Received: 10 May 2023 | Revised: 25 October 2023 | Accepted: 5 December 2023

DOI: 10.1002/sim.9998

RESEARCH ARTICLE

Statistics in Medicine WILEY

Evaluating joint confidence region of hypervolume under ROC manifold and generalized Youden index

Jia Wang[1] | **Jingjing Yin**[2] | **Lili Tian**[1]

[1]Department of Biostatistics, University at Buffalo, Buffalo, New York, USA

[2]Department of Biostatistics, Epidemiology and Environmental Health Sciences, Jiann-Ping Hsu College Public Health, Georgia Southern University, Statesboro, Georgia, USA

Correspondence
Lili Tian, Department of Biostatistics, University at Buffalo, 717 Kimball Tower, 3435 Main Street, Buffalo, NY, 14214, USA.
Email: ltian@buffalo.edu

In biomarker evaluation/diagnostic studies, the hypervolume under the receiver operating characteristic manifold (HUM_K) and the generalized Youden index (J_K) are the most popular measures for assessing classification accuracy under multiple classes. While HUM_K is frequently used to evaluate the overall accuracy, J_K provides direct measure of accuracy at the optimal cut-points. Simultaneous evaluation of HUM_K and J_K provides a comprehensive picture about the classification accuracy of the biomarker/diagnostic test under consideration. This article studies both parametric and non-parametric approaches for estimating the confidence region of HUM_K and J_K for a single biomarker. The performances of the proposed methods are investigated by an extensive simulation study and are applied to a real data set from the Alzheimer's Disease Neuroimaging Initiative.

KEYWORDS

Alzheimer's disease, biomarker evaluation, confidence region, diagnostic studies, generalized inference, ROC analysis

1 | INTRODUCTION

The receiver operating characteristic (ROC) curve is widely used in biomarker evaluation studies as well as diagnostic studies under binary classification (eg, non-diseased vs diseased) and many ROC-related accuracy measures have been extensively studied and reviewed in statistical literature.[1-4] Among them, area under the ROC curve (AUC) is a popular summary index of the discriminating ability of a biomarker/diagnostic test while the Youden index (generally denoted as J) gains popularity because it serves as an accuracy measure as well as a cut-point selection method.[5,6]

In practice, multi-class classification is quite common. Especially in medical diagnosis, we often encounter settings involving multiple ordered stages. For example, in the diagnosis of Alzheimer's disease, mild cognitive impairment (MCI) is a transition stage between the expected cognitive decline of normal aging and the more serious decline of dementia caused by Alzheimer's disease (AD). Therefore, the diagnosis of Alzheimer's disease generally follows ordinal trichotomous classification: healthy, mild cognitive impairment, and fully diseased.[7,8] As another example, for ovarian cancer diagnosis, subjects might be categorized as "benign," "borderline," or "malignant."[9] For general K ($K \geq 3$) classes, there exist abundant statistical research on classification accuracy measures. For example, Scurfield[10] and Mossman[11] extended the two-class ROC to higher-dimensional ROC framework, and AUC to hypervolume under manifold (HUM_K); Nakas and Yiannoutsos[12] proposed inference methods for HUM_K with multiple classes; and Li and Fine[13] presented a rigorous definition of HUM_K in general. Specifically, the three-class setting has received enormous attention due to its popularity in practice and AUC has been extended to the volume under the ROC surface (VUS). Besides

the overall accuracy measure HUM_K,[11,12,14] two-class Youden index has been extended to the generalized Youden index (J_K) for disease classification of $K(K \geq 3)$ stages by Nakas et al[15] and its properties had been thoroughly studied by Luo and Xiong.[16]

Both the hypervolume under manifold (HUM_K) and the generalized Youden index (J_K) are critical accuracy measures for multi-class classification. The former summarizes the discriminating ability of a biomarker over all possible cut-points, while the latter measures the classification accuracy at the optimal cut-points directly. While it is a common practice to rank/select biomarkers based on HUM_K,[17,18] such practice can be misleading as it does not take into consideration the discriminatory accuracy at the optimal cut-points. In this regard, Yin and Tian[19] proposed joint confidence region about AUC and J, which offers a more comprehensive view of the diagnostic accuracy of a biomarker under two-class classification. It is worth noting that joint inference has also been examined in a number of research papers.[20-22] However, to our best knowledge, the joint inference of HUM_K and J_K under multi-class classification has not been explored.

The rest of this article is organized as follows. In Section 2, preliminaries are given and the motivation to construct joint inference for hypervolume under ROC manifold (HUM_K) and generalized Youden index (J_K) is further illustrated. Parametric methods for confidence region estimation are proposed in Section 3 and non-parametric methods are proposed in Section 4. Section 5 presents simulation results. The proposed methods are illustrated using a subset data from Alzheimer's Disease Neuroimaging Initiative (ADNI) study in Section 6. Finally, Section 7 contains summary and discussion.

