120	0.932	0.078
	IFEL	0.931	0.150	0.942	0.115	0.951	0.078
	BIFEL1	0.937	0.153	0.950	0.116	0.949	0.078
	BIFEL2	0.936	0.153	0.950	0.116	0.948	0.078
	BpIFEL1	0.933	0.152	0.946	0.116	0.944	0.078
	BpIFEL2	0.939	0.151	0.944	0.116	0.945	0.078
(50, 100)	NA	0.834	0.363	0.879	0.295	0.917	0.206
	BTI	0.882	0.421	0.912	0.328	0.931	0.213
	BTII	0.890	0.421	0.924	0.328	0.941	0.213
	JEL	0.894	0.188	0.927	0.221	0.909	0.195
	HEL	0.927	0.408	0.934	0.315	0.911	0.190
	BHEL1	0.935	0.384	0.953	0.312	0.955	0.218
	BHEL2	0.922	0.400	0.943	0.314	0.956	0.209
	BpHEL1	0.910	0.406	0.934	0.312	0.946	0.201
	BpHEL2	0.900	0.394	0.928	0.305	0.944	0.198
	IFEL	0.870	0.367	0.910	0.307	0.936	0.208
	BIFEL1	0.845	0.338	0.927	0.302	0.942	0.211
	BIFEL2	0.840	0.340	0.915	0.303	0.945	0.209
	BpIFEL1	0.877	0.400	0.918	0.305	0.944	0.209
	BpIFEL2	0.880	0.399	0.919	0.303	0.944	0.208
(100, 50)	NA	0.893	0.354	0.921	0.295	0.922	0.213
	BTI	0.916	0.393	0.926	0.308	0.926	0.214
	BTII	0.929	0.393	0.936	0.308	0.931	0.214
	JEL	0.911	0.172	0.921	0.200	0.922	0.187
	HEL	0.960	0.388	0.958	0.307	0.913	0.193
	BHEL1	0.964	0.369	0.959	0.304	0.952	0.224
	BHEL2	0.953	0.382	0.955	0.305	0.963	0.218
	BpHEL1	0.952	0.387	0.950	0.307	0.954	0.211
	BpHEL2	0.946	0.377	0.940	0.301	0.952	0.208
	IFEL	0.907	0.363	0.932	0.295	0.925	0.208
	BIFEL1	0.932	0.372	0.950	0.302	0.922	0.211
	BIFEL2	0.928	0.378	0.945	0.304	0.934	0.174
	BpIFEL1	0.916	0.365	0.940	0.300	0.940	0.214
	BpIFEL2	0.919	0.363	0.936	0.299	0.939	0.213

Table (2.3) Coverage probabilities and average lengths of 95% confidence intervals for sensitivity θ at a fixed level of specificity p when $F = Normal(0, 1)$ and $G = Normal(2.5, 1)$

(m,n)	Methods	$p = 0.95$, $\theta = 0.80$		$p = 0.90$, $\theta = 0.90$		$p = 0.80$, $\theta = 0.95$	
		Coverage Probability	Average Length	Coverage Probability	Average Length	Coverage Probability	Average Length
(20, 20)	NA	0.765	0.292	0.892	0.267	0.827	0.268
	BTI	0.820	0.307	0.908	0.281	0.849	0.234
	BTII	0.838	0.307	0.897	0.281	0.864	0.234
	JEL	0.821	0.282	0.797	0.225	0.785	0.220
	HEL	0.747	0.287	0.751	0.239	0.805	0.178
	BHEL1	0.963	0.352	0.963	0.338	0.944	0.312
	BHEL2	0.969	0.338	0.968	0.318	0.964	0.287
	BpHEL1	0.928	0.303	0.910	0.268	0.912	0.216
	BpHEL2	0.923	0.294	0.908	0.261	0.891	0.211
	IF	0.848	0.259	0.854	0.221	0.882	0.165
	BIFEL1	0.844	0.364	0.898	0.368	0.878	0.222
	BIFEL2	0.887	0.367	0.849	0.373	0.895	0.182
	BpIF1	0.926	0.278	0.892	0.251	0.886	0.211
	BpIF2	0.924	0.275	0.890	0.248	0.812	0.209
(50, 50)	NA	0.868	0.119	0.882	0.185	0.869	0.157
	BTI	0.901	0.225	0.906	0.197	0.925	0.166
	BTII	0.915	0.225	0.904	0.197	0.911	0.166
	JEL	0.782	0.213	0.802	0.195	0.892	0.160
	HEL	0.929	0.233	0.923	0.194	0.909	0.151
	BHEL1	0.954	0.238	0.959	0.210	0.956	0.185
	BHEL2	0.951	0.234	0.955	0.202	0.968	0.173
	BpHEL1	0.943	0.230	0.939	0.196	0.933	0.164
	BpHEL2	0.942	0.228	0.936	0.194	0.924	0.161
	IF	0.915	0.220	0.918	0.172	0.916	0.139
	BIFEL1	0.900	0.140	0.916	0.125	0.909	0.070
	BIFEL2	0.915	0.950	0.895	0.129	0.901	0.107
	BpIF1	0.915	0.394	0.919	0.179	0.918	0.152
	BpIF2	0.915	0.338	0.919	0.178	0.918	0.152
(100, 100)	NA	0.902	0.153	0.902	0.135	0.915	0.113
	BTI	0.918	0.167	0.924	0.144	0.921	0.119
	BTII	0.914	0.167	0.927	0.144	0.923	0.119
	JEL	0.842	0.190	0.882	0.122	0.909	0.129
	HEL	0.944	0.169	0.958	0.145	0.916	0.120
	BHEL1	0.950	0.170	0.959	0.149	0.958	0.126
	BHEL2	0.947	0.169	0.955	0.146	0.961	0.121
	BpHEL1	0.944	0.167	0.953	0.144	0.950	0.119
	BpHEL2	0.945	0.167	0.957	0.143	0.947	0.118
	IF	0.916	0.148	0.941	0.131	0.913	0.109
	BIFEL1	0.924	0.148	0.934	0.114	0.928	0.080
	BIFEL2	0.915	0.147	0.945	0.111	0.911	0.082
	BpIF1	0.916	0.148	0.943	0.131	0.931	0.111
	BpIF2	0.915	0.148	0.944	0.131	0.910	0.110

(m,n)	Methods	$p = 0.95$, $\theta = 0.80$		$p = 0.90$, $\theta = 0.90$		$p = 0.80$, $\theta = 0.95$	
		Coverage Probability	Average Length	Coverage Probability	Average Length	Coverage Probability	Average Length
(500, 500)	NA	0.941	0.073	0.943	0.063	0.958	0.052
	BTI	0.949	0.075	0.949	0.065	0.958	0.054
	BTII	0.952	0.075	0.951	0.065	0.956	0.054
	JEL	0.901	0.097	0.929	0.076	0.913	0.085
	HEL	0.959	0.076	0.949	0.065	0.957	0.054
	BHEL1	0.956	0.076	0.949	0.065	0.957	0.055
	BHEL2	0.958	0.075	0.947	0.065	0.954	0.054
	BpHEL1	0.953	0.075	0.942	0.065	0.952	0.054
	BpHEL2	0.954	0.075	0.947	0.064	0.952	0.054
	IF	0.939	0.069	0.937	0.061	0.946	0.051
	BIFEL1	0.950	0.083	0.945	0.060	0.953	0.048
	BIFEL2	0.949	0.083	0.941	0.060	0.950	0.048
	BpIF1	0.939	0.069	0.935	0.061	0.944	0.051
	BpIF2	0.938	0.069	0.933	0.061	0.945	0.051
(50, 100)	NA	0.900	0.848	0.899	0.139	0.911	0.116
	BTI	0.912	0.185	0.922	0.157	0.932	0.128
	BTII	0.932	0.185	0.932	0.157	0.927	0.128
	JEL	0.892	0.163	0.901	0.167	0.916	0.125
	HEL	0.945	0.190	0.954	0.157	0.907	0.126
	BHEL1	0.943	0.192	0.966	0.162	0.961	0.135
	BHEL2	0.946	0.190	0.959	0.158	0.967	0.129
	BpHEL1	0.940	0.188	0.951	0.156	0.954	0.126
	BpHEL2	0.937	0.186	0.950	0.154	0.954	0.125
	IF	0.908	0.180	0.933	0.133	0.909	0.110
	BIFEL1	0.902	0.125	0.911	0.112	0.921	0.106
	BIFEL2	0.901	0.123	0.915	0.109	0.909	0.105
	BpIF1	0.907	0.139	0.932	0.134	0.922	0.112
	BpIF2	0.904	0.089	0.930	0.133	0.900	0.112
(100, 50)	NA	0.878	0.204	0.907	0.184	0.911	0.156
	BTI	0.901	0.214	0.903	0.188	0.928	0.159
	BTII	0.913	0.214	0.916	0.188	0.922	0.159
	JEL	0.901	0.224	0.921	0.188	0.901	0.168
	HEL	0.931	0.218	0.926	0.189	0.918	0.149
	BHEL1	0.956	0.223	0.945	0.200	0.956	0.177
	BHEL2	0.952	0.219	0.952	0.193	0.964	0.166
	BpHEL1	0.941	0.215	0.936	0.189	0.933	0.158
	BpHEL2	0.940	0.213	0.931	0.186	0.930	0.157
	IF	0.922	0.197	0.917	0.171	0.908	0.139
	BIFEL1	0.917	0.172	0.930	0.104	0.926	0.070
	BIFEL2	0.924	0.169	0.935	0.110	0.913	0.107
	BpIF1	0.927	0.199	0.917	0.178	0.930	0.151
	BpIF2	0.925	0.199	0.917	0.178	0.930	0.151

Table (2.4) Coverage probabilities and average lengths of 95% confidence intervals for sensitivity θ at a fixed level of specificity p when $F = exp(1)$ and $G = exp(0.25)$

(m,n)	Methods	$p = 0.95$, $\theta = 0.47$		$p = 0.90$, $\theta = 0.56$		$p = 0.80$, $\theta = 0.67$	
		Coverage Probability	Average Length	Coverage Probability	Average Length	Coverage Probability	Average Length
(20, 20)	NA	0.826	0.485	0.890	0.487	0.880	0.971
	BTI	0.844	0.520	0.890	0.505	0.904	0.458
	BTII	0.846	0.520	0.905	0.505	0.916	0.458
	JEL	0.639	0.289	0.705	0.337	0.893	0.389
	HEL	0.923	0.513	0.959	0.507	0.934	0.451
	BHEL1	0.943	0.476	0.971	0.469	0.976	0.439
	BHEL2	0.926	0.505	0.952	0.494	0.962	0.454
	BpHEL1	0.919	0.524	0.938	0.506	0.953	0.460
	BpHEL2	0.910	0.503	0.931	0.487	0.944	0.444
	IFEL	0.872	0.472	0.902	0.486	0.935	0.445
	BIFEL1	0.901	0.449	0.927	0.471	0.942	0.437
	BIFEL2	0.871	0.470	0.906	0.495	0.936	0.454
	BpIFEL1	0.879	0.483	0.893	0.488	0.941	0.450
	BpIFEL2	0.837	0.564	0.873	0.616	0.914	0.618
(50, 50)	NA	0.904	0.412	0.913	0.333	0.922	0.307
	BTI	0.915	0.376	0.931	0.351	0.933	0.313
	BTII	0.925	0.376	0.939	0.351	0.945	0.313
	JEL	0.396	0.209	0.646	0.246	0.927	0.275
	HEL	0.947	0.382	0.956	0.351	0.956	0.314
	BHEL1	0.955	0.356	0.963	0.337	0.965	0.306
	BHEL2	0.947	0.368	0.956	0.347	0.957	0.312
	BpHEL1	0.942	0.373	0.956	0.353	0.955	0.314
	BpHEL2	0.938	0.366	0.953	0.346	0.950	0.309
	IFEL	0.923	0.395	0.927	0.335	0.936	0.297
	BIFEL1	0.928	0.382	0.933	0.334	0.941	0.294
	BIFEL2	0.925	0.390	0.927	0.341	0.938	0.299
	BpIFEL1	0.904	0.332	0.909	0.310	0.934	0.294
	BpIFEL2	0.898	0.331	0.911	0.309	0.933	0.293
(100, 100)	NA	0.905	0.264	0.927	0.246	0.939	0.222
	BTI	0.932	0.281	0.939	0.256	0.944	0.226
	BTII	0.943	0.281	0.945	0.256	0.950	0.226
	JEL	0.256	0.163	0.649	0.192	0.937	0.204
	HEL	0.946	0.279	0.951	0.256	0.956	0.227
	BHEL1	0.949	0.272	0.955	0.250	0.959	0.223
	BHEL2	0.947	0.277	0.952	0.255	0.957	0.226
	BpHEL1	0.934	0.281	0.937	0.257	0.944	0.226
	BpHEL2	0.928	0.277	0.938	0.255	0.942	0.224
	IFEL	0.923	0.267	0.935	0.246	0.942	0.214
	BIFEL1	0.930	0.266	0.937	0.246	0.946	0.213
	BIFEL2	0.926	0.269	0.934	0.248	0.943	0.215
	BpIFEL1	0.911	0.262	0.925	0.245	0.934	0.214
	BpIFEL2	0.873	0.231	0.901	0.221	0.923	0.208

(m,n)	Methods	$p = 0.95$, $\theta = 0.47$		$p = 0.90$, $\theta = 0.56$		$p = 0.80$, $\theta = 0.67$	
		Coverage Probability	Average Length	Coverage Probability	Average Length	Coverage Probability	Average Length
(500, 500)	NA	0.933	0.123	0.944	0.113	0.950	0.101
	BTI	0.945	0.127	0.948	0.115	0.951	0.102
	BTII	0.951	0.127	0.953	0.115	0.953	0.102
	JEL	0.837	0.139	0.838	0.113	0.842	0.139
	HEL	0.947	0.272	0.944	0.203	0.954	0.209
	BHEL1	0.949	0.272	0.955	0.250	0.959	0.223
	BHEL2	0.947	0.277	0.952	0.255	0.957	0.226
	BpHEL1	0.944	0.127	0.942	0.115	0.946	0.103
	BpHEL2	0.936	0.127	0.938	0.115	0.942	0.102
	IFEL	0.943	0.124	0.947	0.113	0.949	0.099
	BIFEL1	0.935	0.124	0.946	0.113	0.953	0.101
	BIFEL2	0.935	0.124	0.946	0.113	0.952	0.101
	BpIFEL1	0.944	0.124	0.939	0.113	0.951	0.099
	BpIFEL2	0.941	0.123	0.936	0.113	0.950	0.099
(50, 100)	NA	0.869	0.254	0.893	0.275	0.922	0.250
	BTI	0.907	0.329	0.925	0.301	0.936	0.261
	BTII	0.912	0.329	0.934	0.301	0.947	0.261
	JEL	0.349	0.185	0.721	0.226	0.934	0.235
	HEL	0.933	0.328	0.941	0.298	0.951	0.259
	BHEL1	0.939	0.311	0.950	0.289	0.960	0.254
	BHEL2	0.934	0.320	0.942	0.296	0.951	0.258
	BpHEL1	0.923	0.328	0.939	0.300	0.940	0.259
	BpHEL2	0.918	0.323	0.935	0.296	0.937	0.257
	IFEL	0.900	0.338	0.904	0.281	0.925	0.241
	BIFEL1	0.897	0.317	0.913	0.280	0.932	0.241
	BIFEL2	0.894	0.320	0.907	0.283	0.928	0.244
	BpIFEL1	0.859	0.272	0.876	0.247	0.901	0.232
	BpIFEL2	0.861	0.271	0.876	0.246	0.906	0.231
(100, 50)	NA	0.915	0.319	0.928	0.312	0.939	0.286
	BTI	0.929	0.336	0.934	0.315	0.941	0.286
	BTII	0.939	0.336	0.943	0.315	0.948	0.286
	JEL	0.324	0.188	0.621	0.223	0.932	0.254
	HEL	0.950	0.337	0.957	0.317	0.959	0.289
	BHEL1	0.958	0.325	0.963	0.307	0.965	0.283
	BHEL2	0.951	0.334	0.958	0.315	0.961	0.288
	BpHEL1	0.948	0.340	0.957	0.319	0.960	0.298
	BpHEL2	0.940	0.334	0.953	0.315	0.953	0.286
	IFEL	0.928	0.328	0.943	0.308	0.943	0.277
	BIFEL1	0.934	0.327	0.946	0.305	0.954	0.273
	BIFEL2	0.929	0.333	0.943	0.309	0.948	0.277
	BpIFEL1	0.926	0.327	0.944	0.309	0.942	0.278
	BpIFEL2	0.904	0.299	0.931	0.290	0.933	0.274

Table (2.5) Coverage probabilities and average lengths of 95% confidence intervals for sensitivity θ at a fixed level of specificity p when $F = exp(4.25)$ and $G = exp(0.25)$

(m,n)	Methods	$p = 0.95$, $\theta = 0.85$		$p = 0.90$, $\theta = 0.88$		$p = 0.80$, $\theta = 0.92$	
		Coverage	Average	Coverage	Average	Coverage	Average
(20, 20)	NA	0.747	0.409	0.814	0.334	0.734	0.219
	BTI	0.764	0.424	0.864	0.350	0.819	0.218
	BTII	0.773	0.424	0.873	0.350	0.835	0.218
	JEL	0.721	0.411	0.751	0.328	0.758	0.213
	HEL	0.794	0.279	0.796	0.211	0.714	0.151
	BHEL1	0.841	0.271	0.832	0.201	0.853	0.284
	BHEL2	0.898	0.263	0.873	0.079	0.851	0.308
	BpHEL1	0.882	0.382	0.879	0.292	0.874	0.143
	BpHEL2	0.870	0.361	0.864	0.277	0.870	0.148
	IF	0.808	0.275	0.791	0.253	0.769	0.138
	BIFEL1	0.776	0.105	0.858	0.074	0.855	0.100
	BIFEL2	0.886	0.111	0.857	0.174	0.843	0.100
	BpIF1	0.879	0.412	0.875	0.338	0.877	0.208
	BpIF2	0.888	0.407	0.861	0.333	0.842	0.205
(50, 50)	NA	0.844	0.317	0.876	0.231	0.892	0.147
	BTI	0.858	0.365	0.884	0.257	0.823	0.150
	BTII	0.892	0.365	0.889	0.257	0.882	0.150
	JEL	0.794	0.342	0.866	0.257	0.854	0.161
	HEL	0.896	0.348	0.922	0.205	0.902	0.161
	BHEL1	0.939	0.359	0.951	0.261	0.912	0.059
	BHEL2	0.949	0.360	0.948	0.243	0.917	0.036
	BpHEL1	0.926	0.354	0.948	0.237	0.932	0.131
	BpHEL2	0.920	0.343	0.944	0.231	0.927	0.129
	IF	0.906	0.336	0.916	0.229	0.908	0.112
	BIFEL1	0.902	0.416	0.945	0.300	0.923	0.146
	BIFEL2	0.889	0.422	0.935	0.294	0.930	0.136
	BpIF1	0.905	0.345	0.940	0.242	0.901	0.146
	BpIF2	0.899	0.343	0.936	0.241	0.901	0.146
(100, 100)	NA	0.885	0.237	0.917	0.171	0.920	0.105
	BTI	0.874	0.279	0.903	0.184	0.904	0.105
	BTII	0.892	0.279	0.904	0.184	0.902	0.105
	JEL	0.841	0.274	0.848	0.196	0.898	0.120
	HEL	0.927	0.267	0.917	0.169	0.911	0.070
	BHEL1	0.962	0.273	0.957	0.191	0.929	0.101
	BHEL2	0.952	0.272	0.954	0.183	0.953	0.089
	BpHEL1	0.935	0.265	0.938	0.176	0.938	0.100
	BpHEL2	0.932	0.260	0.932	0.173	0.932	0.099
	IF	0.897	0.247	0.933	0.172	0.911	0.093
	BIFEL1	0.907	0.395	0.937	0.242	0.953	0.138
	BIFEL2	0.902	0.398	0.929	0.242	0.954	0.137
	BpIF1	0.919	0.247	0.926	0.173	0.930	0.104
	BpIF2	0.899	0.246	0.923	0.173	0.931	0.104

(m,n)	Methods	$p = 0.95$, $\theta = 0.85$		$p = 0.90$, $\theta = 0.88$		$p = 0.80$, $\theta = 0.92$	
		Coverage	Average	Coverage	Average	Coverage	Average
(500, 500)	NA	0.942	0.116	0.941	0.079	0.949	0.047
	BTI	0.940	0.126	0.944	0.081	0.940	0.047
	BTII	0.948	0.126	0.955	0.081	0.945	0.047
	JEL	0.939	0.165	0.9314	0.078	0.934	0.042
	HEL	0.947	0.126	0.949	0.081	0.949	0.044
	BHEL1	0.947	0.126	0.947	0.082	0.941	0.048
	BHEL2	0.947	0.126	0.948	0.081	0.949	0.047
	BpHEL1	0.934	0.126	0.942	0.081	0.952	0.047
	BpHEL2	0.934	0.125	0.938	0.081	0.947	0.046
	IF	0.943	0.117	0.945	0.079	0.951	0.043
	BIFEL1	0.939	0.137	0.945	0.101	0.946	0.107
	BIFEL2	0.936	0.137	0.945	0.102	0.947	0.071
	BpIF1	0.937	0.117	0.946	0.119	0.946	0.095
	BpIF2	0.942	0.117	0.943	0.109	0.947	0.096
(50, 100)	NA	0.826	0.279	0.899	0.202	0.915	0.124
	BTI	0.885	0.335	0.914	0.230	0.927	0.125
	BTII	0.892	0.335	0.928	0.230	0.937	0.125
	JEL	0.898	0.343	0.829	0.206	0.848	0.126
	HEL	0.918	0.321	0.901	0.188	0.893	0.067
	BHEL1	0.929	0.328	0.955	0.239	0.918	0.113
	BHEL2	0.924	0.329	0.955	0.225	0.952	0.095
	BpHEL1	0.908	0.322	0.934	0.210	0.934	0.112
	BpHEL2	0.902	0.312	0.930	0.206	0.932	0.111
	IF	0.874	0.287	0.919	0.206	0.910	0.107
	BIFEL1	0.845	0.483	0.927	0.239	0.942	0.143
	BIFEL2	0.840	0.485	0.915	0.239	0.945	0.141
	BpIF1	0.877	0.310	0.905	0.208	0.930	0.121
	BpIF2	0.893	0.310	0.904	0.208	0.928	0.121
(100, 50)	NA	0.900	0.279	0.908	0.210	0.905	0.132
	BTI	0.901	0.316	0.915	0.220	0.908	0.131
	BTII	0.904	0.316	0.913	0.220	0.910	0.131
	JEL	0.860	0.315	0.906	0.221	0.846	0.133
	HEL	0.912	0.300	0.933	0.193	0.918	0.099
	BHEL1	0.968	0.313	0.958	0.228	0.936	0.149
	BHEL2	0.942	0.310	0.952	0.215	0.931	0.026
	BpHEL1	0.942	0.302	0.942	0.210	0.946	0.119
	BpHEL2	0.940	0.295	0.940	0.207	0.940	0.117
	IF	0.925	0.286	0.908	0.202	0.903	0.093
	BIFEL1	0.932	0.469	0.950	0.240	0.922	0.172
	BIFEL2	0.928	0.477	0.945	0.240	0.934	0.170
	BpIF1	0.924	0.288	0.947	0.212	0.938	0.129
	BpIF2	0.924	0.287	0.945	0.210	0.927	0.128

Table (2.6) Coverage probabilities and average lengths of 95% confidence intervals for sensitivity θ at a fixed level of specificity p when $F = Normal(1, 1)$ and $G = exp(0.5)$

(m,n)	Methods	$p = 0.95$, $\theta = 0.27$		$p = 0.90$, $\theta = 0.32$		$p = 0.80$, $\theta = 0.40$	
		Coverage Probability	Average Length	Coverage Probability	Average Length	Coverage Probability	Average Length
(20, 20)	NA	0.898	0.430	0.922	0.461	0.926	0.496
	BTI	0.879	0.432	0.903	0.449	0.915	0.468
	BTII	0.878	0.432	0.909	0.449	0.924	0.468
	JEL	0.874	0.272	0.903	0.327	0.939	0.410
	HEL	0.902	0.430	0.938	0.453	0.959	0.475
	BHEL1	0.934	0.421	0.949	0.433	0.968	0.445
	BHEL2	0.940	0.433	0.949	0.450	0.959	0.468
	BpHEL1	0.936	0.437	0.949	0.459	0.954	0.483
	BpHEL2	0.932	0.423	0.944	0.444	0.944	0.467
	IFEL	0.932	0.427	0.943	0.460	0.949	0.496
	BIFEL1	0.943	0.414	0.955	0.448	0.957	0.483
	BIFEL2	0.944	0.435	0.951	0.472	0.953	0.510
	BpIFEL1	0.926	0.430	0.949	0.467	0.943	0.504
	BpIFEL2	0.915	0.425	0.943	0.461	0.939	0.498
(50, 50)	NA	0.926	0.282	0.939	0.297	0.943	0.316
	BTI	0.925	0.289	0.931	0.300	0.940	0.312
	BTII	0.930	0.289	0.936	0.300	0.946	0.312
	JEL	0.775	0.180	0.874	0.224	0.946	0.278
	HEL	0.946	0.286	0.949	0.302	0.955	0.315
	BHEL1	0.949	0.282	0.951	0.294	0.958	0.305
	BHEL2	0.948	0.285	0.949	0.300	0.955	0.312
	BpHEL1	0.948	0.284	0.956	0.302	0.952	0.317
	BpHEL2	0.945	0.280	0.955	0.297	0.952	0.312
	IFEL	0.938	0.293	0.946	0.299	0.952	0.317
	BIFEL1	0.955	0.289	0.951	0.299	0.951	0.316
	BIFEL2	0.952	0.300	0.955	0.308	0.959	0.323
	BpIFEL1	0.951	0.289	0.952	0.298	0.957	0.318
	BpIFEL2	0.949	0.287	0.950	0.297	0.948	0.317
(100, 100)	NA	0.933	0.199	0.940	0.210	0.946	0.224
	BTI	0.940	0.209	0.942	0.214	0.943	0.223
	BTII	0.946	0.209	0.948	0.214	0.950	0.223
	JEL	0.649	0.134	0.869	0.169	0.944	0.203
	HEL	0.948	0.209	0.953	0.215	0.951	0.224
	BHEL1	0.947	0.207	0.951	0.212	0.955	0.221
	BHEL2	0.95	0.208	0.953	0.214	0.952	0.224
	BpHEL1	0.944	0.210	0.950	0.215	0.951	0.225
	BpHEL2	0.947	0.208	0.947	0.214	0.952	0.223
	IFEL	0.940	0.202	0.944	0.211	0.951	0.224
	BIFEL1	0.944	0.205	0.949	0.212	0.959	0.223
	BIFEL2	0.941	0.208	0.946	0.215	0.956	0.226
	BpIFEL1	0.936	0.201	0.942	0.211	0.953	0.224
	BpIFEL2	0.939	0.201	0.941	0.211	0.953	0.224

(m,n)	Methods	$p = 0.95$, $\theta = 0.27$		$p = 0.90$, $\theta = 0.32$		$p = 0.80$, $\theta = 0.40$	
		Coverage Probability	Average Length	Coverage Probability	Average Length	Coverage Probability	Average Length
(500, 500)	NA	0.946	0.090	0.949	0.094	0.947	0.100
	BTI	0.949	0.093	0.952	0.096	0.952	0.100
	BTII	0.953	0.093	0.953	0.096	0.953	0.100
	JEL	0.857	0.168	0.892	0.300	0.944	0.281
	HEL	0.952	0.093	0.952	0.096	0.951	0.101
	BHEL1	0.950	0.093	0.952	0.096	0.953	0.100
	BHEL2	0.952	0.093	0.952	0.096	0.951	0.100
	BpHEL1	0.944	0.093	0.935	0.096	0.934	0.100
	BpHEL2	0.948	0.093	0.942	0.096	0.937	0.100
	IFEL	0.948	0.091	0.950	0.094	0.948	0.100
	BIFEL1	0.943	0.091	0.936	0.095	0.945	0.100
	BIFEL2	0.944	0.092	0.935	0.095	0.944	0.100
	BpIFEL1	0.942	0.091	0.934	0.095	0.933	0.100
	BpIFEL2	0.943	0.091	0.935	0.095	0.932	0.100
(50, 100)	NA	0.923	0.220	0.930	0.233	0.943	0.251
	BTI	0.930	0.234	0.934	0.241	0.941	0.251
	BTII	0.930	0.234	0.942	0.241	0.950	0.251
	JEL	0.665	0.151	0.862	0.190	0.930	0.227
	HEL	0.947	0.231	0.943	0.241	0.950	0.251
	BHEL1	0.945	0.228	0.944	0.237	0.953	0.246
	BHEL2	0.948	0.230	0.943	0.240	0.950	0.250
	BpHEL1	0.944	0.230	0.939	0.241	0.942	0.252
	BpHEL2	0.945	0.228	0.937	0.239	0.939	0.250
	IFEL	0.942	0.234	0.938	0.238	0.949	0.254
	BIFEL1	0.954	0.231	0.946	0.241	0.949	0.256
	BIFEL2	0.944	0.235	0.941	0.245	0.951	0.260
	BpIFEL1	0.943	0.230	0.929	0.236	0.948	0.254
	BpIFEL2	0.941	0.229	0.928	0.235	0.948	0.253
(100, 50)	NA	0.936	0.264	0.942	0.278	0.944	0.293
	BTI	0.933	0.267	0.933	0.276	0.942	0.288
	BTII	0.939	0.267	0.942	0.276	0.948	0.288
	JEL	0.767	0.168	0.872	0.210	0.940	0.261
	HEL	0.947	0.270	0.953	0.280	0.955	0.292
	BHEL1	0.952	0.266	0.954	0.274	0.961	0.285
	BHEL2	0.951	0.269	0.952	0.278	0.956	0.290
	BpHEL1	0.953	0.269	0.951	0.279	0.945	0.293
	BpHEL2	0.951	0.265	0.946	0.276	0.948	0.290
	IFEL	0.940	0.264	0.949	0.276	0.953	0.291
	BIFEL1	0.955	0.262	0.955	0.273	0.957	0.287
	BIFEL2	0.947	0.268	0.952	0.279	0.950	0.293
	BpIFEL1	0.945	0.262	0.950	0.276	0.941	0.292
	BpIFEL2	0.945	0.261	0.946	0.275	0.942	0.291

Table (2.7) Coverage probabilities and average lengths of 95% confidence intervals for sensitivity θ at a fixed level of specificity p when $F = Normal(0, 1)$ and $G = exp(0.1)$

(m,n)	Methods	$p = 0.95$, $\theta = 0.85$		$p = 0.90$, $\theta = 0.89$		$p = 0.80$, $\theta = 0.92$	
		Coverage Probability	Average Length	Coverage Probability	Average Length	Coverage Probability	Average Length
(20, 20)	NA	0.754	0.283	0.872	0.257	0.730	0.205
	BTI	0.790	0.280	0.878	0.259	0.800	0.215
	BTII	0.816	0.280	0.864	0.259	0.826	0.215
	JEL	0.802	0.272	0.800	0.279	0.714	0.291
	HEL	0.806	0.256	0.889	0.210	0.855	0.128
	BHEL1	0.908	0.343	0.902	0.338	0.915	0.327
	BHEL2	0.925	0.295	0.895	0.284	0.911	0.267
	BpHEL1	0.920	0.290	0.912	0.258	0.891	0.195
	BpHEL2	0.919	0.281	0.892	0.250	0.909	0.189
	IF	0.808	0.248	0.831	0.212	0.823	0.138
	BIFEL1	0.822	0.242	0.841	0.211	0.898	0.160
	BIFEL2	0.805	0.259	0.844	0.231	0.867	0.179
	BpIF1	0.915	0.270	0.867	0.244	0.872	0.196
	BpIF2	0.914	0.266	0.860	0.242	0.872	0.193
(50, 50)	NA	0.903	0.199	0.903	0.181	0.874	0.149
	BTI	0.888	0.211	0.906	0.192	0.900	0.161
	BTII	0.904	0.211	0.910	0.192	0.906	0.161
	JEL	0.821	0.191	0.841	0.208	0.836	0.172
	HEL	0.913	0.213	0.906	0.172	0.910	0.125
	BHEL1	0.949	0.226	0.958	0.208	0.930	0.188
	BHEL2	0.947	0.211	0.958	0.190	0.959	0.164
	BpHEL1	0.935	0.217	0.936	0.192	0.931	0.157
	BpHEL2	0.931	0.214	0.928	0.189	0.929	0.155
	IF	0.901	0.193	0.889	0.173	0.854	0.124
	BIFEL1	0.906	0.194	0.866	0.170	0.824	0.130
	BIFEL2	0.876	0.193	0.884	0.171	0.852	0.133
	BpIF1	0.895	0.194	0.887	0.175	0.856	0.144
	BpIF2	0.893	0.193	0.887	0.174	0.856	0.143
(100, 100)	NA	0.893	0.146	0.927	0.131	0.901	0.109
	BTI	0.906	0.156	0.922	0.139	0.914	0.117
	BTII	0.918	0.156	0.924	0.139	0.916	0.117
	JEL	0.860	0.180	0.859	0.140	0.890	0.124
	HEL	0.954	0.158	0.946	0.138	0.918	0.107
	BHEL1	0.957	0.161	0.955	0.145	0.958	0.127
	BHEL2	0.956	0.155	0.953	0.138	0.962	0.118
	BpHEL1	0.951	0.157	0.945	0.140	0.938	0.117
	BpHEL2	0.949	0.157	0.947	0.139	0.939	0.116
	IF	0.932	0.145	0.928	0.130	0.915	0.105
	BIFEL1	0.947	0.143	0.934	0.130	0.897	0.106
	BIFEL2	0.923	0.142	0.910	0.128	0.888	0.105
	BpIF1	0.930	0.142	0.919	0.128	0.914	0.106
	BpIF2	0.931	0.141	0.919	0.127	0.914	0.105

(m,n)	Methods	$p = 0.95$, $\theta = 0.85$		$p = 0.90$, $\theta = 0.89$		$p = 0.80$, $\theta = 0.92$	
		Coverage Probability	Average Length	Coverage Probability	Average Length	Coverage Probability	Average Length
(500, 500)	NA	0.938	0.068	0.937	0.061	0.941	0.051
	BTI	0.934	0.071	0.932	0.064	0.950	0.053
	BTII	0.940	0.071	0.946	0.064	0.954	0.053
	JEL	0.936	0.080	0.931	0.086	0.922	0.072
	HEL	0.947	0.071	0.951	0.064	0.953	0.053
	BHEL1	0.948	0.071	0.949	0.064	0.948	0.054
	BHEL2	0.943	0.071	0.950	0.063	0.955	0.053
	BpHEL1	0.946	0.071	0.952	0.064	0.951	0.053
	BpHEL2	0.943	0.071	0.955	0.063	0.951	0.053
	IF	0.933	0.068	0.946	0.061	0.939	0.051
	BIFEL1	0.941	0.066	0.939	0.059	0.933	0.050
	BIFEL2	0.940	0.066	0.936	0.059	0.934	0.049
	BpIF1	0.938	0.066	0.937	0.059	0.935	0.049
	BpIF2	0.935	0.066	0.936	0.059	0.936	0.049
(50, 100)	NA	0.871	0.150	0.904	0.136	0.890	0.113
	BTI	0.898	0.166	0.906	0.151	0.896	0.127
	BTII	0.902	0.166	0.910	0.151	0.902	0.127
	JEL	0.847	0.209	0.911	0.144	0.892	0.131
	HEL	0.951	0.171	0.929	0.147	0.891	0.108
	BHEL1	0.951	0.175	0.954	0.158	0.960	0.140
	BHEL2	0.955	0.167	0.949	0.149	0.961	0.127
	BpHEL1	0.946	0.171	0.940	0.151	0.935	0.126
	BpHEL2	0.945	0.169	0.932	0.149	0.933	0.125
	IF	0.906	0.152	0.916	0.134	0.895	0.107
	BIFEL1	0.868	0.141	0.894	0.131	0.912	0.108
	BIFEL2	0.885	0.141	0.896	0.130	0.900	0.108
	BpIF1	0.908	0.146	0.890	0.130	0.902	0.108
	BpIF2	0.896	0.145	0.904	0.130	0.917	0.108
(100, 50)	NA	0.914	0.197	0.922	0.179	0.903	0.148
	BTI	0.894	0.202	0.918	0.182	0.922	0.154
	BTII	0.902	0.202	0.908	0.182	0.930	0.154
	JEL	0.876	0.200	0.895	0.207	0.832	0.189
	HEL	0.909	0.200	0.911	0.171	0.885	0.129
	BHEL1	0.954	0.212	0.955	0.197	0.938	0.176
	BHEL2	0.950	0.199	0.951	0.182	0.962	0.156
	BpHEL1	0.934	0.204	0.929	0.185	0.926	0.152
	BpHEL2	0.931	0.202	0.926	0.183	0.921	0.150
	IF	0.912	0.190	0.909	0.172	0.888	0.127
	BIFEL1	0.888	0.192	0.887	0.171	0.902	0.130
	BIFEL2	0.896	0.191	0.896	0.171	0.881	0.133
	BpIF1	0.911	0.192	0.909	0.174	0.908	0.143
	BpIF2	0.910	0.191	0.909	0.174	0.900	0.143

Table (2.8) 95% level confidence intervals and point estimates for the sensitivity θ of different biomarkers at specificity $p = 0.95$.

