Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Isaac Parakati Date

Constructing Confidence Intervals for Sensitivity Under Controlled Specificity in Medical Tests

By

Isaac Parakati
Master of Science in Public Health

Biostatistics and Bioinformatics

Eugene Huang, PhD

(Thesis Advisor)

Limin Peng, PhD

(Reader)

Constructing Confidence Intervals for Sensitivity Under Controlled Specificity in Medical Tests

By

Isaac Parakati

B.S.
University of Chicago
2015

Thesis Committee Chair: Eugene Huang, PhD
Reader: Limin Peng, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
In partial fulfillment of the requirements for the degree of
Master of Science in Public Health
In Biostatistics
2017

Abstract

Constructing Confidence Intervals for Sensitivity
Under Controlled Specificity in Medical Tests
By: Isaac Parakati

Introduction: Medical tests frequently assist health care workers with identifying individuals affected or not affected by a disease. Although medical tests are supposed to correctly identify diseased individuals as diseased and non-diseased individuals as non-diseased, this does not always occur. The test's accuracy is typically measured in terms of sensitivity and specificity. Fixing the specificity of a test at a particular value, the test's corresponding sensitivity can be determined. The goal of this paper is to propose two new approaches for constructing confidence intervals for sensitivity after fixing specificity.

Methods: To estimate sensitivity, both of the two proposed approaches are based on a quadratic inference function but differ by the procedure used to profile out a nuisance parameter. The first approach minimizes the function with respect to the nuisance parameter. The second approach determines an optimal weighted average between two values for the nuisance parameter. To demonstrate the two approaches, confidence intervals were constructed for the sensitivity of a gene expression biomarker using samples of cancerous and non-cancerous tissues, fixing specificity. Simulations were conducted to evaluate the approaches under different distributions with varying sample sizes. Coverage probabilities and average confidence interval length were determined for each simulation.

Results: In the simulations, the two new approaches produced confidence intervals above or near the nominal significance level. The first approach constructed very wide intervals with conservative coverage. The second approach constructed narrower intervals with coverage near the nominal value; this approach performed similarly to the leading existing BTII approach, whose simulation results were extracted from Zhou and Qin's paper[4]. The BTII approach seemed to perform slightly better when the diseased and non-diseased sample sizes differed. With larger sample sizes, average confidence interval length for all approaches narrowed.

Discussion: The second approach proposed in this paper appears to be a suitable non-bootstrap alternative to the BTII approach when constructing confidence intervals.

Constructing Confidence Intervals for Sensitivity Under Controlled Specificity in Medical Tests

By

Isaac Parakati

B.S.
University of Chicago
2015

Thesis Committee Chair: Eugene Huang, PhD
Reader: Limin Peng, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
In partial fulfillment of the requirements for the degree of
Master of Science in Public Health
In Biostatistics
2017

Acknowledgements

I would like to thank the Department of Biostatistics and Bioinformatics at Emory University for their encouragement and advice. Within the Department, I would like to thank Eugene Huang for his invaluable guidance and feedback on my thesis as my thesis advisor. I would also like to thank Limin Peng for reading my thesis and supporting me during the process of writing it. Finally, I would like to thank my family and friends for their cheer, support, and wholehearted love during this process.

Table of Contents

1. INTRODUCTION

Medical tests are used frequently by health care professionals to assist with medical decision-making. A medical test can suggest the presence of a certain disease in an individual. Ideally, this test would accurately determine the presence of disease. However, realistically, accurate assessments do not always occur: this test might label a non-diseased person as diseased or a diseased person as non-diseased. The probability that a medical test correctly labels a non-diseased person as non-diseased is its specificity, and the probability that the test correctly labels a diseased person as diseased is its sensitivity. For a medical test using a single biomarker on a continuous scale, the relationship between the test's sensitivity and specificity is mediated by a decided cutoff point. Assuming higher values for a test suggest the presence of disease, a higher cutoff point gives the test a higher sensitivity and a lower specificity. For a test, it is of interest to determine the maximum sensitivity when the specificity is controlled at a given level, e.g. 0.8 or 0.9, or vice versa.

1.1 Problem Statement and Notation

Let $\mathbf{x_1} = (\mathrm{x}_{11}, \mathrm{x}_{12} \ldots \mathrm{x}_{1n_1})$ represent test values for a medical test from samples of n_1 diseased subjects, and let $\mathbf{x_0} = (\mathrm{x}_{01}, \mathrm{x}_{02} \ldots \mathrm{x}_{0n_0})$ represent test values for a medical test from n_0 non-diseased subjects. Write the distributions of $\mathbf{x_1}$ and $\mathbf{x_0}$ as F_1 and F_0, respectively. For a cutoff value c, the sensitivity and specificity of the test can be represented by

$$\text{Sensitivity} = \Pr(\mathbf{x_1} > c) = 1 - F_1(c)$$

$$\text{Specificity} = \Pr(\mathbf{x_0} \leq c) = F_0(c)$$

Fixing specificity at a desired value Sp, the test's sensitivity θ can be calculated by $\theta = 1 - F_1(F_0^{-1}(Sp))$. However, in practice, the underlying distributions F_1 and F_0 are not

known. Therefore, after fixing specificity, the sensitivity of the test has to be estimated. The point estimate $\hat{\theta}$ for sensitivity can be calculated by

$$\hat{\theta} = \frac{\sum_{i=1}^{n_1} I(\mathrm{x_{1i}} > \hat{c})}{n_1},$$

where the estimated cutoff value $\hat{c}$ is the *Sp*-th quantile of $\mathbf{x_0}$.