2 | PRELIMINARIES: SETTINGS AND MOTIVATION

2.1 | Settings

Consider a setting with K $(K \geq 3)$ independent ordered classes. Without loss of generality, we assume higher marker values indicate greater chance of being diseased. Assume biomarkers are measured on a continuous scale. For a single biomarker, let Y_i, $F_i(.)$, and $f_i(.)$ denote the random variable, the cumulative distribution function (c.d.f.), and the probability density function (p.d.f.) for ith $(i = 1, 2, \dots, K)$ class, respectively. Let P_{ii} $(i = 1, 2, \dots, K)$ denote the correct classification rate for a randomly selected subject in ith class being correctly identified into ith class. For the threshold-based decision rules, we need $K-1$ cut-points to classify K ordered classes. Given the vector of cut-points $\mathbf{c} = (c_1, \dots, c_{K-1})$, where $c_1 < \cdots < c_{K-1}$, the correct classification rates are given by

$$\begin{aligned} P_{11} &= F_1(c_1), \\ P_{ii} &= F_i(c_i) - F_i(c_{i-1}), \quad \text{for} \quad i = 2, \dots, K-1, \\ P_{KK} &= 1 - F_K(c_{K-1}). \end{aligned}$$

For ease of notation, define $t_i = P_{ii}$ $(i = 1, 2, \dots, K)$. The ROC manifold is constructed by plotting the points with coordinates $(t_1, t_2, \dots, t_K)$ in K-dimensional space, while varying the $K-1$ ordered thresholds under $c_1 < \cdots < c_{K-1}$. The hypervolume under the ROC manifold (HUM_K) measures the overall classification accuracy for the biomarker with K ordered stages. A rigorous mathematical definition of HUM_K is given as[23]

$$\text{HUM}_K = \int_0^1 \int_0^{g_1(t_1)} \cdots \int_0^{g_{K-2}(t_1, \dots, t_{K-2})} g_{K-1}(t_1, \dots, t_{K-1}) dt_{K-1} \cdots dt_2 dt_1, \tag{1}$$

where g_{i-1} is a recursive equation defined as $t_i = g_{i-1}(t_1, \dots, t_{i-1})$, $i = 2, \dots, K$. For example, $g_1(t_1) = 1 - F_2(F_1^{-1}(t_1))$, and $g_2(t_1, t_2) = 1 - F_3(F_2^{-1}(t_2 + F_2(F_1^{-1}(t_1))))$. Generally speaking, we can write

$$g_i(t_1, t_2, \dots, t_i) = g_i(t_1, t_2, \dots, t_i; F_1, F_2, \dots, F_i, F_{i+1}), \tag{2}$$

where $i = 1, \dots, K-1$. The HUM_K is equal to the probability that a set of K biomarker values, one from each class, will be in the correct order, that is, $\text{HUM}_K = P(Y_1 < Y_2 < \cdots < Y_K)$.[12] The values of HUM_K vary from $1/K!$ to 1, where $1/K!$

corresponds to a completely uninformative biomarker and 1 a perfect biomarker which separates K classes completely. When $K = 3$, the volume under ROC surface (VUS stands for HUM_3) can be expressed as[14]

$$VUS = \int_{-\infty}^{\infty} F_1(s)(1 - F_3(s))f_2(s)ds.$$

As $K = 4$, the HUM_4 can be simplified as[24]

$$\text{HUM}_4 = \int_0^1 \int_0^{F_2(F_3^{-1}(u_3))} F_1(F_2^{-1}(u_2)) \left[1 - F_4(F_3^{-1}(u_3))\right] du_2 du_3.$$

The generalized Youden index, on the other hand, evaluates the accuracy of a biomarker under K ordinal classes based on the total correct classification rate at the optimal cut-points.[15] We define the generalized Youden index by introducing a weight of $1/(K-1)$ as follows:

$$J_K = \frac{1}{K-1} \max_{c_1, \dots, c_{K-1}} \left[\sum_{i=1}^{K} P_{ii} - 1 \right]. \tag{3}$$

The generalized Youden index defined in (3) falls into range 0 to 1, making it practically convenient. When K ordinal classes overlap completely, J_K becomes zero, indicating a completely useless marker. When a biomarker perfectly separates K classes, J_K equals to one, indicating a perfect biomarker. Note that J_K can be further written as

$$J_K = \frac{1}{K-1} \left\{ \sum_{i=1}^{K-1} \max_{c_i} [F_i(c_i) - F_{i+1}(c_i)] \right\}. \tag{4}$$

2.2 | Under normality and gamma distribution

For parametric assumptions, we consider normal and gamma as both distributions are widely used for modeling data in applied fields. In the following, we present some formulas of HUM_K and J_K under both normality and gamma distribution.