	Confidence interval			Point estimate $\hat{\theta}$		
Biomarker	TAU/ABETA	FDG	ADAS11	TAU/ABETA	FDG	ADAS11
HEL	(0.001, 0.024)	(0.537, 0.825)	(0.688, 0.964)	0.012	0.694	0.865
BHEL1	(0.005, 0.067)	(0.529, 0.813)	(0.654, 0.941)	0.027	0.679	0.821
BHEL2	(0.001, 0.049)	(0.539, 0.821)	(0.688, 0.958)	0.017	0.689	0.850
BpHEL1	(0.000, 0.073)	(0.511, 0.848)	(0.721, 0.962)	0.012	0.694	0.864
BpHEL2	(0.000, 0.071)	(0.522, 0.841)	(0.720, 0.960)	0.012	0.694	0.865
IFEL	(0.000, 0.039)	(0.617, 0.804)	(0.813, 0.959)	0.017	0.719	0.896
BIFEL1	(0.004 ,0.037)	(0.610, 0.799)	(0.804, 0.953)	0.021	0.711	0.886
BIFEL2	(0.001, 0.036)	(0.613, 0.801)	(0.809, 0.957)	0.017	0.714	0.891
BpELIF1	(0.001, 0.042)	(0.625, 0.806)	(0.818, 0.962)	0.017	0.719	0.896
BpELIF2	(0.001, 0.043)	(0.617, 0.805)	(0.818, 0.959)	0.017	0.718	0.896

Table (2.9) 95% level confidence intervals and point estimates for the sensitivity θ of different biomarkers at specificity $p = 0.9$.

	Confidence interval			Point estimate $\hat{\theta}$		
Biomarker	TAU/ABETA	FDG	ADAS11	TAU/ABETA	FDG	ADAS11
HEL	(0.010, 0.128)	(0.666, 0.847)	(0.913, 0.986)	0.059	0.765	0.959
BHEL1	(0.024, 0.207)	(0.659, 0.840)	(0.894, 0.980)	0.095	0.755	0.945
BHEL2	(0.012, 0.175)	(0.666, 0.845)	(0.908, 0.985)	0.071	0.762	0.955
BpHEL1	(0.005, 0.166)	(0.594, 0.899)	(0.864, 0.998)	0.059	0.766	0.959
BpHEL2	(0.006, 0.167)	(0.597, 0.898)	(0.859, 0.998)	0.058	0.764	0.959
IFEL	(0.000, 0.128)	(0.658, 0.844)	(0.919, 0.995)	0.051	0.758	0.963
BIFEL1	(0.004, 0.130)	(0.651, 0.839)	(0.913, 0.989)	0.059	0.751	0.956
BIFEL2	(0.004, 0.127)	(0.654, 0.842)	(0.919, 0.993)	0.056	0.754	0.961
BpELIF1	(0.003, 0.131)	(0.661, 0.844)	(0.921, 0.995)	0.057	0.758	0.963
BpELIF2	(0.003, 0.127)	(0.659, 0.846)	(0.921, 0.995)	0.056	0.757	0.963

Table (2.10) 95% level confidence intervals and point estimates for the sensitivity θ of different biomarkers at specificity $p = 0.8$.

	Confidence interval			Point estimate $\hat{\theta}$		
Biomarker	TAU/ABETA	FDG	ADAS11	TAU/ABETA	FDG	ADAS11
HEL	(0.082, 0.401)	(0.808, 0.929)	(0.988, 1.000)	0.212	0.876	0.994
BHEL1	(0.100, 0.427)	(0.802, 0.921)	(0.925, 0.997)	0.246	0.867	0.973
BHEL2	(0.086, 0.402)	(0.810, 0.925)	(0.952, 1.000)	0.223	0.873	0.988
BpHEL1	(0.085, 0.377)	(0.739, 0.969)	(0.953, 1.000)	0.213	0.877	0.994
BpHEL2	(0.088, 0.374)	(0.739, 0.966)	(0.954, 1.000)	0.211	0.877	0.994
IFEL	(0.044, 0.356)	(0.800, 0.937)	(0.987, 1.000)	0.209	0.874	1.000
BIFEL1	(0.051, 0.355)	(0.795, 0.933)	(0.987, 0.999)	0.206	0.868	0.994
BIFEL2	(0.047, 0.353)	(0.799, 0.937)	(0.987, 1.000)	0.202	0.872	0.995
BpELIF1	(0.050, 0.363)	(0.802, 0.938)	(0.981, 1.000)	0.210	0.874	1.000
BpELIF2	(0.049, 0.359)	(0.798, 0.937)	(0.980, 1.000)	0.211	0.872	1.000

PART 3

BAYESIAN AND INFLUENCE FUNCTION BASED EMPIRICAL LIKELIHOODS FOR INFERENCE OF SENSITIVITY TO THE EARLY DISEASED STAGE IN DIAGNOSTIC TESTS

3.1 Introduction

In practice, a disease process might be more complicated and involve three diagnostic stages: the non-diseased stage, the early diseased stage, and the fully diseased stage. For example, mild cognitive impairment (MCI) is a transitional stage between the cognitive changes of normal aging and the more severe Alzheimer's Disease (AD)[18]. To be more specific, let Y_1, Y_2, and Y_3 denote the continuous test results of a diagnostic test from the non-diseased, the early diseased, and the fully diseased groups respectively, F_1, F_2, and F_3 represent the corresponding cumulative distribution functions of the test results, and n_1, n_2, and n_3 denote the corresponding sample sizes. Assume that the higher values of the test results indicate greater severity of the disease. Given a pair of threshold values c_1 and c_2 ($c_1 < c_2$), the subject is identified as non-diseased if the test result is smaller than c_1, as fully diseased if the test result is greater than c_2, and as early diseased if the test result is between c_1 and c_2. The specificity P_1, which is the correct classification rate of individuals in the non-diseased stage, the sensitivity P_2 to the early diseased stage, and the sensitivity P_3 to the fully diseased stage, are defined as

$$
\begin{aligned}
P_1 &= F_1(c_1), \\
P_2 &= F_2(c_2) - F_2(c_1) = F_2[F_3^{-1}(1 - P_3)] - F_2[F_1^{-1}(P_1)], \\
P_3 &= 1 - F_3(c_2),
\end{aligned}
$$

respectively. Given P_1 and P_3, c_1 and c_2 can be determined if F_1 and F_3 are known/specified. P_2, the sensitivity to the early diseased stage given the specificity P_1 and the sensitivity to

the fuly diseased stage P_3, can be formulated as a function of P_1 and P_3, which also defines a surface in the three-dimensional space (P_1, P_3, P_2) with $0 \leq P_1, P_3 \leq 1$, namely, the ROC surface of the test. The ROC surface was introduced by Scurfield [45]. A few years later, Mossman [46] independently proposed a similar construction, implemented in Mathematical by Heckerling[47]. A non-parametric estimation of ROC surface was proposed by Nakas and Yiannoutsos[48] and was reshaped later by Xiong et al.[49], and Li and Zhou[50].

The probability associated with the early detection of the disease, such as AD, is a very important accuracy measure for the diagnostic test of the disease with three ordinal stages. Dong et al.[19] first provided parametric and non-parametric confidence intervals for P_2, the sensitivity to the early diseased stage, depending on either normality assumption or Box–Cox transformation to normality. However, their approaches fail if the normal assumption can not be satisfied. Dong and Tian[20] proposed two empirical likelihood (EL) based confidence intervals (ELP and ELB) for P_2, which can overcome the normal assumption. However, the empirical likelihood ratio follows a scaled chi-square distribution asymptotically. Thus an extra step, density estimation, or bootstrap procedure is required to estimate this scale, respectively. Similar to the study in two classification problems, we proposed an influence function empirical likelihood-based and Bayesian empirical likelihood (BEL and BpEL) based confidence intervals for P_2 at a given value of the pair (P_1, P_3) in this part.

The part is organized as follows. In Section 3.2, we review Dong and Tian[20]'s EL methods for interval estimation of sensitivity to the early stage. In Section 3.3, we introduce a new EL ratio statistic for sensitivity to the early stage based on the influence function. In Section 3.4, we propose Bayesian EL methods based on influence function and Dong and Tian[20]'s EL methods. In Section 3.5, we conduct simulation studies to compare the performance of the proposed methods with existing methods. In Section 3.6, we apply the new methods to a real data set to assess the diagnostic accuracy of two biomarkers in the detection of Alzheimer's disease.

3.2 Existing methods for sensitivity to early stage inference

Without normal assumption, Dong and Tian[20] proposed two empirical likelihood-based confidence intervals for P_2. They defined an indicator function Φ:

$$\Phi(X,Y,Z) = \begin{cases} 1, & X < Y < Z, \\ \frac{1}{2}, & X = Y < Z \text{ or } X < Y = Z, \\ \frac{1}{6}, & X = Y = Z, \\ 0, & otherwise, \end{cases}$$

and a random variable U:

$$U(Y) = \Phi[F_1^{-1}(P_1), Y, F_3^{-1}(1-P_3)].$$

The value of $U(Y_2)$ can be interpreted as the placement value of Y_2 in the healthy and fully diseased populations.

Let $\{Y_{1,j} : j = 1, 2, \cdots, n_1\}$, $\{Y_{2,j} : j = 1, 2, \cdots, n_2\}$, and $\{Y_{3,j} : j = 1, 2, \cdots, n_3\}$ denote the n_1, n_2, and n_3 test results from the non-diseased, early stage, and diseased groups respectively. From the following relationship between $U(Y_2)$ and P_2,

$$\begin{aligned} E(U(Y_2)) &= E\{\Phi[F_1^{-1}(P_1), Y_2, F_3^{-1}(1-P_3)]\} \\ &= P[F_1^{-1}(P_1) < Y_2 < F_3^{-1}(1-P_3)] \\ &= P[F_1^{-1}(P_1) < Y_2 \leq F_3^{-1}(1-P_3)] \\ &= P_2, \end{aligned}$$

a profile empirical likelihood for P_2 can be defined as

$$L(P_2) = \sup_{\boldsymbol{p}} \left\{ \prod_{j=1}^{n_2} p_j : \sum_{j=1}^{n_2} p_j = 1, \sum_{j}^{n_2} p_j(\hat{U}_j - P_2) = 0 \right\}, \tag{3.1}$$

where $\hat{U}_j = \Phi[\hat{F}_1^{-1}(P_1), Y_{2,j}, \hat{F}_3^{-1}(1-P_3)]$, $j = 1, 2, \ldots, n_2$, $\hat{F}_1$ and $\hat{F}_3$ are the empirical

distributions of F_1 and F_3, respectively. Using the Lagrange multiplier method, we can easily obtain the expression of p_j:

$$\tilde{p}_j = \frac{1}{n_2}\{1 + \tilde{\lambda}(\hat{U}_j - P_2)\}^{-1}$$

where $\tilde{\lambda}$ is the solution of

$$\frac{1}{n_2}\sum_{j=1}^{n_2}\frac{\hat{U}_j - P_2}{1 + \tilde{\lambda}(\hat{U}_j - P_2)} = 0,$$

and the corresponding profile empirical log-likelihood ratio for P_2:

$$l(P_2) = 2\sum_{j=1}^{n_2}\log\{1 + \tilde{\lambda}(\hat{U}_j - P_2)\}.$$

Dong and Tian[20] shown that the asymptotic distribution of the log-EL ratio is a scaled chi-square distribution with one degree of freedom. Thus, a $100(1-\alpha)\%$ level empirical likelihood-based confidence interval (ELP) for P_2 can be constructed as follows:

$$CI_1(P_2) = \{P_2 : \hat{r}_{P_1,P_2,P_3} l(P_2) \leq \chi_1^2(1-\alpha)\},$$

where $\chi_1^2(1-\alpha)$ is the $(1-\alpha)$th quantile of χ_1^2, and $\hat{r}_{P_1,P_2,P_3}$ is an estimate for the scale constant:

$$\hat{r}_{P_1,P_2,P_3} = \frac{\hat{\bar{P}}_2(1-\hat{\bar{P}}_2)}{n_2\hat{\sigma}^2_{\hat{\bar{P}}_2}} \tag{3.2}$$

where

$$\hat{\bar{P}}_2 = \frac{\sum_{j=1}^{n_2} I[\hat{F}_1^{-1}(P_1) < Y_{2,j} \leq \hat{F}_3^{-1}(1-P_3)]}{n_2}, \tag{3.3}$$

$$\hat{\sigma}^2_{\hat{\bar{P}}_2} = \frac{\hat{\bar{P}}_2(1-\hat{\bar{P}}_2)}{n_2} + \frac{P_1(1-P_1)}{n_1}\frac{\hat{f}_2^2[\hat{F}_1^{-1}(P_1)]}{\hat{f}_1^2[\hat{F}_1^{-1}(P_1)]} + \frac{P_3(1-P_3)}{n_3}\frac{\hat{f}_2^2[\hat{F}_3^{-1}(1-P_3)]}{\hat{f}_3^2[\hat{F}_3^{-1}(1-P_3)]}, \tag{3.4}$$

and $\hat{f}_i$ is a kernel density estimate for the probability density function f_i of Y_i, $i = 1, 2, 3$.

In order to estimate the density function f_i, Dong and Tian[20] used the "over-smoothed bandwidth selector" by Wand and Jones [51] to select the bandwidth with a Gaussian kernel

function. The performance of this method highly depends on the kernel density estimates with the Gaussian kernel, whose bandwidth is chosen without a well recognized standard. Therefore, they proposed the following bootstrap procedure to obtain a bootstrap estimate $\hat{\sigma}^{2*}_{\hat{\hat{P}}_2}$ for the variance instead of $\hat{\sigma}^{2}_{\hat{\hat{P}}_2}$ in Equation (3.2):

Step 1: Draw bootstrap resamples of sizes n_1, n_2, and n_3 with replacement from the non-diseased sample Y_{1j}'s, the early diseased sample Y_{2j}'s, and the fully diseased sample Y_{3j}'s, respectively. Denote the bootstrap samples as $\{Y^b_{ij}\}$, $i = 1, 2, 3$, $j = 1, 2, \ldots, n_i$.

Step 2: Calculate the bootstrap version $\hat{\hat{P}}^b_2$ of $\hat{\hat{P}}_2$ according to Equation (3.3).

Step 3: Repeat the first two steps B times to obtain the bootstrap variance estimate of $\hat{\hat{P}}_2$, which is defined as

$$\hat{\sigma}^{2*}_{\hat{\hat{P}}_2} = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\hat{P}}^b_2 - \bar{\hat{\hat{P}}}^*_2)^2,$$

where $\bar{\hat{\hat{P}}}^*_2 = \frac{1}{B}\sum_{b=1}^{B}\hat{\hat{P}}^b_2$.

This leads to the second $100(1-\alpha)\%$ level empirical likelihood confidence interval (ELB) for P_2:

$$CI_2(P_2) = \{P_2 : r^*_{P_1,P_2,P_3} l(P_2) \leq \chi^2_1(1-\alpha)\},$$

where

$$r^*_{P_1,P_2,P_3} = \frac{\bar{\hat{\hat{P}}}^*_2(1-\bar{\hat{\hat{P}}}^*_2)}{n_2\hat{\sigma}^{2*}_{\hat{\hat{P}}_2}} \tag{3.5}$$

3.3 Influence function-based empirical likelihood (IF) method

The application of the existing empirical likelihood-based ELP and ELB intervals for P_2 needs estimation of a scale constant by density estimation and bootstrap process which is time consuming. The finite sample performance of ELP and ELB intervals depends on

the estimation accuracy of the estimators for the scale constant. To get rid of the estimation for the unknown scale constant, we propose a new influence function-based EL method to construct confidence intervals for sensitivity to the early diseased stage in this section.

We combine the samples $\{Y_{i,j} : i = 1, 2, 3; j = 1, 2, \cdots, n_i\}$ as:

$$Z_l = \begin{cases} Y_{1,l}, & l = 1, \ldots, n_1, \\ Y_{2,l-n_1}, & l = 1 + n_1, \ldots, n_1 + n_2, \\ Y_{3,l-n_1-n_2}, & l = 1 + n_1 + n_2, \ldots, N, \end{cases}$$

where $N = n_1 + n_2 + n_3$.

Let $c_1 = F_1^{-1}(P_1)$, and $c_2 = F_3^{-1}(1 - P_3)$. Define $\hat{F}_2$ as the empirical distribution of F_2, $\hat{c}_1 = \hat{F}_1^{-1}(P_1)$ (i.e., the P_1-th sample quantile of $Y_{1,j}$'s), and $\hat{c}_2 = \hat{F}_3^{-1}(1 - P_3)$ (i.e., the $(1-P_3)$-th sample quantile of $Y_{3,j}$'s). Then the sensitivity P_2 of the test to the early diseased stage can be consistently estimated by

$$\tilde{P}_2 = \hat{F}_2(\hat{c}_2) - \hat{F}_2(\hat{c}_1) = \hat{F}_2[\hat{F}_3^{-1}(1 - P_3)] - \hat{F}_2[\hat{F}_1^{-1}(P_1)] \equiv \hat{P}_2(P_1, P_3).$$

We have the following decomposition:

$$\begin{aligned} \tilde{P}_2 - P_2 &= [\hat{F}_2(\hat{c}_2) - \hat{F}_2(\hat{c}_1)] - [F_2(c_2) - F_2(c_1)] \\ &= [\hat{F}_2(\hat{c}_2) - \hat{F}_2(c_2)] - [\hat{F}_2(\hat{c}_1) - \hat{F}_2(c_1)] + \big\{[\hat{F}_2(c_2) - \hat{F}_2(c_1)] - [F_2(c_2) - F_2(c_1)]\big\} \\ &\equiv I_1 - I_2 + I_3. \end{aligned} \tag{3.6}$$

The third term of Equation (3.6) can be written as

$$\begin{aligned} I_3 &= [\hat{F}_2(c_2) - \hat{F}_2(c_1)] - [F_2(c_2) - F_2(c_1)] \\ &= \frac{1}{n_2} \sum_{j=1}^{n_2} [I(c_1 < Y_{2,j} \le c_2) - P_2] \\ &= \frac{1}{N} \sum_{l=n_1+1}^{n_1+n_2} \frac{N}{n_2} [I(c_1 < Z_l \le c_2) - P_2]. \end{aligned} \tag{3.7}$$

From the Bahadur representation for the sample quantiles $\hat{c}_1$ and $\hat{c}_2$ [37],

$$\hat{c}_1 - c_1 = \frac{P_1 - \frac{1}{n_1}\sum_{j=1}^{n_1} I(Y_{1,j} \le c_1)}{f_1(c_1)} + o_p({n_1}^{-\frac{1}{2}}),$$

$$\hat{c}_2 - c_2 = \frac{\frac{1}{n_3}\sum_{j=1}^{n_3} I(Y_{3,j} > c_2) - P_3}{f_3(c_2)} + o_p({n_3}^{-\frac{1}{2}}),$$

it follows that

$$\begin{aligned}
I_2 &= \hat{F}_2(\hat{c}_1) - \hat{F}_2(c_1) = \int [I(y \le \hat{c}_1) - I(y \le c_1)] d\hat{F}_2(y) \\
&= \int [I(y \le \hat{c}_1) - I(y \le c_1)] dF_2(y) + o_p(n_2^{-1/2}) \\
&= f_2(c_1)(\hat{c}_1 - c_1) + o_p(n_2^{-1/2} + n_1^{-1/2}) \\
&= -\frac{1}{n_1}\frac{f_2(c_1)}{f_1(c_1)}\sum_{j=1}^{n_1}[I(Y_{1,j} \le c_1) - P_1] + o_p(n_2^{-1/2} + n_1^{-1/2}) \\
&= -\frac{1}{N}\sum_{l=1}^{n_1}\frac{N}{n_1}\frac{f_2(c_1)}{f_1(c_1)}[I(Z_l \le c_1) - P_1] + o_p(N^{-1/2}),
\end{aligned} \tag{3.8}$$

$$\begin{aligned}
I_1 &= \hat{F}_2(\hat{c}_2) - \hat{F}_2(c_2) = \int [I(y \le \hat{c}_2) - I(y \le c_2)] d\hat{F}_2(y) \\
&= \int [I(y \le \hat{c}_2) - I(y \le c_2)] dF_2(y) + o_p(n_2^{-1/2}) \\
&= f_2(c_2)(\hat{c}_2 - c_2) + o_p(n_2^{-1/2} + n_3^{-1/2}) \\
&= \frac{1}{n_3}\frac{f_2(c_2)}{f_3(c_2)}\sum_{j=1}^{n_3}[I(Y_{3,j} > c_2) - P_3] + o_p(n_2^{-1/2} + n_3^{-1/2}) \\
&= \frac{1}{N}\sum_{l=n_1+n_2+1}^{N}\frac{N}{n_3}\frac{f_2(c_2)}{f_3(c_2)}[I(Z_l > c_2) - P_3] + o_p(N^{-1/2}).
\end{aligned} \tag{3.9}$$

Therefore,

$$\tilde{P}_2 - P_2 = \frac{1}{N}\sum_{l=1}^{N} W_l(P_1, P_2, P_3) + o_p(N^{-1/2}), \tag{3.10}$$

where

$$W_l(P_1,P_2,P_3)=\begin{cases}\frac{N}{n_1}\frac{f_2(c_1)}{f_1(c_1)}[I(Z_l\le c_1)-P_1], & l=1,\ldots,n_1,\\ \frac{N}{n_2}[I(c_1<Z_l\le c_2)-P_2], & l=n_1+1,\ldots,n_1+n_2,\\ \frac{N}{n_3}\frac{f_2(c_2)}{f_3(c_2)}[I(Z_l>c_2)-P_3], & l=n_1+n_2+1,\ldots,N,\end{cases} \tag{3.11}$$

is called the influence function of P_2.

From Equation (3.10), we can get the following asymptotic distribution of the empirical estimator $\tilde{P}_2$ for P_2.

Proposition 3.1: Assume that F_1, F_2, and F_3 are continuous distribution functions with density functions f_1, f_2 and f_3, respectively, $f_2'(x)$ is bounded in neighborhoods of $c_1=F_1^{-1}(P_1)$ and $c_2=F_3^{-1}(1-P_3)$, $f_1(c_1)$ and $f_3(c_2)$ are strictly positive, $\frac{f_2(x)}{f_1(x)}$ is bounded in a neighborhood of $c_1=F_1^{-1}(P_1)$, and $\frac{f_2(x)}{f_3(x)}$ is bounded in a neighborhood of $c_2=F_3^{-1}(1-P_3)$. If $\lim\frac{n_1}{n_2}=\rho_1$ $(0<\rho_1<\infty)$, $\lim\frac{n_3}{n_2}=\rho_2$ $(0<\rho_2<\infty)$ and $\lim\frac{n_1}{n_3}=\rho_3$ $(0<\rho_3<\infty)$, then

$$\sqrt{N}(\tilde{P}_2-P_2)\xrightarrow{d}N(0,\sigma^2), \tag{3.12}$$

where $N=n_1+n_2+n_3$, and

$$\begin{aligned}\sigma^2=\ & (1+\rho_1^{-1}+\rho_3^{-1})P_1(1-P_1)\frac{f_2^2[F_1^{-1}(P_1)]}{f_1^2[F_1^{-1}(P_1)]}+(1+\rho_1+\rho_2)P_2(1-P_2)\\ & +(1+\rho_2^{-1}+\rho_3)P_3(1-P_3)\frac{f_2^2[F_3^{-1}(1-P_3)]}{f_3^2[F_3^{-1}(1-P_3)]}.\end{aligned}$$

Dong and Tian[20] derived a similar conclusion to Proposition 1, but they didn't explicitly give the proof. Proposition 1 can be used to construct a normal approximation-based confidence interval for P_2 if we can get a good estimate for σ^2. But estimating σ^2 involves an estimation of unknown densities and quantiles. To avoid the complex variance estimation, we propose the following influence function-based EL method for inference of P_2.

Based on the influence function in (3.11), an EL for the sensitivity P_2 to the early

diseased stage at a given pair of (P_1, P_3) can be defined as follows:

$$L_{IF}(P_1, P_2, P_3) = \sup_{\boldsymbol{p}} \left\{ \prod_{l=1}^{N} p_l : \sum_{l=1}^{N} p_l = 1, \sum_{l=1}^{N} p_l \hat{W}_l(P_1, P_2, P_3) = 0 \right\}, \tag{3.13}$$

where $\boldsymbol{p} = (p_1, \cdots, p_N)$ is a probability vector, and $\hat{W}_l(P_1, P_2, P_3)$ is the estimated influence function of P_2 given as follows

$$\hat{W}_l(P_1, P_2, P_3) = \begin{cases} \frac{N}{n_1} \frac{\hat{f}_2(\hat{c}_1)}{\hat{f}_1(\hat{c}_1)} [I(Z_l \leq \hat{c}_1) - P_1], & l = 1, \ldots, n_1, \\ \frac{N}{n_2} [I(\hat{c}_1 < Z_l \leq \hat{c}_2) - P_2], & l = n_1 + 1, \ldots, n_1 + n_2, \\ \frac{N}{n_3} \frac{\hat{f}_2(\hat{c}_2)}{\hat{f}_3(\hat{c}_2)} [I(Z_l > \hat{c}_2) - P_3], & l = n_1 + n_2 + 1, \ldots, N, \end{cases} \tag{3.14}$$

where $\hat{f}_i$ is the density estimators for f_i, $i = 1, 2, 3$. We use the "over-smoothed bandwidth selector" to select the bandwidth for the Gaussian kernel function for f_i as described in Dong and Tian's study[20].

By the Lagrange multiplier, the maximization of Equation (3.13) is achieved at

$$p_l = \frac{1}{N} [1 + \lambda \hat{W}_l(P_1, P_2, P_3)]^{-1}, l = 1, \ldots, N,$$

where λ is the solution of

$$\frac{1}{N} \sum_{l=1}^{N} \frac{\hat{W}_l(P_1, P_2, P_3)}{1 + \lambda \hat{W}_l(P_1, P_2, P_3)} = 0. \tag{3.15}$$

The corresponding empirical log-likelihood ratio statistic is

$$l_{IF}(P_1, P_2, P_3) = -\sum_{l=1}^{N} \log\{1 + \lambda \hat{W}_l(P_1, P_2, P_3)\}. \tag{3.16}$$

When test results $Y_{i,j}$'s are not all greater/smaller than $\hat{c}_i$, the empirical log-likelihood ratio $l_{IF}(P_1, P_2, P_3)$ is well defined on $(0, 1)$. The following theorem establishes the asymptotic distribution of $l_{IF}(P_1, P_2, P_3)$, and the proof is given in the Appendix B.

Theorem 3.1: Assume that F_1, F_2, and F_3 are continuous distribution functions with

density functions f_1, f_2 and f_3, respectively, $f_2'(x)$ is bounded in neighborhoods of $c_1 = F_1^{-1}(P_1)$ and $c_2 = F_3^{-1}(1-P_3)$, $f_1(c_1)$ and $f_3(c_2)$ are strictly positive, $\frac{f_2(x)}{f_1(x)}$ is bounded in a neighborhood of $c_1 = F_1^{-1}(P_1)$ and $\frac{f_2(x)}{f_3(x)}$ is bounded in a neighborhood of $c_2 = F_3^{-1}(1-P_3)$. If $\lim \frac{n_1}{n_2} = \rho_1$ $(0 < \rho_1 < \infty)$, $\lim \frac{n_3}{n_2} = \rho_2$ $(0 < \rho_2 < \infty)$ and $\lim \frac{n_1}{n_3} = \rho_3$ $(0 < \rho_3 < \infty)$, and P_2^0 is the true value of sensitivity P_2 to the early diseased stage at a fixed level P_1 of specificity and P_3 of sensitivity to the fully diseased stage, then the asymptotic distribution of $-2l_{IF}(P_1, P_2^0, P_3)$ is a standard chi-squared distribution with one degree of freedom as $n_1, n_2, n_3 \to \infty$.

From Theorem 3.1, a $100(1-\alpha)\%$ level influence function-based EL confidence interval (IFEL) for P_2 can be constructed as follows:

$$CI_{IF}(P_1, P_2, P_3) = \{P_2 : -2l_{IF}(P_1, P_2, P_3) \leq \chi_1^2(1-\alpha)\}.$$

3.4 Bayesian Empirical Likelihood (BEL) Method

Bayesian empirical likelihood has the potential to be used as the basis for Bayesian inference. As Lazar[17] pointed out that EL has many asymptotic properties obtained from parametric models, Bayesian EL methods are naturally used to quantify uncertainty and can have good small sample properties. In this section, we propose two types of Bayesian EL methods to construct reliable intervals to improve sensitivity to early disease.

3.4.1 Bayesian empirical likelihood-based on sensitivity

We follow Lazar[17]'s idea to combine empirical likelihood $L(P_2)$ with prior $\pi(P_2)$ on P_2 by the Bayesian theorem to obtain a posterior:

$$\pi(P_2|data) \propto L(P_2)\pi(P_2).$$

We consider reference priors, originally introduced by [31], and further developed in [32333233], on P_2 in this study. Reference priors only depend on the assumed model and the

available data. In our problem, we do not have a parametric model. Therefore, we follow [30] to derive reference priors for EL based on different measurements. The following proposition gives the reference priors for the Bayesian EL method where $[L(P_2)]^{\hat{r}_{P_1,P_2,P_3}}$ in ELP is used as the likelihood.

Proposition 3.2: The reference prior based on the relative entropy for ELP is

$$\pi_{EL,1}(P_2) = \beta(\frac{3}{2}, \frac{3}{2}),$$

and the reference prior based on Hellinger distance is

$$\pi_{EL,2}(P_2) = \beta(\frac{1}{2}, \frac{1}{2}),$$

where $\beta(a, b)$ is the beta distribution with parameters a and b.

The corresponding posterior is

$$\pi_{EL}(P_2|Y) \propto \prod_{j=1}^{n_2} [1 + \tilde{\lambda}(\hat{U}_j - P_2)]^{-\hat{r}_{P_1,P_2,P_3}} \pi_{EL}(P_2),$$

where $\pi_{EL}(P_2) = \pi_{EL,1}(P_2)$, or $\pi_{EL,2}(P_2)$, and $\hat{U}_j = \Phi[\hat{F}_1^{-1}(P_1), Y_{2,j}, \hat{F}_3^{-1}(1 - P_3)]$, $j = 1, 2, \ldots, n_2$. Based on these posteriors, we can calculate equal-tail credible intervals for P_2. These two new methods are called as Bayesian Empirical likelihood 1 (BEL1) and Bayesian Empirical likelihood 2 (BEL2).

Similarly, to construct Bayesian credible intervals for P_2 based on the IFEL using influence function $W_l(P_1, P_2, P_3)$ in Equation (3.11), we propose the following reference priors:

$$\pi_{IF,1}(P_2) \propto \Big[\frac{1}{n_1} P_1(1 - P_1) \frac{f_2^2[F_1^{-1}(P_1)]}{f_1^2[F_1^{-1}(P_1)]} + \frac{1}{n_2} P_2(1 - P_2) + \frac{1}{n_3} P_3(1 - P_3) \frac{f_2^2[F_3^{-1}(1 - P_3)]}{f_3^2[F_3^{-1}(1 - P_3)]}\Big]^{\frac{1}{2}},$$

and

$$\pi_{IF,2}(P_2) \propto \Big[\frac{1}{n_1}P_1(1-P_1)\frac{f_2^2[F_1^{-1}(P_1)]}{f_1^2[F_1^{-1}(P_1)]} + \frac{1}{n_2}P_2(1-P_2) + \frac{1}{n_3}P_3(1-P_3)\frac{f_2^2[F_3^{-1}(1-P_3)]}{f_3^2[F_3^{-1}(1-P_3)]}\Big]^{-\frac{1}{2}}.$$

These two priors are both proper since $\pi_{IF,1}(P_2)$ is bounded by a constant and $\pi_{IF,2}(P_2)$ is bounded by a beta distribution. In practice, we use $\hat{W}_l(P_1, P_2, P_3)$ to estimate the influence function $W_l(P_1, P_2, P_3)$, and replace f_1, f_2, f_3, c_1 and c_2 with their estimates since they are generally unknown. The posterior based on this approach is then

$$\pi_{IF}(P_2|Z) \propto \prod_{l=1}^{N}[1 + \tilde{\lambda}\hat{W}_l(P_1, P_2, P_3)]^{-1}\pi_{IF}(P_2).$$

where $\pi_{IF}(P_2) = \pi_{IF,1}(P_2),$ or $\pi_{IF,2}(P_2)$. Based on these posteriors, we can calculate equal-tail credible intervals for P_2. Therefore, we have two more new methods called as Bayesian influence function based Empirical likelihood 1 (BIF1) and Bayesian influence function based Empirical likelihood 2 (BIF2).