1.2 Purpose Statement

The purpose of this paper is to construct confidence intervals for $\hat{\theta}$ after fixing specificity. Two confidence intervals will be proposed. Simulation studies will be conducted comparing the two intervals to the well-performing BTII interval proffered by Zhou and Qin[4]. The results from simulations studies on the BTII interval are extracted from Zhou and Qin's paper[4].

1.3 Significance Statement

This paper will introduce two new confidence intervals for the sensitivity of a medical test controlling for its specificity. Because constructing these intervals does not involve bootstrap approaches, the intervals may be less computationally intensive to generate while still performing well, compared to earlier intervals. Thus, this paper will also provide evidence for the case of adopting the proposed intervals.

2. BACKGROUND/LITERATURE REVIEW

2.1 Naive Interval

The variance of the estimated sensitivity $\mathrm{Var}(\hat{\theta})$ might be represented as

$$\widehat{\mathrm{Var}_{\mathrm{N}}}(\hat{\theta}) = \frac{\hat{\theta} \times (1 - \hat{\theta})}{n_1}$$

The resulting naive (1- α)% confidence interval would be

$$[\hat{\theta} + \mathrm{z}_{1-\frac{\alpha}{2}}\sqrt{\widehat{\mathrm{Var}_{\mathrm{N}}}(\hat{\theta})},\ \hat{\theta} - \mathrm{z}_{1-\frac{\alpha}{2}}\sqrt{\widehat{\mathrm{Var}_{\mathrm{N}}}(\hat{\theta})}],$$

where z_{β} represents the β-th quantile of the normal distribution. Linnet[2] demonstrated that the coverage probability of this confidence interval, or the probability that the interval contains the true sensitivity, falls below the nominal coverage probability of 1- α. This result may be expected since $\widehat{\mathrm{Var}_{\mathrm{N}}}(\hat{\theta})$ does not account for the variability of $\hat{c}$.

2.2 Linnet Interval

To account for the variability of $\hat{c}$ when determining the variance of the estimated sensitivity, Linnet[2] proposed another approach, articulated by Platt et al[3] and Zhou and Qin[4]. In this approach, the variance of the estimated sensitivity can be determined by

$$\widehat{\mathrm{Var}_{\mathrm{L}}}(\hat{\theta}) = \frac{\hat{\theta} \times (1 - \hat{\theta})}{n_1} + f_1^2(\hat{c}\,) \times \mathrm{Var}(\hat{c}\,)$$

where f_1 represents the probability density function of $\mathbf{x_1}$. Let f_0 represent the probability function of $\mathbf{x_0}$. Then, because approximately

$$\hat{c} - c \sim N\left(0, \frac{(1 - Sp)Sp}{n_0 f_0^2(\hat{c})}\right),$$

the estimate $\widehat{\mathrm{Var}_{\mathrm{L}}}(\hat{\theta})$ can be obtained as

$$\widehat{\mathrm{Var}_{\mathrm{L}}}(\hat{\theta}) = \frac{\hat{\theta} \times (1 - \hat{\theta})}{n_1} + \widehat{f_1^2}(\hat{c}\,) \times \frac{(1 - Sp)Sp}{n_0 \widehat{f_0^2}(\hat{c})}$$

According to Platt's coding of Linnet's interval[3], Linnet uses the probability density function of the normal distribution for both $\widehat{f_1}$ and $\widehat{f_0}$, and he uses the *Sp*-th quantile of the normal distribution for $\hat{c}$. Linnet demonstrated that the interval

$$[\hat{\theta} + \mathrm{z}_{1-\frac{\alpha}{2}}\sqrt{\widehat{\mathrm{Var}_{\mathrm{L}}}(\hat{\theta})}, \hat{\theta} - \mathrm{z}_{1-\frac{\alpha}{2}}\sqrt{\widehat{\mathrm{Var}_{\mathrm{L}}}(\hat{\theta})}]$$

performed well with regard to statistical power. However, Platt et al[3] showed that Linnet's interval did not perform well when $\mathbf{x_1}$ and $\mathbf{x_0}$ are not normally distributed.