2.2.1 | Under normality

Assume the biomarker measurements follow normal distributions, that is, $Y_i \sim N(\mu_i, \sigma_i^2)$ for ith class ($i = 1, 2 \dots K$). Given a vector of cut-points $\mathbf{c} = (c_1, \dots, c_{K-1})$, where $c_1 < \cdots < c_{K-1}$, the correct classification rates in (1) are given by

$$\begin{aligned}
P_{11} &= \Phi\left(\frac{c_1 - \mu_1}{\sigma_1}\right), \\
P_{ii} &= \Phi\left(\frac{c_i - \mu_i}{\sigma_i}\right) - \Phi\left(\frac{c_{i-1} - \mu_i}{\sigma_i}\right), \quad \text{for} \quad i = 2, \dots, K-1, \\
P_{KK} &= 1 - \Phi\left(\frac{c_{K-1} - \mu_K}{\sigma_K}\right),
\end{aligned}$$

where $\Phi(.)$ is the c.d.f. for standard normal distribution.

The HUM_K in (1) can be obtained by expressing g_i's in (2) using normal c.d.f. and p.d.f. of K ordered classes. For example, $g_1(t_1) = 1 - \Phi((\Phi^{-1}((t_1 - \mu_1)/\sigma_1) - \mu_2)/\sigma_2)$. More specifically, we have

$$VUS = \int_{-\infty}^{\infty} \Phi(as - b)\Phi(-cs + d)\phi(s)ds, \tag{5}$$

where $a = \sigma_2/\sigma_1$, $b = (\mu_1 - \mu_2)/\sigma_1$, $c = \sigma_2/\sigma_3$, and $d = (\mu_3 - \mu_2)/\sigma_3$,[14] and

$$\text{HUM}_4 = \int_0^1 \int_0^l \Phi\left(\frac{\sigma_2\Phi^{-1}(u_2) + \mu_2 - \mu_1}{\sigma_1}\right)\left[1 - \Phi\left(\frac{\sigma_3\Phi^{-1}(u_3) + \mu_3 - \mu_4}{\sigma_3}\right)\right] du_2 du_3, \tag{6}$$

where $l = \Phi\left(\frac{\sigma_3\Phi^{-1}(u_3)+\mu_3-\mu_2}{\sigma_2}\right)$.

Under normality, the generalized Youden index defined in (3) can be written as

$$J_K = \frac{1}{K-1}\sum_{i=1}^{K-1}\left\{\Phi\left(\frac{c_i - \mu_i}{\sigma_i}\right) - \Phi\left(\frac{c_i - \mu_{i+1}}{\sigma_{i+1}}\right)\right\}, \tag{7}$$

where c_i is the optimal cut-point as

$$c_i = \frac{\left(\mu_{i+1}\sigma_i^2 - \mu_i\sigma_{i+1}^2\right) - \sigma_i\sigma_{i+1}\sqrt{(\mu_i - \mu_{i+1})^2 + \left(\sigma_i^2 - \sigma_{i+1}^2\right)\ln\left(\sigma_i^2/\sigma_{i+1}^2\right)}}{\sigma_i^2 - \sigma_{i+1}^2}, \tag{8}$$

for $i = 1, 2, \ldots, K-1$.[25] If all the variances σ_i^2's are equal, $c_i = (\mu_i + \mu_{i+1})/2$.

2.2.2 | Under gamma distribution

For ith class ($i = 1, 2, \ldots, K$), assume the biomarker measurement $Y_i \sim G(\alpha_i, \beta_i)$, where $\alpha_i > 0$ (shape parameter), and $\beta_i > 0$ (rate parameter). The HUM_K and J_K defined in (1) and (4) can be calculated using gamma p.d.f., $f_i(z|\alpha_i, \beta_i) = \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)}z^{\alpha_i-1}e^{-\beta_i z}$, and gamma c.d.f., $F_i(z|\alpha_i, \beta_i) = \frac{\gamma(\alpha_i,\beta_i x)}{\Gamma(\alpha_i)}$, where $\Gamma(\alpha_i)$ stands for gamma function $\int_0^\infty t^{\alpha_i-1}e^{-t}dt$, and $\gamma(\alpha_i, \beta_i x)$ stands for lower gamma function $\int_0^{\beta_i x} t^{\alpha_i-1}e^{-t}dt$. Specifically, as $K = 3$, we have

$$VUS = \frac{\beta_2^{\alpha_2}}{\prod_{i=1}^3\Gamma(\alpha_i)}\int_{-\infty}^{\infty}\gamma(\alpha_1, \beta_1 s)(\Gamma(\alpha_3) - \gamma(\alpha_3, \beta_3 s))s^{\alpha_2-1}e^{-\beta_2 s}ds, \tag{9}$$

$$J_3 = \left[\frac{\gamma(\alpha_1, \beta_1 c_1)}{\Gamma(\alpha_1)} - \frac{\gamma(\alpha_2, \beta_2 c_1)}{\Gamma(\alpha_2)} + \frac{\gamma(\alpha_2, \beta_2 c_2)}{\Gamma(\alpha_2)} - \frac{\gamma(\alpha_3, \beta_3 c_2)}{\Gamma(\alpha_3)}\right]\Big/2, \tag{10}$$

where c_1 and c_2 are the pair of optimal cut-points with $c_1 < c_2$.