3.4.2 Bayesian pseudo empirical likelihood (BpEL) based on probability vector

In previous section, we apply Bayesian framework to P_2, the sensitivity to early diseased stage, as the interested parameter. In this section, we apply Rao and Wu's method [34] to obtain Bayesian EL based on probability vector $(p_1, \ldots, p_l)$ instead of P_2. We treat $(p_1, \ldots, p_l)$ as unknown parameters and the EL function is:

$$L_{EL}(p_1, \ldots, p_l) = \prod_{i=1}^{l} p_i,$$

where $l = n_2$ for ELP, and $l = N$ for influence function based EL. Consider the Dirichlet prior $D(\alpha_1, \ldots, \alpha_l)$ on $(p_1, \ldots, p_l)$:

$$\pi(p_1, \ldots, p_l) = c(\alpha_1, \ldots, \alpha_l)\prod_{i=1}^{l} p_i^{\alpha_i - 1},$$

where $c(\alpha_1, \ldots, \alpha_l) = \Gamma(\sum_{i=1}^{l} \alpha_i) / \prod_{i=1}^{l} \Gamma(\alpha_i)$. The posterior distribution of $(p_1, \ldots, p_l)$ given the data is Dirichlet $D(1+\alpha_1, \ldots, 1+\alpha_l)$ and is given by:

$$\pi(p_1, \ldots, p_l | data) = c(1+\alpha_1, \ldots, 1+\alpha_n) \prod_{i=1}^{l} p_i^{\alpha_i}.$$

The posterior of the early stage sensitivity P_2 satisfies the following equation:

$$\sum_{i=1}^{l} p_i \hat{Q}_i(P_2) = 0, \tag{3.17}$$

where $\hat{Q}_i(P_2)$ is an estimating/influence function and $(p_1, \ldots, p_l)$ follows the Dirichlet distribution $D(1+\alpha_1, \ldots, 1+\alpha_l)$. In practice, we can generate samples of $(p_1, \ldots, p_l)$ from $D(1+\alpha_1, \ldots, 1+\alpha_l)$, and by solving Equation (2.12), we get the posterior samples of P_2. Based on these posterior samples, we can calculate the equal-tail credible intervals for sensitivity P_2.

Similar to Section 3.4.1, we consider two types of EL: EL in Equation (3.1) as in ELP method and influence function EL in Equation (3.13). We call them Bayesian pseudo EL (BpEL) and Bayesian pseudo influence function based EL (BpIF), respectively. For BpEL, we use $\hat{W}_{ELP}(P_1, P_2, P_3)$ to replace $\hat{Q}_i(P_2)$ in Equation (3.17), and consider $D(r^*, \ldots, r^*)$ and $D(r^* + \frac{1}{n_2}, \ldots, r^* + \frac{1}{n_2})$ as the priors (labeled BpEL1 and BpEL2, respectively), where $r^* = \hat{r}_{P_1,P_2,P_3}$ is the estimate defined in Equation (3.2) in section 2 for the scale constant. For BpIFEL, similarly, we use $\hat{W}_l(P_1, P_2, P_3)$ to replace $\hat{Q}_i(P_2)$ in Equation (3.17), and consider $D(1, \ldots, 1)$ and $D(1+\frac{1}{N}, \ldots, 1+\frac{1}{N})$ as the priors (labeled as BpIF1 and BpIF2, respectively).

3.5 Simulation Study

Simulation studies are conducted to examine the finite sample performance of the proposed approaches: influence function-based empirical likelihood (IF), Bayesian influence function empirical likelihood methods (BIF1 and BIF2) with reference priors $\pi_{IF,1}(P_2)$ and $\pi_{IF,2}(P_2)$, Bayesian EL methods (BEL1 and BEL2) with reference priors $\pi_{ELP,1}(P_2)$ and

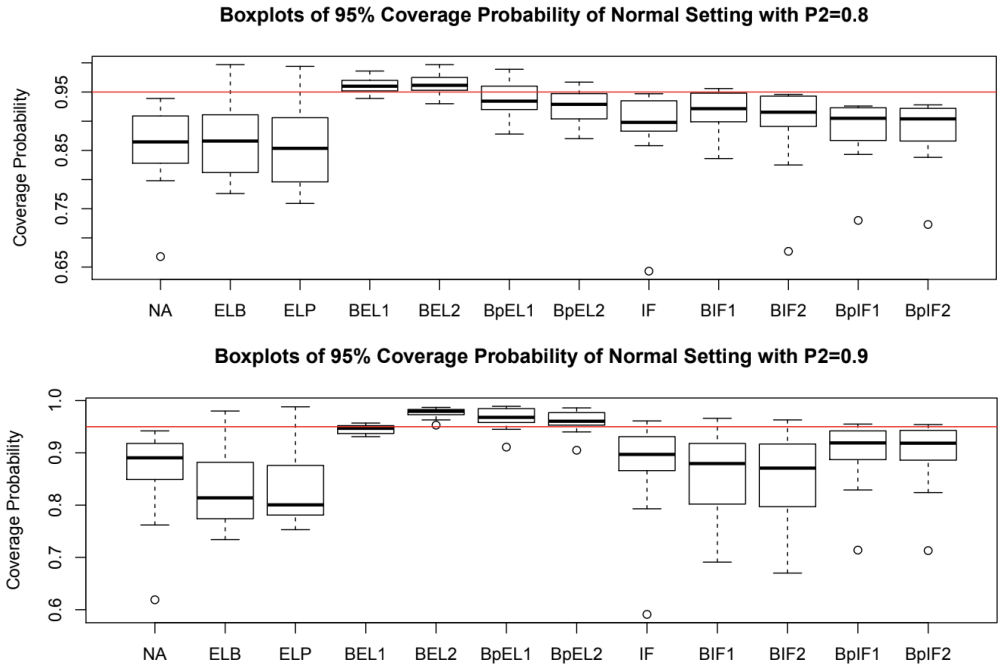
$\pi_{ELP,2}(P_2)$, Bayesian pseudo influence function empirical likelihood methods (BpIF1 and BpIF2), and Bayesian pseudo ELP methods (BpEL1 and BpEL2). We compare them with the existing approaches ELP and ELB proposed by [20]. For comparison purpose, we also include the normal approximation (NA) method by using $\hat{\hat{P}}_2$ in Equation (3.2) as the point estimate and $\hat{\sigma}^2_{\hat{\hat{P}}_2}$ Equation (3.3) as the variance estimate.

We evaluate these approaches under scenarios when the underlying distributions are normal distrutions, beta distributions, and the combined scenario where the normality assumptions cannot be met, that is, gamma distribution for the non-diseased, log-normal distribution for the early diseased and Weibull distribution for the fully diseased groups. Sample sizes (n_1, n_2, n_3) are set as (10, 10, 10), (30, 30, 30), (50, 30, 30), (50, 50, 50), (100, 100, 100), (100, 50, 50) and (100, 100, 50). With a fixed 80% or 90% specificity and a fixed 80% or 90% sensitivity to the fully diseased stage, the parameters for the distributions are chosen correspondingly so that P_2 equals to 80% or 90%. Under each distribution scenario, there are four settings corresponding to different levels of P_1 and P_3, and true value of P_2: (i) P_1=P_3=0.8 and P_2=0.8, (ii) P_1=P_3=0.9 and P_2=0.8, (iii) P_1=P_3=0.8 and P_2=0.9, (iv) P_1=P_3=0.9 and P_2=0.9. Under each setting, 5000 random samples are generated. The simulation results are presented in Tables 3.1–3.2 and Figures 3.1–3.3.

Place Tables 3.1–3.2 here

Under the normal scenario (Table 3.1–3.2 and Figure 3.1), we observe that the confidence intervals of IF related methods(IF, BIF1, BIF2, BpIF1, and BpIF2) are generally conservative. The confidence intervals of the existing methods(NA, ELP, and ELB) are also conservative except the small sample size (10, 10, 10). New methods always have better performance comparing with NA, ELP, and ELB methods for all settings consider here. Especially, BEL1 has the best overall performance in terms of coverage probability closed to 95%. Bayesian and Bayesian pseudo approaches generally have similar or improved performance over ELP, ELB, or IF. IF function related methods have poor performance when the sample size is small. The possible reason is limited small sample results in the poor

Figure (3.1) Boxplots of coverage probabilities under normal distribution setting.



density estimation involved in IF related methods. Comparing the results from the normal distribution setting (i) and (iii) with those from the normal distribution setting (ii) and (iv), which have higher specificity(P_1) and sensitivity(P_3) to the fully diseased group, we can see that the performance of NA, ELB, ELP and IF related methods all depends on the degree of separation of test outcomes in the diseased, early diseased and non-diseased groups. Under the higher specificity and sensitivity to full diseased group, they have lower coverage probability. However, the performance of BEL1, BEL2, BpEL1, and BpEL2 does not obviously change with different P_1 ans P_3. Comparing the results from the normal distribution setting (i) with (iii) or normal distribution setting (ii) with (iv), which have fixed P_1 and P_3 but higher true value of P_2, we note that ELP, ELB, BEL2, BpEL1, IF, BIF1, and BIF2 generally have similar or poorer finite sample performance with higher true value of P_2 and other methods are similar or slightly better. For example, when the sample size is (100,100) with P_1=P_3=0.8, the coverage probability of ELB dropped from 0.944 to 0.906 when the true value of P_2 is changed from 0.8 to 0.9.

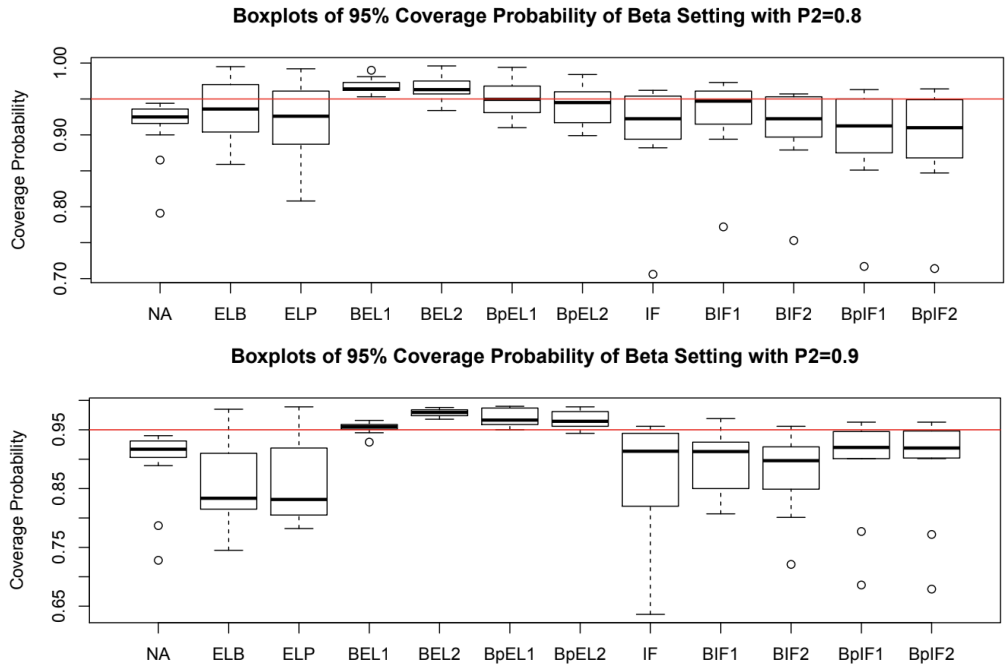
The simulation results under Beta distribution setting are reported in Table 3.3 – 3.4

and Figure 3.2. Similar as the normal distribution setting, we observe that the confidence intervals of NA, ELP, ELB, and FI-related methods(IF, BIF1, BIF2, BpIF1, and BpIF2) are generally conservative. New methods always have better performance and BEL1 has the best overall performance because the coverage probability is closed to 95% except the first setting where IF related methods work well. The performance of NA, ELB, ELP, BpEL1, BpEL2, and IF related methods all depend on the degree of separation of test outcomes in the diseased, early diseased and non-diseased groups. Under the higher specificity and sensitivity to full diseased group, they have slightly lower coverage probability. However, BEL1 and BEL2 do not obviously change when P_1 ans P_3 increase. Comparing the results from the beta distribution setting (i) with (iii) or beta distribution setting (ii) with (iv), which have fixed P_1 ans P_3 but higher true value of P_2, we note that BEL1, BpEL1, and BpEL2 generally have similar or better finite sample performance with higher true value of P_2 and other methods are similar or slightly worse except BpIF1 and BpIF2, which have no obvious trend. Specifically, BpIF1 and BpIF2 perform very well when P_1=P_3=0.8 and true $P_2 = 0.8$ and has similar or slightly lower coverage probability than that of $P_2 = 0.9$. However the coverage probability of BpIF1 and BpIF2 work better when the true value of P_2 is changed from 0.8 to 0.9.

Place Table 3.3–3.4 here

The simulation results under the combined distribution setting are reported in Table 3.5–3.6 and Figure 3.3. Clearly, the new methods always have better performance comparing with ELB and ELP. However, BEL1 does not always have the best overall performance. Bayesian pseudo empirical likelihood methods(BpEL1, BpEL2, BpIF1, and BpIF2) also perform well in most settings considered here. IF related methods also work well when true $P_2 = 0.8$. The performance of ELB, ELP, and IF related methods all depend on the degree of separation of test outcomes in the fully diseased, early diseased, and non-diseased groups. Under the higher specificity and sensitivity to fully diseased group, they have slightly lower coverage probability. NA, BEL1, BEL2, BpEL1, and BpEL2 do not obviously change when P_1 ans

Figure (3.2) Boxplots of 95% coverage probabilities under Beta distribution setting.

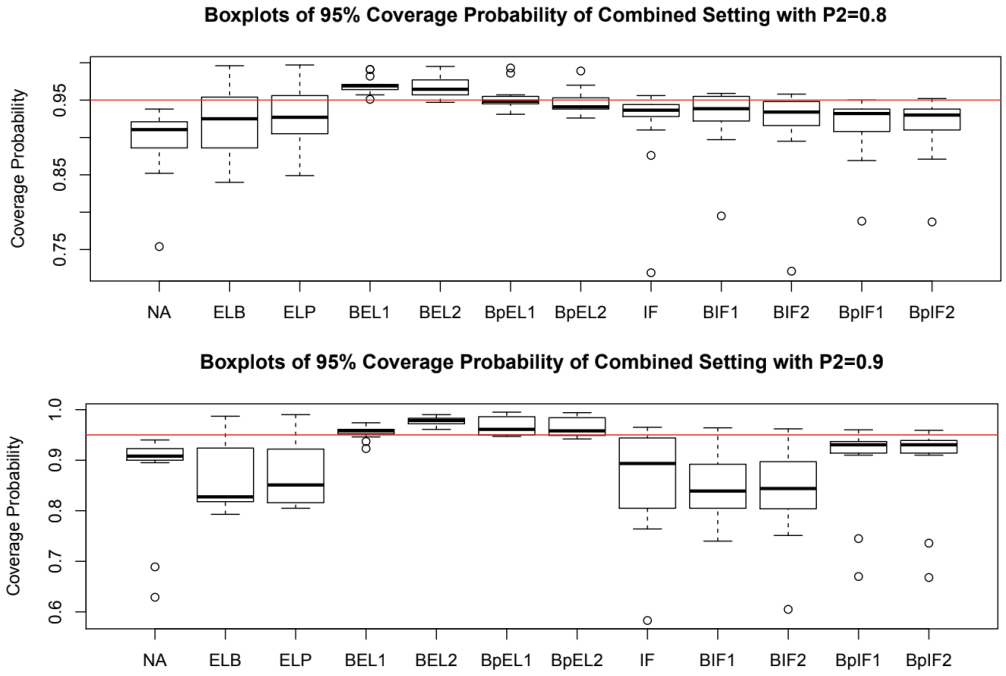


P_3 increase. Comparing the results from the beta distribution setting (i) with (iii) or beta distribution setting (ii) with (iv) which have fixed P_1 ans P_3 but higher true value of P_2, we can see that BEL1, BEL2, BpEL1, BpEL2, BpIF1, and BpIF2 generally have similar or better finite sample performance with higher true value of P_2 and other methods are obvious worse when the true value of P_2 is changed from 0.8 to 0.9, especially IF and Bayesian IF methods. The possible reason might be it is more difficult to preform density estimation when the true P_2 value is higher(higher degree of separation of early diseased test outcomes from other groups).

Place Table 3.5–3.6 here

In summary, new Bayesian and Bayesian pseudo empirical likelihood methods, especially BEL1, are consistent and have coverage probabilities closer to the nominal confidence level than other methods in all settings. The performance of IF and Bayesian IF methods are acceptable in some settings due to poor density estimation.

Figure (3.3) Boxplots of 95% coverage probabilities under the combined distribution setting.

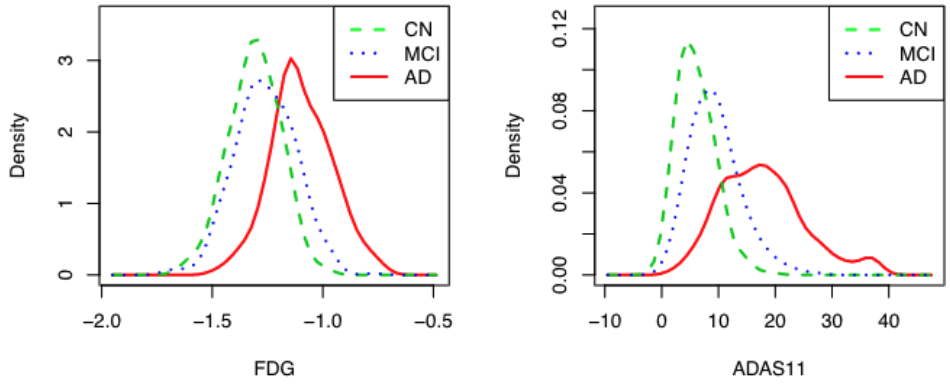


3.6 A Real Example in the Detection of Alzheimer's Disease

In this section, we illustrate the application of the proposed methods to assess the diagnostic accuracy of biomarkers in detecting Alzheimer's disease (AD). The data used in this section were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`adni.loni.usc.edu`). The goal of the ADNI study is to track the progression of the diseases, mild cognitive impairment (MCI) and AD, using biomarkers and clinical measures.

We apply the proposed methods to a small subset of a data-freeze named "QT-PAD Project Data" downloaded on June 29th 2017. It is available in the "Test Data/Data for Challenges" section of the LONI website (ADNI database). Here we only consider non-missing records based on two commonly used biomarkers: fluorodeoxyglucose (FDG), and Alzheimer's Disease Assessment Scale 11 (ADAS11). The dataset we used consists of 203 control subjects (CN) , 389 MCI, and 237 AD patients. We consider MCI as the early stage of AD. Figure 3.4 presents the estimated density curves of FDG and ADAS11 for these three

Figure (3.4) Estimated densities for FDG and ADAS11 in the ADNI data.



groups, respectively. We note that the results of MCI patients are very similar with that of control groups. Therefore, we can expect that the sensitivity of MCI will not be high enough.

Place Table 3.7 here

The point estimates and confidence intervals for the sensitivity to the early stage MCI for these two biomarkers when $P_1 = P_3 = 0.6$ or 0.7 are reported in Table 3.7. As expecting, the pointing estimates of FDG and ADS11 are all small. FDG sensitivity for FDG drops from 0.37 to 0.2 (around 50 percentage points) if the P_1 and P_3 are increased from 0.6 to 0.7. Specifically, BEL1 and BEL2 methods provide larger point estimates of P_2 using FDG and other methods have similar points estimates. It is consistent with the longer length of the confidence interval comparing with other methods. ADAS11 achieves moderate (0.61) sensitivity when the P_1 and P_3 are 0.60, suggesting it has higher diagnostic accuracy in detecting Alzheimer's Disease compared with FDG. However, BEL1 and BEL2 methods provide similar or smaller point estimates of P_2 using ADAS11 as other methods, although they have a longer length of the confidence interval comparing with other methods.

Table (3.1) Coverage probabilities and average lengths of 95% level confidence intervals for P_2 under Normal distributions with $P_2 = 0.8$

Size		(10,10,10)	(30,30,30)	(50,30,30)	(50,50,50)	(100,100,100)	(100,50,50)	(100,100,50)
Setting (i): $(\mu_1,\sigma_1)=(0,1)$,$(\mu_2,\sigma_2)=(3,1.2)$,$(\mu_3,\sigma_3)=(5.858,2), P_1=P_3=0.8, P_2=0.8$								
NA	CP	0.798	0.896	0.872	0.909	0.939	0.914	0.912
	AL	0.796	0.493	0.491	0.388	0.283	0.393	0.359
ELB	CP	0.99	0.853	0.827	0.87	0.944	0.867	0.911
	AL	0.744	0.443	0.437	0.37	0.285	0.373	0.348
ELP	CP	0.991	0.845	0.827	0.862	0.944	0.867	0.905
	AL	0.749	0.439	0.428	0.354	0.276	0.362	0.344
BEL1	CP	0.982	0.970	0.957	0.964	0.950	0.959	0.960
	AL	0.708	0.472	0.462	0.375	0.277	0.375	0.345
BEL2	CP	0.988	0.985	0.960	0.963	0.953	0.971	0.962
	AL	0.774	0.477	0.467	0.375	0.277	0.375	0.346
BpEL1	CP	0.983	0.960	0.960	0.939	0.948	0.937	0.932
	AL	0.691	0.453	0.447	0.361	0.273	0.368	0.343
BpEL2	CP	0.966	0.949	0.947	0.935	0.944	0.932	0.926
	AL	0.627	0.432	0.426	0.350	0.269	0.356	0.334
IF	CP	0.872	0.935	0.914	0.933	0.947	0.942	0.936
	AL	0.635	0.476	0.479	0.392	0.290	0.397	0.370
BIF1	CP	0.927	0.939	0.923	0.953	0.948	0.952	0.956
	AL	0.663	0.484	0.484	0.393	0.290	0.370	0.398
BIF2	CP	0.919	0.939	0.925	0.943	0.946	0.945	0.944
	AL	0.669	0.482	0.483	0.392	0.291	0.370	0.397
BpIF1	CP	0.843	0.926	0.918	0.925	0.926	0.919	0.919
	AL	0.872	0.526	0.512	0.402	0.287	0.402	0.366
BpIF2	CP	0.838	0.928	0.918	0.927	0.922	0.917	0.919
	AL	0.872	0.526	0.512	0.402	0.287	0.402	0.366
Setting (ii): $(\mu_1,\sigma_1)=(0,1)$,$(\mu_2,\sigma_2)=(4,1.2)$,$(\mu_3,\sigma_3)=(7.625,2), P_1=P_3=0.9, P_2=0.8$								
Size		(10,10,10)	(30,30,30)	(50,30,30)	(50,50,50)	(100,100,100)	(100,50,50)	(100,100,50)
NA	CP	0.668	0.828	0.806	0.852	0.881	0.857	0.857
	AL	0.650	0.512	0.498	0.416	0.314	0.429	0.400
ELB	CP	0.997	0.776	0.781	0.799	0.906	0.812	0.865
	AL	0.732	0.461	0.481	0.401	0.328	0.399	0.394
ELP	CP	0.994	0.783	0.759	0.796	0.906	0.791	0.842
	AL	0.770	0.441	0.442	0.361	0.301	0.368	0.359
BEL1	CP	0.986	0.974	0.965	0.960	0.939	0.952	0.948
	AL	0.735	0.485	0.486	0.391	0.298	0.395	0.368
BEL2	CP	0.997	0.975	0.961	0.948	0.930	0.953	0.934
	AL	0.809	0.491	0.494	0.390	0.298	0.396	0.369
BpEL1	CP	0.989	0.930	0.920	0.914	0.925	0.892	0.878
	AL	0.677	0.454	0.457	0.377	0.299	0.379	0.363
BpEL2	CP	0.967	0.913	0.903	0.904	0.914	0.884	0.870
	AL	0.616	0.430	0.432	0.363	0.293	0.365	0.350
IF	CP	0.643	0.883	0.858	0.885	0.902	0.889	0.894
	AL	0.417	0.469	0.465	0.413	0.318	0.428	0.406
BIF1	CP	0.839	0.852	0.836	0.908	0.920	0.899	0.901
	AL	0.490	0.427	0.421	0.398	0.319	0.369	0.409
BIF2	CP	0.677	0.850	0.825	0.891	0.912	0.891	0.893
	AL	0.436	0.427	0.422	0.398	0.318	0.369	0.409
BpIF1	CP	0.730	0.877	0.856	0.889	0.923	0.892	0.867
	AL	0.757	0.537	0.519	0.419	0.314	0.429	0.396
BpIF2	CP	0.723	0.874	0.854	0.887	0.923	0.891	0.866
	AL	0.757	0.537	0.519	0.419	0.314	0.429	0.396

Table (3.2) Coverage probabilities and average lengths of 95% level confidence intervals for P_2 under Normal distributions with $P_2 = 0.9$

Size		(10,10,10)	(30,30,30)	(50,30,30)	(50,50,50)	(100,100,100)	(100,50,50)	(100,100,50)
Setting (iii): $(\mu_1, \sigma_1) = (0, 1)$,$(\mu_2, \sigma_2) = (3, 1.2)$,$(\mu_3, \sigma_3) = (6.515, 2), P_1 = P_3 = 0.8, P_2 = 0.9$								
NA	CP	0.762	0.914	0.912	0.918	0.942	0.926	0.924
	AL	0.562	0.334	0.334	0.257	0.182	0.258	0.223
ELB	CP	0.980	0.827	0.805	0.816	0.908	0.812	0.882
	AL	0.743	0.351	0.346	0.233	0.175	0.230	0.203
ELP	CP	0.982	0.795	0.781	0.809	0.905	0.806	0.876
	AL	0.753	0.345	0.340	0.226	0.172	0.226	0.203
BEL1	CP	0.931	0.950	0.931	0.952	0.950	0.946	0.956
	AL	0.726	0.434	0.429	0.303	0.197	0.294	0.244
BEL2	CP	0.979	0.985	0.977	0.982	0.970	0.980	0.977
	AL	0.792	0.424	0.420	0.285	0.187	0.277	0.231
BpEL1	CP	0.975	0.985	0.985	0.966	0.963	0.970	0.954
	AL	0.614	0.329	0.322	0.243	0.177	0.245	0.217
BpEL2	CP	0.958	0.977	0.981	0.963	0.961	0.964	0.953
	AL	0.559	0.314	0.308	0.237	0.175	0.239	0.212
IF	CP	0.866	0.906	0.871	0.931	0.961	0.931	0.956
	AL	0.465	0.304	0.310	0.249	0.186	0.251	0.226
BIF1	CP	0.915	0.918	0.868	0.915	0.966	0.942	0.919
	AL	0.529	0.320	0.323	0.253	0.186	0.227	0.254
BIF2	CP	0.929	0.908	0.867	0.914	0.963	0.934	0.917
	AL	0.527	0.311	0.315	0.246	0.184	0.223	0.248
BpIF1	CP	0.829	0.932	0.936	0.944	0.955	0.942	0.945
	AL	0.646	0.363	0.345	0.268	0.189	0.269	0.234
BpIF2	CP	0.824	0.932	0.935	0.943	0.954	0.944	0.945
	AL	0.633	0.361	0.344	0.267	0.189	0.268	0.234
Setting (iv): $(\mu_1, \sigma_1) = (0, 1)$,$(\mu_2, \sigma_2) = (4, 1.2)$,$(\mu_3, \sigma_3) = (8.189, 2), P_1 = P_3 = 0.9, P_2 = 0.9$								
NA	CP	0.619	0.849	0.845	0.864	0.899	0.882	0.878
	AL	0.473	0.355	0.348	0.287	0.212	0.299	0.271
ELB	CP	0.973	0.769	0.774	0.734	0.845	0.763	0.804
	AL	0.738	0.439	0.436	0.268	0.196	0.279	0.238
ELP	CP	0.988	0.778	0.781	0.757	0.867	0.753	0.789
	AL	0.815	0.405	0.421	0.267	0.193	0.268	0.225
BEL1	CP	0.944	0.948	0.937	0.952	0.944	0.935	0.957
	AL	0.762	0.496	0.489	0.344	0.227	0.352	0.291
BEL2	CP	0.987	0.983	0.982	0.984	0.953	0.973	0.963
	AL	0.841	0.498	0.489	0.326	0.214	0.337	0.274
BpEL1	CP	0.977	0.989	0.986	0.963	0.945	0.958	0.911
	AL	0.635	0.353	0.362	0.269	0.204	0.274	0.248
BpEL2	CP	0.960	0.986	0.979	0.955	0.940	0.953	0.905
	AL	0.578	0.334	0.342	0.260	0.200	0.265	0.241
IF	CP	0.591	0.804	0.793	0.888	0.929	0.887	0.916
	AL	0.297	0.310	0.308	0.272	0.211	0.284	0.265
BIF1	CP	0.880	0.692	0.691	0.802	0.879	0.870	0.794
	AL	0.395	0.279	0.276	0.260	0.208	0.240	0.269
BIF2	CP	0.670	0.708	0.713	0.808	0.875	0.866	0.797
	AL	0.331	0.281	0.282	0.257	0.205	0.238	0.265
BpIF1	CP	0.714	0.887	0.864	0.901	0.938	0.906	0.891
	AL	0.569	0.388	0.370	0.293	0.217	0.305	0.273
BpIF2	CP	0.713	0.886	0.864	0.905	0.936	0.905	0.891
	AL	0.559	0.385	0.368	0.292	0.217	0.304	0.272

Table (3.3) Coverage probabilities and average lengths of 95% level confidence intervals for P_2 under Beta distributions with $P_2 = 0.8$

Size		(10,10,10)	(30,30,30)	(50,30,30)	(50,50,50)	(100,100,100)	(100,50,50)	(100,100,50)
Setting (i): $(\alpha_1,\beta_1)=(2,6)$,$(\alpha_2,\beta_2)=(8,6)$,$(\alpha_3,\beta_3)=(21.3,6), P_1=P_3=0.8, P_2=0.8$								
NA	CP	0.865	0.931	0.925	0.930	0.944	0.936	0.940
	AL	0.625	0.378	0.356	0.295	0.210	0.275	0.229
ELB	CP	0.990	0.907	0.904	0.926	0.972	0.946	0.970
	AL	0.656	0.377	0.359	0.303	0.219	0.284	0.238
ELP	CP	0.986	0.894	0.887	0.934	0.962	0.938	0.961
	AL	0.624	0.362	0.334	0.287	0.209	0.268	0.226
BEL1	CP	0.973	0.968	0.959	0.964	0.963	0.963	0.964
	AL	0.618	0.367	0.351	0.289	0.208	0.270	0.225
BEL2	CP	0.987	0.976	0.975	0.961	0.957	0.969	0.957
	AL	0.655	0.365	0.349	0.287	0.207	0.268	0.224
BpEL1	CP	0.988	0.968	0.970	0.953	0.955	0.964	0.946
	AL	0.580	0.355	0.339	0.285	0.207	0.266	0.224
BpEL2	CP	0.980	0.960	0.960	0.946	0.954	0.957	0.944
	AL	0.541	0.345	0.331	0.280	0.205	0.262	0.221
IF	CP	0.885	0.961	0.962	0.948	0.954	0.955	0.948
	AL	0.541	0.370	0.353	0.299	0.213	0.278	0.232
BIF1	CP	0.938	0.956	0.956	0.961	0.962	0.957	0.973
	AL	0.560	0.364	0.349	0.294	0.213	0.232	0.275
BIF2	CP	0.912	0.953	0.953	0.951	0.957	0.949	0.957
	AL	0.563	0.365	0.352	0.297	0.215	0.234	0.278
BpIF1	CP	0.855	0.950	0.942	0.941	0.963	0.954	0.961
	AL	0.678	0.392	0.363	0.301	0.212	0.277	0.230
BpIF2	CP	0.847	0.949	0.943	0.942	0.964	0.953	0.961
	AL	0.678	0.392	0.363	0.301	0.212	0.277	0.230
Setting (ii): $(\alpha_1,\beta_1)=(1,6)$,$(\alpha_2,\beta_2)=(6,6)$,$(\alpha_3,\beta_3)=(15.2,6)=(8.189,2), P_1=P_3=0.9, P_2=0.8$								
NA	CP	0.791	0.900	0.916	0.918	0.940	0.925	0.919
	AL	0.748	0.522	0.491	0.415	0.301	0.382	0.331
ELB	CP	0.995	0.859	0.877	0.882	0.963	0.924	0.962
	AL	0.656	0.428	0.397	0.356	0.267	0.314	0.278
ELP	CP	0.992	0.808	0.852	0.844	0.924	0.902	0.928
	AL	0.692	0.394	0.368	0.338	0.261	0.297	0.262
BEL1	CP	0.990	0.981	0.979	0.965	0.953	0.962	0.959
	AL	0.676	0.427	0.388	0.345	0.259	0.298	0.260
BEL2	CP	0.996	0.965	0.965	0.938	0.934	0.960	0.940
	AL	0.730	0.427	0.386	0.344	0.258	0.296	0.259
BpEL1	CP	0.994	0.934	0.931	0.910	0.921	0.937	0.916
	AL	0.598	0.397	0.369	0.336	0.257	0.293	0.258
BpEL2	CP	0.984	0.918	0.917	0.899	0.915	0.926	0.913
	AL	0.551	0.380	0.356	0.327	0.253	0.287	0.254
IF	CP	0.706	0.894	0.914	0.882	0.911	0.931	0.908
	AL	0.421	0.380	0.374	0.349	0.273	0.309	0.274
BIF1	CP	0.772	0.894	0.915	0.918	0.917	0.913	0.964
	AL	0.456	0.369	0.361	0.330	0.264	0.265	0.300
BIF2	CP	0.753	0.879	0.897	0.884	0.904	0.906	0.933
	AL	0.449	0.366	0.362	0.330	0.267	0.268	0.303
BpIF1	CP	0.717	0.851	0.876	0.875	0.907	0.918	0.907
	AL	0.673	0.456	0.408	0.367	0.270	0.311	0.271
BpIF2	CP	0.714	0.848	0.878	0.868	0.901	0.916	0.904
	AL	0.673	0.456	0.408	0.367	0.270	0.311	0.271

Table (3.4) Coverage probabilities and average lengths of 95% level confidence intervals for P_2 under Normal distributions with $P_2 = 0.9$