2.3 Bootstrap Intervals

Platt et al[3] developed another approach to constructing confidence intervals for sensitivity using a bootstrap approach. A bootstrap procedure proceeds by first drawing samples $\mathbf{x_1^*} = (\mathrm{x}_{11}^*, \mathrm{x}_{12}^*, \dots, \mathrm{x}_{1n_1}^*)$ and $\mathbf{x_0^*} = (\mathrm{x}_{01}^*, \mathrm{x}_{02}^*, \dots, \mathrm{x}_{0n_0}^*)$ of sizes n_1 and n_0 from the observed samples $\mathbf{x_1}$ and $\mathbf{x_0}$, respectively. Fixing specificity at a value *Sp*, let $\widehat{c^*}$ represent the *Sp*-th quantile of $\mathbf{x_0^*}$. Then, the bootstrap sensitivity estimate $\widehat{\theta^*}$ can be represented by

$$\widehat{\theta_P^*} = \frac{\sum_{i=1}^{n_1} I(\mathrm{x}_{1i}^* > \widehat{c^*})}{n_1}$$

Repeating these steps N times will create a set of bootstrap replications $(\theta_{P1}^*, \theta_{P2}^*, \dots, \theta_{PN}^*)$. The mean and variance of $(\widehat{\theta_{P1}^*}, \widehat{\theta_{P2}^*}, \dots, \widehat{\theta_{PN}^*})$ provide the estimated sensitivity $\widehat{\theta_P}$ and its variance.

$$\widehat{\theta_P} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\theta_{Pi}^*}$$

$$\widehat{\mathrm{Var}_{\mathrm{P}}}(\widehat{\theta_P}) = \frac{1}{N-1}\sum_{i=1}^{N}(\widehat{\theta_{Pi}^*} - \widehat{\theta_P})^2$$

To construct confidence intervals from bootstrap replications, Platt used Efron and Tibshirani's[5] bias-corrected and accelerated ($\mathrm{BC_a}$) intervals. Through simulations, Platt et al[3]

demonstrated that the BC_a interval has a higher coverage probability compared to Linnet's[2] interval for non-normally distributed data. When the data is normally distributed, the BC_a interval and Linnet's interval have similar coverage probabilities.

Later, Zhou and Qin also proposed a bootstrap approach called BTII to generate confidence intervals for sensitivity. Similar to the bootstrap procedure used by Platt et al[3], first, samples $\mathbf{x_1^*}$and $\mathbf{x_0^*}$ of sizes n_1 and n_0, respectively, are drawn. Then, the sensitivity is determined by

$$\widehat{\theta_{ZQ}^*} = \frac{\sum_{i=1}^{n_1} I(\mathrm{x}_{1i}^* > \widehat{c^*}) + \frac{1}{2} z_{1-\frac{\alpha}{2}}^2}{n_1 + z_{1-\frac{\alpha}{2}}^2}$$

The term $z_{1-\frac{\alpha}{2}}^2$ is a finite-sample correction adapted from Agresti and Coull's paper[6]. In their paper, Agresti and Coull found through simulations that adding $z_{1-\frac{\alpha}{2}}^2$ improves coverage accuracy for Wald intervals. Since $z_{1-\frac{\alpha}{2}}^2 \approx 2$ when $\alpha = 0.05$, including this term equates to adding 2 diseased and 2 non-diseased observations to the sample.

Repeating the $\widehat{\theta_{ZQ}^*}$ draws N times creates a set of bootstrap replications $(\widehat{\theta_{ZQ1}^*}, \widehat{\theta_{ZQ2}^*}, \dots, \widehat{\theta_{ZQN}^*})$. The estimated sensitivity $\widehat{\theta_{ZQ}}$ and its estimated variance $\widehat{\mathrm{Var}}_{\mathrm{ZQ}}\left(\widehat{\theta_{ZQ}}\right)$ are

$$\widehat{\theta_{ZQ}} = \frac{1}{N}\sum_{i=1}^{N} \widehat{\theta_{ZQi}^*}$$

$$\widehat{\mathrm{Var}}_{\mathrm{ZQ}}(\widehat{\theta_{ZQ}}) = \frac{1}{N-1}\sum_{i=1}^{N} (\widehat{\theta_{ZQi}^*} - \widehat{\theta_{ZQ}})^2$$

From the estimated sensitivity and its estimated variance, the $\left(1-\frac{\alpha}{2}\right)$% BTII confidence interval is $\widehat{\theta_{ZQ}}+\mathrm{z}_{1-\frac{\alpha}{2}}\sqrt{\widehat{\mathrm{Var}_{\mathrm{ZQ}}}\left(\widehat{\theta_{ZQ}}\right)},\widehat{\theta_{ZQ}}-\mathrm{z}_{1-\frac{\alpha}{2}}\sqrt{\widehat{\mathrm{Var}_{\mathrm{ZQ}}}\left(\widehat{\theta_{ZQ}}\right)}]$. Through simulations, Zhou and Qin[4] demonstrated that the BTII interval had better coverage intervals and narrower length compared to the BC_a interval.

Compared to the naïve interval and Linnet's interval, the BTII and BC_a intervals had better coverage. However, the BTII and BC_a intervals use resampling approaches, making them more computationally intensive.