2.3 | Motivation: A numerical study on correlation

In this section, we numerically investigate the correlation between HUM_K and J_K to provide justification for estimating joint confidence region of HUM_K and J_K when $K = 3$. Assume $Y_1 \sim N(0, 1^2)$, $Y_2 \sim N(1, 1^2)$ and $Y_3 \sim N(\mu_3, \sigma_3^2)$ where μ_3 ranges from 1.2 to 7.0 with step size of 0.2 and σ_3 from 1.2 to 2.8 with step size of 0.2. The VUS and J_3 are calculated following (5) and (7) and the correlation is estimated using non-parametric bootstrap method (500 bootstrap samples) at each setting. The results are presented in a heat map (Figure 1), which demonstrates the pattern of correlation between VUS and J_3 under different settings of μ_3 and σ_3^2.

From Figure 1, we observe that the correlation between VUS and J_3 is relatively strong (ranges from 0.68 to 0.85) under the settings investigated. However, the trend of the correlation follows a more complicated pattern than monotone. Specifically, when σ_3 is fixed, the correlation tends (but not always) to increase as μ_3 increases; similarly, when μ_3 is fixed, correlation tends to decrease (but not always) as σ_3 increases. We also investigate another setting where $Y_1 \sim N(0, 1^2)$, $Y_2 \sim N(\mu_2, \sigma_2^2)$, and $Y_3 \sim N(5, 3^2)$. The correlation between VUS and J_3 is estimated by varying μ_2 from 1.2 to 4.8 with step size of 0.2 and σ_2 from 1 to 3 with step size of 0.2. We observe similar patterns regarding the correlation between HUM_K and J_K.

Generally speaking, VUS and J_3 are correlated and their correlation is not ignorable, hence it is necessary to consider VUS (or HUM_K) and J_3 (or J_K) simultaneously for the purpose of providing a comprehensive picture of a biomarker's discriminatory ability.

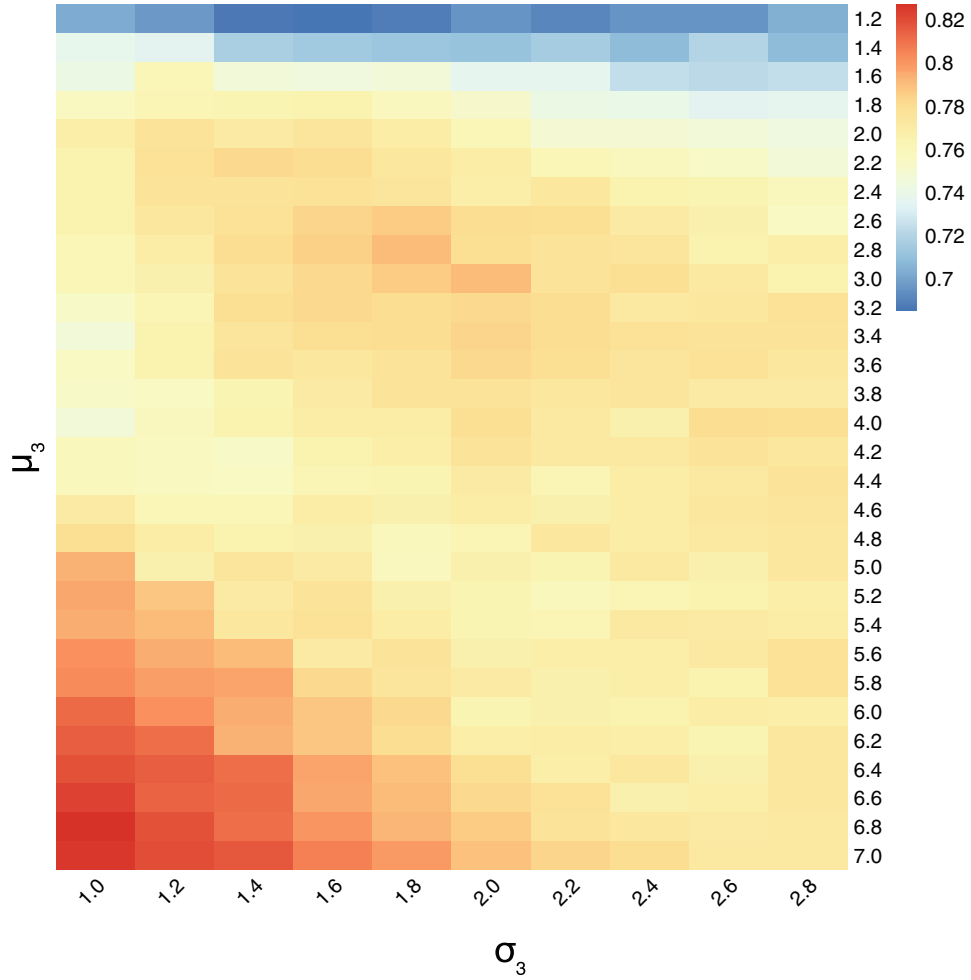

FIGURE 1 The estimated correlation between VUS and J_3 under normality ($K = 3$). The rows correspond to different values of μ_3 from 1.2 (top) to 7.0 (bottom) and the column σ_3 from 1.0 (left) to 2.8 (right).