Size		(10,10,10)	(30,30,30)	(50,30,30)	(50,50,50)	(100,100,100)	(100,50,50)	(100,100,50)
	Setting (iii): $(\alpha_1,\beta_1)=(2,6)$,$(\alpha_2,\beta_2)=(8,6)$,$(\alpha_3,\beta_3)=(31.8,6), P_1=P_3=0.8, P_2=0.9$							
NA	CP	0.728	0.920	0.923	0.914	0.940	0.940	0.937
	AL	0.452	0.283	0.262	0.224	0.160	0.199	0.164
ELB	CP	0.985	0.817	0.824	0.815	0.922	0.843	0.910
	AL	0.695	0.310	0.289	0.211	0.162	0.190	0.164
ELP	CP	0.981	0.822	0.805	0.815	0.919	0.836	0.917
	AL	0.702	0.322	0.283	0.206	0.159	0.192	0.161
BEL1	CP	0.929	0.955	0.947	0.951	0.955	0.954	0.953
	AL	0.687	0.376	0.342	0.259	0.171	0.224	0.175
BEL2	CP	0.988	0.984	0.974	0.982	0.969	0.980	0.975
	AL	0.740	0.363	0.329	0.243	0.164	0.211	0.167
BpEL1	CP	0.988	0.987	0.983	0.969	0.959	0.961	0.964
	AL	0.529	0.277	0.258	0.217	0.158	0.193	0.161
BpEL2	CP	0.983	0.987	0.981	0.965	0.956	0.956	0.964
	AL	0.490	0.268	0.252	0.213	0.156	0.190	0.160
IF	CP	0.813	0.850	0.860	0.918	0.956	0.936	0.952
	AL	0.357	0.247	0.240	0.214	0.163	0.196	0.167
BIF1	CP	0.921	0.850	0.848	0.909	0.962	0.967	0.923
	AL	0.407	0.247	0.238	0.210	0.161	0.164	0.193
BIF2	CP	0.893	0.850	0.849	0.902	0.954	0.950	0.921
	AL	0.393	0.241	0.234	0.206	0.161	0.164	0.191
BpIF1	CP	0.777	0.921	0.914	0.919	0.963	0.947	0.963
	AL	0.526	0.303	0.269	0.232	0.165	0.205	0.168
BpIF2	CP	0.772	0.920	0.916	0.918	0.963	0.948	0.961
	AL	0.515	0.302	0.268	0.232	0.164	0.204	0.168
	Setting (iv): $(\alpha_1,\beta_1)=(1,6)$,$(\alpha_2,\beta_2)=(6,6)$,$(\alpha_3,\beta_3)=(20.4,6), P_1=P_3=0.9, P_2=0.9$							
NA	CP	0.787	0.889	0.906	0.903	0.931	0.911	0.929
	AL	0.612	0.440	0.406	0.349	0.253	0.311	0.268
ELB	CP	0.966	0.786	0.789	0.745	0.877	0.816	0.874
	AL	0.738	0.421	0.361	0.277	0.197	0.234	0.197
ELP	CP	0.989	0.800	0.782	0.803	0.924	0.827	0.913
	AL	0.708	0.331	0.296	0.210	0.164	0.191	0.167
BEL1	CP	0.956	0.960	0.945	0.958	0.961	0.959	0.966
	AL	0.704	0.391	0.354	0.271	0.177	0.235	0.183
BEL2	CP	0.986	0.986	0.974	0.980	0.973	0.979	0.968
	AL	0.763	0.382	0.342	0.257	0.169	0.221	0.174
BpEL1	CP	0.987	0.990	0.981	0.950	0.962	0.957	0.954
	AL	0.521	0.285	0.265	0.224	0.163	0.200	0.168
BpEL2	CP	0.972	0.989	0.977	0.944	0.958	0.954	0.955
	AL	0.483	0.274	0.257	0.219	0.161	0.196	0.166
IF	CP	0.636	0.816	0.820	0.909	0.954	0.929	0.944
	AL	0.295	0.243	0.239	0.215	0.169	0.201	0.173
BIF1	CP	0.896	0.808	0.807	0.894	0.969	0.929	0.917
	AL	0.381	0.242	0.237	0.210	0.165	0.166	0.197
BIF2	CP	0.721	0.801	0.811	0.885	0.956	0.910	0.915
	AL	0.342	0.237	0.234	0.207	0.165	0.166	0.194
BpIF1	CP	0.686	0.901	0.901	0.918	0.951	0.929	0.941
	AL	0.492	0.312	0.279	0.242	0.172	0.213	0.177
BpIF2	CP	0.679	0.901	0.902	0.916	0.950	0.929	0.939
	AL	0.483	0.310	0.278	0.242	0.171	0.213	0.177

Table (3.5) Coverage probabilities and average lengths of 95% level confidence intervals for P_2 under combined distributions with $P_2 = 0.8$

Size		(10,10,10)	(30,30,30)	(50,30,30)	(50,50,50)	(100,100,100)	(100,50,50)	(100,100,50)
Setting (i): $Gamma(\alpha,\beta)=(4,10)$,$LN(\mu,\sigma)=(1,0.5)$,$Weibull(a,b)=(4.07,6)$, $P_1=P_3=0.8, P_2=0.8$								
NA	CP	0.852	0.920	0.921	0.926	0.938	0.928	0.914
	AL	0.757	0.359	0.362	0.284	0.200	0.235	0.280
ELB	CP	0.983	0.902	0.903	0.930	0.957	0.924	0.946
	AL	0.594	0.348	0.350	0.285	0.203	0.280	0.237
ELP	CP	0.982	0.911	0.910	0.938	0.960	0.941	0.956
	AL	0.607	0.343	0.346	0.279	0.202	0.280	0.240
BEL1	CP	0.957	0.968	0.970	0.951	0.961	0.964	0.964
	AL	0.607	0.358	0.360	0.284	0.202	0.280	0.237
BEL2	CP	0.979	0.977	0.970	0.959	0.957	0.952	0.959
	AL	0.642	0.356	0.358	0.283	0.201	0.279	0.237
BpEL1	CP	0.986	0.952	0.947	0.948	0.957	0.941	0.954
	AL	0.563	0.348	0.349	0.281	0.201	0.277	0.236
BpEL2	CP	0.970	0.940	0.938	0.942	0.953	0.939	0.957
	AL	0.526	0.338	0.339	0.276	0.199	0.272	0.234
IF	CP	0.876	0.944	0.940	0.948	0.956	0.939	0.953
	AL	0.551	0.366	0.368	0.293	0.206	0.290	0.247
BIF1	CP	0.914	0.936	0.929	0.958	0.959	0.959	0.950
	AL	0.578	0.368	0.370	0.294	0.206	0.249	0.289
BIF2	CP	0.895	0.934	0.934	0.948	0.958	0.955	0.941
	AL	0.582	0.369	0.371	0.295	0.207	0.250	0.291
BpIF1	CP	0.869	0.936	0.938	0.945	0.950	0.938	0.948
	AL	0.629	0.372	0.374	0.293	0.205	0.288	0.244
BpIF2	CP	0.871	0.934	0.932	0.944	0.948	0.937	0.952
	AL	0.629	0.372	0.374	0.293	0.205	0.288	0.244
Setting (ii): $Gamma(\alpha,\beta)=(4,10)$,$LN(\mu,\sigma)=(0.5,0.5)$,$Weibull(a,b)=(2.8,6)$, $P_1=P_3=0.9, P_2=0.8$								
NA	CP	0.754	0.886	0.870	0.907	0.917	0.891	0.887
	AL	0.701	0.483	0.487	0.396	0.278	0.349	0.383
ELB	CP	0.996	0.850	0.840	0.886	0.954	0.871	0.926
	AL	0.678	0.445	0.442	0.378	0.286	0.368	0.348
ELP	CP	0.997	0.852	0.849	0.905	0.952	0.890	0.916
	AL	0.740	0.433	0.430	0.366	0.281	0.369	0.353
BEL1	CP	0.991	0.991	0.982	0.971	0.970	0.968	0.970
	AL	0.703	0.463	0.462	0.379	0.278	0.370	0.337
BEL2	CP	0.995	0.981	0.973	0.967	0.962	0.953	0.947
	AL	0.768	0.471	0.471	0.383	0.278	0.372	0.340
BpEL1	CP	0.993	0.955	0.946	0.945	0.949	0.931	0.932
	AL	0.659	0.446	0.447	0.379	0.277	0.367	0.340
BpEL2	CP	0.989	0.949	0.934	0.939	0.946	0.926	0.929
	AL	0.602	0.424	0.424	0.366	0.272	0.355	0.330
IF	CP	0.719	0.928	0.910	0.934	0.944	0.928	0.930
	AL	0.454	0.474	0.474	0.418	0.295	0.404	0.377
BIF1	CP	0.795	0.922	0.897	0.949	0.955	0.939	0.938
	AL	0.482	0.439	0.438	0.407	0.298	0.347	0.391
BIF2	CP	0.721	0.916	0.896	0.940	0.949	0.930	0.925
	AL	0.450	0.439	0.438	0.407	0.298	0.348	0.391
BpIF1	CP	0.788	0.908	0.901	0.928	0.937	0.921	0.924
	AL	0.808	0.520	0.519	0.420	0.291	0.406	0.368
BpIF2	CP	0.787	0.910	0.897	0.928	0.938	0.920	0.924
	AL	0.808	0.520	0.519	0.420	0.291	0.406	0.368

Table (3.6) Coverage probabilities and average lengths of 95% level confidence intervals for P_2 under combined distributions with $P_2 = 0.9$

Size		(10,10,10)	(30,30,30)	(50,30,30)	(50,50,50)	(100,100,100)	(100,50,50)	(100,100,50)
Setting (iii): $Gamma(\alpha,\beta)=(4,10)$,$LN(\mu,\sigma)=(1,0.5)$,$Weibull(a,b)=(4.07,7.49)$, $P_1=P_3=0.8, P_2=0.9$								
NA	CP	0.689	0.911	0.910	0.923	0.940	0.928	0.900
	AL	0.398	0.253	0.256	0.204	0.142	0.164	0.198
ELB	CP	0.977	0.820	0.793	0.833	0.946	0.818	0.917
	AL	0.638	0.275	0.276	0.195	0.142	0.187	0.159
ELP	CP	0.990	0.810	0.829	0.853	0.922	0.849	0.917
	AL	0.678	0.277	0.272	0.195	0.143	0.191	0.164
BEL1	CP	0.923	0.974	0.952	0.946	0.967	0.956	0.961
	AL	0.681	0.342	0.343	0.234	0.151	0.229	0.178
BEL2	CP	0.978	0.990	0.982	0.968	0.974	0.982	0.972
	AL	0.732	0.330	0.331	0.222	0.145	0.216	0.170
BpEL1	CP	0.976	0.995	0.986	0.950	0.965	0.952	0.957
	AL	0.507	0.256	0.259	0.202	0.142	0.196	0.164
BpEL2	CP	0.970	0.994	0.984	0.949	0.960	0.948	0.956
	AL	0.473	0.249	0.252	0.199	0.141	0.193	0.163
IF	CP	0.764	0.838	0.821	0.914	0.965	0.903	0.955
	AL	0.365	0.240	0.243	0.205	0.146	0.197	0.171
BIF1	CP	0.836	0.831	0.805	0.892	0.964	0.925	0.884
	AL	0.432	0.243	0.245	0.204	0.146	0.172	0.197
BIF2	CP	0.805	0.827	0.804	0.897	0.962	0.924	0.884
	AL	0.409	0.237	0.240	0.201	0.146	0.171	0.194
BpIF1	CP	0.745	0.932	0.927	0.936	0.960	0.933	0.948
	AL	0.444	0.265	0.268	0.211	0.145	0.204	0.171
BpIF2	CP	0.736	0.933	0.928	0.939	0.959	0.932	0.948
	AL	0.435	0.263	0.267	0.210	0.145	0.204	0.170
Setting (iv): $Gamma(\alpha,\beta)=(4,10)$,$LN(\mu,\sigma)=(1,0.5)$,$Weibull(a,b)=(4.25,6)$, $P_1=P_3=0.9, P_2=0.9$								
NA	CP	0.629	0.906	0.900	0.904	0.926	0.910	0.895
	AL	0.369	0.252	0.254	0.209	0.145	0.163	0.199
ELB	CP	0.987	0.798	0.812	0.822	0.924	0.822	0.905
	AL	0.634	0.298	0.304	0.204	0.151	0.195	0.169
ELP	CP	0.989	0.816	0.805	0.807	0.929	0.819	0.918
	AL	0.700	0.286	0.288	0.192	0.145	0.189	0.166
BEL1	CP	0.937	0.959	0.958	0.958	0.958	0.960	0.973
	AL	0.691	0.355	0.351	0.242	0.153	0.230	0.175
BEL2	CP	0.990	0.983	0.984	0.979	0.961	0.977	0.965
	AL	0.749	0.345	0.342	0.228	0.147	0.217	0.167
BpEL1	CP	0.985	0.990	0.991	0.957	0.948	0.949	0.947
	AL	0.503	0.254	0.257	0.203	0.143	0.195	0.162
BpEL2	CP	0.981	0.988	0.988	0.952	0.942	0.949	0.947
	AL	0.469	0.247	0.250	0.200	0.142	0.192	0.160
IF	CP	0.583	0.805	0.801	0.904	0.944	0.884	0.946
	AL	0.282	0.235	0.238	0.206	0.148	0.197	0.169
BIF1	CP	0.765	0.745	0.740	0.852	0.932	0.842	0.834
	AL	0.355	0.227	0.229	0.201	0.149	0.156	0.191
BIF2	CP	0.605	0.751	0.751	0.855	0.930	0.849	0.839
	AL	0.296	0.223	0.226	0.199	0.148	0.156	0.189
BpIF1	CP	0.670	0.910	0.914	0.925	0.940	0.929	0.937
	AL	0.418	0.263	0.264	0.213	0.147	0.203	0.168
BpIF2	CP	0.668	0.910	0.914	0.924	0.939	0.929	0.939
	AL	0.411	0.261	0.263	0.213	0.147	0.203	0.168

Table (3.7) 95% level confidence intervals and point estimates for the sensitivity P_2 for MCI.

Biomarker	FDG		ADAS11	
	Point estimate	CI	Point estimate	CI
		$P_1 = P_3 = 0.6$		
BEL1	0.395	(0.163 , 0.655)	0.583	(0.325 , 0.819)
BEL2	0.377	(0.135 , 0.658)	0.598	(0.319 , 0.845)
BpEL1	0.365	(0.283 , 0.451)	0.608	(0.523 , 0.689)
BpEL2	0.365	(0.282 , 0.451)	0.606	(0.525 , 0.687)
IF	0.365	(0.278 , 0.450)	0.614	(0.520 , 0.687)
BIF1	0.364	(0.278 , 0.451)	0.603	(0.520 , 0.687)
BIF2	0.364	(0.277 , 0.451)	0.604	(0.520 , 0.688)
BpIF1	0.364	(0.277 , 0.452)	0.604	(0.524 , 0.686)
BpIF2	0.365	(0.277 , 0.448)	0.614	(0.522 , 0.688)
		$P_1 = P_3 = 0.7$		
BEL1	0.262	(0.075 , 0.515)	0.459	(0.216 , 0.712)
BEL2	0.222	(0.044 , 0.488)	0.452	(0.194 , 0.725)
BpEL1	0.196	(0.118 , 0.287)	0.448	(0.347 , 0.547)
BpEL2	0.196	(0.121 , 0.290)	0.447	(0.348 , 0.546)
IF	0.195	(0.112 , 0.282)	0.463	(0.352 , 0.558)
BIF1	0.198	(0.113 , 0.284)	0.455	(0.351 , 0.558)
BIF2	0.196	(0.111 , 0.282)	0.455	(0.351 , 0.558)
BpIF1	0.195	(0.111 , 0.284)	0.454	(0.352 , 0.554)
BpIF2	0.195	(0.109 , 0.282)	0.463	(0.348 , 0.557)

PART 4

DIRECT ESTIMATION OF THE AREA UNDER THE RECEIVER OPERATING CHARACTERISTIC CURVE WITH VERIFICATION BIASED DATA

4.1 Introduction

For a diagnostic test that yields a continuous test result, the receiver operating characteristic (ROC) curve is a popular tool for evaluating the ability of the test to discriminate between diseased and non-diseased subjects. The continuous test results can be separated at a specified cutoff point, and the sensitivity and specificity can be computed. If we use T denote the continuous-scale test result of a subject in both diseased and non-diseased groups, and let D be a disease indicator with 1 as the diseased subject and 0 as the non-diseased subject, then, for a given cut-off point c, the sensitivity and the specificity of the test can be defined as follows:

$$\text{Sensitivity(c)} = P(T > c|D = 1), \quad \text{Specificity(c)} = P(T \leq c|D = 0),$$

respectively.

When we vary the cut-off point throughout the entire real line, the resulting pairs (1-specificity, sensitivity) form the ROC curve. The area under the ROC curve (AUC) is commonly used as a summary index of the accuracy of the diagnostic test. The AUC can be interpreted as $P(T_1 > T_2)$, where T_1 is the test result from a randomly selected diseased subject and T_2 is the test result from a randomly selected non-diseased subject. The AUC of a test of interest is bounded between 0.5 and 1. A larger AUC value represents a better separation in test/biomarker values between the diseased and non-diseased populations. In particular, a perfect test would achieve an AUC of 1.0, whereas an uninformative test would

have an AUC of 0.5.

In medical diagnostics, not all subjects with given screening test results ultimately have their true disease status verified through a very accurate gold standard test. That is to say, the labels referred to as true disease status of the subjects are partially missing. One reason for the missing is that the gold standard test is usually costly and invasive. So the common practice is to apply it only on high-risk subjects based on screening test results. Because patients at low-risk are more likely to have their true disease status missing, simply ignoring this missingness and using only subjects with verified disease status may lead to biased results. Such bias is called verification bias.[21] The missing at random (MAR) assumption[22] will be adopted to deal with missing labels. Under the MAR assumption, the probability of a subject being verified does not depend on the true disease status. The application of existing completed data approaches to MAR problems may result in biased inference and loss of efficiency. Such bias may mislead in real data analyses. To correct verification bias, Alonzo and Pepe[23] proposed several methods for estimating the sensitivity and the specificity of a test by using parametric models (e.g. Probit model or Logistic regression model) for the probability that a subject is diseased/verified. He et al.[24] developed an inverse probability weighting (IPW) based method for directly estimating the AUC in the setting of verification bias. Adimari and Chiogna[25] proposed a fully non-parametric method for the AUC estimator based on K nearest-neighbor imputation. To apply this non-parametric method for accurate inference of the AUC, sufficient information from the data needs to be provided, and a suitable value for K needs to be selected in practice.

This part aims to develop new methods directly estimating the AUC in the presence of verification bias when the test result is continuous under the assumption that the true disease status, if missing, is missing at random. We derive new closed-form expressions for the AUC estimators. The proposed AUC estimators can be easily computed and directly applied in practice. Our simulation results show that the newly proposed AUC estimators have outstanding finite sample performance. In section 4.2, we give a brief review of existing methods to estimate the AUC with verification biased data. The new estimators and their

properties are introduced in section 4.3. A numeric simulation study is presented in section 4.4. In section 4.5, the competing methods are illustrated using a data set from a study of neonatal hearing screening.

4.2 Existing Methods with Verification Bias

As mentioned in section 1, Alonzo and Pepe[23] proposed several methods, including the inverse probability weighting (IPW) method, the full imputations (FI) method, the mean score imputation (MSI) method and the semi-parametric efficient (SPE) method, to estimate sensitivity and specificity. We briefly introduce these methods in this section.

Let T_i denote the continuous test result and let D_i denote the true disease status for the i-th subject, $i = 1, 2, \ldots, n$, where $D_i = 1$ indicates that the subject has a disease and $D_i = 0$ indicates that subject does not have the disease. Only a subset of the subjects have their disease status verified. Let $V_i = 1$ if the i-th subject has the true disease status verified, and $V_i = 0$ otherwise. Let A_i be a vector of observed covariates for the i-th subject that may be associated with both D_i and V_i. The cumulative distribution functions of $T|D = 1$ and $T|D = 0$ are F_1 and F_0, respectively. Without loss of generality, suppose that larger values of T are more indicative of disease. All of the methods reviewed in this section are based on the assumption that disease status verification is conditionally independent of the true disease status given the test result. The decision to verify the subject's true disease status depends on the true disease status only through A and T. If all subjects have their disease status verified, i.e., $V_i = 1$, $i = 1, 2, \ldots, n$, we have a complete data set. For any cutoff point c, the sensitivity, $Se(c)$, and specificity, $Sp(c)$, of the test can be easily estimated by

$$\hat{Se}_{Full}(c) = \frac{\sum_{i=1}^{n} I(T_i \geq c) D_i}{\sum_{i=1}^{n} D_i},$$

$$\hat{Sp}_{Full}(c) = \frac{\sum_{i=1}^{n} I(T_i < c)(1 - D_i)}{\sum_{i=1}^{n} (1 - D_i)},$$

respectively. These estimators are unbiased for $Se(c)$ and $Sp(c)$, respectively.

4.2.1 Estimation Methods for Sensitivity and Specificity with Verification Bias

Full Imputation One approach to estimating the prevalence of disease in a two-phase design is to use full imputation (FI) over the distribution $P(D|T, A)$, [23][52] i.e., FI imputes the probability of disease for all subjects in the study as a function of (T, A). The FI estimator of the disease prevalence is

$$\hat{P}(D = 1) = \frac{1}{n} \sum_{i=1}^{n} \hat{\rho}_i,$$

where $\hat{\rho}_i$ is an estimate of $\rho_i = P(D_i = 1|T_i, A_i)$ that is obtained by using, for example, a logistic regression model. By the MAR assumption, the disease model $P(D = 1|T, A)$ can be estimated by using the verified sample.

The ratio of FI estimators of $P(T \geq c, D = 1)$ and $P(D = 1)$ yields the following FI estimators for $Se(c)$ and $Sp(c)$:

$$\hat{Se}_{FI}(c) = \frac{\sum_{i=1}^{n} I(T_i \geq c)\hat{\rho}_i}{\sum_{i=1}^{n} \hat{\rho}_i}, \tag{4.1}$$

$$\hat{Sp}_{FI}(c) = \frac{\sum_{i=1}^{n} I(T_i < c)(1 - \hat{\rho}_i)}{\sum_{i=1}^{n} (1 - \hat{\rho}_i)}. \tag{4.2}$$

Mean Score Imputation Mean score imputation (MSI) is another approach used for estimating the prevalence of disease in two-phase studies.[53][54] In contrast with FI, MSI only imputes disease status for subjects who are not in the verification sample and uses the observed disease status for those who are in the verification sample. The ratio of MSI estimators of $P(T \geq c, D = 1)$ and $P(D = 1)$ yields the following MSI estimators for $Se(c)$ and $Sp(c)$:

$$\hat{Se}_{MSI}(c) = \frac{\sum_{i=1}^{n} I(T_i \geq c)\{V_i D_i + (1 - V_i)\hat{\rho}_i\}}{\sum_{i=1}^{n} \{V_i D_i + (1 - V_i)\hat{\rho}_i\}}, \tag{4.3}$$

$$\hat{Sp}_{MSI}(c) = \frac{\sum_{i=1}^{n} I(T_i < c)\{V_i(1 - D_i) + (1 - V_i)(1 - \hat{\rho}_i)\}}{\sum_{i=1}^{n} \{V_i(1 - D_i) + (1 - V_i)(1 - \hat{\rho}_i)\}}. \tag{4.4}$$

Again, the MAR assumption implies data from the verification sample can be used to obtain a valid estimate $\hat{\rho}_i$ for ρ_i.

Inverse Probability Weighting An inverse probability weighting (IPW) estimator [55] that weights each observation in the verification sample by the inverse of the sampling fraction (i.e. the probability that the subject was selected for verification) is another approach used to estimate the prevalence of disease in a two-phase design. The ratio of IPW estimators of $P(T \geq c, D = 1)$ and $P(D = 1)$ yields the following IPW estimators for $Se(c)$ and $Sp(c)$:

$$\hat{Se}_{IPW}(c) = \frac{\sum_{i=1}^{n} I(T_i \geq c) V_i D_i \hat{\pi}_i^{-1}}{\sum_{i=1}^{n} V_i D_i \hat{\pi}_i^{-1}}, \tag{4.5}$$

$$\hat{Sp}_{IPW}(c) = \frac{\sum_{i=1}^{n} I(T_i < c) V_i (1 - D_i) \hat{\pi}_i^{-1}}{\sum_{i=1}^{n} V_i (1 - D_i) \hat{\pi}_i^{-1}} \tag{4.6}$$

where $\hat{\pi}_i$ is an estimate for $\pi_i = P(V_i = 1 | T_i, A_i)$. IPW corrects for the biased sampling by weighting the observed value by inverting the probability that the subject was verified.

Semi-parametric Efficient Approach Following the semi-parametric efficient (SPE) approach of Alonzo and Pepe,[23] the ratio of SPE estimators of $P(T \geq c, D = 1)$ and $P(D = 1)$ yields the following SPE estimators for $Se(c)$ and $Sp(c)$:

$$\hat{Se}_{SPE}(c) = \frac{\sum_{i=1}^{n} I(T_i \geq c) \{V_i D_i \hat{\pi}_i^{-1} - (V_i - \hat{\pi}_i) \hat{\rho}_i \hat{\pi}_i^{-1}\}}{\sum_{i=1}^{n} \{V_i D_i \hat{\pi}_i^{-1} - (V_i - \hat{\pi}_i) \hat{\rho}_i \hat{\pi}_i^{-1}\}}, \tag{4.7}$$

$$\hat{Sp}_{SPE}(c) = \frac{\sum_{i=1}^{n} I(T_i < c) \{V_i (1 - D_i) \hat{\pi}_i^{-1} - (V_i - \hat{\pi}_i)(1 - \hat{\rho}_i) \hat{\pi}_i^{-1}\}}{\sum_{i=1}^{n} \{V_i (1 - D_i) \hat{\pi}_i^{-1} - (V_i - \hat{\pi}_i)(1 - \hat{\rho}_i) \hat{\pi}_i^{-1}\}}. \tag{4.8}$$

For each of the above approaches, when c varies throughout the real line, an empirical bias corrected ROC curve can be obtained by plotting the pairs $(1 - \hat{Sp}(c), \hat{Se}(c))$, and the associated AUC needs to be calculated numerically. A limitation to these methods is that there is no closed-form expression for the AUC estimators, which motivates the search of direct estimation methods for the AUC with verification biased data.

4.2.2 Direct Estimation of the AUC based on IPW

When the verification probability $\pi_i = P(V_i = 1|T_i, A_i)$ is assumed to be known, He et al.[24] proposed the following estimator of the AUC in the presence of verification bias:

$$\begin{aligned} \hat{AUC} =& \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i^{-1}\pi_j^{-1}V_iV_jI(T_i > T_j)I(D_i > D_j)}{\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i^{-1}\pi_j^{-1}V_iV_jI(D_i > D_j)} \\ =& \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{2}\pi_i^{-1}\pi_j^{-1}V_iV_j[I(T_i > T_j)I(D_i > D_j) + I(T_i < T_j)I(D_i < D_j)]}{\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{2}\pi_i^{-1}\pi_j^{-1}V_iV_j[I(D_i > D_j) + I(D_i < D_j)]}. \end{aligned} \tag{4.9}$$

Note that the reason for writing the estimator in the second form (symmetric form) is to express it as a function of U-statistics. This estimator uses IPW to correct verification bias, where the weight $\pi_i^{-1}\pi_j^{-1}$ is attached to all possible pairs of verified subjects. This estimator is not unbiased, but it is consistent.

According to Equations (4.5) and (4.6), we define two weights:

$$r_{IPW,j} = V_jD_j\hat{\pi}_j^{-1},\ j = 1, \cdots, n, \tag{4.10}$$

$$w_{IPW,i} = V_i(1 - D_i)\hat{\pi}_i^{-1},\ i = 1, \cdots, n. \tag{4.11}$$

When π_i is known, we take $\hat{\pi}_i = \pi_i$. Then we can rewrite the AUC estimate from Equation (4.9) as:

$$\hat{\delta}_{IPW} = \frac{\sum_{i=1}^{n}\hat{g}_{IPW}(T_i)w_{IPW,i}}{\sum_{i=1}^{n}w_{IPW,i}}, \tag{4.12}$$

where

$$\hat{g}_{IPW}(T_i) = \frac{\sum_{j=1}^{n}I(T_j \geq T_i)r_{IPW,j}}{\sum_{j=1}^{n}r_{IPW,j}}. \tag{4.13}$$

4.2.3 Non-parametric Verification Bias-Corrected AUC Estimation based on K-NN Method

Adimari and Chiogna [25] proposed a nonparametric verification bias-corrected AUC estimator using a K-nearest-neighbor (K-NN) imputation method. For a finite positive integer K and a suitable distance measure, a nearest neighbor imputation estimate of $\rho_i = P(D_i = 1|T_i, A_i)$, for a subject with true disease status not verified, is defined as

$$\hat{\rho}_{Ki} = \frac{1}{K}\sum_{j=1}^{K} D_{i(j)},$$

where $\{(Y_{i(j)}, D_{i(j)}) : V_{i(j)} = 1, j = 1, \ldots, K\}$ is a set of K observed data pairs and $Y_{i(j)}$ denotes the j-th nearest neighbor to $Y_i = (T_i, A_i)$ among all Y_h's corresponding to the verified patients, i.e., to those D_h's with $V_h = 1$. Then, the estimate $\hat{\rho}_{Ki}$ could be used as an imputation value for the missing label D_i. This leads to Adimari and Chiogna's non-parametric estimator for the AUC:

$$\hat{AUC} = \frac{\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} I(T_i > T_j)\hat{D}_{Ki}(1-\hat{D}_{Ki})}{\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} \hat{D}_{Ki}(1-\hat{D}_{Ki})}, \tag{4.14}$$

where

$$\hat{D}_{Ki} = V_i D_i + (1 - V_i)\hat{\rho}_{Ki}.$$

This method is based on the K-nearest-neighbor imputation, which requires the choice of a value for K. Simulation results in the study of Adimari and Chiogna[25] suggested that a value for K around 3 could to be adequate. In this paper, we choose $K = 3$ in our simulation study and real data analysis.

4.3 New Direct Estimation Methods for the AUC with Verification Bias

Observe that the AUC can be expressed as

$$\begin{aligned}\delta &= P(T_j \geq T_i | D_j = 1, D_i = 0) \\ &= E_{F_0}\big\{E_{F_1}[I(T_j \geq T_i | T_i, D_j = 1, D_i = 0)]\big\} \\ &= E_{F_0}[g(T_i)] = \int_{-\infty}^{\infty} g(t) dF_0(t), \end{aligned} \tag{4.15}$$

where

$$\begin{aligned} g(T_i) &= E_{F_1}[I(T_j \geq T_i | T_i, D_j = 1, D_i = 0)], \\ g(t) &= E_{F_1}[I(T_j \geq t | D_j = 1)] = P(T_j \geq t | D_j = 1) = Se(t), \\ F_0(t) &= P(T_i < t | D_i = 0)] = Sp(t). \end{aligned}$$

In other words, the AUC can be directly derived based on sensitivity $Se(t)$ and specificity $Sp(t)$. Therefore, we can directly get an estimate for the AUC

$$\hat{\delta} = \int_{-\infty}^{\infty} \hat{g}(t) d\hat{F}_0(t) \tag{4.16}$$

when an estimate $\hat{g}(t)$ for $g(t)$ and an estimate $\hat{F}_0(t)$ for $F_0(t)$ are available.

4.3.1 AUC Estimates based on SPE, FI and MSI

Alonzo and Pepe[23] proposed estimating bias-corrected AUC by applying the trapezoidal rule to bias-corrected estimates of true positive fractions and false positive fractions. But they didn't give the explicit formulas of the AUC estimators. Motivated by He et al.[24] and Alonzo and Pepe[23], we derive three new closed-form AUC estimators with verification bias by applying the estimation methods (SPE, FI, and MSI) for sensitivity and specificity mentioned in section 4.2.

According to Equation (4.8), the SPE estimate for $F_0(t)$ is

$$\hat{F}_0(t) = \hat{Sp}_{SPE}(t) = \frac{\sum_{i=1}^{m} I(T_i < t) w_{SPE,i}}{\sum_{i=1}^{m} w_{SPE,i}}, \tag{4.17}$$

where the weight $w_{SPE,i}$ is defined as

$$w_{SPE,i} = V_i(1 - D_i)\hat{\pi}_i^{-1} - (V_i - \hat{\pi}_i)(1 - \hat{\rho}_i)\hat{\pi}_i^{-1} \tag{4.18}$$

with $\hat{\pi}_i = \hat{P}(V_i = 1|T_i, X_i)$, $\hat{\rho}_i = \hat{P}(D_i = 1|T_i, X_i)$, $i = 1, \cdots, n$.

From Equation (4.7), we get the SPE estimate for $g(t)$:

$$\hat{g}(t) = \hat{g}_{SPE}(t) = \hat{Se}_{SPE}(t) = \frac{\sum_{j=1}^{n} I(T_j \geq t) r_{SPE,j}}{\sum_{j=1}^{n} r_{SPE,j}}, \tag{4.19}$$

where the weight $r_{SPE,j}$ is defined as

$$r_{SPE,j} = V_j D_j \hat{\pi}_i^{-1} - (V_j - \hat{\pi}_j)\hat{\rho}_j \hat{\pi}_j^{-1}, \ j = 1, \cdots, n. \tag{4.20}$$

From Equations (4.16), (4.17) and (4.19), we get an SPE-based estimate for the AUC:

$$\begin{aligned} \hat{\delta}_{SPE} &= \int_{-\infty}^{\infty} \hat{g}_{SPE}(t) d\hat{F}_0(t) \\ &= \frac{\sum_{i=1}^{n} \hat{g}_{SPE}(T_i) w_{SPE,i}}{\sum_{i=1}^{n} w_{SPE,i}}. \end{aligned} \tag{4.21}$$

The direct AUC estimate $\hat{\delta}_{IPW}$ defined in equation (4.12) can be derived by using equations (4.5) and (4.6) and the similar approach above.

Similarly, we can get an MSI-based estimate for the AUC using MSI weights as following:

$$\hat{\delta}_{MSI} = \frac{\sum_{i=1}^{n} \hat{g}_{MSI}(T_i) w_{MSI,i}}{\sum_{i=1}^{n} w_{MSI,i}}, \tag{4.22}$$

where

$$w_{MSI,i} = V_i(1 - D_i) + (1 - V_i)(1 - \hat{\rho}_i),\ i = 1, \cdots, n, \tag{4.23}$$

$$\hat{g}_{MSI}(T_i) = \frac{\sum_{j=1}^{n} I(T_j \geq T_i) r_{MSI,j}}{\sum_{j=1}^{n} r_{MSI,j}} \tag{4.24}$$

with

$$r_{MSI,j} = V_j D_j + (1 - V_j)\hat{\rho}_j,\ j = 1, \cdots, n, \tag{4.25}$$

and a FI-based estimate for the AUC using FI weights as follows

$$\hat{\delta}_{FI} = \frac{\sum_{i=1}^{n} \hat{g}_{FI}(T_i) w_{FI,i}}{\sum_{i=1}^{n} w_{FI,i}}, \tag{4.26}$$

where

$$w_{FI,i} = 1 - \hat{\rho}_i,\ i = 1, \cdots, n, \tag{4.27}$$

$$\hat{g}_{FI}(T_i) = \frac{\sum_{j=1}^{n} I(T_j \geq T_i) r_{FI,j}}{\sum_{j=1}^{n} r_{FI,j}} \tag{4.28}$$

with

$$r_{FI,j} = \hat{\rho}_j,\ j = 1, \cdots, n. \tag{4.29}$$

4.3.2 AUC Estimates based on Combined SPE, MSI, FI and IPW Approaches

Since AUC can be directly derived from (4.15) based on sensitivity $Se(t)$ and specificity $Sp(t)$ that can be estimated using w-weights (i.e., $w_{SPE,i}$'s, $w_{MSI,i}$'s, $w_{FI,i}$'s, $w_{IPW,i}$'s) and r-weights (i.e., $r_{SPE,j}$'s, $r_{MSI,j}$'s, $r_{FI,j}$'s, $r_{IPW,j}$'s), one question is if an optimal/better

combination(s) of w-weights and r-weights can be identified to produce an optimal/better estimate(s) for the AUC, which motivates us propose twelve new estimators for the AUC by using different combinations of w-weights and r-weights in this section.