3. METHODS

This paper proposes two new approaches for constructing confidence intervals for sensitivity, controlling for specificity. Unlike BTII and BC_a, the approaches proposed in this paper do not use resampling approaches. Meanwhile, our methods do not impose assumptions on F_1 and F_0. Our proposed confidence intervals are constructed by inverting hypothesis tests.

Consider the following quadratic inference function

$$\phi(c,\theta)=\frac{\left\{\frac{1}{n_1}\sum_{j=1}^{n_1}I\left(x_{1j}>c\right)-\theta\right\}^2}{\left(\frac{1}{n_1}\right)\times\theta\times(1-\theta)}+\frac{\left\{\frac{1}{n_1}\sum_{k=1}^{n_0}I(x_{0k}\le c)-Sp\right\}^2}{\left(\frac{1}{n_0}\right)\times Sp\times(1-Sp)}$$

Clearly, $\phi(c,\theta)$ may be used as a test statistic for the values of (c,θ) as the true one. Let σ_1 and σ_0 represent the standard deviations of the populations of diseased and non-diseased test values, respectively. Because $\sqrt{n_1}(n^{-1}\sum_{j=1}^{n_1}I\left(x_{1j}>c\right)-\theta)\xrightarrow{d}N(0,\sigma_1^2)$ and $\sqrt{n_0}\left(n^{-1}\sum_{k=1}^{n_0}I(x_{0k}\le c)-Sp\right)\xrightarrow{d}N(0,\sigma_0^2)$ by the Central Limit Theorem, $\phi(c,\theta)$ can be approximated by a chi-square distribution with 2 degrees of freedom, as the diseased and

nondiseased samples are independent. However, the cutoff value c is a nuisance parameter. Therefore, we need to profile it out for our inference on θ, which is of interest. It is worthwhile to point out that one unique challenge here lies in the fact that $\phi(c,\theta)$ is not differentiable with respect to c.

Our first method, denoted as M1, takes the function $\eta(\theta) = \min_{c|\theta} \phi(c,\theta)$ as a quadratic inference function for θ. When θ is the true value, it may be shown that $\eta(\theta)$ can be approximated by $\chi^2(1)$. Computationally, the minimization to obtain $\eta(\theta)$ needs only to consider a finite number of c values since c is involved in $\phi(c,\theta)$ only through indicator functions. Precisely, with given θ, $\phi(c,\theta)$ takes at most $n_1 + n_0 + 1$ distinct values corresponding to c taking values in $\mathbf{x_1}$, $\mathbf{x_0}$, and $-\infty$. Furthermore, the search for a minimizer c may be restricted to those values between the empirical *Sp*-th quantile of $\mathbf{x_0}$ and the empirical θ-th quantile of $\mathbf{x_1}$, since $\phi(c,\theta)$ increases beyond that range. A $(1-\alpha)\%$ confidence interval for sensitivity can be constructed at the two values θ_p, θ_q for which $\eta(\theta_p) = \eta(\theta_q) = \chi^2_{1-\alpha}(1)$. The two values θ_p and θ_q are found using the bisection method described in the Appendix to this paper.

Referred to as M2, the second approach constructs a confidence interval θ by inverting a test based on $\phi(\widehat{c_w},\theta)$ with $\widehat{c_w}$ being the optimal linear combination of two cutoff values. It can be shown that this test statistic can also be approximated by $\chi^2(1)$ distribution when θ is the true value. The first cutoff value $\widehat{c_0}$ is found by solving the following equation for $\widehat{c_0}$:

$$\frac{1}{n_0}\sum_{i=1}^{n_0} I(\mathrm{x}_{0i} \le \widehat{c_0}) - Sp = 0$$

From the equation, $\widehat{c_0}$ is equivalent to the Sp-th quantile of $\mathbf{x_0}$. Similarly, for a value θ, the second cutoff value $\widehat{c_1}$ is obtained from the equation

$$\frac{1}{n_1}\sum_{j=1}^{n_1} I\left(\mathrm{x}_{1j} > \widehat{c_1}\right) - \theta = 0$$

The value $\widehat{c_1}$ is equivalent to the $(1-\theta)$-th quantile of $\mathbf{x_1}$. When θ is the true value, both estimated cutoff's target the same estimand and an optimal estimate can be obtained by using a weighted average of the two to achieve better efficiency. The optimal weight is proportional to reciprocal of the variance for the corresponding cut-off estimate. Borrowing methodology developed by Huang[7], an estimated quantity proportional to the standard deviation of $\widehat{c_0}$ is determined by "perturbing" the above equation as follows:

$$\frac{1}{n_0}\sum_{i=1}^{n_0} I(\mathrm{x}_{0i} \le \widehat{c_0}) = F_{bin}^{-1}(0.025;\ \mathrm{n}_0, Sp)/\mathrm{n}_0,$$

$$\frac{1}{n_0}\sum_{i=1}^{n_0} I(\mathrm{x}_{0i} \le \widehat{c_0}) = F_{bin}^{-1}(0.975;\ \mathrm{n}_0, Sp)/\mathrm{n}_0,$$