3 | PARAMETRIC CONFIDENCE REGION ESTIMATION

In this section, we propose generalized inference approach for joint confidence region estimation of hypervolume under ROC manifold (HUM_K) and generalized Youden index (J_K) under normality and gamma distribution. The generalized variables and generalized pivots were introduced by Tsui and Weerahandi[26] and Weerahandi.[27] More details can be found in the book by Weerahandi.[28] When standard solutions do not exist for confidence intervals and hypothesis testing, the generalized inference methods can be applied to different practical settings and have been shown to have satisfactory performance, even at small sample sizes.[19,29-31] Generalized confidence intervals have been shown to coincide with fiducial confidence intervals by Hannig.[32]

Section 3.1 presents methods under normality, and Section 3.2 for handling non-normal data by Box-Cox transformation. Due to the popularity of gamma distribution in analyzing right-skewed data, fiducial methods under gamma distribution are presented in Section 3.3.

3.1 | Under normality: The generalized inference method

The generalized pivotal quantities for normal variance and means are well known as[29]

$$R_{\sigma_i^2} = \frac{(n_i - 1)s_i^2}{V_i}, \quad R_{\mu_i} = \bar{y}_i - Z_i\sqrt{R_{\sigma_i^2}/n_i}, \tag{11}$$

where $V_i = \frac{(n_i-1)S_i^2}{\sigma_i^2} \sim \chi^2_{n_i-1}$ and $Z_i = \frac{\sqrt{n_i}(\bar{Y}_i-\mu_i)}{\sigma_i} \sim N(0,1)$ $(i = 1, \ldots, K)$. By substituting μ_i and σ_i^2 with their corresponding R_{μ_i} and $R_{\sigma_i^2}$ in (6) and (7) for HUM_K and J_K, we can obtain their generalized pivotal quantities R_{HUM_K} and R_{J_K}.

It can be easily checked that $R_\eta = (R_{\text{HUM}_K}, R_{J_K})^T$ is a bona fide generalized pivot for $\eta = (\text{HUM}_K, J_K)^T$ satisfying the following two conditions: (1) The distribution of R_η is independent of any unknown parameters, and (2) the observed value of R_η equals to η for given $\bar{y}_i$ and s_i^2 $(i = 1, 2, \ldots, K)$.

As $K = 3$, substituting μ_i and σ_i^2 in (5) and (7) using R_{μ_i} and $R_{\sigma_i^2}$ respectively, the generalized pivotal quantity for VUS can be obtained as

$$R_{VUS} = \int_{-\infty}^{\infty} \Phi(R_a s - R_b)\Phi(-R_c s + R_d)\phi(s)ds, \tag{12}$$

where $R_a = \frac{R_{\sigma_2}}{R_{\sigma_1}}, R_b = \frac{R_{\mu_1}-R_{\mu_2}}{R_{\sigma_1}}, R_c = \frac{R_{\sigma_2}}{R_{\sigma_3}}, R_d = \frac{R_{\mu_3}-R_{\mu_2}}{R_{\sigma_3}}$. The generalized pivot for the generalized Youden Index (J_3) is

$$R_{J_3} = \frac{\Phi\left(\frac{R_{c_1}-R_{\mu_1}}{R_{\sigma_1}}\right) - \Phi\left(\frac{R_{c_1}-R_{\mu_2}}{R_{\sigma_2}}\right) + \Phi\left(\frac{R_{c_2}-R_{\mu_2}}{R_{\sigma_2}}\right) - \Phi\left(\frac{R_{c_2}-R_{\mu_3}}{R_{\sigma_3}}\right)}{2}, \tag{13}$$

where R_{c_1} and R_{c_2} are the generalized pivots for optimal cut-points value c_1 and c_2 obtained by substituting R_{μ_i} and $R_{\sigma_i^2}$ in (8), that is, $R_{c_i} = \left[\left(R_{\mu_{i+1}}R^2_{\sigma_i} - R_{\mu_i}R^2_{\sigma_{i+1}}\right) - R_{\sigma_i}R_{\sigma_{i+1}}\sqrt{\left(R_{\mu_i} - R_{\mu_{i+1}}\right)^2 + \left(R^2_{\sigma_i} - R^2_{\sigma_{i+1}}\right) ln\left(\frac{R^2_{\sigma_i}}{R^2_{\sigma_{i+1}}}\right)}\right] / \left(R^2_{\sigma_i} - R^2_{\sigma_{i+1}}\right)$, where $i = 1, 2$. When all 3 groups have equal variances, $R_{c_i} = \left(R_{\mu_i} + R_{\mu_{i+1}}\right)/2$, where $i = 1, 2$.