According to Equations (4.10), (4.13), (4.17) and (4.18), r_{IPW} and w_{SPE} weights can be used to estimate sensitivity $g(t)$ and specificity $F_0(t)$, respectively. Hence, we get the following IPW-SPE estimator of the AUC by combining weights $r_{IPW,j}$'s and $w_{SPE,i}$'s:

$$\hat{\delta}_{IPW-SPE} = \frac{\sum_{i=1}^{n} \hat{g}_{IPW}(T_i) w_{SPE,i}}{\sum_{i=1}^{n} w_{SPE,i}}. \tag{4.30}$$

Similarly, according to Equations (4.13), (4.23) and (4.24), we get the following IPW-MSI estimator of the AUC by combining weights $r_{IPW,j}$'s and $w_{MSI,i}$'s:

$$\hat{\delta}_{IPW-MSI} = \frac{\sum_{i=1}^{n} \hat{g}_{IPW}(T_i) w_{MSI,i}}{\sum_{i=1}^{n} w_{MSI,i}}. \tag{4.31}$$

According to Equations (4.13), (4.27) and (4.28), we get the following IPW-FI estimator of the AUC by combining weights $r_{IPW,j}$'s and $w_{FI,i}$'s:

$$\hat{\delta}_{IPW-FI} = \frac{\sum_{i=1}^{n} \hat{g}_{IPW}(T_i) w_{FI,i}}{\sum_{i=1}^{n} w_{FI,i}}. \tag{4.32}$$

According to Equations (4.19), (4.20) and (4.11), we get the following SPE-IPW estimator of the AUC by combining weights $r_{SPE,j}$'s and $w_{IPW,i}$'s:

$$\hat{\delta}_{SPE-IPW} = \frac{\sum_{i=1}^{n} \hat{g}_{SPE}(T_i) w_{IPW,i}}{\sum_{i=1}^{n} w_{IPW,i}}. \tag{4.33}$$

According to Equations (4.19), (4.20) and (4.23), we get the following SPE-MSI estimator of the AUC by combining weights $r_{SPE,j}$'s and $w_{MSI,i}$'s:

$$\hat{\delta}_{SPE-MSI} = \frac{\sum_{i=1}^{n} \hat{g}_{SPE}(T_i) w_{MSI,i}}{\sum_{i=1}^{n} w_{MSI,i}}. \tag{4.34}$$

According to Equations (4.19), (4.20) and (4.27), we get the following SPE-FI estimator

of the AUC by combining weights $r_{SPE,j}$'s and $w_{FI,i}$'s:

$$\hat{\delta}_{SPE-FI} = \frac{\sum_{i=1}^{n} \hat{g}_{SPE}(T_i) w_{FI,i}}{\sum_{i=1}^{n} w_{FI,i}}. \tag{4.35}$$

According to Equations (4.24), (4.25) and (4.11), we get the following MSI-IPW estimator of the AUC by combining weights $r_{MSI,j}$'s and $w_{IPW,i}$'s:

$$\hat{\delta}_{MSI-IPW} = \frac{\sum_{i=1}^{n} \hat{g}_{MSI}(T_i) w_{IPW,i}}{\sum_{i=1}^{n} w_{IPW,i}}. \tag{4.36}$$

According to Equations (4.24), (4.25) and (4.18), we get the following MSI-SPE estimator of the AUC by combining weights $r_{MSI,j}$'s and $w_{SPE,i}$'s:

$$\hat{\delta}_{MSI-SPE} = \frac{\sum_{i=1}^{n} \hat{g}_{MSI}(T_i) w_{SPE,i}}{\sum_{i=1}^{n} w_{SPE,i}}. \tag{4.37}$$

According to Equations (4.24), (4.25) and (4.27), we get the following MSI-FI estimator of the AUC by combining weights $r_{MSI,j}$'s and $w_{FI,i}$'s:

$$\hat{\delta}_{MSI-FI} = \frac{\sum_{i=1}^{n} \hat{g}_{MSI}(T_i) w_{FI,i}}{\sum_{i=1}^{n} w_{FI,i}}. \tag{4.38}$$

According to Equations (4.28), (4.29) and (4.11), we get the following FI-IPW estimator of the AUC by combining weights $r_{FI,j}$'s and $w_{IPW,i}$'s:

$$\hat{\delta}_{FI-IPW} = \frac{\sum_{i=1}^{n} \hat{g}_{FI}(T_i) w_{IPW,i}}{\sum_{i=1}^{n} w_{IPW,i}}. \tag{4.39}$$

According to Equations (4.28), (4.29) and (4.18), we get the following FI-SPE estimator of the AUC by combining weights $r_{FI,j}$'s and $w_{SPE,i}$'s:

$$\hat{\delta}_{FI-SPE} = \frac{\sum_{i=1}^{n} \hat{g}_{FI}(T_i) w_{SPE,i}}{\sum_{i=1}^{n} w_{SPE,i}}. \tag{4.40}$$

According to Equations (4.28), (4.29) and (4.23), we get the following FI-MSI estimator

of the AUC by combining weights $r_{FI,j}$'s and $w_{MSI,i}$'s:

$$\hat{\delta}_{FI-MSI} = \frac{\sum_{i=1}^{n} \hat{g}_{FI}(T_i) w_{MSI,i}}{\sum_{i=1}^{n} w_{MSI,i}}. \tag{4.41}$$

Obviously, all the new estimators for the AUC proposed above have closed-form expressions and can be easily computed. These estimators have the following property.

Theorem: All the new estimators for the AUC proposed in section 3 are consistent.

Proof: From the results on bias-corrected estimators of sensitivity and specificity in Alonzo et al.[54] (see also Alonzo et al.[23]), it follows that the estimators $\hat{g}(t)$ for $g(t)$ (which are the SPE, FI, MSI, IPW estimators for $Se(t)$) and the estimators $\hat{F}_0(t)$ for $F_0(t)$ (which are SPE, FI, MSI, IPW estimators for $Sp(t)$) are consistent. Then the theorem follows from (4.16) and consistency of $\hat{g}(t)$ and $\hat{F}_0(t)$ right away.

He et al.[24] proposed an asymptotic variance estimate for the IPW-based AUC estimator under the assumption that the probability π_i of verification is known. However, in practice the true value of π_i is unknown, and explicitly estimating the variances of the new estimators for the AUC is still an open question. However, bootstrap method can be used to estimate the variances of the new estimators.

4.4 Simulation Studies

In this section, simulation studies are conducted to evaluate the finite sample performance and robustness of the proposed various bias-corrected estimators of the AUC in terms of mean squared error (MSE) and absolute bias. We also compare our newly proposed methods with the existing IPW-based method and the nonparametric K-NN method. The

simulation set-up is similar to that of Alonzo and Pepe's study.[23] Briefly, D is generated as a dichotomous variable indicating whether a random variable $Z = Z_1 + Z_2 \sim N(0,1)$ is greater than a threshold h, where $Z_1 \sim N(0, 0.5)$, and $Z_2 \sim N(0, 0.5)$. In this study, we select h to make the disease prevalence equal 0.3 and 0.5, respectively. T and A are generated from $T = \nu_1 Z_1 + \tau_1 Z_2 + \epsilon_1$ and $A = \nu_2 Z_1 + \tau_2 Z_2 + \epsilon_2$, where $\epsilon_1 \sim N(0, 0.25)$ and $\epsilon_2 \sim N(0, 0.25)$, and ϵ_1 and ϵ_2 are independent. The inherent accuracy of T or A can be altered by changing the values of ν_1 or ν_2 and τ_1 or τ_2 respectively. We fix ν_2 and τ_2 and then select different values of ν_1 and τ_1 to generate reasonable values of the AUC.

4.4.1 Correct Models

To introduce the verification bias under the MAR assumption, V is generated by using a Bernoulli random variable with $P(V = 1) = 1$ for subjects with $T > t^{(0.8)}$ and $P(V = 1) = 0.2$ for the rest, where $t^{(0.8)}$ is the 80-th quantile of the distribution of T. This verification mechanism results in an average of 36% of the subjects receiving disease verification. Empirical estimates of the verification probabilities yield:

$$\hat{\pi}_i = \begin{cases} 1.0, & T_i > t^{(0.8)} \\ \frac{\sum_{i=1}^{n} V_i I(T_i \le t^{(0.8)})}{\sum_{i=1}^{n} I(T_i \le t^{(0.8)})}, & T_i \le t^{(0.8)} \end{cases} \tag{4.42}$$

To apply FI, MSI, and SPE methods, a parametric model for disease probabilities, ρ_i's, must be specified. It was shown in Alonzo and Pepe [23] that a probit model that was linear in T and A was a true model under above settings. We apply the same probit model for our correct model study. To apply the K-NN method, we choose $K = 3$ and Euclidean distance based on both T and A. 2000 random samples are drawn from underlying distributions with the sample sizes $n = 100$, 200, 500, and 1000, respectively, to evaluate the performance of the proposed estimators and the K-NN estimator for the AUC in terms of MSE and bias. At this moment, we fix $\nu_2 = \tau_2 = 1$, and select h to make the prevalence of disease equal to 0.3 and 0.5, respectively. Different values of ν_1, τ_1 are selected to generate the true AUC,

which equals 0.7, 0.8, 0.9, and 0.96, respectively. We report the numerical results in Table 4.2 (with 0.3 disease prevalence) and Table 4.3 (with 0.5 disease prevalence) and boxplots in Figure 4.1 .

Place Table 4.2–4.3 here

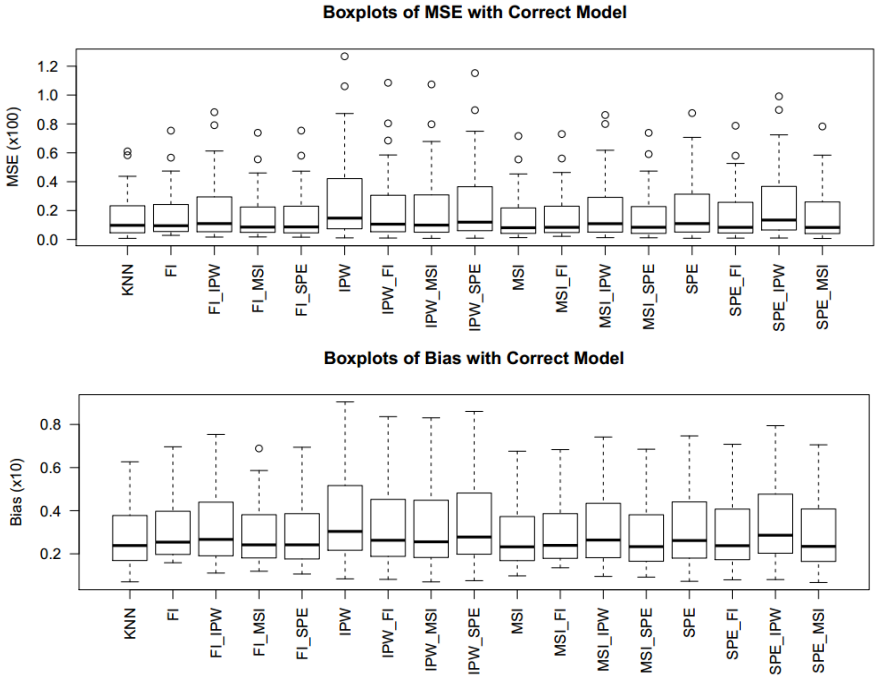
From Tables 4.2–4.3, we can see that all methods always have lower MSE and bias with larger sample sizes and higher AUC than those with smaller sample size and lower AUC. Since the correct disease and verification models are used with the FI, MSI, IPW, and SPE estimators, it is not surprising that little MSE and absolute bias are observed for their estimators of the AUC when the sample size is bigger. Obviously, more samples with higher level separation of disease stages associated with higher AUC can achieve higher accuracy of the AUC estimators. We also note that all methods almost always have lower MSE and absolute bias with disease prevalence 0.5 than that with disease prevalence 0.3, especially when sample size is small (n=100 and 200). It indicates that the balanced diseased and non-diseased small samples provide higher accuracy of AUC estimates. The K-NN method works well, especially when the disease prevalence is 0.5. The overall top six methods, MSI, MSI-FI, FI-MSI, FI-SPE, SPE-FI and SPE-MSI, except the K-NN method, are all MSI and FI related approaches. We also note that SPE does not work very well but becomes accurate when it is combined with FI and MSI. MSI method almost always has the best AUC estimate in terms of lowest MSE and absolute bias under all settings considered here.

4.4.2 Misspecified Models

In this subsection, the misspecification of underlying models will be introduced. We use similar simulation settings in Alonzo and Pepe's study [23] included both misspecified disease models and misspecified verification models.

Misspecified Diseased Models We consider two misspecified diseased models here. Recall that disease status is simulated as $D = I(Z_1 + Z_2 > h)$, and T and A are generated as

Figure (4.1) Boxplots of MSE and Bias with Correct Models.



linear combinations of Z_1 and Z_2. In the setting where T contains information only about Z_1 (i.e. $\nu_1 = 1$ and $\tau_1 = 0$) and A contains information only about Z_2 (i.e. $\nu_2 = 0$ and $\tau_2 = 1$), the disease model $P(D|T)$ only linear in T is misspecified because the true disease process depends on aspects of A that are not included in T. By borrowing this idea and generating V in the same way as section 4.1, we have the first misspecified diseased model. We keep $\tau_1 = 0$, $\nu_2 = 0$, $\tau_2 = 1$ and select reasonable values of ν_1 to generate AUC which equals 0.7, 0.8, and 0.84, respectively. To apply the K-NN method, we use K-nearest neighbors with $K = 3$ and Euclidean distance based on only T. As mentioned in Section 4.1, only FI, MSI, and SPE depend on diseased probabilities ρ_i's, so we expect there is no much change for IPW method in this model compared with the correct model. We report the numerical results in Table 4.4 and boxplots in Figure 4.2 .

Place Table 4.4 here

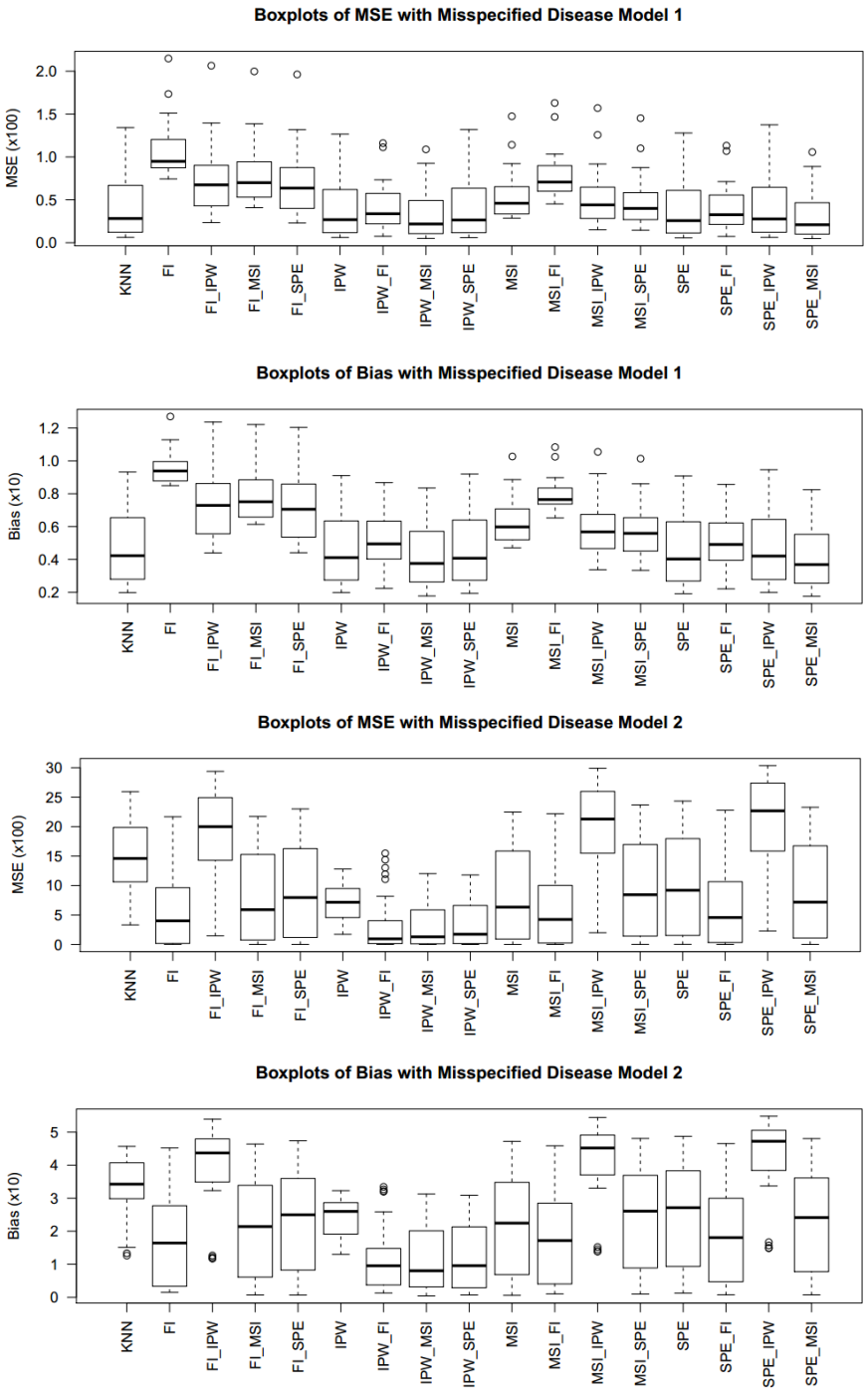
Similar to the correct models, generally all methods have lower MSE and bias with larger

sample size and higher AUC than that with smaller sample size and lower AUC. However, the MSE and bias have no big change as the disease prevalence changes from 0.3 to 0.5. The reason is that we are using an incorrect disease model and the accuracy of AUC estimator depends on the estimation of disease probability. Although using the correct verification process in misspecified disease setting, FI, MSI, and their related estimators of the AUC that use the incorrect disease model for $P(D|T)$ are less efficient than the corresponding estimators that use the correct model for $P(D|T, A)$. Specifically, FI and MSI estimators that use this incorrect disease model have bigger MSE and absolute bias, especially with small sample size and small AUC. As expected, there is no much change for IPW method comparing with the correct model. However, other methods, for example, SPE and SPE related methods, perform much better than IPW. The overall top six methods are SPE-MSI, IPW-MSI, SPE, IPW-SPE, IPW, and SPE-IPW. For some settings MSI method also performs well, especially when the sample size is small ($n = 100$). The K-NN method does not perform well except when the disease prevalence is 0.5, and sample size is big (n=1000).

The previous simulations all generated verification as a function of only T. Now, consider the setting where V is simulated such that the true verification model is specified as a function of T and A. we set $log(\frac{\pi}{1-\pi}) = -0.7 + T + A$ with $\pi = P(V = 1|T, A)$. In the presence of verification bias, D is only available for those patients with $V = 1$. Therefore, disease status results are available for roughly 40% of patients when $\nu_2 = 1$ and $\tau_2 = 1$. The disease model is misspecified if $P(D|T)$ linear in T is used because A contains information about the disease process not captured by T. For the K-NN method, we still only consider T to calculate the distance to estimate the probability of disease. This is the second misspecified disease model. Similar as section 4.1, we fix $\nu_2 = \tau_2 = 1$ and select different values of ν_1 and τ_1 to generate AUC which equals 0.7, 0.8, 0.9 and 0.96, respectively. We report the numerical results in Tables 4.5–4.6 and boxplots in Figure 4.2 . *Place Table 4.5–4.6 here*

Using the misspecified disease model 2, all methods always have lower MSE and bias with larger sample size and higher AUC than that with smaller sample size and lower AUC.

Figure (4.2) Boxplots of MSE and Bias with misspecified disease Models.



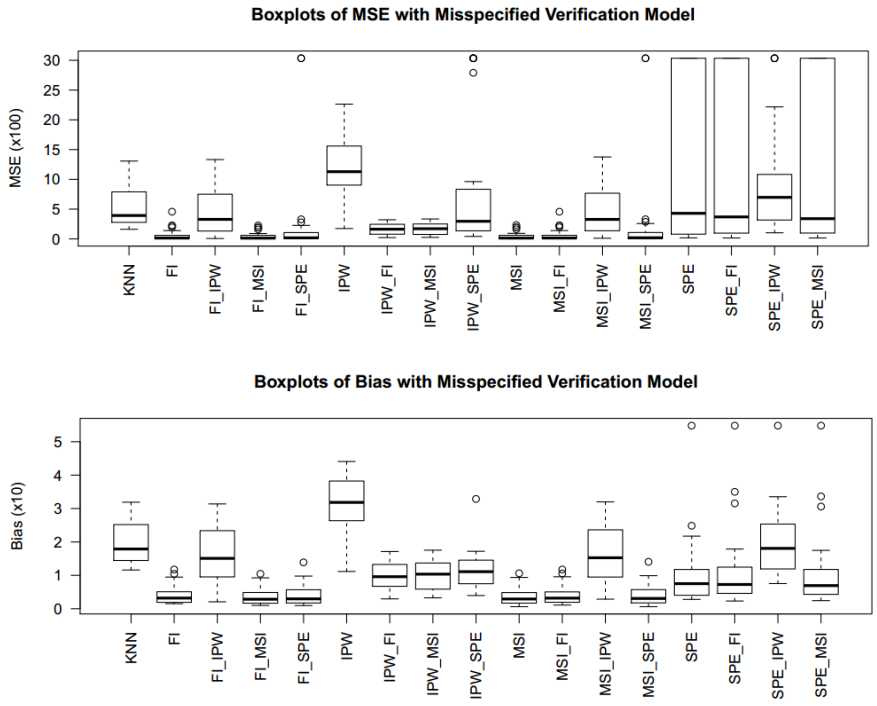
However, the performance of the AUC estimates is even worse as the disease prevalence changes from 0.3 to 0.5 because of the misspecified disease model. Due to the new underlying model of verification, the performance of IPW is worse compared with the correct model and misspecified disease model 1. Moreover, the performance of all these methods is worse than that in previous settings when AUC is less than 0.9. The overall top three methods are IPW-FI, IPW-MSI, and IPW-SPE. It is because the IPW weight is not depending on disease probability. Beside these three IPW related methods, FI, MSI-FI, and SPE-FI methods also perform well comparing with other methods. Note that FI-MSI method and MSI method are comparable with top IPW related methods when $AUC = 0.96$.The performance of the K-NN estimators is not acceptable in this setting.

Misspecified Verification Model We keep the correct diseased model as in section 4.1 and generate V by a logit model as the second misspecified disease model. However, we estimate the verification probabilities, $\pi_i = P(V|T)$, by a logistic regression instead of a correct way. Then a misspecified verification model happens because A is omitted from logistic regression when we estimate π_i but verification depends on both T and A in this setting. Similar as section 4.1, we keep $\nu_2 = \tau_2 = 1$ and select different values of ν_1 and τ_1 to generate AUC which equals 0.7, 0.8, 0.9 and 0.96, respectively. We report the numerical results in Tables 4.7–4.8 and boxplots in Figure 4.3 .

Place Table 4.7–4.8 here

From Tables 4.7–4.8 , we observe that the overall performance trend is similar to that with misspecified disease models. In addition, some methods, for example, SPE methods and SPE related methods, always have big MSE through all settings of this model. Especially when the sample size is small ($n = 100$), the MSE of SPE related methods is around thousand time of the best ones under the same setting. In order to show the mean difference between methods we set 29.531, the maximum value of MSE in misspecified disease model 2, as a upper bound to adjust the MSE for the boxplots in Figure 4.3 . The overall top four methods

Figure (4.3) Boxplots of MSE and Bias with Misspecified Verification Models.



are MSI, FI-MSI, MSI-FI, and FI. It is obviously due to MSI and FI weights not depending on verification probability π_i. Beside these four methods, FI-SPE and MSI-SPE are the best methods expected as SPE method is double robust (see [23]). IPW-FI and IPW-MSI also work well compared with other methods. The K-NN estimators are still not acceptable in this setting.

4.5 Study of Neonatal Hearing Screening Data

We apply the proposed methods to estimate AUC in the presence of verification bias to a neonatal hearing screening data set. Congenital hearing loss (hearing loss present at birth) is one of the most prevalent chronic conditions in children. Early detection and intervention will prevent delays in speech and language development and have long-lasting beneficial effects on social and emotional development and quality of life.[56] The Identification of Neonatal Hearing Impairment (INHI) is a study designed to evaluate the accuracy of two passive electronic devices, the Distortion Product Otoacoustic Emissions (DPOAE) test

and the Transient Evoked Otoacoustic Emissions (TEOAE) test.[23] These two tests can be administered soon after birth compared with the gold standard test for determining neonatal hearing loss, the Visual Reinforcement Audiometry (VRA) test, which cannot be administered until infants are eight to 12 months old. Therefore, the evaluation of these two screening tests is necessary for the earliest possible diagnosis.

The subset of the INHI data used in this section was generated by Alonzo and Pepe[23] following a two-phase design. In the first phase, DPOAE and TEOAE test results are available for all infants. In the second phase, all infants with DPOAE test results greater than the 80-th quantile of the distribution of DPOAE test results for at least one ear are sent to be verified, and remaining infants are verified with probability 0.4. The subset includes TEOAE and DPOAE test results on 5103 ears, corresponding to 2763 infants. Also, verification statuses V_i's and VRA results, D_i's, for 1571 verified infants are available. We follow the same argument in Alonzo and Pepe's study[23] to let T =DPOAE and A =TEOAE. Logistic regression is used to fit $P(D|T, A)$ to obtain ρ_i's in the final model. We use 5103 observations in our real case study, and the AUC is 0.632 based on the full data. For infants with DPOAE test results greater than the 80-th quantile of DPOAE results, the verification proportion $\pi_i = 1$; for infants with DPOAE test results below the threshold, $\pi_i = 0.4$. We find the point estimates of the AUC and the corresponding confidence interval using bootstrap variances. The results are provided in Table 4.1 . Comparing to the full data, all these methods overestimate the AUC. IPW method, which gives an AUC estimate closest to 0.632, has the best performance. IPW-related methods also work better than other methods. FI and FI-related methods have a shorter half length of confidence interval. The sample R code to estimate the AUC for the real example is provided in appendix.

Table (4.1) Real data AUC estimates and confidence intervals with verification bias.

Methods	Estimated AUC	Confidence Interval
Full data	0.632	(0.585, 0.679)
3-NN	0.640	(0.568, 0.711)
FI	0.646	(0.597, 0.694)
MSI	0.650	(0.597, 0.703)
SPE	0.644	(0.575, 0.714)
IPW	0.637	(0.577, 0.697)
SPE-FI	0.645	(0.576, 0.713)
SPE-MSI	0.645	(0.576, 0.714)
SPE-IPW	0.642	(0.572, 0.712)
FI-IPW	0.643	(0.592, 0.695)
FI-MSI	0.646	(0.598, 0.694)
FI-SPE	0.645	(0.597, 0.694)
IPW-SPE	0.639	(0.582, 0.696)
IPW-FI	0.639	(0.582, 0.696)
IPW-MSI	0.639	(0.582, 0.696)
MSI-FI	0.650	(0.598, 0.703)
MSI-SPE	0.650	(0.597, 0.703)
MSI-IPW	0.648	(0.593, 0.703)

Table (4.2) Correct models: MSE (×100) and Bias (×10) of AUC estimates with verification bias. Disease prevalence =0.3.

AUC	Prevalence	Methods	n=100		n=200		n=500		n=1000	
			MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.7	0.3	KNN	0.610	0.627	0.323	0.455	0.148	0.302	0.073	0.214
		FI	0.754	0.696	0.332	0.461	0.139	0.298	0.075	0.220
		MSI	0.716	0.676	0.307	0.444	0.125	0.284	0.065	0.204
		SPE	0.875	0.747	0.377	0.488	0.153	0.312	0.076	0.222
		IPW	1.268	0.905	0.576	0.612	0.211	0.365	0.117	0.272
		SPE-FI	0.788	0.708	0.335	0.461	0.134	0.293	0.067	0.208
		SPE-MSI	0.783	0.706	0.333	0.459	0.134	0.293	0.067	0.208
		SPE-IPW	0.991	0.794	0.442	0.528	0.170	0.329	0.090	0.241
		FI-IPW	0.882	0.754	0.402	0.508	0.160	0.323	0.089	0.238
		FI-MSI	0.739	0.689	0.324	0.456	0.135	0.294	0.072	0.216
		FI-SPE	0.754	0.694	0.328	0.458	0.136	0.295	0.071	0.214
		IPW-SPE	1.152	0.860	0.512	0.576	0.195	0.350	0.103	0.256
		IPW-FI	1.085	0.836	0.480	0.560	0.181	0.338	0.096	0.246
		IPW-MSI	1.074	0.831	0.475	0.557	0.179	0.337	0.095	0.245
		MSI-FI	0.730	0.683	0.314	0.449	0.129	0.287	0.067	0.208
		MSI-SPE	0.738	0.685	0.316	0.449	0.128	0.287	0.065	0.205
		MSI-IPW	0.862	0.742	0.388	0.498	0.151	0.314	0.082	0.229
0.8	0.3	KNN	0.426	0.521	0.236	0.387	0.116	0.270	0.058	0.192
		FI	0.474	0.555	0.239	0.392	0.099	0.254	0.058	0.194
		MSI	0.454	0.540	0.221	0.376	0.083	0.232	0.043	0.166
		SPE	0.613	0.628	0.296	0.433	0.117	0.273	0.062	0.198
		IPW	0.872	0.756	0.415	0.515	0.168	0.327	0.086	0.232
		SPE-FI	0.527	0.585	0.257	0.405	0.099	0.252	0.050	0.180
		SPE-MSI	0.529	0.585	0.258	0.406	0.102	0.255	0.053	0.185
		SPE-IPW	0.725	0.687	0.342	0.464	0.135	0.292	0.071	0.210
		FI-IPW	0.568	0.614	0.282	0.427	0.109	0.266	0.059	0.197
		FI-MSI	0.460	0.545	0.229	0.383	0.092	0.244	0.050	0.181
		FI-SPE	0.473	0.553	0.234	0.388	0.092	0.244	0.050	0.180
		IPW-SPE	0.750	0.695	0.375	0.487	0.152	0.310	0.078	0.221
		IPW-FI	0.685	0.668	0.344	0.467	0.135	0.294	0.067	0.206
		IPW-MSI	0.679	0.662	0.342	0.466	0.137	0.295	0.069	0.209
		MSI-FI	0.464	0.547	0.228	0.383	0.089	0.239	0.047	0.174
		MSI-SPE	0.474	0.552	0.230	0.383	0.086	0.235	0.044	0.168
		MSI-IPW	0.573	0.613	0.279	0.422	0.104	0.258	0.053	0.185
0.9	0.3	KNN	0.231	0.368	0.125	0.276	0.058	0.190	0.032	0.141
		FI	0.222	0.383	0.112	0.275	0.059	0.201	0.043	0.178
		MSI	0.215	0.368	0.098	0.251	0.039	0.159	0.022	0.121
		SPE	0.331	0.448	0.160	0.311	0.061	0.196	0.034	0.145
		IPW	0.428	0.518	0.209	0.359	0.080	0.225	0.043	0.164
		SPE-FI	0.269	0.413	0.127	0.282	0.047	0.174	0.024	0.125
		SPE-MSI	0.278	0.414	0.134	0.286	0.052	0.182	0.029	0.135
		SPE-IPW	0.393	0.489	0.192	0.343	0.069	0.208	0.038	0.155
		FI-IPW	0.261	0.414	0.129	0.295	0.054	0.190	0.033	0.150
		FI-MSI	0.213	0.371	0.100	0.258	0.046	0.174	0.030	0.143
		FI-SPE	0.218	0.376	0.105	0.262	0.046	0.175	0.030	0.142
		IPW-SPE	0.364	0.478	0.178	0.330	0.073	0.213	0.038	0.156
		IPW-FI	0.312	0.452	0.150	0.307	0.059	0.193	0.029	0.137
		IPW-MSI	0.317	0.449	0.155	0.309	0.064	0.200	0.034	0.146
		MSI-FI	0.221	0.378	0.105	0.264	0.048	0.178	0.031	0.146
		MSI-SPE	0.226	0.376	0.105	0.259	0.041	0.163	0.023	0.123
		MSI-IPW	0.273	0.418	0.132	0.294	0.049	0.179	0.027	0.132

continued from previous page

AUC	Prevalence	Methods	n=100		n=200		n=500		n=1000	
			MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.96	0.3	KNN	0.070	0.200	0.038	0.151	0.014	0.094	0.008	0.070
		FI	0.065	0.216	0.040	0.176	0.031	0.159	0.028	0.159
		MSI	0.063	0.205	0.030	0.145	0.016	0.108	0.013	0.098
		SPE	0.102	0.251	0.044	0.163	0.017	0.104	0.008	0.072
		IPW	0.136	0.288	0.060	0.192	0.022	0.118	0.011	0.084
		SPE-FI	0.075	0.225	0.035	0.155	0.015	0.102	0.009	0.079
		SPE-MSI	0.080	0.226	0.035	0.150	0.014	0.095	0.007	0.067
		SPE-IPW	0.134	0.280	0.056	0.183	0.021	0.114	0.010	0.080
		FI-IPW	0.077	0.226	0.038	0.164	0.021	0.125	0.016	0.114
		FI-MSI	0.062	0.207	0.032	0.154	0.020	0.125	0.017	0.119
		FI-SPE	0.064	0.210	0.032	0.153	0.019	0.119	0.016	0.112
		IPW-SPE	0.103	0.258	0.047	0.171	0.018	0.107	0.009	0.075
		IPW-FI	0.079	0.236	0.040	0.164	0.017	0.106	0.010	0.081
		IPW-MSI	0.083	0.236	0.040	0.160	0.015	0.099	0.008	0.070
		MSI-FI	0.064	0.212	0.036	0.164	0.024	0.138	0.022	0.135
		MSI-SPE	0.067	0.211	0.031	0.146	0.016	0.105	0.012	0.092
		MSI-IPW	0.082	0.231	0.038	0.160	0.018	0.112	0.012	0.095

Table (4.3) Correct models:MSE (×100) and Bias (×10) of AUC estimates with verification bias. Disease prevalence =0.5.