where $F_{bin}^{-1}(0.025;\ \mathrm{n}_0, Sp)$ and $F_{bin}^{-1}(0.975;\ \mathrm{n}_0, Sp)$ are the 2.5-th and 97.5-th percentiles of the binomial distribution $\mathrm{Bin}(n_0, Sp)$, respectively. These two equations give the values $\widehat{c_{0+}}$ and $\widehat{c_{0-}}$; similarly, $\widehat{c_{1+}}$ and $\widehat{c_{1-}}$ are found from the diseased sample in parallel. The lengths of the intervals $(\widehat{c_{0+}}, \widehat{c_{0-}})$ and $(\widehat{c_{1+}}, \widehat{c_{1-}})$ are proportional to the standard deviations of $\widehat{c_0}$ and $\widehat{c_1}$, respectively. Then, take $w_0 = (\widehat{c_{0+}} - \widehat{c_{0-}})^{-2}$ and $w_1 = (\widehat{c_{1+}} - \widehat{c_{1-}})^{-2}$. The value $\widehat{c_w}$ is determined from $\widehat{c_w} = \frac{w_1 \times \widehat{c_1} + w_0 * \widehat{c_0}}{w_1 + w_0}$. A $(1-\alpha)$% confidence interval can be generated at the two values θ_s, θ_t for which $\phi(\widehat{c_w}, \theta_s) = \phi(\widehat{c_w}, \theta_t) = \chi^2_{1-\alpha}(1)$, found using the bisection method described in the Appendix.

4. PRACTICAL APPLICATION TO CANCER TISSUE

To demonstrate the confidence interval constructed by the two approaches proposed in this paper, the following cancer tissue data is extracted from Pepe's textbook[1]. The data comes from a gene-expression experiment conducted by the Institute for Systems Biology in Seattle, WA. The expression of a gene was analyzed from 30 ovarian cancer tissues and 23 non-diseased ovarian tissues. Under a fixed specificity, researchers are interested in the sensitivity of the gene expression in detecting patients with ovarian cancer. Thus, it is of interest to construct a confidence interval for the sensitivity of the medical test when specificity is fixed.

Fixing specificity at a specificity of 0.80, the approaches M1 and M2 can be used to calculate 95% confidence intervals for the sensitivity of the gene expression. The 95% confidence intervals generated from M1 and M2 are [0.0936, 0.8737] and [0.3082, 0.8275], respectively. These confidence intervals are illustrated in Figures 1 and 2 by the vertical lines. The horizontal line in each figure indicates $\chi^2_{0.05}(1) \approx 3.84$. As illustrated, M2 produces a narrower confidence interval length compared to M1; M1 is more conservative.

5. SIMULATIONS

Fixing specificity, we performed simulations to evaluate the approaches M1 and M2 and compare them to the BTII approach proposed by Zhou and Qin[4]. For the BTII approach, simulation results were extracted from Zhou and Qin's paper[4]. For M1 and M2, simulations were conducted under different distributions with varying sample sizes for the test values $\mathbf{x_1}$ and $\mathbf{x_0}$. The sample sizes for $\mathbf{x_1}$ and $\mathbf{x_0}$ are (n_1, n_0)=(20, 20), (50, 50), and (40, 20). (20, 20) represents small samples, (50, 50) represents larger samples, and (40, 20) represents samples of differing sizes. In some applications, however, the sample sizes of $\mathbf{x_1}$ and $\mathbf{x_0}$ can be much

larger. The sample size combinations listed here are only used for simulations, though meant to reflect practical situations.

In each simulation, 5,000 pairs of random samples $\mathbf{x_1}$ and $\mathbf{x_0}$ are drawn from F_1 and F_0 following either a beta or normal distribution setup. 5 beta distribution setups and 5 normal distributions setups are used. Under each setup, simulations are repeated with the different sample size combinations described earlier. To compare and evaluate the confidence intervals constructed under each simulation setup, average confidence interval length and coverage probability are calculated.

For the first set of simulations using beta distributions, Table 1 lists the parameters used, the fixed value for specificity, and the corresponding true sensitivity for each setup. The setups follow the ones used by Zhou and Qin to demonstrate the BTII approach. Table 2 depicts coverage probabilities and average 95% confidence interval lengths after applying the BTII approach and the approaches proposed in this thesis, M1 and M2, under each of the setups introduced in Table 1, with $F_1 \sim Beta\ (a_1, b_1)$ and $F_0 \sim Beta\ (a_0, b_0)$.

In Table 2, the coverage probability and average confidence interval length for M1 and M2 differed from the BTII approach. In general, compared to the BTII approach, M1 and M2 had better coverage probabilities and a wider average confidence length. M1 consistently had a coverage probability higher than the nominal level of 0.95, yet its average confidence interval length was much larger compared to the lengths from BTII and M2. In most setups, M2 performed similarly to BTII: M2 produced slightly wider average confidence intervals and coverage probabilities slightly better or worse than BTII. With larger sample sizes, for all three approaches, average confidence interval length decreased and coverage probability

improved or diminished, depending on setup. BTII may produce better coverage probabilities and narrower confidence intervals when (n_1, n_0) = (40, 20).