Given a data set with $n_1, n_2, \ldots, n_K$ subjects from $1, 2, \ldots, K$ classes, respectively, the confidence region for $\eta = (\text{HUM}_K, J_K)^T$ using generalized inference approach can be obtained via the following steps: (1) for $i = 1, 2, \ldots, K$, generate $V_i \sim \chi^2_{n_i-1}, Z_i \sim N(0,1)$, then calculate $R_{\sigma_i^2}$ and R_{μ_i} following (11); (2) calculate $R_\eta = \left(R_{\text{HUM}_K}, R_{J_K}\right)^T$, for example, following (12) and (13) when $K = 3$; (3) repeat steps 1 and 2 for B = 2500 times to obtain a set of values $R_\eta^b = \left(R^b_{\text{HUM}_K}, R^b_{J_K}\right)^T$ for $b = 1, 2, \ldots, B$; (4) calculate the sample mean vector $\hat{\eta}_{\text{GPQ}} = \frac{1}{B}\sum_{b=1}^B R_\eta^b$ and sample covariance matrix $\hat{\Sigma}_{\text{GPQ}} = \frac{1}{B-1}\sum_{b=1}^B \left(R_\eta^b - \hat{\eta}_{\text{GPQ}}\right)\left(R_\eta^b - \hat{\eta}_{\text{GPQ}}\right)^T$; (5) calculate $\tilde{R}_\eta^b = \hat{\Sigma}_{\text{GPQ}}^{-1/2}\left(R_\eta^b - \hat{\eta}_{\text{GPQ}}\right)$, the standardized version of R_η^b, and its length $||\tilde{R}_\eta^b||$ for $b = 1, 2, \ldots, B$.

Denote $q_{\{\|\tilde{R}_\eta\|;1-\alpha\}}$ as the $100(1-\alpha)$th percentile of the set $||\tilde{R}_\eta^b||$ $(b = 1, \ldots, B)$. The $100(1-\alpha)\%$ generalized confidence region of $\eta = (\text{HUM}_K, J_K)^T$ is

$$\left\{\eta : \left(\eta - \hat{\eta}_{\text{GPQ}}\right)^T \hat{\Sigma}_{\text{GPQ}}^{-1}\left(\eta - \hat{\eta}_{\text{GPQ}}\right) \le q^2_{\{\|\tilde{R}_\eta\|;1-\alpha\}}\right\}.$$

The area of confidence region is estimated by $A_{\text{GPQ}} = \pi\left(q^2_{\{\|\tilde{R}_\eta\|;1-\alpha\}}\right)\sqrt{|\hat{\Sigma}_{\text{GPQ}}|}$ where $|\hat{\Sigma}_{\text{GPQ}}|$ is the determinant of $\hat{\Sigma}_{\text{GPQ}}$ obtained in step 4. Such confidence region is denoted as **GI** (ie, the generalized inference approach).

To improve the performance of proposed confidence region, we can use some monotonic transformations such as arcsine-square-root and the logit transformations. More details can be found in Appendix A. However, simulation results indicate that these transformations are not beneficial for generalized inference methods under normality. Hence such transformations are not pursued further under normality.

3.2 | Without normality: Box-Cox transformation

Box-Cox transformation, widely used in ROC analysis,[6,33,34] is a standard approach to transform original data to achieve normality when normality assumption is not satisfied. Due to that fact that both HUM_K and J_K are invariant under monotonic transformations, Box-Cox transformation can also be used here.

For the jth $(j = 1, \ldots, n_i)$ subject in the ith group $(i = 1, 2, \ldots, K)$, let Y_{ij} denote the variable and $Y_{ij}^{(\lambda)}$ the transformed variable. The Box-Cox transformation is constructed as:

$$Y_{ij}^{(\lambda)} = \begin{cases} \frac{Y_{ij}^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(Y_{ij}), & \lambda = 0, \end{cases}$$

where it is assumed that $Y_{ij}^{(\lambda)} \overset{iid}{\sim} N\left(\mu_i, \sigma_i^2\right)$. The log-likelihood function can be written as:

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left[-\frac{1}{2}\log\left(2\pi\sigma_i^2\right) - \frac{\left(Y_{ij}^{(\lambda)} - \mu_i\right)^2)}{2\sigma_i^2} + (\lambda - 1)\log(Y_{ij})\right].$$

The maximum likelihood estimate of λ can be obtained by maximizing the above log-likelihood function. After transforming data using Box-Cox approach, the generalized inference method for normal data presented in Section 3.1 can be applied on the transformed data $Y_{ij}^{(\lambda)}$ for confidence region estimation of (HUM_K, J_K). Such obtained confidence region is denoted as **BCGI** (ie, the generalized inference approach with Box-Cox transformation).