AUC	Prevalence	Methods	n=100		n=200		n=500		n=1000	
			MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.7	0.5	KNN	0.583	0.606	0.287	0.422	0.134	0.294	0.079	0.225
		FI	0.567	0.593	0.275	0.420	0.102	0.254	0.053	0.181
		MSI	0.555	0.586	0.268	0.415	0.101	0.251	0.055	0.185
		SPE	0.707	0.662	0.346	0.467	0.139	0.293	0.083	0.231
		IPW	1.061	0.827	0.532	0.585	0.204	0.356	0.118	0.279
		SPE-FI	0.580	0.601	0.278	0.422	0.104	0.254	0.057	0.190
		SPE-MSI	0.584	0.601	0.282	0.424	0.110	0.262	0.064	0.202
		SPE-IPW	0.897	0.744	0.460	0.538	0.174	0.328	0.101	0.255
		FI-IPW	0.791	0.703	0.406	0.510	0.150	0.307	0.081	0.229
		FI-MSI	0.555	0.586	0.269	0.415	0.100	0.250	0.052	0.181
		FI-SPE	0.580	0.600	0.282	0.426	0.106	0.257	0.058	0.192
		IPW-SPE	0.895	0.755	0.429	0.524	0.173	0.328	0.105	0.261
		IPW-FI	0.804	0.717	0.377	0.495	0.145	0.301	0.081	0.227
		IPW-MSI	0.798	0.714	0.378	0.495	0.149	0.304	0.087	0.237
		MSI-FI	0.561	0.590	0.271	0.417	0.100	0.251	0.052	0.180
		MSI-SPE	0.591	0.604	0.289	0.430	0.111	0.263	0.063	0.200
		MSI-IPW	0.800	0.705	0.411	0.512	0.153	0.309	0.085	0.234
0.8	0.5	KNN	0.437	0.523	0.213	0.369	0.089	0.238	0.046	0.171
		FI	0.438	0.531	0.246	0.402	0.116	0.279	0.083	0.240
		MSI	0.415	0.513	0.211	0.370	0.083	0.233	0.050	0.180
		SPE	0.528	0.580	0.248	0.397	0.088	0.237	0.048	0.176
		IPW	0.806	0.719	0.379	0.491	0.142	0.302	0.073	0.217
		SPE-FI	0.434	0.525	0.219	0.375	0.085	0.235	0.051	0.182
		SPE-MSI	0.430	0.520	0.206	0.362	0.074	0.219	0.040	0.161
		SPE-IPW	0.694	0.659	0.327	0.456	0.118	0.276	0.063	0.199
		FI-IPW	0.613	0.624	0.308	0.452	0.120	0.281	0.068	0.209
		FI-MSI	0.420	0.516	0.221	0.379	0.092	0.246	0.059	0.197
		FI-SPE	0.442	0.531	0.227	0.385	0.088	0.240	0.053	0.185
		IPW-SPE	0.667	0.653	0.309	0.442	0.115	0.273	0.061	0.199
		IPW-FI	0.585	0.617	0.293	0.436	0.116	0.275	0.064	0.203
		IPW-MSI	0.574	0.608	0.275	0.421	0.104	0.260	0.053	0.186
		MSI-FI	0.430	0.524	0.232	0.390	0.103	0.261	0.070	0.217
		MSI-SPE	0.446	0.532	0.222	0.379	0.082	0.231	0.047	0.173
		MSI-IPW	0.618	0.625	0.304	0.446	0.115	0.274	0.062	0.199

continued from previous page

AUC	Prevalence	Methods	n=100		n=200		n=500		n=1000	
			MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.9	0.5	KNN	0.255	0.393	0.121	0.278	0.052	0.184	0.027	0.133
		FI	0.262	0.419	0.156	0.331	0.091	0.259	0.076	0.245
		MSI	0.250	0.404	0.129	0.294	0.060	0.201	0.042	0.171
		SPE	0.336	0.458	0.154	0.313	0.054	0.185	0.028	0.133
		IPW	0.474	0.541	0.222	0.375	0.077	0.218	0.040	0.162
		SPE-FI	0.258	0.410	0.130	0.292	0.056	0.194	0.036	0.155
		SPE-MSI	0.261	0.410	0.123	0.280	0.046	0.173	0.025	0.127
		SPE-IPW	0.450	0.521	0.198	0.352	0.070	0.209	0.037	0.154
		FI-IPW	0.366	0.479	0.173	0.340	0.071	0.217	0.046	0.176
		FI-MSI	0.253	0.408	0.137	0.304	0.068	0.217	0.051	0.193
		FI-SPE	0.263	0.411	0.134	0.297	0.057	0.195	0.037	0.158
		IPW-SPE	0.367	0.486	0.182	0.339	0.064	0.200	0.032	0.145
		IPW-FI	0.300	0.453	0.158	0.324	0.066	0.208	0.041	0.166
		IPW-MSI	0.301	0.448	0.150	0.312	0.056	0.189	0.030	0.140
		MSI-FI	0.256	0.413	0.146	0.317	0.079	0.239	0.063	0.220
		MSI-SPE	0.269	0.414	0.132	0.292	0.052	0.185	0.031	0.143
		MSI-IPW	0.375	0.482	0.172	0.338	0.067	0.209	0.040	0.164
0.96	0.5	KNN	0.106	0.238	0.047	0.168	0.019	0.109	0.010	0.079
		FI	0.081	0.243	0.052	0.200	0.040	0.180	0.037	0.179
		MSI	0.079	0.237	0.045	0.181	0.030	0.151	0.025	0.142
		SPE	0.119	0.276	0.056	0.189	0.023	0.121	0.011	0.082
		IPW	0.153	0.305	0.075	0.215	0.029	0.134	0.014	0.095
		SPE-FI	0.083	0.240	0.043	0.172	0.023	0.129	0.016	0.106
		SPE-MSI	0.086	0.242	0.042	0.169	0.021	0.120	0.013	0.093
		SPE-IPW	0.152	0.303	0.068	0.207	0.026	0.129	0.013	0.090
		FI-IPW	0.111	0.268	0.053	0.191	0.026	0.134	0.017	0.111
		FI-MSI	0.080	0.240	0.048	0.189	0.034	0.163	0.030	0.158
		FI-SPE	0.086	0.244	0.045	0.177	0.023	0.127	0.016	0.106
		IPW-SPE	0.124	0.283	0.061	0.198	0.025	0.125	0.012	0.087
		IPW-FI	0.089	0.250	0.048	0.183	0.025	0.132	0.017	0.111
		IPW-MSI	0.092	0.252	0.047	0.180	0.023	0.124	0.014	0.098
		MSI-FI	0.079	0.239	0.049	0.191	0.035	0.167	0.031	0.162
		MSI-SPE	0.088	0.246	0.044	0.174	0.021	0.121	0.013	0.095
		MSI-IPW	0.114	0.270	0.053	0.189	0.024	0.128	0.015	0.101

Table (4.4) Misspecified-disease model 1: MSE (×100) and Bias (×10) of AUC estimates with verification bias.

AUC	Prevalence	Methods	n=100		n=200		n=500		n=1000	
			MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.7	0.3	KNN	1.342	0.933	0.644	0.646	0.248	0.399	0.123	0.279
		FI	2.149	1.270	1.511	1.071	1.198	0.998	1.117	0.999
		MSI	1.474	1.027	0.922	0.814	0.659	0.713	0.585	0.702
		SPE	1.280	0.908	0.588	0.616	0.225	0.381	0.113	0.268
		IPW	1.266	0.910	0.578	0.607	0.223	0.378	0.116	0.272
		SPE-FI	1.132	0.857	0.509	0.575	0.203	0.359	0.108	0.263
		SPE-MSI	1.057	0.824	0.471	0.552	0.183	0.341	0.094	0.245
		SPE-IPW	1.375	0.946	0.602	0.617	0.231	0.385	0.120	0.277
		FI-IPW	2.065	1.237	1.357	1.006	1.021	0.904	0.919	0.889
		FI-MSI	1.998	1.221	1.388	1.023	1.094	0.951	1.015	0.950
		FI-SPE	1.963	1.204	1.319	0.989	1.003	0.897	0.915	0.892
		IPW-SPE	1.320	0.920	0.625	0.636	0.237	0.388	0.118	0.273
		IPW-FI	1.162	0.867	0.538	0.593	0.211	0.368	0.112	0.268
		IPW-MSI	1.089	0.835	0.501	0.571	0.192	0.349	0.099	0.251
		MSI-FI	1.630	1.084	1.034	0.866	0.744	0.762	0.663	0.752
		MSI-SPE	1.452	1.013	0.876	0.786	0.594	0.664	0.513	0.645
		MSI-IPW	1.570	1.055	0.918	0.807	0.613	0.676	0.520	0.646
0.8	0.3	KNN	0.941	0.777	0.428	0.524	0.165	0.326	0.084	0.232
		FI	1.271	1.019	1.089	0.967	0.970	0.950	0.941	0.951
		MSI	0.795	0.765	0.567	0.654	0.422	0.593	0.383	0.585
		SPE	0.806	0.727	0.391	0.502	0.145	0.305	0.075	0.217
		IPW	0.836	0.739	0.417	0.518	0.153	0.316	0.080	0.226
		SPE-FI	0.712	0.693	0.365	0.490	0.148	0.312	0.085	0.236
		SPE-MSI	0.653	0.655	0.322	0.456	0.119	0.278	0.061	0.198
		SPE-IPW	0.890	0.754	0.427	0.524	0.155	0.318	0.081	0.227
		FI-IPW	1.102	0.931	0.860	0.835	0.708	0.790	0.666	0.787
		FI-MSI	1.068	0.918	0.871	0.851	0.744	0.821	0.712	0.821
		FI-SPE	1.056	0.910	0.837	0.826	0.702	0.791	0.663	0.788
		IPW-SPE	0.823	0.733	0.403	0.508	0.149	0.309	0.076	0.220
		IPW-FI	0.729	0.700	0.374	0.496	0.153	0.318	0.086	0.239
		IPW-MSI	0.671	0.663	0.333	0.463	0.124	0.282	0.063	0.200
		MSI-FI	0.974	0.862	0.743	0.767	0.595	0.721	0.554	0.716
		MSI-SPE	0.797	0.765	0.547	0.637	0.396	0.565	0.351	0.552
		MSI-IPW	0.863	0.797	0.581	0.658	0.406	0.570	0.357	0.554
0.84	0.3	KNN	0.693	0.662	0.316	0.446	0.126	0.282	0.062	0.199
		FI	0.927	0.882	0.838	0.864	0.763	0.849	0.744	0.850
		MSI	0.564	0.647	0.415	0.566	0.317	0.515	0.286	0.508
		SPE	0.632	0.644	0.290	0.433	0.117	0.272	0.058	0.192
		IPW	0.661	0.660	0.319	0.456	0.123	0.278	0.062	0.200
		SPE-FI	0.553	0.613	0.283	0.430	0.125	0.287	0.073	0.221
		SPE-MSI	0.510	0.583	0.243	0.396	0.097	0.249	0.048	0.175
		SPE-IPW	0.690	0.670	0.323	0.458	0.124	0.279	0.062	0.199
		FI-IPW	0.771	0.784	0.625	0.721	0.526	0.687	0.493	0.682
		FI-MSI	0.745	0.774	0.632	0.733	0.548	0.709	0.525	0.708
		FI-SPE	0.737	0.766	0.611	0.715	0.520	0.686	0.494	0.685
		IPW-SPE	0.647	0.648	0.297	0.437	0.120	0.275	0.059	0.194
		IPW-FI	0.568	0.621	0.290	0.436	0.128	0.291	0.075	0.223
		IPW-MSI	0.527	0.589	0.250	0.401	0.100	0.253	0.049	0.177
		MSI-FI	0.719	0.750	0.581	0.691	0.481	0.656	0.452	0.653
		MSI-SPE	0.567	0.644	0.403	0.551	0.299	0.495	0.267	0.486
		MSI-IPW	0.616	0.673	0.426	0.566	0.308	0.501	0.268	0.484

continued from previous page

AUC	Prevalence	Methods	n=100		n=200		n=500		n=1000	
			MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.7	0.5	KNN	1.173	0.856	0.573	0.606	0.214	0.372	0.109	0.261
		FI	1.735	1.129	1.238	0.970	0.951	0.884	0.894	0.895
		MSI	1.141	0.886	0.648	0.666	0.365	0.505	0.292	0.470
		SPE	1.134	0.854	0.523	0.573	0.194	0.353	0.097	0.250
		IPW	1.087	0.837	0.519	0.573	0.195	0.354	0.100	0.255
		SPE-FI	1.070	0.842	0.602	0.638	0.343	0.494	0.277	0.462
		SPE-MSI	0.888	0.753	0.424	0.519	0.172	0.333	0.102	0.261
		SPE-IPW	1.296	0.901	0.592	0.605	0.215	0.370	0.112	0.268
		FI-IPW	1.397	0.988	0.788	0.737	0.425	0.543	0.324	0.487
		FI-MSI	1.346	0.977	0.834	0.769	0.540	0.633	0.469	0.619
		FI-SPE	1.254	0.931	0.712	0.696	0.401	0.523	0.311	0.481
		IPW-SPE	1.169	0.864	0.561	0.600	0.214	0.369	0.105	0.259
		IPW-FI	1.113	0.863	0.637	0.654	0.359	0.503	0.288	0.468
		IPW-MSI	0.927	0.774	0.460	0.546	0.189	0.350	0.111	0.272
		MSI-FI	1.467	1.025	0.977	0.846	0.692	0.737	0.630	0.736
		MSI-SPE	1.100	0.861	0.574	0.616	0.272	0.422	0.181	0.352
		MSI-IPW	1.258	0.922	0.654	0.659	0.298	0.448	0.196	0.365
0.8	0.5	KNN	0.855	0.728	0.397	0.501	0.143	0.302	0.072	0.214
		FI	1.211	0.994	1.028	0.938	0.937	0.937	0.913	0.939
		MSI	0.741	0.728	0.495	0.602	0.354	0.533	0.316	0.524
		SPE	0.792	0.709	0.357	0.478	0.130	0.291	0.069	0.213
		IPW	0.782	0.705	0.385	0.501	0.139	0.298	0.073	0.217
		SPE-FI	0.703	0.703	0.441	0.558	0.309	0.492	0.265	0.476
		SPE-MSI	0.564	0.609	0.281	0.428	0.128	0.291	0.084	0.238
		SPE-IPW	0.913	0.751	0.420	0.516	0.149	0.309	0.078	0.224
		FI-IPW	0.886	0.793	0.548	0.633	0.348	0.513	0.291	0.485
		FI-MSI	0.873	0.808	0.638	0.702	0.504	0.657	0.469	0.655
		FI-SPE	0.796	0.749	0.497	0.599	0.330	0.503	0.282	0.482
		IPW-SPE	0.796	0.709	0.375	0.491	0.136	0.295	0.072	0.217
		IPW-FI	0.734	0.715	0.461	0.568	0.315	0.493	0.269	0.475
		IPW-MSI	0.592	0.626	0.302	0.446	0.134	0.297	0.088	0.243
		MSI-FI	1.027	0.898	0.823	0.823	0.719	0.810	0.690	0.810
		MSI-SPE	0.713	0.695	0.397	0.522	0.221	0.396	0.171	0.355
		MSI-IPW	0.817	0.746	0.456	0.565	0.241	0.414	0.181	0.364
0.84	0.5	KNN	0.720	0.670	0.328	0.459	0.121	0.281	0.062	0.198
		FI	0.947	0.891	0.852	0.873	0.772	0.856	0.763	0.863
		MSI	0.592	0.656	0.425	0.571	0.313	0.510	0.288	0.509
		SPE	0.651	0.642	0.289	0.425	0.113	0.268	0.056	0.191
		IPW	0.677	0.661	0.313	0.444	0.119	0.277	0.061	0.199
		SPE-FI	0.559	0.631	0.368	0.520	0.255	0.450	0.225	0.442
		SPE-MSI	0.462	0.553	0.236	0.396	0.114	0.276	0.077	0.229
		SPE-IPW	0.768	0.695	0.340	0.456	0.126	0.283	0.063	0.201
		FI-IPW	0.683	0.701	0.434	0.569	0.278	0.463	0.234	0.440
		FI-MSI	0.688	0.723	0.538	0.658	0.430	0.614	0.408	0.618
		FI-SPE	0.604	0.657	0.400	0.548	0.265	0.455	0.230	0.441
		IPW-SPE	0.652	0.643	0.293	0.427	0.117	0.273	0.059	0.195
		IPW-FI	0.582	0.645	0.375	0.524	0.259	0.450	0.229	0.443
		IPW-MSI	0.484	0.571	0.244	0.401	0.118	0.280	0.080	0.234
		MSI-FI	0.812	0.808	0.695	0.775	0.606	0.751	0.592	0.756
		MSI-SPE	0.550	0.616	0.323	0.479	0.184	0.364	0.145	0.334
		MSI-IPW	0.642	0.667	0.364	0.508	0.199	0.376	0.150	0.337

Table (4.5) Misspecified-disease model 2: MSE (×100) and Bias (×10) of AUC estimates with verification bias. Disease prevalence=0.3.

AUC	Prevalence	Methods	n=100		n=200		n=500		n=1000	
			MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.7	0.3	KNN	11.420	3.026	10.983	3.079	10.287	3.090	10.188	3.127
		FI	10.513	2.776	9.364	2.768	8.367	2.759	8.124	2.775
		MSI	11.235	2.987	10.605	3.049	9.964	3.069	9.875	3.093
		SPE	13.433	3.372	13.100	3.464	12.625	3.490	12.628	3.519
		IPW	7.787	2.601	7.399	2.626	7.004	2.610	6.954	2.617
		SPE-FI	11.677	2.981	10.751	3.016	9.865	3.024	9.678	3.045
		SPE-MSI	12.363	3.163	11.748	3.229	11.091	3.247	11.011	3.272
		SPE-IPW	20.992	4.479	21.229	4.562	21.148	4.583	21.212	4.597
		FI-IPW	18.166	4.143	18.633	4.264	18.719	4.308	18.823	4.329
		FI-MSI	10.717	2.893	10.044	2.949	9.383	2.969	9.279	2.994
		FI-SPE	11.492	3.072	11.180	3.172	10.753	3.209	10.747	3.239
		IPW-SPE	4.311	1.706	3.355	1.570	2.605	1.475	2.432	1.480
		IPW-FI	4.564	1.688	3.004	1.378	1.840	1.121	1.488	1.055
		IPW-MSI	4.165	1.637	2.964	1.418	2.063	1.254	1.822	1.236
		MSI-FI	10.813	2.839	9.790	2.853	8.866	2.853	8.654	2.872
		MSI-SPE	12.129	3.181	11.825	3.277	11.394	3.310	11.395	3.339
		MSI-IPW	19.159	4.270	19.527	4.372	19.550	4.405	19.638	4.423
0.8	0.3	KNN	13.467	3.278	13.247	3.323	12.313	3.303	12.097	3.339
		FI	6.452	1.957	4.725	1.794	3.456	1.673	3.099	1.657
		MSI	8.402	2.377	6.956	2.317	5.729	2.246	5.406	2.244
		SPE	11.128	2.864	9.822	2.858	8.587	2.806	8.292	2.811
		IPW	8.237	2.684	7.800	2.702	7.412	2.687	7.340	2.690
		SPE-FI	7.581	2.182	5.886	2.065	4.571	1.969	4.206	1.958
		SPE-MSI	9.349	2.539	7.840	2.486	6.545	2.415	6.208	2.413
		SPE-IPW	25.187	4.851	25.685	4.991	25.678	5.037	25.820	5.065
		FI-IPW	21.166	4.412	21.927	4.597	22.110	4.669	22.303	4.705
		FI-MSI	7.767	2.256	6.327	2.183	5.128	2.110	4.809	2.108
		FI-SPE	9.205	2.547	8.012	2.537	6.898	2.493	6.625	2.500
		IPW-SPE	2.686	1.206	1.637	0.980	0.933	0.797	0.739	0.752
		IPW-FI	2.045	1.030	0.918	0.714	0.313	0.441	0.157	0.317
		IPW-MSI	2.221	1.063	1.145	0.777	0.490	0.532	0.311	0.438
		MSI-FI	6.876	2.048	5.193	1.910	3.919	1.803	3.563	1.791
		MSI-SPE	9.991	2.685	8.774	2.680	7.623	2.634	7.344	2.640
		MSI-IPW	22.936	4.616	23.598	4.779	23.710	4.839	23.884	4.871

continued from previous page

AUC	Prevalence	Methods	n=100		n=200		n=500		n=1000	
			MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.9	0.3	KNN	11.707	2.946	11.136	2.906	10.250	2.845	9.544	2.787
		FI	1.250	0.723	0.564	0.527	0.223	0.359	0.147	0.305
		MSI	2.111	0.980	1.193	0.822	0.672	0.697	0.561	0.679
		SPE	3.149	1.280	2.018	1.145	1.302	1.025	1.154	1.012
		IPW	5.985	2.257	5.649	2.285	5.238	2.253	5.167	2.254
		SPE-FI	1.567	0.821	0.801	0.645	0.387	0.497	0.295	0.464
		SPE-MSI	2.383	1.063	1.396	0.907	0.819	0.783	0.697	0.767
		SPE-IPW	17.298	3.826	16.684	3.902	15.469	3.855	15.290	3.867
		FI-IPW	13.748	3.342	13.500	3.478	12.656	3.476	12.581	3.502
		FI-MSI	1.868	0.904	1.005	0.736	0.530	0.602	0.430	0.580
		FI-SPE	2.461	1.081	1.489	0.942	0.901	0.828	0.782	0.817
		IPW-SPE	0.552	0.460	0.194	0.309	0.069	0.202	0.043	0.166
		IPW-FI	0.327	0.431	0.156	0.347	0.129	0.335	0.123	0.337
		IPW-MSI	0.385	0.387	0.108	0.243	0.034	0.148	0.018	0.111
		MSI-FI	1.391	0.764	0.671	0.580	0.297	0.424	0.214	0.382
		MSI-SPE	2.792	1.181	1.750	1.048	1.104	0.934	0.972	0.922
		MSI-IPW	15.633	3.616	15.237	3.720	14.231	3.695	14.110	3.714
0.96	0.3	KNN	5.681	1.763	4.458	1.509	3.605	1.332	3.326	1.263
		FI	0.065	0.211	0.038	0.173	0.029	0.156	0.027	0.155
		MSI	0.099	0.207	0.037	0.137	0.012	0.083	0.006	0.059
		SPE	0.172	0.271	0.073	0.193	0.031	0.140	0.022	0.124
		IPW	2.279	1.327	1.991	1.318	1.777	1.298	1.735	1.299
		SPE-FI	0.067	0.191	0.027	0.132	0.011	0.089	0.008	0.074
		SPE-MSI	0.117	0.223	0.045	0.150	0.016	0.095	0.009	0.073
		SPE-IPW	4.014	1.667	3.014	1.571	2.431	1.498	2.302	1.485
		FI-IPW	2.715	1.266	1.996	1.215	1.561	1.173	1.477	1.175
		FI-MSI	0.085	0.203	0.033	0.140	0.012	0.089	0.007	0.068
		FI-SPE	0.117	0.223	0.044	0.149	0.014	0.090	0.007	0.066
		IPW-SPE	0.048	0.164	0.024	0.119	0.011	0.082	0.007	0.068
		IPW-FI	0.035	0.162	0.025	0.140	0.020	0.129	0.018	0.128
		IPW-MSI	0.031	0.137	0.014	0.095	0.005	0.060	0.003	0.043
		MSI-FI	0.062	0.190	0.028	0.141	0.016	0.109	0.013	0.101
		MSI-SPE	0.144	0.244	0.058	0.169	0.023	0.115	0.015	0.096
		MSI-IPW	3.487	1.523	2.631	1.450	2.115	1.390	2.008	1.383

Table (4.6) Misspecified-disease model 2: MSE (×100) and Bias (×10) of AUC estimates with verification bias. Disease prevalence=0.5.

AUC	Prevalence	Methods	n=100		n=200		n=500		n=1000	
			MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.7	0.5	KNN	20.639	4.211	18.541	4.057	16.736	3.962	15.846	3.904
		FI	20.623	3.965	21.031	4.135	21.672	4.418	21.568	4.521
		MSI	20.987	4.233	21.679	4.485	22.351	4.680	22.480	4.720
		SPE	22.493	4.442	22.889	4.652	23.225	4.784	23.205	4.801
		IPW	12.810	3.229	11.528	3.199	10.827	3.221	10.531	3.210
		SPE-FI	21.390	4.054	21.964	4.242	22.786	4.545	22.767	4.654
		SPE-MSI	22.038	4.345	22.609	4.583	23.191	4.768	23.288	4.805
		SPE-IPW	25.239	4.877	24.589	4.889	24.338	4.910	24.135	4.901
		FI-IPW	22.903	4.635	22.661	4.691	22.644	4.736	22.519	4.734
		FI-MSI	20.100	4.136	20.860	4.395	21.577	4.597	21.725	4.640
		FI-SPE	20.380	4.216	21.040	4.455	21.559	4.608	21.605	4.632
		IPW-SPE	11.798	3.028	10.676	3.016	10.052	3.078	9.816	3.089
		IPW-FI	15.490	3.344	14.385	3.264	13.032	3.212	11.933	3.194
		IPW-MSI	12.029	3.023	10.975	3.030	10.347	3.099	10.118	3.125
		MSI-FI	20.939	4.003	21.451	4.184	22.199	4.479	22.143	4.586
		MSI-SPE	21.345	4.321	21.904	4.549	22.356	4.693	22.377	4.714
		MSI-IPW	23.972	4.748	23.563	4.785	23.456	4.820	23.299	4.815
0.8	0.5	KNN	24.382	4.523	21.688	4.314	19.299	4.164	18.015	4.087
		FI	15.919	3.103	13.210	2.804	9.926	2.532	7.544	2.324
		MSI	20.311	3.868	20.429	4.066	21.086	4.379	21.355	4.518
		SPE	22.497	4.139	22.916	4.434	23.919	4.755	24.337	4.874
		IPW	12.327	3.085	10.745	3.027	9.929	3.061	9.564	3.050
		SPE-FI	16.519	3.174	13.819	2.886	10.571	2.635	8.164	2.441
		SPE-MSI	21.157	3.952	21.124	4.140	21.706	4.446	21.946	4.582
		SPE-IPW	29.531	5.253	29.067	5.305	29.225	5.380	29.094	5.382
		FI-IPW	27.075	5.022	27.169	5.128	27.623	5.231	27.601	5.242
		FI-MSI	19.491	3.784	19.721	3.990	20.432	4.307	20.722	4.449
		FI-SPE	20.643	3.953	21.361	4.273	22.524	4.611	23.004	4.738
		IPW-SPE	10.629	2.713	8.931	2.558	7.869	2.576	7.463	2.611
		IPW-FI	11.078	2.587	8.190	2.201	4.319	1.577	2.146	1.123
		IPW-MSI	10.327	2.627	8.642	2.467	7.237	2.385	6.628	2.387
		MSI-FI	16.201	3.137	13.513	2.845	10.253	2.584	7.859	2.384
		MSI-SPE	21.551	4.046	22.142	4.355	23.239	4.686	23.693	4.809
		MSI-IPW	28.280	5.137	28.125	5.218	28.447	5.308	28.375	5.315

continued from previous page

AUC	Prevalence	Methods	n=100		n=200		n=500		n=1000	
			MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.9	0.5	KNN	25.945	4.569	24.058	4.444	21.073	4.239	18.851	4.020
		FI	6.831	1.628	3.376	1.023	0.613	0.465	0.147	0.269
		MSI	11.423	2.377	9.962	2.206	5.633	1.725	3.484	1.474
		SPE	12.979	2.562	12.228	2.618	8.291	2.314	6.187	2.145
		IPW	9.438	2.598	8.169	2.553	6.994	2.531	6.606	2.516
		SPE-FI	7.004	1.654	3.496	1.047	0.653	0.480	0.162	0.279
		SPE-MSI	11.743	2.413	10.183	2.235	5.755	1.751	3.571	1.498
		SPE-IPW	28.964	5.047	29.543	5.250	30.160	5.433	30.345	5.482
		FI-IPW	26.942	4.855	28.129	5.121	29.077	5.334	29.378	5.394
		FI-MSI	11.059	2.336	9.705	2.172	5.490	1.696	3.384	1.447
		FI-SPE	12.162	2.470	11.632	2.541	7.919	2.250	5.901	2.087
		IPW-SPE	5.779	1.680	4.230	1.416	1.854	0.934	0.916	0.664
		IPW-FI	3.750	1.336	2.169	1.106	0.979	0.943	0.943	0.965
		IPW-MSI	5.130	1.497	3.815	1.327	1.460	0.829	0.593	0.560
		MSI-FI	6.921	1.641	3.440	1.035	0.634	0.473	0.155	0.274
		MSI-SPE	12.593	2.518	11.950	2.582	8.120	2.284	6.054	2.118
		MSI-IPW	27.998	4.956	28.883	5.190	29.665	5.388	29.908	5.442
0.96	0.5	KNN	20.442	3.953	20.708	4.018	17.521	3.699	15.739	3.510
		FI	4.578	1.195	1.442	0.502	0.050	0.174	0.028	0.149
		MSI	2.743	1.014	1.732	0.621	0.198	0.227	0.059	0.137
		SPE	2.045	0.859	1.779	0.656	0.338	0.309	0.109	0.212
		IPW	3.990	1.569	3.662	1.572	2.744	1.507	2.482	1.506
		SPE-FI	4.586	1.196	1.444	0.501	0.049	0.170	0.027	0.143
		SPE-MSI	2.766	1.019	1.745	0.625	0.200	0.229	0.060	0.139
		SPE-IPW	15.874	3.369	16.727	3.642	15.848	3.778	15.633	3.851
		FI-IPW	14.834	3.230	16.152	3.571	15.588	3.749	15.489	3.836
		FI-MSI	2.710	1.008	1.713	0.617	0.195	0.224	0.058	0.136
		FI-SPE	1.967	0.843	1.735	0.645	0.327	0.301	0.104	0.205
		IPW-SPE	0.891	0.577	0.728	0.446	0.119	0.268	0.075	0.251
		IPW-FI	0.889	0.694	0.429	0.489	0.158	0.397	0.158	0.397
		IPW-MSI	0.534	0.365	0.588	0.400	0.125	0.315	0.106	0.311
		MSI-FI	4.583	1.195	1.443	0.501	0.049	0.172	0.027	0.145
		MSI-SPE	2.011	0.852	1.761	0.651	0.333	0.306	0.107	0.209
		MSI-IPW	15.401	3.306	16.475	3.611	15.743	3.767	15.583	3.846

Table (4.7) Misspecified-verification model: MSE (×100) and Bias (×10) of AUC estimates with verification bias. Disease prevalence=0.3.

AUC	Prevalence	Methods	n=100		n=200		n=500		n=1000	
			MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.7	0.3	KNN	2.394	1.319	2.062	1.253	1.712	1.164	1.615	1.160
		FI	0.735	0.666	0.337	0.462	0.138	0.297	0.074	0.220
		MSI	0.704	0.648	0.330	0.458	0.137	0.299	0.078	0.227
		SPE	208.639	1.526	3.784	0.859	0.931	0.692	0.710	0.664
		IPW	12.007	3.148	11.057	3.123	9.919	3.057	9.716	3.069
		SPE-FI	479.406	1.788	6.385	0.902	0.918	0.680	0.662	0.637
		SPE-MSI	458.758	1.747	6.161	0.882	0.883	0.665	0.638	0.624
		SPE-IPW	22.175	2.119	3.996	1.574	1.724	1.100	1.125	0.904
		FI-IPW	4.485	1.711	3.183	1.490	2.171	1.310	1.900	1.283
		FI-MSI	0.718	0.656	0.329	0.456	0.132	0.291	0.071	0.214
		FI-SPE	1.183	0.671	0.337	0.463	0.145	0.310	0.085	0.237
		IPW-SPE	9.302	1.509	2.627	1.449	2.303	1.434	2.263	1.460
		IPW-FI	2.683	1.400	2.527	1.441	2.368	1.466	2.382	1.505
		IPW-MSI	2.743	1.421	2.585	1.462	2.412	1.482	2.423	1.519
		MSI-FI	0.719	0.658	0.338	0.464	0.143	0.306	0.082	0.234
		MSI-SPE	1.168	0.666	0.343	0.468	0.155	0.321	0.096	0.255
		MSI-IPW	4.556	1.728	3.168	1.485	2.086	1.281	1.794	1.243
0.8	0.3	KNN	3.531	1.634	3.091	1.556	2.728	1.488	2.448	1.443
		FI	0.489	0.526	0.210	0.363	0.089	0.241	0.053	0.188
		MSI	0.474	0.507	0.200	0.349	0.079	0.227	0.046	0.173
		SPE	100.222	2.341	13.014	1.167	6.603	0.961	5.486	0.828
		IPW	14.081	3.308	12.930	3.298	11.430	3.219	11.035	3.229
		SPE-FI	244.420	3.187	20.466	1.253	10.471	1.029	7.744	0.871
		SPE-MSI	230.040	3.115	19.751	1.226	10.184	1.001	7.545	0.839
		SPE-IPW	148.301	1.000	7.266	1.764	4.079	1.295	3.563	0.980
		FI-IPW	3.850	1.527	2.512	1.282	1.530	1.064	1.231	1.001
		FI-MSI	0.483	0.518	0.203	0.354	0.081	0.228	0.045	0.171
		FI-SPE	0.501	0.525	0.237	0.363	0.087	0.239	0.052	0.186
		IPW-SPE	4.969	1.605	5.303	1.660	3.320	1.666	3.282	1.720
		IPW-FI	3.098	1.443	3.121	1.540	3.123	1.637	3.213	1.713
		IPW-MSI	3.241	1.489	3.266	1.588	3.260	1.680	3.348	1.753
		MSI-FI	0.477	0.516	0.207	0.359	0.088	0.242	0.055	0.192
		MSI-SPE	0.508	0.521	0.245	0.363	0.088	0.241	0.055	0.190
		MSI-IPW	4.006	1.568	2.565	1.299	1.517	1.061	1.199	0.988

continued from previous page

AUC	Prevalence	Methods	n=100		n=200		n=500		n=1000	
			MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.9	0.3	KNN	4.530	1.820	4.018	1.747	3.451	1.659	3.316	1.668
		FI	0.196	0.340	0.086	0.241	0.048	0.184	0.037	0.167
		MSI	0.190	0.314	0.070	0.209	0.029	0.138	0.018	0.110
		SPE	12.216	1.232	221.075	1.469	46.229	1.185	22.283	0.997
		IPW	10.999	2.696	9.886	2.688	8.348	2.610	7.785	2.610
		SPE-FI	17.510	1.269	31.858	1.242	110.345	1.295	40.878	1.085
		SPE-MSI	16.577	1.246	30.799	1.222	108.162	1.276	40.227	1.067
		SPE-IPW	6.727	1.899	9.899	1.719	19.154	1.510	11.166	1.283
		FI-IPW	2.235	1.074	1.258	0.860	0.635	0.655	0.448	0.576
		FI-MSI	0.196	0.329	0.077	0.223	0.035	0.153	0.023	0.125
		FI-SPE	0.258	0.344	1.718	0.259	0.043	0.159	0.023	0.126
		IPW-SPE	9.101	1.232	193.422	1.552	3.908	1.327	2.504	1.362
		IPW-FI	1.758	0.964	1.835	1.048	1.888	1.163	1.931	1.250
		IPW-MSI	1.913	1.024	1.995	1.117	2.050	1.233	2.099	1.317
		MSI-FI	0.186	0.324	0.077	0.226	0.041	0.168	0.030	0.150
		MSI-SPE	0.270	0.339	2.238	0.257	0.041	0.149	0.020	0.113
		MSI-IPW	2.439	1.145	1.363	0.911	0.687	0.692	0.484	0.607
0.96	0.3	KNN	3.517	1.448	2.832	1.273	2.343	1.210	2.093	1.171
		FI	0.043	0.179	0.031	0.157	0.027	0.152	0.026	0.154
		MSI	0.034	0.143	0.015	0.101	0.007	0.072	0.005	0.062
		SPE	1.919	0.390	2.394	0.391	0.750	0.301	4.802	0.315
		IPW	3.454	1.246	2.726	1.180	2.046	1.121	1.738	1.116
		SPE-FI	3.485	0.376	3.865	0.368	0.195	0.246	0.154	0.232
		SPE-MSI	3.285	0.365	3.757	0.358	0.205	0.247	0.168	0.243
		SPE-IPW	2.395	0.940	2.107	0.870	1.249	0.766	1.021	0.754
		FI-IPW	0.576	0.486	0.253	0.355	0.104	0.250	0.065	0.208
		FI-MSI	0.038	0.160	0.021	0.124	0.014	0.104	0.012	0.101
		FI-SPE	0.039	0.160	0.026	0.125	0.015	0.101	0.026	0.096
		IPW-SPE	0.652	0.406	0.926	0.408	0.861	0.413	4.793	0.457
		IPW-FI	0.284	0.326	0.272	0.311	0.256	0.310	0.222	0.330
		IPW-MSI	0.328	0.352	0.313	0.345	0.296	0.358	0.264	0.386
		MSI-FI	0.036	0.157	0.022	0.127	0.016	0.112	0.014	0.111
		MSI-SPE	0.039	0.149	0.024	0.107	0.010	0.073	0.028	0.062
		MSI-IPW	0.721	0.578	0.341	0.438	0.161	0.329	0.110	0.287

Table (4.8) Misspecified-verification model: MSE (×100) and Bias (×10) of AUC estimates with verification bias. Disease prevalence=0.5.