In the second set of simulations conducted, rather than beta distributions, normal distributions were used to model F_1 and F_0. Table 3 lists the normal distribution parameters used. Throughout these simulations, F_0 followed a standard normal $N(0,1)$ distribution and F_1 abided by a standard deviation of 1, with varying values for mean between 1.6832 and 2.9264, i.e. $F_0 \sim N(0,1)$ and $F_1 \sim (\mu_1, 1)$. Table 3 also lists the values fixed for specificity and the corresponding true sensitivities.

Coverage probabilities and average confidence interval lengths with normal distributions are shown in Table 4. BTII performed perform similarly to M2. Across different parameter settings, BTII and M2 alternated between producing slightly better coverage and wider confidence intervals or poorer coverage yet narrower confidence intervals, compared to the other. For instance, compared to BTII, M2 had better coverage and wider confidence intervals under Setup 3, and poorer coverage yet narrower confidence intervals under Setup 5. M1 demonstrated better coverage than M2 and BTII due to the wider confidence intervals M1 produces. Sample size trends observed among beta distribution setups persisted under the normal distribution setups. As sample size increased, average confidence intervals length for all three approaches decreased. Similar to results from using beta distributions, BTII seemed to have slightly better coverage and produce narrower confidence intervals when (n_1, n_0) = (40, 20).

5. DISCUSSION

Fixing specificity, the approaches M1 and M2 proposed in this paper construct $(1-\alpha)$% confidence intervals with coverage probability above or near the nominal level α,

without using bootstrap approaches or estimated the density functions of the diseased sample $\mathbf{x_1}$ and the non-diseased sample $\mathbf{x_0}$, unlike previous approaches. Compared to the existing leading approach BTII by Zhou and Qin[4], M1 constructs confidence intervals with better coverage yet wider length, and M2 construct confidence intervals with similar coverage and average length. However, BTII might perform better when $\mathbf{x_1}$ and $\mathbf{x_0}$ have different sample sizes. All in all, based on these results shown, M2 could be a suitable non-bootstrap alternative to BTII.

There are several limitations to this paper. Notably, the simulation results from the BTII approach were extracted from Zhou and Qin's paper[4]. Therefore, Zhou and Qin's simulation results for BTII were not replicated. In addition, to allow comparison between BTII, M1, and M2, simulations using M1 and M2 could only be conducted under the distribution and sample size settings established by Zhou and Qin in their paper[4].

Nevertheless, despite having room for improvement, approaches M1 and M2 demonstrate that reliable confidence intervals can be constructed without using resampling methods. Moreover, M1 and M2 represent only two ways of constructing confidence intervals in this manner; other approaches may exist as well. Future work will explore additional approaches and fine-tune M1 and M2.

6. REFERENCES

1. Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction*. Oxford: Oxford University Press.
2. Linnet, K. (1987). Comparison of quantitative diagnostic tests: Type I error, power, and sample size. *Statistics in Medicine, 6*(2), 147-158.
3. Platt, R. W., Hanley, J. A., & Yang, H. (2000). Bootstrap confidence intervals for the sensitivity of a quantitative diagnostic test. *Statistics in Medicine, 19*(3), 313-322.
4. Zhou, X., & Qin, G. (2005). Improved confidence intervals for the sensitivity at a fixed level of specificity of a continuous-scale diagnostic test. *Statistics in Medicine, 24*(3), 465-477.
5. Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.
6. Agresti, A., & Coull, B. A. (1998). Approximate Is Better than "Exact" for Interval Estimation of Binomial Proportions. *The American Statistician, 52*(2), 119-126.
7. Huang, Y. (2002). Calibration Regression of Censored Lifetime Medical Cost. *Journal of the American Statistical Association, 97*(457), 318-327.

7. APPENDIX

7.1 Bisection Method

Used to find roots for a function f on a range $[a, b]$, the bisection method works by first by finding the midpoint $c = \frac{a+b}{2}$ of the range $[a, b]$. Then, then if the sign of $f(b)$ is the same as the sign of $f(c)$, b is updated by c; otherwise, a is updated by c. This procedure is repeated until $b - a$ is sufficiently small. Finally, the value $f(c)$ is the approximated root. In this paper, the bisection method is used to find the left and right bounds of $(1-\alpha)\%$ confidence intervals for sensitivity constructed by the approaches proposed in this paper. For approach M1, the roots are approximated at $\eta(\theta) - \chi^2_{1-\alpha}(1)$. For M2, the roots are approximated at $\phi(\theta, \widehat{c_w}) - \chi^2_{1-\alpha}(1)$. The initial range for the lower bound of the interval is between 0 and the point estimate $\hat{\theta}$, and the initial range for the upper bound is between $\hat{\theta}$ and 1. Although numerically $\phi(0, c)$ and $\phi(1, c)$ cannot be evaluated, they are reasonably assumed to produce values greater than $\phi(\theta, \tilde{c}) - \chi^2_{1-\alpha}(1)$. Making this assumption allows the bisection method to function well.