3.3 | Under gamma distribution

In practice, biomarker measurements often can be continuous and positively skewed. It is well known that gamma distribution is a popular option for modeling positively skewed data. Hence we present some direct generalized inference methods for constructing joint inference of (HUM_K, J_K) under gamma distribution. While the generalized inference method based on Box-Cox transformed data presented in Section 3.2 is an option for handling gamma data, direct methods that can handle gamma data is more desirable as it is more convenient. Since the exact fiducial quantity for shape parameter (α) and rate parameter (β) in gamma distribution are not available, three approximate pivotal quantities for α and β have been proposed in the literature.[35-37] In the following, we will briefly review them.

Let $Y_{i,1}, \ldots, Y_{i,n_i}$ be an iid random sample from $G(\alpha_i, \beta_i)$ for the ith class $(i = 1, 2, \ldots, K)$. Let $\overline{Y}_i$ and $\tilde{Y}_i$ stand for the arithmetic mean and geometric mean, and $\bar{y}_i$ and $\tilde{y}_i$ be the observed values of $\overline{Y}_i$ and $\tilde{Y}_i$, respectively.

Chen and Ye's method[35]: It is known that $2n\alpha_i \log\left(\overline{Y}_i/\tilde{Y}_i\right) \sim \alpha_i\chi^2_{v_i}$ approximately, where $v_i = 2E^2(W_i)/\mathrm{var}(W_i)$ and $c_i = E(W_i/v_i)$. The detailed formulas for $E(W_i)$ and $\mathrm{var}(W_i)$ can be found in Chen and Ye.[35] Using this result, an approximate generalized pivotal quantity for α_i can be written as

$$R_{\alpha_i} = \frac{W_i}{2n_i\log(\bar{y}_i/\tilde{y}_i)},$$

where $W_i \sim c_i\chi^2_{(v_i)}$. Furthermore, utilizing a well-known result regarding gamma distribution, that is, $2n_i\beta_i\overline{Y}_i \sim \chi^2_{2n_i\alpha_i}$, the generalized pivot quantity for β_i can be written as

$$R_{\beta_i} = \frac{U_i}{2n_i\bar{y}_i}, \tag{14}$$

where $U_i \sim \chi^2_{2n_iR_{\alpha_i}}$.

Wang and Wu's method[36]: Let $T_i = \log\left(\tilde{Y}_i/\overline{Y}_i\right), i = 1, 2, \ldots, K$. Note that $U_i = F_i(.) \sim U(0, 1)$, where $F_i(.)$ is the c.d.f of T_i. On the basis of Cornish-Fisher expansion, the U_i percentile of T can be approximated by $\kappa_1(\alpha_i) + [\kappa_2(\alpha_i)]^{1/2}Q(\alpha_i, U_i)$, where $\kappa_j(\alpha_i)$ is the kth cumulant of T and $Q(\alpha_i, U_i)$ is a function of $\kappa_j(\alpha_i)$. The detailed formulas can be found in Wang and Wu.[36] Let t denote the observed value of T. An approximate generalized pivotal quantity for α_i, that is, R_{α_i}, can be obtained by solving $t = \kappa_1(\alpha_i) + [\kappa_2(\alpha_i)]^{1/2}Q(\alpha_i, u)$. Similar to Chen and Ye's method, the approximate generalized pivotal quantity for rate parameter, R_{β_i}, can be obtained by (14).

Krishnamoorthy and Wang's method[37,38]: By applying the Wilson-Hilferty normal approximation, that is, $Y_{i,j}^{1/3} \sim N(\mu_i, \sigma_i)$, $j = 1, \ldots, n_i$, generalized pivotal quantities for normal mean and variance, R_{μ_i} and R_{σ_i} can be obtained for transformed data. The GPQs for α_i and β_i can be further expressed as:

$$R_{\alpha_i} = \frac{1}{9}\left\{\left(1 + 0.5\frac{R^2_{\mu_i}}{R^2_{\sigma_i}}\right) + \left[\left(1 + 0.5\frac{R^2_{\mu_i}}{R^2_{\sigma_i}}\right)^2 - 1\right]^{1/2}\right\},$$

$$R_{\beta_i} = \frac{1}{27(R_{\alpha_i})^{1/2}(R_{\sigma_i^2})^{3/2}}.$$

Under gamma distribution, the generalized pivot for HUM_K and J_K in (1) and (4) can be obtained by substituting α_i's and β_i's in the p.d.f. and c.d.f. of gamma distribution (see Section 2.2) with the corresponding R_{α_i}'s and R_{β_i}'s obtained from three approximate generalized inference methods presented above. Take $K = 3$ as an example, replacing α_i's and β_i's in (9) and (10) with R_{α_i}'s and R_{β_i}'s, we obtain approximate generalized pivots R_{VUS} and R_{J_3} for VUS and J_3, respectively. Then following similar steps as presented in Section 3.1 for normal distribution, we can obtain the confidence region for $\eta = (\text{HUM}_K, J_K)^T$ under gamma distribution. We refer these three generalized inference methods for gamma distribution as: **GammaGI1** (based on Chen and Ye's method), **GammaGI2** (based on Wang and Wu's method) and **GammaGI3** (based on Krishnamoorthy and Wang's method).