AUC	Prevalence	Methods	n=100		n=200		n=500		n=1000	
			MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.7	0.5	KNN	5.810	2.006	4.419	1.849	3.835	1.811	3.451	1.766
		FI	2.205	1.044	0.831	0.665	0.254	0.393	0.115	0.271
		MSI	2.320	1.062	0.929	0.694	0.291	0.411	0.130	0.285
		SPE	138.694	1.457	1.107	0.760	0.365	0.463	0.171	0.327
		IPW	20.145	4.195	17.713	3.979	15.761	3.856	14.908	3.794
		SPE-FI	3.513	1.129	1.014	0.733	0.328	0.445	0.156	0.315
		SPE-MSI	3.506	1.132	1.073	0.747	0.344	0.451	0.158	0.316
		SPE-IPW	14.151	3.267	11.245	2.958	8.821	2.740	7.695	2.631
		FI-IPW	13.325	3.138	10.782	2.887	8.718	2.741	7.765	2.663
		FI-MSI	2.263	1.047	0.894	0.680	0.273	0.399	0.120	0.274
		FI-SPE	138.059	1.390	0.933	0.692	0.288	0.408	0.129	0.283
		IPW-SPE	651.943	2.287	1.838	1.142	1.330	1.069	1.258	1.078
		IPW-FI	2.749	1.253	1.535	1.023	1.075	0.945	1.010	0.955
		IPW-MSI	2.919	1.298	1.727	1.103	1.254	1.035	1.190	1.047
		MSI-FI	2.245	1.056	0.857	0.676	0.266	0.400	0.119	0.276
		MSI-SPE	138.122	1.405	0.972	0.708	0.308	0.422	0.141	0.295
		MSI-IPW	13.762	3.199	11.104	2.935	8.941	2.777	7.947	2.694
0.8	0.5	KNN	8.121	2.402	6.457	2.234	5.420	2.171	4.886	2.099
		FI	1.936	0.948	0.696	0.607	0.229	0.395	0.139	0.317
		MSI	1.938	0.937	0.748	0.613	0.219	0.371	0.106	0.267
		SPE	119.058	2.174	1.144	0.711	0.385	0.461	0.218	0.355
		IPW	22.634	4.408	19.499	4.107	16.800	3.915	15.431	3.809
		SPE-FI	461.225	3.153	1.222	0.721	0.430	0.495	0.278	0.413
		SPE-MSI	428.042	3.059	1.222	0.714	0.394	0.464	0.224	0.359
		SPE-IPW	49.884	3.353	10.496	2.732	7.224	2.355	5.603	2.143
		FI-IPW	13.127	2.998	10.082	2.658	7.242	2.382	5.851	2.226
		FI-MSI	1.895	0.926	0.726	0.603	0.211	0.366	0.104	0.267
		FI-SPE	2.787	0.978	0.780	0.614	0.217	0.369	0.105	0.266
		IPW-SPE	9.622	1.421	1.898	1.069	1.411	1.032	1.370	1.073
		IPW-FI	2.557	1.164	1.509	0.935	1.125	0.887	1.079	0.926
		IPW-MSI	2.684	1.187	1.693	1.015	1.324	0.993	1.290	1.037
		MSI-FI	1.963	0.957	0.712	0.615	0.233	0.396	0.137	0.312
		MSI-SPE	2.868	0.990	0.803	0.624	0.226	0.374	0.107	0.267
		MSI-IPW	13.551	3.056	10.357	2.698	7.396	2.406	5.954	2.244

continued from previous page

AUC	Prevalence	Methods	n=100		n=200		n=500		n=1000	
			MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
0.9	0.5	KNN	11.853	2.987	10.554	2.894	8.421	2.703	7.672	2.637
		FI	2.270	0.892	0.485	0.466	0.139	0.322	0.106	0.293
		MSI	1.615	0.785	0.528	0.464	0.120	0.285	0.072	0.229
		SPE	50.738	2.485	120.114	0.832	0.791	0.421	0.484	0.375
		IPW	20.798	2.089	18.204	3.838	14.696	3.543	12.724	3.348
		SPE-FI	118.810	2.500	1.815	0.610	1.644	0.500	1.275	0.481
		SPE-MSI	116.920	2.363	1.789	0.603	1.581	0.456	1.201	0.413
		SPE-IPW	48.455	2.343	9.142	2.438	5.013	1.851	3.166	1.498
		FI-IPW	11.147	2.593	8.599	2.294	4.960	1.837	3.375	1.580
		FI-MSI	1.592	0.779	0.517	0.457	0.115	0.280	0.069	0.225
		FI-SPE	1.958	0.704	119.228	0.694	0.114	0.275	0.066	0.217
		IPW-SPE	27.902	1.201	80.042	1.045	1.107	0.797	1.034	0.846
		IPW-FI	2.073	0.982	1.068	0.708	0.808	0.655	0.771	0.690
		IPW-MSI	1.850	0.884	1.200	0.753	0.949	0.741	0.930	0.795
		MSI-FI	2.280	0.896	0.493	0.471	0.143	0.326	0.108	0.294
		MSI-SPE	2.030	0.711	119.242	0.700	0.119	0.281	0.069	0.221
		MSI-IPW	11.497	2.642	8.786	2.319	5.017	1.844	3.384	1.578
0.96	0.5	KNN	12.859	3.146	13.073	3.189	10.660	2.960	9.098	2.800
		FI	4.564	1.175	1.392	0.492	0.054	0.194	0.040	0.189
		MSI	1.939	0.820	0.673	0.392	0.040	0.165	0.027	0.149
		SPE	3.721	0.767	128.663	0.743	26.871	0.416	0.797	0.280
		IPW	11.147	2.748	10.333	2.663	7.793	2.375	6.354	2.202
		SPE-FI	5.636	1.249	88.462	0.772	72.814	0.477	2.326	0.357
		SPE-MSI	2.972	0.897	85.983	0.673	73.649	0.454	2.328	0.324
		SPE-IPW	7.352	2.044	6.431	1.858	3.169	1.332	1.641	1.014
		FI-IPW	6.297	1.772	5.437	1.630	2.620	1.160	1.384	0.906
		FI-MSI	1.933	0.819	0.670	0.391	0.039	0.162	0.026	0.146
		FI-SPE	3.292	0.689	2.265	0.359	0.111	0.158	0.022	0.130
		IPW-SPE	7.578	0.671	143.390	0.701	6.403	0.448	0.420	0.396
		IPW-FI	1.727	0.837	0.723	0.477	0.242	0.305	0.217	0.296
		IPW-MSI	0.569	0.434	0.373	0.363	0.273	0.327	0.257	0.337
		MSI-FI	4.565	1.176	1.393	0.493	0.055	0.196	0.041	0.191
		MSI-SPE	3.298	0.690	2.573	0.363	0.114	0.161	0.023	0.134
		MSI-IPW	6.519	1.815	5.536	1.647	2.630	1.159	1.371	0.897

PART 5

CONCLUSIONS

In the medical diagnostic study, the sensitivity/ sensitivity to the early stage, ROC curve/ROC surface and AUC/VUS(Area Under the ROC surface) are meaningful measures of the diagnostic accuracy. In this dissertation, we developed Bayesian and influence function-based empirical likelihood methods to construct confidence intervals for sensitivity and sensitivity to early stage in two and three ordinal diagnostic classes case. We also derive several closed-form expressions of the AUC estimator with verification bias data under missing at random (MAR) assumption.

5.1 Inference of Sensitivity in Diagnostic Test

In Part 2, we focused on the two classes diagnostic test that classifies individuals with/ without targeted disease. We reviewed the existing methods for inference on sensitivity given specificity that include normal approximation based methods (NA, BTI, BTII) and empirical likelihood methods (JEL, HEL) for comparison. We proposed various influence function empirical likelihood (IFEL) and Bayesian empirical likelihood (BEL and BpEL) confidence intervals for sensitivity given the fixed specificity (80%, 90%, and 95%). The idea of IFEL is to replace the estimating function in the EL with an influence function of the parameter of interest (sensitivity). The corresponding empirical log-likelihood ratio statistic converges to a standard chi-square distribution with one degree of freedom, making inference for sensitivity more convenient. The proposed Bayesian EL approaches (BEL and BpEL) include two types according to unknown parameters we consider. We considered building EL and assigning priors on either the sensitivity parameter itself or the probability vector $(p_1, \ldots, p_n)$ in building EL. Then we used the empirical likelihood in Bayesian inference like parametric likelihoods. Numerical studies are performed to compare the finite sample

performance of the proposed approaches with existing methods. The simulation studies show that the existing HEL interval performs well and the proposed intervals have similar or better coverage accuracy than the existing intervals. The BHEL and BpHEL intervals have good small sample performance and do not require density estimation. However, they involve bootstrap process, which is computationally expensive and might be undesirable. We note that influence function-based intervals perform slightly worse than the hybrid EL intervals, probably because of the poor density estimation. IFEL, BIFEL, and BpIFEL methods are good alternative methods when computational cost is a concern.

In Part 3, we extended a similar study as in Part 2 to three ordinal classes diagnostic test that classifies individuals into three stage: normal healthy stage, early diseased stage and full diseased stage. We reviewed two empirical likelihood methods (ELP and ELB) and proposed various influence function empirical likelihood (IF) and Bayesian empirical likelihood (BEL and BpEL) confidence intervals for sensitivity to the early stage given the fixed specificity (80% and 90%) and sensitivity (80% and 90%) to the fully diseased stage. Similarly, the corresponding empirical log-likelihood ratio statistic of IF empirical likelihood converges to a standard chi-square distribution with one degree of freedom and both two types of Bayesian EL are considered. The simulation studies show that the proposed intervals have better coverage accuracy than the existing EL intervals. The BEL and BpEL intervals are consistent and always have the best performance estimation. We note that influence function-based intervals perform slightly worse than the BEL and BpEL intervals. In addition, IF, BIF1 and BIF2 methods do not work well when P_2, the true sensitivity to the early stage, is high ($P_2 = 0.9$). However, BpIF1 and BpIF2 has better or similar performance when P_2 is high. The possible reason is that the BpEL is constructed by assigning prior to the probability vector, which does not depend on the true sensitivity. Unexpectedly, NA methods work acceptably in most of the settings consider here. The performance of ELP and ELB is similar and not acceptable.

In practice, clinicians sometimes need to compare two tests in terms of their sensitivities/sensitivities to early stage at the same specificity/(specificity and sensitivity to full dis-

eased stage). The inference procedure is simpler with the proposed Bayesian approach. We can generate posterior samples of these two sensitivities separately to obtain posterior samples of the difference of them. Based on these posterior samples, Bayesian credible intervals can be constructed. In addition, the influence function techniques can extend immediately to the difference between the sensitivities of two tests at a fixed specificity since the influence function of the difference is the difference between respective influence functions. Similar ideas can also applied to three diagnostic classes cases for the inference of the difference between sensitivities to the early stage. The EL methods considered in this dissertation could be extended for the difference between the corresponding sensitivities $\theta(p_1) - \theta(p_2)$ by constructing suitable estimating functions. Alternatively, we can consider two-dimensional estimating functions to apply EL method on $(\theta(p_1), \theta(p_2))$ and then construct a confidence region. Furthermore, credible intervals for $\theta(p_1) - \theta(p_2)$ can be constructed based on posterior samples of $(\theta(p_1), \theta(p_2))$.

5.2 Inference for the AUC with Verification Biased Data

Medical diagnostic procedure usually involves two-phrase tests, diagnostic/screening test (non-invasive) and gold standard test that verifies the true disease status. A diagnostic test need to be evaluated by a study with the selected true non-diseased and true diseased samples determined according to the gold standard test. However, in many situations, not all individuals with given diagnostic/screening test results ultimately have their true disease status verified through a very accurate gold standard test. That is to say, the labels referred to as true disease status of the individuals are partially missing. Because the estimates of AUC based on partially validated subjects are usually biased, it is usually necessary to estimate AUC from a bias-corrected ROC curve. He et al.[24] proposed an estimator for the AUC that is based on IPW and has a simple closed-form expression. Innovated by He et al.[24] and Alonzo and Pepe's[23] work, in Part 4, we proposed various direct estimations of the AUC based on combination of w and r weights of FI, MSI, IPW, and SPE with verification biased data when the test result is continuous under the assumption that the

true disease status, if missing, is missing at random (MAR). He et al.[24] also proposed a closed form expression for the asymptotic variance of the IPW-based estimator for the AUC under the assumption that the verification probability is known. However, in practice, the verification mechanism for true disease status is unknown. Hence it is very challenging to provide explicit variance estimates for the proposed AUC estimators. Alonzo and Pepe[23] suggested that bootstrap resampling can be used to obtain confidence intervals for the AUC based on their bias-corrected ROC curves. Similarly, bootstrap method can be used to estimate the variances of the new AUC estimators and construct bootstrap-based confidence intervals for the AUC.

Our simulation results show that the newly proposed AUC estimators are accurate for the biased sampling if the disease and verification models are correctly specified. Especially, FI and MSI related methods almost always perform best. Misspecifying the disease model yields biased FI and MSI estimates, and misspecifying the verification model yields biased IPW and SPE related estimates.

However, MSI and IPW related estimators, SPE-MSI, IPW-MSI , FI-MSI, IPW-FI and IWP-MSI still perform well although using incorrect disease models.

FI-SPE, MSI-SPE, IPW-FI and IWP-MSI methods work well even under the misspecified verification model. The K-NN methods work well if the model is correct and the prevalence rate is high (0.5). However, when model misspecification is involved, the K-NN methods have poor performances. Based on our simulation study results, we recommend the use of IPW-FI and IWP-MSI estimators for the AUC to assess the accuracy of a continuous-scale test if the disease model and verification model are unknown. In the situation with the disease model being correctly specified or well approximated, it is preferable to use MSI and FI estimators for the AUC.

The proposed AUC estimators have closed-form expressions and can be easily computed and directly applied in practice under the common MAR assumption for the verification biased data. However, disease validations in practice may depend on some unobserved covariates associated with the disease, which results in nonignorable verification biased data.

Liu and Zhou[57] proposed four types of estimates (i.e, FI, MSI, IPW, and PDR estimators) for sensitivity and specificity as well as the AUC under nonignorable missing data mechanism using imputation and reweighting methods. Similarly, the direct approaches proposed in this dissertation can be extended to find explicit estimators for the AUC with nonignorable verification biased data.

REFERENCES

[1] D. Berger, "A brief history of medical diagnosis and the birth of the clinical laboratory. part 4–fraud and abuse, managed-care, and lab consolidation." *MLO: Medical Laboratory Observer*, vol. 31, no. 12, p. 38, 1999.

[2] P. L. Epner, J. E. Gans, and M. L. Graber, "When diagnostic testing leads to harm: a new outcomes-based approach for laboratory medicine," *BMJ quality & safety*, vol. 22, no. Suppl 2, pp. ii6–ii10, 2013.

[3] C. M. Umemneku Chikere, K. Wilson, S. Graziadio, L. Vale, and A. J. Allen, "Diagnostic test evaluation methodology: A systematic review of methods employed to evaluate diagnostic tests in the absence of gold standard–an update," *PloS one*, vol. 14, no. 10, p. e0223832, 2019.

[4] P. M. Bossuyt, J. B. Reitsma, K. Linnet, and K. G. Moons, "Beyond diagnostic accuracy: the clinical utility of diagnostic tests," *Clinical chemistry*, vol. 58, no. 12, pp. 1636–1643, 2012.

[5] W. Burke, "Genetic tests: clinical validity and clinical utility," *Current protocols in human genetics*, vol. 81, no. 1, pp. 9–15, 2014.

[6] H. J. Baskin, R. H. Cobin, D. S. Duick, H. Gharib, R. B. Guttler, M. M. Kaplan, R. L. Segal, J. R. Garber, C. R. Hamilton Jr, Y. Handelsman *et al.*, "American association of clinical endocrinologists medical guidelines for clinical practice for the evaluation and treatment of hyperthyroidism and hypothyroidism: Aace thyroid task force," *Endocrine practice*, vol. 8, no. 6, pp. 457–469, 2002.

[7] G. Grunau and S. Linn, "Detection and diagnostic overall accuracy measures of medical tests," *Rambam Maimonides medical journal*, vol. 9, no. 4, 2018.

[8] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (roc) curve." *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[9] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.

[10] R. W. Platt, J. A. Hanley, and H. Yang, "Bootstrap confidence intervals for the sensitivity of a quantitative diagnostic test," *Statistics in Medicine*, vol. 19, no. 3, pp. 313–322, 2000.

[11] D. Saslow, P. E. Castle, J. T. Cox, D. D. Davey, M. H. Einstein, D. G. Ferris, S. J. Goldie, D. M. Harper, W. Kinney, A.-B. Moscicki *et al.*, "American cancer society guideline for human papillomavirus (hpv) vaccine use to prevent cervical cancer and its precursors," *CA: a cancer journal for clinicians*, vol. 57, no. 1, pp. 7–28, 2007.

[12] S. W. Greenhouse and N. Mantel, "The evaluation of diagnostic tests," *Biometrics*, vol. 6, no. 4, pp. 399–412, 1950.

[13] K. Linnet, "Comparison of quantitative diagnostic tests: type i error, power, and sample size," *Statistics in Medicine*, vol. 6, no. 2, pp. 147–158, 1987.

[14] X.-H. Zhou and G. Qin, "Improved confidence intervals for the sensitivity at a fixed level of specificity of a continuous-scale diagnostic test," *Statistics in Medicine*, vol. 24, no. 3, pp. 465–477, 2005.

[15] Y. Gong, L. Peng, and Y. Qi, "Smoothed jackknife empirical likelihood method for roc curve," *Journal of Multivariate Analysis*, vol. 101, no. 6, pp. 1520–1531, 2010.

[16] G. Qin, A. E. Davis, and B.-Y. Jing, "Empirical likelihood-based confidence intervals for the sensitivity of a continuous-scale diagnostic test at a fixed level of specificity," *Statistical Methods in Medical Research*, vol. 20, no. 3, pp. 217–231, 2011.

[17] N. A. Lazar, "Bayesian empirical likelihood," *Biometrika*, vol. 90, no. 2, pp. 319–326, 2003.

[18] R. C. Petersen, "Mild cognitive impairment as a diagnostic entity," *Journal of internal medicine*, vol. 256, no. 3, pp. 183–194, 2004.

[19] T. Dong, L. Tian, A. Hutson, and C. Xiong, "Parametric and non-parametric confidence intervals of the probability of identifying early disease stage given sensitivity to full disease and specificity with three ordinal diagnostic groups," *Statistics in medicine*, vol. 30, no. 30, pp. 3532–3545, 2011.

[20] T. Dong and L. Tian, "Confidence interval estimation for sensitivity to the early diseased stage based on empirical likelihood," *Journal of biopharmaceutical statistics*, vol. 25, no. 6, pp. 1215–1233, 2015.

[21] C. B. Begg and R. A. Greenes, "Assessment of diagnostic tests when disease verification is subject to selection bias," *Biometrics*, pp. 207–215, 1983.

[22] R. J. Little and D. B. Rubin, *Statistical analysis with missing data*. John Wiley & Sons, 2019, vol. 793.

[23] T. A. Alonzo and M. S. Pepe, "Assessing accuracy of a continuous screening test in the presence of verification bias," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 54, no. 1, pp. 173–190, 2005.

[24] H. He, J. M. Lyness, and M. P. McDermott, "Direct estimation of the area under the receiver operating characteristic curve in the presence of verification bias," *Statistics in medicine*, vol. 28, no. 3, pp. 361–376, 2009.

[25] G. Adimari and M. Chiogna, "Nonparametric verification bias-corrected inference for the area under the roc curve of a continuous-scale diagnostic test," *Statistics and its Interface*, vol. 10, no. 4, pp. 629–641, 2017.

[26] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap*. CRC press, 1994.

[27] A. Agresti and B. A. Coull, "Approximate is better than "exact" for interval estimation of binomial proportions," *The American Statistician*, vol. 52, no. 2, pp. 119–126, 1998.

[28] A. B. Owen, "Empirical likelihood ratio confidence intervals for a single functional," *Biometrika*, vol. 75, no. 2, pp. 237–249, 1988.

[29] W. Yu, Z. Zhao, and M. Zheng, "Empirical likelihood methods based on influence functions," *Statistics and Its Interface*, vol. 5, no. 3, pp. 355–366, 2012.

[30] B. Clarke and A. Yuan, "Reference priors for empirical likelihoods," in *Frontiers of Statistical Decision Making and Bayesian Analysis: In honor of James O. Berger*, M.-H. Chen, P. Müller, D. Sun, K. Ye, and D. K. Dey, Eds. Springer, 2010, pp. 56–68.

[31] J. M. Bernardo, "Reference posterior distributions for bayesian inference," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 41, no. 2, pp. 113–147, 1979.

[32] J. O. Berger and J. M. Bernardo, "On the development of reference priors," *Bayesian statistics*, vol. 4, no. 4, pp. 35–60, 1992.

[33] J. O. Berger, J. M. Bernardo, D. Sun *et al.*, "The formal definition of reference priors," *The Annals of Statistics*, vol. 37, no. 2, pp. 905–938, 2009.

[34] J. Rao and C. Wu, "Bayesian pseudo-empirical-likelihood intervals for complex surveys," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 72, no. 4, pp. 533–544, 2010.

[35] B.-Y. Jing, J. Yuan, and W. Zhou, "Jackknife empirical likelihood," *Journal of the American Statistical Association*, vol. 104, no. 487, pp. 1224–1232, 2009.

[36] M. Sullivan Pepe and T. Cai, "The analysis of placement values for evaluating discriminatory measures," *Biometrics*, vol. 60, no. 2, pp. 528–535, 2004.

[37] J. K. Ghosh, "A new proof of the bahadur representation of quantiles and an application," *The Annals of Mathematical Statistics*, pp. 1957–1961, 1971.

[38] B. W. Silverman, *Density estimation for statistics and data analysis*. CRC press, 1986.

[39] K. H. Zou, W. Hall, and D. E. Shapiro, "Smooth non-parametric receiver operating characteristic (roc) curves for continuous diagnostic tests," *Statistics in Medicine*, vol. 16, no. 19, pp. 2143–2156, 1997.

[40] B. W. Silverman, "Weak and strong uniform consistency of the kernel estimate of a density and its derivatives," *The Annals of Statistics*, pp. 177–184, 1978.

[41] A. Vexler and J. Yu, *Empirical Likelihood Methods in Biomedicine and Health*. CRC Press, 2018.

[42] A. Association, "2019 alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 15, no. 3, pp. 321–387, 2019.

[43] K. B. Smith, P. Kang, M. N. Sabbagh, A. D. N. Initiative *et al.*, "The effect of statins on rate of cognitive decline in mild cognitive impairment," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 3, no. 2, pp. 149–156, 2017.

[44] M. N. Samtani, N. Raghavan, G. Novak, P. Nandy, and V. A. Narayan, "Disease progression model for clinical dementia rating–sum of boxes in mild cognitive impairment and alzheimer's subjects from the alzheimer's disease neuroimaging initiative," *Neuropsychiatric disease and treatment*, vol. 10, p. 929, 2014.

[45] B. K. Scurfield, "Multiple-event forced-choice tasks in the theory of signal detectability," *Journal of Mathematical Psychology*, vol. 40, no. 3, pp. 253–269, 1996.

[46] D. Mossman, "Three-way rocs," *Medical Decision Making*, vol. 19, no. 1, pp. 78–89, 1999.

[47] P. S. Heckerling, "Parametric three-way receiver operating characteristic surface analysis using mathematica," *Medical Decision Making*, vol. 21, no. 5, pp. 409–417, 2001.

[48] C. T. Nakas and C. T. Yiannoutsos, "Ordered multiple-class roc analysis with continuous measurements," *Statistics in medicine*, vol. 23, no. 22, pp. 3437–3449, 2004.

[49] C. Xiong, G. van Belle, J. P. Miller, and J. C. Morris, "Measuring and estimating diagnostic accuracy when there are three ordinal diagnostic groups," *Statistics in Medicine*, vol. 25, no. 7, pp. 1251–1273, 2006.

[50] J. Li and X.-H. Zhou, "Nonparametric and semiparametric estimation of the three way receiver operating characteristic surface," *Journal of Statistical Planning and Inference*, vol. 139, no. 12, pp. 4133–4142, 2009.

[51] M. P. Wand and M. C. Jones, *Kernel smoothing*. Crc Press, 1994.

[52] S. Gao, S. L. Hui, K. S. Hall, and H. C. Hendrie, "Estimating disease prevalence from two-phase surveys with non-response at the second phase," *Statistics in medicine*, vol. 19, no. 16, pp. 2101–2114, 2000.

[53] D. Clayton, D. Spiegelhalter, G. Dunn, and A. Pickles, "Analysis of longitudinal binary data from multiphase sampling," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 60, no. 1, pp. 71–87, 1998.

[54] T. A. Alonzo, M. S. Pepe, and T. Lumley, "Estimating disease prevalence in two-phase studies," *Biostatistics*, vol. 4, no. 2, pp. 313–326, 2003.

[55] D. G. Horvitz and D. J. Thompson, "A generalization of sampling without replacement from a finite universe," *Journal of the American statistical Association*, vol. 47, no. 260, pp. 663–685, 1952.

[56] A. M. Korver, R. J. Smith, G. Van Camp, M. R. Schleiss, M. A. Bitner-Glindzicz, L. R. Lustig, S.-i. Usami, and A. N. Boudewyns, "Congenital hearing loss," *Nature reviews Disease primers*, vol. 3, no. 1, pp. 1–17, 2017.

[57] D. Liu and X.-H. Zhou, "A model for adjusting for nonignorable verification bias in estimation of the roc curve and its area with likelihood-based approach," *Biometrics*, vol. 66, no. 4, pp. 1119–1128, 2010.

[58] A. Owen, "Empirical likelihood ratio confidence regions," *The Annals of Statistics*, vol. 18, no. 1, pp. 90–120, 1990.

Appendix A

PROOFS IN PART 2

Proof of Proposition 2.1.

From (2.6), we only need to prove that

$$\frac{1}{\sigma\sqrt{m+n}}\sum_{k=1}^{m+n} W_k(\theta,p) \xrightarrow{d} N(0,1). \tag{A.1}$$

From (2.7), we have that

$$\frac{1}{\sqrt{m+n}}\sum_{k=1}^{m+n} W_k(\theta,p) = \sqrt{m+n}\Big\{\frac{1}{n}\sum_{j=1}^{n}[I(Y_j>\eta)-\theta] + \frac{g(\eta)}{f(\eta)}\frac{1}{m}\sum_{i=1}^{m}[I(X_i\le\eta)-p]\Big\}.$$

Since $I(Y_j > \eta)$'s $\overset{iid}{\sim}$ $Binomial(1,\theta)$, and $I(X_i \le \eta)$'s $\overset{iid}{\sim}$ $Binomial(1,p)$, by Central Limit Theorem, we have that

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}[I(Y_j>\eta)-\theta] \xrightarrow{d} N(0,\theta(1-\theta)),$$
$$\frac{1}{\sqrt{m}}\sum_{i=1}^{m}[I(X_i\le\eta)-p] \xrightarrow{d} N(0,p(1-p)).$$

Hence, (A.1) and Proposition 2.1 follows immediately from (2.6) and the independence of Y_j's and X_i's.

We need the following lemma 2.1 for the proof of Theorem 2.1.

Lemma 2.1. Under the conditions in Theorem 2.1, we have that

(i) $\frac{1}{\sigma\sqrt{m+n}}\sum_{k=1}^{m+n}\hat{W}_k(\theta,p) \xrightarrow{d} N(0,1)$.

(ii) $\frac{1}{m+n}\sum_{k=1}^{m+n}\hat{W}_k^2(\theta,p) \xrightarrow{p} \sigma^2$.

Proof.

(i) From (A.1), we only need to prove that

$$\frac{1}{\sqrt{m+n}}\sum_{k=1}^{m+n}\hat{W}_k(\theta,p)=\frac{1}{\sqrt{m+n}}\sum_{k=1}^{m+n}W_k(\theta,p)+o_p(1).$$

We have the following decomposition:

$$\begin{aligned}\frac{1}{\sqrt{m+n}}\sum_{k=1}^{m+n}\hat{W}_k(\theta,p)=&\frac{1}{\sqrt{m+n}}\sum_{k=1}^{m+n}W_k(\theta,p)+\sqrt{m+n}\Big\{\frac{1}{n}\sum_{j=1}^{n}[I(Y_j>\hat{\eta})-I(Y_j>\eta)]\Big\}\\&+\sqrt{m+n}\Big\{\frac{\hat{g}(\hat{\eta})}{\hat{f}(\hat{\eta})}\frac{1}{m}\sum_{i=1}^{m}[I(X_i\le\hat{\eta})-p]-\frac{g(\eta)}{f(\eta)}\frac{1}{m}\sum_{i=1}^{m}[I(X_i\le\eta)-p]\Big\}.\end{aligned}$$

Using the Bahadur representation of the sample quantile $\hat{\eta}$,

$$\hat{\eta}-\eta=\frac{p-\frac{1}{m}\sum_{i=1}^{m}I(X_i\le\eta)}{f(\eta)}+o_p(m^{-\frac{1}{2}}),$$

we get that

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n}[I(Y_j>\hat{\eta})-I(Y_j>\eta)]=&\int[I(y\le\eta)-I(y\le\hat{\eta})]d\hat{G}(y)\\=&\int[I(y\le\eta)-I(y\le\hat{\eta})]dG(y)+o_p(n^{-1/2})\\=&g(\eta)(\eta-\hat{\eta})+o_p(m^{-1/2}+n^{-1/2})\\=&\frac{g(\eta)}{f(\eta)}\frac{1}{m}\sum_{i=1}^{m}[I(X_i\le\eta)-p]+o_p((m+n)^{-1/2}),\end{aligned}\tag{A.2}$$

and

$$\begin{aligned}
\frac{1}{m}\sum_{i=1}^{m}[I(X_i \le \hat{\eta}) - p] =& \frac{1}{m}\sum_{j=1}^{n}[I(X_i \le \hat{\eta}) - I(X_i \le \eta)] + \frac{1}{m}\sum_{j=1}^{n}[I(X_i \le \eta) - p] \\
=& \int [I(x \le \hat{\eta}) - I(x \le \eta)]d\hat{F}(x) + \frac{1}{m}\sum_{j=1}^{n}[I(X_i \le \eta) - p] \\
=& \int [I(x \le \hat{\eta}) - I(x \le \eta)]dF(x) + o_p(m^{-1/2}) + \frac{1}{m}\sum_{j=1}^{n}[I(X_i \le \eta) - p] \\
=& f(\eta)(\hat{\eta} - \eta) + \frac{1}{m}\sum_{j=1}^{n}[I(X_i \le \eta) - p] + o_p(m^{-1/2}) \\
=& o_p(m^{-1/2}).
\end{aligned} \tag{A.3}$$

Therefore,

$$\begin{aligned}
\frac{1}{\sqrt{m+n}}\sum_{k=1}^{m+n}\hat{W}_k(\theta, p) =& \frac{1}{\sqrt{m+n}}\sum_{k=1}^{m+n}W_k(\theta, p) + \sqrt{m+n}\Big\{\frac{g(\eta)}{f(\eta)}\frac{1}{m}\sum_{i=1}^{m}[I(X_i \le \eta) - p]\Big\} \\
&+ \sqrt{m+n}\Big\{\frac{\hat{g}(\hat{\eta})}{\hat{f}(\hat{\eta})}\frac{1}{m}\sum_{i=1}^{m}[I(X_i \le \hat{\eta}) - p] - \frac{g(\eta)}{f(\eta)}\frac{1}{m}\sum_{i=1}^{m}[I(X_i \le \eta) - p]\Big\} + o_p(1) \\
=& \frac{1}{\sqrt{m+n}}\sum_{k=1}^{m+n}W_k(\theta, p) + \sqrt{m+n}\frac{\hat{g}(\hat{\eta})}{\hat{f}(\hat{\eta})}\frac{1}{m}\sum_{i=1}^{m}[I(X_i \le \hat{\eta}) - p] + o_p(1) \\
=& \frac{1}{\sqrt{m+n}}\sum_{k=1}^{m+n}W_k(\theta, p) + o_p(1).
\end{aligned}$$

The last equality holds by the uniform consistency of the density estimates $\hat{g}$ and $\hat{f}$ [40], and $\frac{g(\eta)}{f(\eta)} = O(1)$. Lemma 2.1 (i) is thus proved.