7.2 Tables & Figures

Figure 1: M1 Interval for sensitivity of biomarker gene in detecting ovarian cancer, with fixed specificity

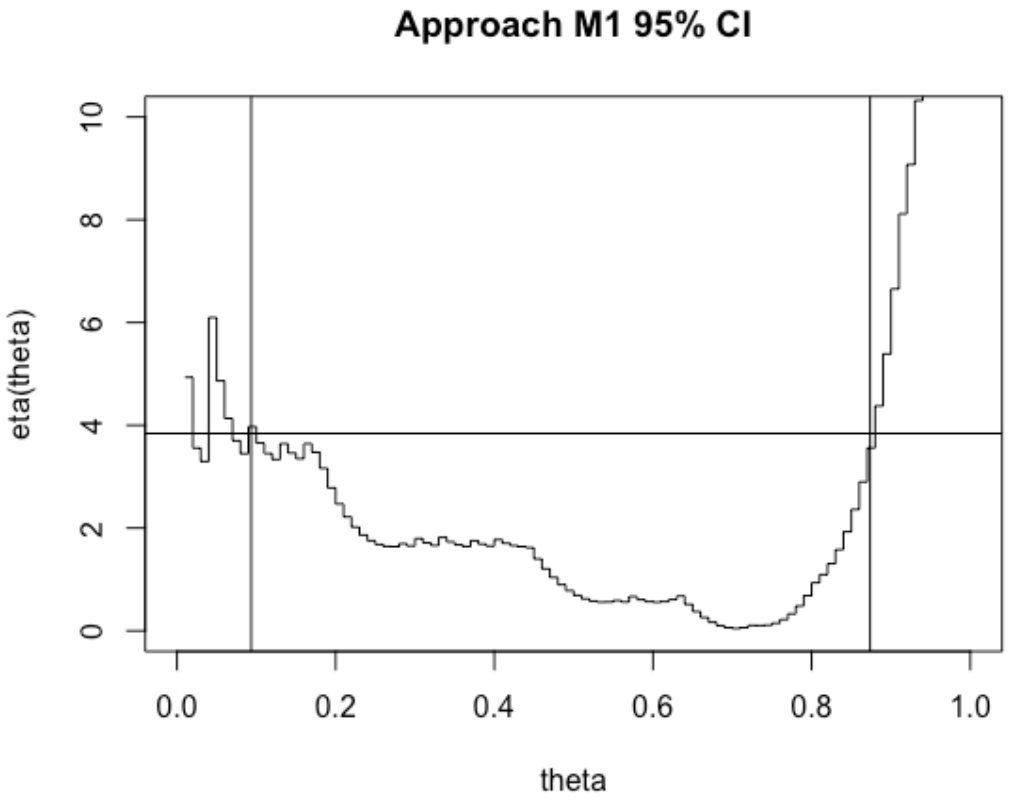

Figure 2: M2 Interval for sensitivity of biomarker gene in detecting ovarian cancer, with fixed specificity