(ii) Since

$$\begin{aligned}\frac{1}{m+n}\sum_{k=1}^{m+n} W_k^2(\theta,p) &= \frac{m+n}{n^2}\sum_{j=1}^{n}[(I(Y_j \geq \eta)-\theta)]^2 + \frac{m+n}{m^2}\frac{g^2(\eta)}{f^2(\eta)}\sum_{i=1}^{m}[I(X_i \leq \eta)-p]^2 \\ &= (1+\rho)E[I(Y_j \geq \eta)-\theta]^2 + (1+\rho^{-1})\frac{g^2(\eta)}{f^2(\eta)}E[I(X_i \leq \eta)-p]^2 + o_p(1) \\ &= (1+\rho)\theta(1-\theta) + (1+\rho^{-1})p(1-p)\frac{g^2(\eta)}{f^2(\eta)} + o_p(1) = \sigma^2 + o_p(1),\end{aligned}$$

we only need to prove that

$$\frac{1}{m+n}\sum_{k=1}^{m+n}\hat{W}_k^2(\theta,p) = \frac{1}{m+n}\sum_{k=1}^{m+n} W_k^2(\theta,p) + o_p(1).$$

Under the assumptions in Theorem 2.1, using the uniform consistency of the density estimate $\hat{f}$ [40] and the strong consistency of the sample quantile $\hat{\eta}$, we get that

$$\begin{aligned}\left|\hat{f}(\hat{\eta}) - f(\eta)\right| &\leq \left|\hat{f}(\hat{\eta}) - f(\hat{\eta})\right| + \left|f(\hat{\eta}) - f(\eta)\right| \\ &\leq \sup_x \left|\hat{f}(x) - f(x)\right| + o_p(1) = o_p(1).\end{aligned}$$

So, $\hat{f}(\hat{\eta}) = f(\eta) + o_p(1)$. Similarly, we have $\hat{g}(\hat{\eta}) = g(\eta) + o_p(1)$. By Slutsky's Theorem, we have that $\frac{\hat{g}^2(\hat{\eta})}{\hat{f}^2(\hat{\eta})} = \frac{g^2(\eta)}{f^2(\eta)} + o_p(1)$.

From (A.2), it follows that

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n}[I(Y_j \geq \eta) - I(Y_j \geq \hat{\eta})] &= -\frac{g(\eta)}{f(\eta)}[\frac{1}{m}\sum_{i=1}^{m} I(X_i \leq \eta) - p] + o_p((m+n)^{-1/2}) \\ &= O_p(m^{-1/2}) + o_p((m+n)^{-1/2}) = O_p((m+n)^{-1/2}).\end{aligned}$$

Similarly, we have

$$\frac{1}{m}\sum_{i=1}^{m}[I(X_i \leq \eta) - I(X_i \leq \hat{\eta})] = O_p((m+n)^{-1/2}).$$

Therefore,

$$\begin{aligned}
&\Big|\frac{1}{m+n}\sum_{k=1}^{m+n} W_k^2(\theta,p) - \frac{1}{m+n}\sum_{k=1}^{m+n}\hat{W}_k^2(\theta,p)\Big| \\
&= (m+n)\Big|\frac{1-2\theta}{n^2}\sum_{j=1}^{n}[I(Y_j \geq \eta) - I(Y_j \geq \hat{\eta})] + \frac{p^2}{m}\big[\frac{g^2(\eta)}{f^2(\eta)} - \frac{\hat{g}^2(\hat{\eta})}{\hat{f}^2(\hat{\eta})}\big] \\
&\quad + \frac{1-2p}{m^2}\sum_{i=1}^{m}\big[\big(\frac{g^2(\eta)}{f^2(\eta)} - \frac{\hat{g}^2(\hat{\eta})}{\hat{f}^2(\hat{\eta})}\big)I(X_i \leq \eta) + \frac{\hat{g}^2(\hat{\eta})}{\hat{f}^2(\hat{\eta})}(I(X_i \leq \eta) - I(X_i \leq \hat{\eta}))\big]\Big|. \\
&\leq o_p(1),
\end{aligned}$$

and Lemma 2.1 (ii) is proved.

Proof of Theorem 2.1:

By the definition of $\hat{W}_k(\theta,p)$, we have

$$\max_k |\hat{W}_k(\theta,p)| \leq 2\max\{\frac{m+n}{m}\frac{\hat{g}(\hat{\eta})}{\hat{f}(\hat{\eta})}, \frac{m+n}{n}\} = O_p(1).$$

Moreover, from Lemma (ii), it follows that

$$\frac{1}{m+n}\sum_{k=1}^{m+n}|\hat{W}_k(\theta,p)|^3 \leq \max_k |\hat{W}_k(\theta,p)|\frac{1}{m+n}\sum_{k=1}^{m+n}|\hat{W}_k(\theta,p)|^2 = O_p(1). \tag{A.4}$$

Using arguments similar to Owen [58], we can prove that

$$|\lambda| = O_p((m+n)^{-1/2}). \tag{A.5}$$

Hence, we have

$$\max_k |\lambda\hat{W}_k(\theta,p)| = O_p((m+n)^{-1/2}). \tag{A.6}$$

Recall (2.10)

$$\begin{aligned} 0 &= \frac{1}{m+n}\sum_{k=1}^{m+n}\frac{\hat{W}_k(\theta,p)}{1+\lambda\hat{W}_k(\theta,p)} \\ &= \frac{1}{m+n}\sum_{k=1}^{m+n}\hat{W}_k(\theta,p)[1-\lambda\hat{W}_k(\theta,p)+\frac{(\lambda\hat{W}_k(\theta,p))^2}{1+\lambda\hat{W}_k(\theta,p)}] \\ &= \frac{1}{m+n}\sum_{k=1}^{m+n}\hat{W}_k(\theta,p) - \frac{1}{m+n}\lambda\sum_{k=1}^{m+n}\hat{W}_k^2(\theta,p) + \frac{1}{m+n}\sum_{k=1}^{m+n}\frac{\hat{W}_k(\theta,p)(\lambda\hat{W}_k(\theta,p))^2}{1+\lambda\hat{W}_k(\theta,p)}. \end{aligned} \tag{A.7}$$

From Lemma 2.1 (ii), (A.4), (A.5) and (A.6), the final term in (A.7) is bounded by

$$|\lambda|^2\frac{1}{m+n}\sum_{k=1}^{m+n}|\hat{W}_k(\theta,p)|^3\max|(1+\lambda\hat{W}_k(\theta,p))^{-1}| \ = O_p((m+n)^{-1})O_p(1)O_p(1) = O_p((m+n)^{-1}),$$

which implies that

$$\lambda = \frac{\sum_{k=1}^{m+n}\hat{W}_k(\theta,p)}{\sum_{k=1}^{m+n}\hat{W}_k^2(\theta,p)} + O_p((m+n)^{-1}). \tag{A.8}$$

Further, multiplying both side of Equation (2.10) by λ, we get that

$$\begin{aligned} 0 &= \frac{1}{m+n}\sum_{k=1}^{m+n}\frac{\lambda\hat{W}_k(\theta,p)}{1+\lambda\hat{W}_k(\theta,p)} \\ &= \frac{1}{m+n}\sum_{k=1}^{m+n}\lambda\hat{W}_k(\theta,p) - \frac{1}{m+n}\sum_{k=1}^{m+n}\lambda^2\hat{W}_k^2(\theta,p) + \frac{1}{m+n}\sum_{k=1}^{m+n}\frac{(\lambda\hat{W}_k(\theta,p))^3}{1+\lambda\hat{W}_k(\theta,p)}. \end{aligned} \tag{A.9}$$

Similarly, the final term in (A.9) is bounded by

$$|\lambda|^3\frac{1}{m+n}\sum_{k=1}^{m+n}|\hat{W}_k(\theta,p)|^3\max|(1+\lambda\hat{W}_k(\theta,p))^{-1}| = O_p((m+n)^{-3/2})O_p(1)O_p(1) = O_p((m+n)^{-3/2}).$$

Hence, we have

$$\sum_{k=1}^{m+n}\lambda\hat{W}_k(\theta,p) = \sum_{k=1}^{m+n}\lambda^2\hat{W}_k^2(\theta,p) + O_p((m+n)^{-1/2}). \tag{A.10}$$

By Taylor's expansion of Equation (2.11) and using (A.8), (A.10), and Lemma, we have that

$$\begin{aligned}
-2l_{IF}(\theta,p) &= 2\sum_{k=1}^{m+n}\log\{1+\lambda\hat{W}_k(\theta,p)\} \\
&= 2\lambda\sum_{k=1}^{m+n}\hat{W}_k(\theta,p) - \lambda^2\sum_{k=1}^{m+n}\hat{W}_k^2(\theta,p) + r_n \\
&= \frac{[\frac{1}{(m+n)^{1/2}}\sum_{k=1}^{m+n}\hat{W}_k(\theta,p)]^2}{\frac{1}{m+n}\sum_{k=1}^{m+n}\hat{W}_k^2(\theta,p)} + \sum_{k=1}^{m+n}\hat{W}_k(\theta,p)O_p((m+n)^{-1}) + O_p((m+n)^{-1/2}) + r_n,
\end{aligned}$$

where

$$|r_n| \le C\sum_{k=1}^{m+n}|\lambda\hat{W}_k(\theta,p)|^3 \le C|\lambda|^3\sum_{k=1}^{m+n}|\hat{W}_k(\theta,p)|^3 = O_p((m+n)^{-3/2})(m+n)O_p(1) = O_p((m+n)^{-1/2}).$$

From Lemma, it follows that

$$\sum_{k=1}^{m+n}\hat{W}_k(\theta,p)O_p((m+n)^{-1}) = O_p((m+n)^{1/2})O_p((m+n)^{-1}) = O_p((m+n)^{-1/2}),$$

and

$$-2l_{IF}(\theta,p) = \frac{[\frac{1}{(m+n)^{1/2}}\sum_{k=1}^{m+n}\hat{W}_k(\theta,p)]^2}{\frac{1}{m+n}\sum_{k=1}^{m+n}\hat{W}_k^2(\theta,p)} + o_p(1) \xrightarrow{d} \chi_1^2.$$

Proof of Proposition 2.2:

We first briefly introduce Clarke and Yuan's approach [30]. Define the outer product matrix $\Omega = E[g(Z_j,\theta)g'(Z_j,\theta)]$, Jacobian matrix $D(\theta) = E[\partial g(Z_j,\theta)/\partial\theta]$ and the matrix $\Lambda(\theta) = D'(\theta)\Omega^{-1}(\theta)D(\theta)$, where $g(Z_j,\theta)$ is an estimating function.

For our Bayesian hybrid EL approach, $g(Z_j,\theta) = W_{Hj}(\theta,p) = I(F(Y_j)\ge p) - \theta$, and

$$\Omega(\theta) = E[g(Z_j,\theta)]^2 = E[I(F(Y_j)\ge p) - \theta]^2 = \theta(1-\theta).$$

Thus we have

$$\Lambda(\theta) = D'(\theta)\Omega^{-1}(\theta)D(\theta) = \frac{1}{\theta(1-\theta)}.$$

So the reference prior for the hybrid EL under the relative entropy is

$$\pi_{H,1}(\theta) \propto |\Lambda^{-1}(\theta)|^{1/2} = \sqrt{\theta(1-\theta)},$$

i.e.,

$$\pi_{H,1}(\theta) = \beta(\frac{3}{2}, \frac{3}{2}),$$

and the reference prior for the hybrid EL under Hellinger distance is

$$\pi_{H,2}(\theta) \propto |\Lambda(\theta)|^{1/2} = \frac{1}{\sqrt{\theta(1-\theta)}},$$

i.e.,

$$\pi_{H,2}(\theta) = \beta(\frac{1}{2}, \frac{1}{2}).$$

Appendix B

PROOFS IN PART 3

Proof of Proposition 3.1.

From (3.10), we only need to prove that

$$\frac{1}{\sigma\sqrt{N}}\sum_{l=1}^{N} W_l(P_1,P_2,P_3) \xrightarrow{d} N(0,1). \tag{B.1}$$

From (3.11), we have that

$$\frac{1}{\sqrt{N}}\sum_{l=1}^{N} W_l(P_1,P_2,P_3) = \sqrt{N}\Big\{\frac{1}{n_1}\frac{f_2(c_1)}{f_1(c_1)}\sum_{j=1}^{n_1}[I(Y_{1,j}\le c_1)-P_1] + \frac{1}{n_2}\sum_{j=1}^{n_2}[I(c_1<Y_{2,j}\le c_2)-P_2] + \frac{1}{n_3}\frac{f_2(c_2)}{f_3(c_2)}\sum_{j=1}^{n_3}[I(Y_{3,j}>c_2)-P_3]\Big\}.$$

Since $I(Y_{1,j}\le c_1)$'s $\overset{iid}{\sim} Binomial(1,P_1)$, $I(c_1<Y_{2,j}\le c_2)$'s $\overset{iid}{\sim} Binomial(1,P_2)$, and $I(Y_{3,j}>c_2)$'s $\overset{iid}{\sim} Binomial(1,P_3)$, by Central Limit Theorem, we have that

$$\frac{1}{\sqrt{n_1}}\sum_{j=1}^{n_1}[I(Y_{1,j}\le c_1)-P_1] \xrightarrow{d} N(0,P_1(1-P_1)),$$
$$\frac{1}{\sqrt{n_2}}\sum_{j=1}^{n_2}[I(c_1<Y_{2,j}\le c_2)-P_2] \xrightarrow{d} N(0,P_2(1-P_2)),$$
$$\frac{1}{\sqrt{n_3}}\sum_{i=1}^{n_3}[I(Y_{3,j}>c_2)-P_3] \xrightarrow{d} N(0,P_3(1-P_3)).$$

Hence, (B.1) and Proposition 3.1 follows immediately from (3.10) and the independence of $Y_{1,i}$'s, $Y_{2,j}$'s and $Y_{3,k}$'s.

We need the following lemma for the proof of Theorem 3.1.

Lemma 3.1. Under the conditions in Theorem 3.1, we have that

(i) $\frac{1}{\sigma\sqrt{N}}\sum_{l=1}^{N}\hat{W}_l(P_1,P_2,P_3) \xrightarrow{d} N(0,1)$.

(ii) $\frac{1}{N}\sum_{l=1}^{N}\hat{W}_l^2(P_1,P_2,P_3)\xrightarrow{p}\sigma^2$.

Proof.

(i) From (B.1), we only need to prove that

$$\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\hat{W}_l(P_1,P_2,P_3)=\frac{1}{\sqrt{N}}\sum_{l=1}^{N}W_l(P_1,P_2,P_3)+o_p(1).$$

We have the following decomposition:

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\hat{W}_l(1,P_2,P_3)\\
&\quad=\frac{1}{\sqrt{N}}\sum_{l=1}^{N}W_l(P_1,P_2,P_3)+\sqrt{N}\Big\{\frac{1}{n_2}\sum_{j=1}^{n_2}[I(\hat{c}_1<Y_{2,j}\le\hat{c}_2)-I(c_1<Y_{2,j}\le c_2)]\Big\}\\
&\qquad+\sqrt{N}\Big\{\frac{1}{n_1}\frac{\hat{f}_2(\hat{c}_1)}{\hat{f}_1(\hat{c}_1)}\sum_{j=1}^{n_1}[I(Y_{1,j}\le\hat{c}_1)-P_1]-\frac{1}{n_1}\frac{f_2(c_1)}{f_1(c_1)}\sum_{j=1}^{n_1}[I(Y_{1,j}\le c_1)-P_1]\Big\}\\
&\qquad+\sqrt{N}\Big\{\frac{1}{n_3}\frac{\hat{f}_2(\hat{c}_2)}{\hat{f}_3(\hat{c}_2)}\sum_{j=1}^{n_3}[I(Y_{3,j}>\hat{c}_2)-P_3]-\frac{1}{n_3}\frac{f_2(c_2)}{f_3(c_2)}\sum_{j=1}^{n_3}[I(Y_{3,j}>c_2)-P_3]\Big\}.
\end{aligned}$$

As we mentioned in Section 3.3, using the Bahadur representation of the sample quantiles $\hat{c}_1$ and $\hat{c}_2$[37],

$$\hat{c}_1-c_1=\frac{P_1-\frac{1}{n_1}\sum_{j=1}^{n_1}I(Y_{1,j}\le c_1)}{f_1(c_1)}+o_p(n_1^{-\frac{1}{2}}),$$

$$\hat{c}_2-c_2=\frac{\frac{1}{n_3}\sum_{j=1}^{n_3}I(Y_{3,j}>c_2)-P_3}{f_3(c_2)}+o_p(n_3^{-\frac{1}{2}}),$$

and Equation 3.8 and 3.9, we get that

$$\begin{aligned}
&\frac{1}{n_2}\sum_{j=1}^{n_2}[I(\hat{c}_1 < Y_{2,j} \le \hat{c}_2) - I(c_1 < Y_{2,j} \le c_2)] \\
&= \frac{1}{n_2}\sum_{j=1}^{n_2}[I(Y_{2,j} \le \hat{c}_2) - I(Y_{2,j} \le \hat{c}_1)] - \frac{1}{n_2}\sum_{j=1}^{n_2}[I(Y_{2,j} \le c_2) - I(Y_{2,j} \le c_1)] \\
&= \frac{1}{n_2}\sum_{j=1}^{n_2}[I(Y_{2,j} \le \hat{c}_2) - I(Y_{2,j} \le c_2)] - \frac{1}{n_2}\sum_{j=1}^{n_2}[I(Y_{2,j} \le \hat{c}_1) - I(Y_{2,j} \le c_1)] \\
&= [\hat{F}_2(\hat{c}_2) - \hat{F}_2(c_2)] - [\hat{F}_2(\hat{c}_1) - \hat{F}_2(c_1)] \\
&= \frac{1}{n_3}\frac{f_2(c_2)}{f_3(c_2)}\sum_{j=1}^{n_3}[I(Y_{3,j} > c_2) - P_3] + \frac{1}{n_1}\frac{f_2(c_1)}{f_1(c_1)}\sum_{j=1}^{n_1}[I(Y_{1,j} \le c_1) - P_1] + o_p(N^{-1/2}).
\end{aligned} \tag{B.2}$$

In addition, we have

$$\begin{aligned}
&\frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j} \le \hat{c}_1) - P_1] \\
&= \frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j} \le \hat{c}_1) - I(Y_{1,j} \le c_1)] + \frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j} \le c_1) - P_1] \\
&= \int [I(y \le \hat{c}_1) - I(y \le c_1)]d\hat{F}_1(y) + \frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j} \le c_1) - P_1] \\
&= \int [I(y \le \hat{c}_1) - I(y \le c_1)]dF_1(y) + o_p(n_1^{-1/2}) + \frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j} \le c_1) - P_1] \\
&= f_1(c_1)(\hat{c}_1 - c_1) + \frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j} \le c_1) - P_1] + o_p(n_1^{-1/2}) \\
&= o_p(n_1^{-1/2}),
\end{aligned} \tag{B.3}$$

and

$$\begin{aligned}
&\frac{1}{n_3}\sum_{j=1}^{n_3}[I(Y_{3,j} > \hat{c}_2) - P_3] \\
&= \frac{1}{n_3}\sum_{j=1}^{n_3}[I(Y_{3,j} > \hat{c}_2) - I(Y_{3,j} > c_2)] + \frac{1}{n_3}\sum_{j=1}^{n_3}[I(Y_{3,j} > c_2) - P_3] \\
&= -\int [I(y \le \hat{c}_2) - I(y \le c_2)]d\hat{F}_3(y) + \frac{1}{n_3}\sum_{j=1}^{n_3}[I(Y_{3,j} > c_2) - P_3] \\
&= -\int [I(y \le \hat{c}_2) - I(y \le c_2)]dF_3(y) + o_p(n_3^{-1/2}) + \frac{1}{n_3}\sum_{j=1}^{n_3}[I(Y_{3,j} > c_2) - P_3] \\
&= -f_3(c_2)(\hat{c}_2 - c_2) + \frac{1}{n_3}\sum_{j=1}^{n_3}[I(Y_{3,j} > c_2) - P_3] + o_p(n_3^{-1/2}) \\
&= o_p(n_3^{-1/2}),
\end{aligned} \tag{B.4}$$

Therefore,

$$\begin{aligned}
\frac{1}{\sqrt{N}}\sum_{l=1}^{N}\hat{W}_l(_1, P_2, P_3) &= \frac{1}{\sqrt{N}}\sum_{l=1}^{N} W_l(P_1, P_2, P_3) \\
&+\frac{\sqrt{N}}{n_1}\frac{\hat{f}_2(\hat{c}_1)}{\hat{f}_1(\hat{c}_1)}\sum_{j=1}^{n_1}[I(Y_{1,j} \le \hat{c}_1) - P_1] + \frac{\sqrt{N}}{n_3}\frac{\hat{f}_2(\hat{c}_2)}{\hat{f}_3(\hat{c}_2)}\sum_{j=1}^{n_3}[I(Y_{3,j} > \hat{c}_2) - P_3] + o_p(1) \\
&= \frac{1}{\sqrt{N}}\sum_{l=1}^{N} W_l(P_1, P_2, P_3) + o_p(1).
\end{aligned}$$

The last equality holds by the uniform consistency of the density estimates $\hat{f}_1$, $\hat{f}_2$ and $\hat{f}_3$ [40], and $\frac{f_2(c_1)}{f_1(c_1)} = O(1)$, $\frac{f_2(c_2)}{f_3(c_2)} = O(1)$. Lemma 3.1 (i) is thus proved.

(ii) Since

$$
\begin{aligned}
&\frac{1}{N}\sum_{l=1}^{N} W_l^2(P_1,P_2,P_3)\\
&= \frac{N}{n_1^2}\frac{f_2^2(c_1)}{f_1^2(c_1)}\sum_{j=1}^{n_1}[I(Y_{1,j}\le c_1)-P_1]^2 + \frac{N}{n_2^2}\sum_{j=1}^{n_2}[I(c_1<Y_{2,j}\le c_2)-P_2]^2\\
&\quad + \frac{N}{n_3^2}\frac{f_2^2(c_2)}{f_3^2(c_2)}\sum_{j=1}^{n_3}[I(Y_{3,j}>c_2)-P_3]^2\\
&= (1+\rho_1^{-1}+\rho_3^{-1})\frac{f_2^2(c_1)}{f_1^2(c_1)}E[I(Y_{1,j}\le c_1)-P_1]^2 + (1+\rho_1+\rho_2)E[I(c_1<Y_{2,j}\le c_2)-P_2]^2\\
&\quad + (1+\rho_2^{-1}+\rho_3)\frac{f_2^2(c_2)}{f_3^2(c_2)}E[I(Y_{3,j}>c_2)-P_3]^2 + o_p(1)\\
&= (1+\rho_1^{-1}+\rho_3^{-1})P_1(1-P_1)\frac{f_2^2(c_1)}{f_1^2(c_1)} + (1+\rho_1+\rho_2)P_2(1-P_2)\\
&\quad + (1+\rho_2^{-1}+\rho_3)\frac{f_2^2(c_2)}{f_3^2(c_2)}P_3(1-P_3) + o_p(1) = \sigma^2 + o_p(1),
\end{aligned}
$$

we only need to prove that

$$
\frac{1}{N}\sum_{l=1}^{N}\hat{W}_l^2(P_1,P_2,P_3) = \frac{1}{N}\sum_{l=1}^{N} W_l^2(P_1,P_2,P_3) + o_p(1).
$$

Under the assumptions in Theorem 3.1, using the uniform consistency of the density estimate $\hat{f}_1$ [40] and the strong consistency of the sample quantile $\hat{c}_1$, we get that

$$
\begin{aligned}
\left|\hat{f}_1(\hat{c}_1)-f_1(c_1)\right| &\le \left|\hat{f}_1(\hat{c}_1)-f_1(\hat{c}_1)\right| + \left|f_1(\hat{c}_1)-f_1(c_1)\right|\\
&\le \sup_x\left|\hat{f}_1(x)-f_1(x)\right| + o_p(1) = o_p(1).
\end{aligned}
$$

So, $\hat{f}_1(\hat{c}_1) = f_1(c_1)+o_p(1)$. Similarly, we have $\hat{f}_3(\hat{c}_2) = f_3(c_2)+o_p(1)$, $\hat{f}_2(\hat{c}_1) = f_2(c_1)+o_p(1)$, and $\hat{f}_2(\hat{c}_2) = f_2(c_2)+o_p(1)$. By Slutsky's Theorem, we have that $\frac{\hat{f}_2^2(c_1)}{\hat{f}_1^2(c_1)} = \frac{f_2^2(c_1)}{f_1^2(c_1)} + o_p(1)$ and $\frac{\hat{f}_2^2(c_2)}{\hat{f}_3^2(c_2)} = \frac{f_2^2(c_2)}{f_3^2(c_2)} + o_p(1)$.

From Equation (B.2) and CLT, it follows that

$$\begin{aligned}
&\frac{1}{n_2}\sum_{j=1}^{n_2}\Big[I(c_1 < Y_{2,j} \le c_2) - I(\hat{c}_1 < Y_{2,j} \le \hat{c}_2)\Big] \\
&= -\frac{1}{n_3}\frac{f_2(c_2)}{f_3(c_2)}\sum_{j=1}^{n_3}[I(Y_{3,j} > c_2) - P_3] - \frac{1}{n_1}\frac{f_2(c_1)}{f_1(c_1)}\sum_{j=1}^{n_1}[I(Y_{1,j} \le c_1) - P_1] + o_p(N^{-1/2}) \\
&= O_p(n_1^{-1/2}) + O_p(n_3^{-1/2}) + o_p(N^{-1/2}) \\
&= O_p(N^{-1/2}).
\end{aligned}$$

Similarly, from Equation (3.8) and Bahadur representation of the sample quantile, we have that

$$\begin{aligned}
\frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j} \le c_1) - I(Y_{1,j} \le \hat{c}_1)] &= f_1(c_1)(c_1 - \hat{c}_1) + o_p(n_1^{-1/2}) \\
&= \frac{1}{n_1}\sum_{j=1}^{n_1}[I(Y_{1,j} \le c_1) - P_1] + o_p(n_1^{-1/2}) \\
&= O_p(n_1^{-1/2})
\end{aligned}$$

and

$$\frac{1}{n_3}\sum_{j=1}^{n_3}[I(Y_{3,j} > c_2) - I(Y_{3,j} > \hat{c}_2)] = O_p(n_3^{-1/2}).$$

Therefore,

$$
\begin{aligned}
&\Big|\frac{1}{N}\sum_{l=1}^{N} W_l^2(P_1,P_2,P_3) - \frac{1}{N}\sum_{l=1}^{N}\hat{W}_l^2(P_1,P_2,P_3)\Big| \\
&\quad = (N)\Big|\frac{1-2P_2}{n_2^2}\sum_{j=1}^{n_2}\Big[I(c_1 < Y_{2,j} \le c_2) - I(\hat{c}_1 < Y_{2,j} \le \hat{c}_2)\Big] + \frac{P_1^2}{n^2}\Big[\frac{f_2^2(c_1)}{f_1^2(c_1)} - \frac{\hat{f}_2^2(c_1)}{\hat{f}_1^2(c_1)}\Big] \\
&\qquad + \frac{1-2P_1}{n_1^2}\sum_{j=1}^{n_1}\Big\{\big(\frac{f_2^2(c_1)}{f_1^2(c_1)} - \frac{\hat{f}_2^2(c_1)}{\hat{f}_1^2(c_1)}\big)I(Y_{1,j} \le c_1) + \frac{\hat{f}_2^2(c_1)}{\hat{f}_1^2(c_1)}\big[I(Y_{1,j} \le c_1) - I(Y_{1,j} \le \hat{c}_1)\big]\Big\} \\
&\qquad + \frac{1-2P_3}{n_3^2}\sum_{j=1}^{n_3}\Big\{\big(\frac{f_2^2(c_2)}{f_3^2(c_2)} - \frac{\hat{f}_2^2(c_2)}{\hat{f}_3^2(c_2)}\big)I(Y_{3,j} > c_2) + \frac{\hat{f}_2^2(c_2)}{\hat{f}_3^2(c_2)}\big[I(Y_{3,j} > c_2) - I(Y_{3,j} > \hat{c}_2)\big]\Big\} \\
&\qquad + \frac{P_3^2}{n^3}\Big[\frac{f_2^2(c_2)}{f_3^2(c_2)} - \frac{\hat{f}_2^2(c_2)}{\hat{f}_3^2(c_2)}\Big]\Big| \\
&\quad \le o_p(1),
\end{aligned}
$$

and Lemma 3.1 (ii) is proved.

Proof of Theorem 3.1:

By the definition of $\hat{W}_l(P_1,P_2,P_3)$, we have

$$\max_l |\hat{W}_l(P_1,P_2,P_3)| \le 2\max\{\frac{N}{n_1}\frac{\hat{f}_2(\hat{c}_1)}{\hat{f}_1(\hat{c}_1)}, \frac{N}{n_2}, \frac{N}{n_3}\frac{\hat{f}_2(\hat{c}_2)}{\hat{f}_3(\hat{c}_2)}\} = O_p(1).$$

Moreover, from Lemma 3.1 (ii), it follows that

$$\frac{1}{N}\sum_{l=1}^{N}|\hat{W}_l(P_1,P_2,P_3)|^3 \le \max_l |\hat{W}_l(P_1,P_2,P_3)|\frac{1}{N}\sum_{l=1}^{N}|\hat{W}_l(P_1,P_2,P_3)|^2 = O_p(1). \tag{B.5}$$

Using arguments similar to Owen [58], we get

$$|\lambda| = O_p(N^{-1/2}). \tag{B.6}$$

Hence, we have

$$\max_l |\lambda\hat{W}_l(P_1,P_2,P_3)| = O_p(N^{-1/2}). \tag{B.7}$$

Recall (3.15)

$$\begin{aligned} 0 &= \frac{1}{N}\sum_{l=1}^{N}\frac{\hat{W}_l(P_1,P_2,P_3)}{1+\lambda\hat{W}_l(P_1,P_2,P_3)} \\ &= \frac{1}{N}\sum_{l=1}^{N}\hat{W}_l(P_1,P_2,P_3)\big[1-\lambda\hat{W}_l(P_1,P_2,P_3)+\frac{(\lambda\hat{W}_l(P_1,P_2,P_3))^2}{1+\lambda\hat{W}_l(P_1,P_2,P_3)}\big] \\ &= \frac{1}{N}\sum_{l=1}^{N}\hat{W}_l(P_1,P_2,P_3)-\frac{1}{N}\lambda\sum_{l=1}^{N}\hat{W}_l^2(P_1,P_2,P_3) \\ &\quad +\frac{1}{N}\sum_{l=1}^{N}\frac{\hat{W}_l(P_1,P_2,P_3)(\lambda\hat{W}_l(P_1,P_2,P_3))^2}{1+\lambda\hat{W}_l(P_1,P_2,P_3)}. \end{aligned} \tag{B.8}$$

From Lemma 3.1 (ii), (A.4), (A.5) and (A.6), it follows that the final term in (A.7) is bounded by

$$|\lambda|^2\frac{1}{N}\sum_{l=1}^{N}|\hat{W}_l(P_1,P_2,P_3)|^3\max|(1+\lambda\hat{W}_l(P_1,P_2,P_3))^{-1}|=O_p(N^{-1}),$$

which implies that

$$\lambda=\frac{\sum_{l=1}^{N}\hat{W}_l(P_1,P_2,P_3)}{\sum_{l=1}^{N}\hat{W}_l^2(P_1,P_2,P_3)}+O_p(N^{-1}). \tag{B.9}$$

Further, multiplying both side of Equation (3.15) by λ, we get that

$$\begin{aligned} 0 &= \frac{1}{N}\sum_{l=1}^{N}\frac{\lambda\hat{W}_l(P_1,P_2,P_3)}{1+\lambda\hat{W}_l(P_1,P_2,P_3)} \\ &= \frac{1}{N}\sum_{l=1}^{N}\lambda\hat{W}_l(P_1,P_2,P_3)-\frac{1}{N}\sum_{l=1}^{N}\lambda^2\hat{W}_l^2(P_1,P_2,P_3) \\ &\quad +\frac{1}{N}\sum_{l=1}^{N}\frac{(\lambda\hat{W}_l(P_1,P_2,P_3))^3}{1+\lambda\hat{W}_l(P_1,P_2,P_3)}. \end{aligned} \tag{B.10}$$

Similarly, the final term in (A.9) is bounded by

$$|\lambda|^3\frac{1}{N}\sum_{k=1}^{N}|\hat{W}_l(P_1,P_2,P_3)|^3\max|(1+\lambda\hat{W}_l(P_1,P_2,P_3))^{-1}|=O_p(N^{-3/2}).$$

Hence, we have

$$\sum_{l=1}^{N} \lambda \hat{W}_l(P_1, P_2, P_3) = \sum_{l=1}^{N} \lambda^2 \hat{W}_l^2(P_1, P_2, P_3) + O_p(N^{-1/2}). \tag{B.11}$$

By Taylor's expansion of Equation (3.16) and using (B.9), (B.11), we have that

$$\begin{aligned} -2l_{IF}(P_1, P_2, P_3) &= 2\sum_{l=1}^{N} \log\{1 + \lambda \hat{W}_l(P_1, P_2, P_3)\} \\ &= 2\lambda \sum_{l=1}^{N} \hat{W}_l(P_1, P_2, P_3) - \lambda^2 \sum_{l=1}^{N} \hat{W}_l^2(P_1, P_2, P_3) + r_n \\ &= \lambda \sum_{l=1}^{N} \hat{W}_l(P_1, P_2, P_3) + O_p(N^{-1/2}) + r_n \end{aligned}$$

where

$$\begin{aligned} |r_n| &\le C \sum_{l=1}^{N} |\lambda \hat{W}_l(P_1, P_2, P_3)|^3 \le C|\lambda|^3 \sum_{l=1}^{N} |\hat{W}_l(P_1, P_2, P_3)|^3 \\ &= O_p(N^{-3/2}) O_p(N) = O_p(N^{-1/2}). \end{aligned}$$

From (B.9) and Lemma, it follows that

$$-2l_{IF}(P_1, P_2, P_3) = \frac{\sigma^2 \left[\frac{1}{\sigma N^{1/2}} \sum_{l=1}^{N} \hat{W}_l(P_1, P_2, P_3)\right]^2}{\frac{1}{N} \sum_{l=1}^{N} \hat{W}_l^2(P_1, P_2, P_3)} + o_p(1) \xrightarrow{d} \chi_1^2.$$

Proof of Proposition 3.2:

We first briefly introduce Clarke and Yuan's approach [30]. Define the outer product matrix $\Omega = E[g(Z_j, P_2)g'(Z_j, P_2)]$, Jacobian matrix $D(P_2) = E[\partial g(Z_j, P_2)/\partial P_2]$ and the matrix $\Lambda(P_2) = D'(P_2)\Omega^{-1}(P_2)D(P_2)$, where $g(Z_j, P_2)$ is an estimating function.

For our Bayesian EL approach, $g(Z_j, P_2) = U(Y_{2,j}) - P_2$, where $U(Y)$ is defined in Section 3.2, and

$$\Omega(P_2) = E[g(Z_j, P_2)]^2 = E[U(Y_{2,j}) - P_2]^2 = P_2(1 - P_2).$$

Thus we have

$$\Lambda(P_2) = D'(P_2)\Omega^{-1}(P_2)D(P_2) = \frac{1}{P_2(1-P_2)}.$$

So the reference prior for the EL under the relative entropy is

$$\pi_{EL,1}(P_2) \propto |\Lambda^{-1}(P_2)|^{1/2} = \sqrt{P_2(1-P_2)},$$

i.e.,

$$\pi_{EL,1}(P_2) = \beta(\frac{3}{2}, \frac{3}{2}),$$

and the reference prior for the hybrid EL under Hellinger distance is

$$\pi_{EL,2}(P_2) \propto |\Lambda(P_2)|^{1/2} = \frac{1}{\sqrt{P_2(1-P_2)}},$$

i.e.,

$$\pi_{EL,2}(P_2) = \beta(\frac{1}{2}, \frac{1}{2}).$$