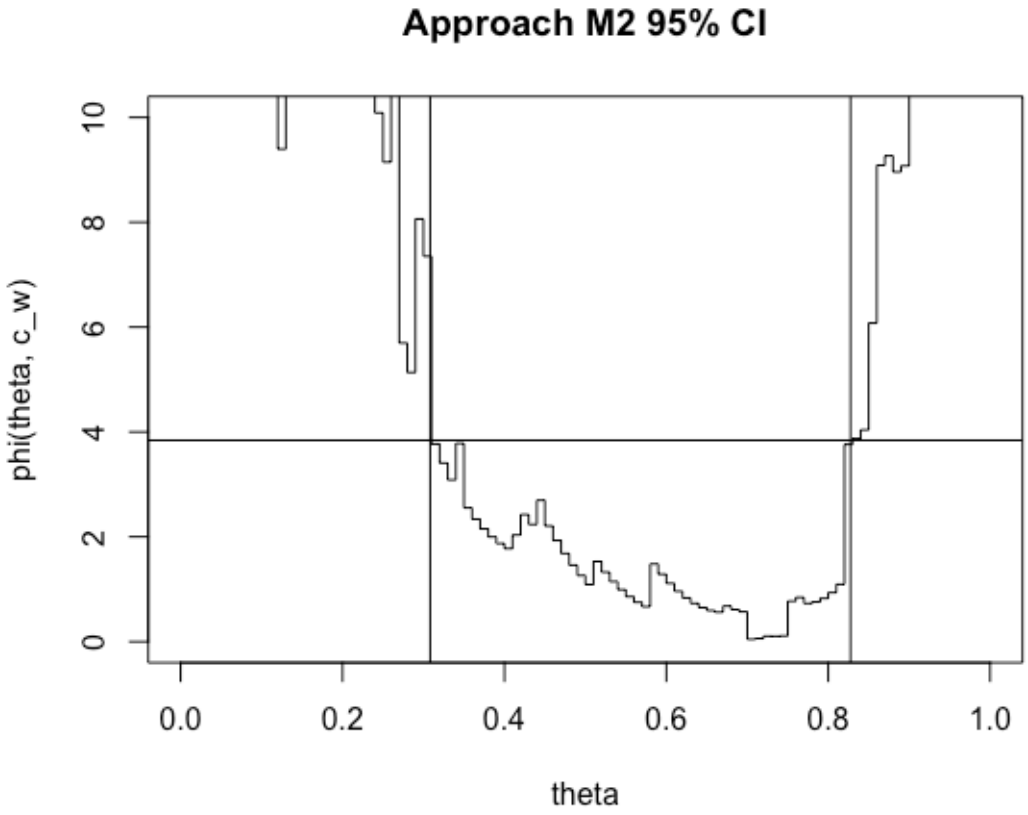

Table 1: Beta distributions setups for simulations

Setup	$(\boldsymbol{a_1}, \boldsymbol{b_1})$	$(\boldsymbol{a_0}, \boldsymbol{b_0})$	**Specificity**	**True sensitivity**
1	(4,1)	(1,3.5)	0.90	0.95
2	(3,1)	(1,3)	0.80	0.93
3	(3,1)	(1,3)	0.90	0.85
4	(4,2)	(2,4)	0.80	0.82
5	(3,2)	(2,3)	0.80	0.55

Table 2: Simulation results under Beta distribution setups

		$n_1 = 20, n_0 = 20$		$n_1 = 50, n_0 = 50$		$n_1 = 40, n_0 = 20$	
Setup	**Approach**	**Coverage Prob**	**Avg Length**	**Coverage Prob**	**Avg Length**	**Coverage Prob**	**Avg Length**
1	BTII	0.8395	0.2278	0.9590	0.1678	0.9595	0.2229
	M1	0.9868	0.9508	0.9874	0.3123	0.9816	0.9713
	M2	0.9546	0.4330	0.9454	0.2065	0.9324	0.4605
2	BTII	0.9255	0.2457	0.9530	0.1783	0.9715	0.2342
	M1	0.9888	0.4513	0.9868	0.2323	0.9874	0.4044
	M2	0.9414	0.3309	0.9364	0.1886	0.9406	0.2866
3	BTII	0.9460	0.3741	0.9565	0.2829	0.9330	0.3518
	M1	0.9932	0.9367	0.9896	0.4564	0.9870	0.9490
	M2	0.9454	0.5225	0.9476	0.3215	0.9286	0.5350
4	BTII	0.9545	0.3875	0.9485	0.2812	0.9620	0.3608
	M1	0.9890	0.6029	0.9848	0.3523	0.9832	0.5640
	M2	0.9452	0.4434	0.9506	0.2844	0.9380	0.4097
5	BTII	0.9440	0.5223	0.9540	0.3858	0.9335	0.4939
	M1	0.9900	0.6968	0.9852	0.4601	0.9850	0.6511
	M2	0.9494	0.5392	0.9504	0.3789	0.9430	0.5057

Table 3: Normal distributions setups for simulations

Setup	**(μ_1, σ_1)**	**(μ_0, σ_0)**	**Specificity**	**True Sensitivity**
1	(2.9264,1)	(0,1)	0.90	0.95
2	(2.5631,1)	(0,1)	0.90	0.93
3	(2.1231,1)	(0,1)	0.90	0.85
4	(2.4865,1)	(0,1)	0.80	0.82
5	(1.6832,1)	(0,1)	0.80	0.55

Table 4: Simulation results under Normal distribution setups

		n_1 = **20,** n_0 = **20**		n_1 = **50,** n_0 = **50**		n_1 = **40,** n_0 = **20**	
Setup	**Method**	**Coverage**	**Length**	**Coverage**	**Length**	**Coverage**	**Length**
1	BTII	0.7835	0.2196	0.9600	0.1572	0.9350	0.2001
	M1	0.9874	0.9510	0.9894	0.3104	0.9870	0.9717
	M2	0.9538	0.3294	0.9474	0.1957	0.9354	0.3141
2	BTII	0.9405	0.3061	0.9675	0.2293	0.9595	0.2974
	M1	0.9872	0.9451	0.9874	0.4144	0.9874	0.9626
	M2	0.9390	0.4030	0.9368	0.2699	0.9432	0.3881
3	BTII	0.9445	0.4277	0.9485	0.3201	0.9430	0.4080
	M1	0.9902	0.9265	0.9850	0.5257	0.9868	0.9345
	M2	0.9456	0.4856	0.9354	0.3582	0.9280	0.4692
4	BTII	0.7980	0.1990	0.9575	0.1443	0.9640	0.1942
	M1	0.9884	0.9430	0.9820	0.4363	0.9886	0.9594
	M2	0.9432	0.4178	0.9332	0.2863	0.9442	0.4047
5	BTII	0.9640	0.4159	0.9585	0.2977	0.9555	0.3835
	M1	0.9872	0.9473	0.9876	0.3803	0.9890	0.9664
	M2	0.9344	0.3783	0.9414	0.2441	0.9406	0.3640