4 | NON-PARAMETRIC CONFIDENCE REGION ESTIMATION

In Section 3, we considered parametric inference methods for confidence region estimation under normal distribution and gamma distribution. When no distributional assumptions can be made or data transformation is not able to achieve normality under both groups, HUM_K and generalized Youden index (J_K) can be estimated using non-parametric approaches. In the following, we investigate some non-parametric bootstrapping approaches for estimating confidence region of $\eta = (\text{HUM}_K, J_K)^T$.

Given an observed data set, let y_{i,j_i} $(i = 1, 2, \dots, K, j_i = 1, 2, \dots, n_i)$ stand for the measure for the j_ith subject in ith group. The empirical estimate of HUM_K can be given as

$$\widehat{\text{HUM}}_K = \frac{\sum_{j_1=1}^{n_1}\sum_{j_2=1}^{n_2}\cdots\sum_{j_K=1}^{n_K} I(y_{1,j_1} \le y_{2,j_2} \le \cdots \le y_{K,j_K})}{n_1 n_2 \cdots n_K}. \tag{15}$$

And the empirical estimate of generalized Youden index can be given by

$$\hat{J}_K = \frac{1}{K-1}\sum_{i=1}^{K-1}\left[\frac{1}{n_i}\sum_{j_i=1}^{n_i} I(y_{i,j_i} \le \hat{c}_i) - \frac{1}{n_{i+1}}\sum_{j_{i+1}=1}^{n_{i+1}} I(y_{i+1,j_{i+1}} \le \hat{c}_i)\right], \tag{16}$$

where $\hat{c}_i$ $(i = 1, 2, \dots, K-1)$ is the optimal empirical cut-point that gives the maximum of $\frac{1}{n_i}\sum_{j_i=1}^{n_i} I(y_{i,j_i} \le \hat{c}_i) - \frac{1}{n_{i+1}}\sum_{j_{i+1}=1}^{n_{i+1}} I(y_{i+1,j_{i+1}} \le \hat{c}_i)$.

Denote $\hat{\eta}_0 = (\widehat{\text{HUM}}_K, \hat{J}_K)^T$ for the observed data. Given a specific data set, the non-parametric bootstrap joint confidence region for $\eta = (\text{HUM}_K, J_K)^T$ can be obtained through following steps: (1) empirically search for the optimal cut-points, $\hat{c}_i$, where $i = 1, 2, \dots, K-1$; (2) draw a bootstrap sample of size n_i $(i = 1, 2 \dots, K)$ from ith sample, and calculate $\widehat{\text{HUM}}_K$ and $\hat{J}_K$ following (15) and (16); (3) repeat step 2 for $B = 500$ times to acquire a set of $\hat{\eta}^b = \left(\widehat{\text{HUM}}_K^b, \hat{J}_K^b\right)^T (b = 1, 2, \dots, B)$, then calculate the bootstrap sample mean $\bar{\hat{\eta}}^B = \frac{1}{B}\sum_{b=1}^{B}\hat{\eta}^b$ and sample covariance matrix $\hat{\Sigma}^B = \frac{1}{B-1}\left(\hat{\eta}^b - \bar{\hat{\eta}}^b\right)\left(\hat{\eta}^b - \bar{\hat{\eta}}^b\right)$; 4) Calculate $\tilde{\hat{\eta}}^b = \left(\hat{\Sigma}^B\right)^{-1/2}\left(\hat{\eta}^b - \bar{\hat{\eta}}^B\right)$, and its length $\|\tilde{\hat{\eta}}^b\|$ for $b = 1, 2, \dots, B$.

Denote $q_{\{\|\tilde{\hat{\eta}}\|;1-\alpha\}}$ as the $100(1-\alpha)$th percentile of the set $\|\tilde{\hat{\eta}}^b\|$ $(b = 1, 2, \dots, B)$. The $100(1-\alpha)\%$ generalized confidence region of $\eta = (\text{HUM}_K, J_K)^T$ is

$$\left\{\eta : \left(\eta - \bar{\hat{\eta}}^B\right)^T\left(\hat{\Sigma}^B\right)^{-1}\left(\eta - \bar{\hat{\eta}}^B\right) \le q^2_{\{\|\tilde{\hat{\eta}}\|;1-\alpha\}}\right\}.$$

The area of confidence region is estimated by $A_{BTI} = \pi\left(q^2_{\{\|\tilde{\hat{\eta}}\|;1-\alpha\}}\right)\sqrt{|\hat{\Sigma}^B|}$ where $|\hat{\Sigma}^B|$ is the determinant of $\hat{\Sigma}^B$ in step 3. This approach is referred as **BTI**.

Using $\hat{\eta}_0 = \left(\widehat{\text{HUM}}_K, \hat{J}_K\right)^T$ to replace $\bar{\hat{\eta}}^B$ in **BTI** leads to another bootstrap method, that is, **BTII**. Simulation studies indicate that the performance of **BTI** and **BTII** are very similar with **BTI** being slightly better. Hence we will only focus on **BTI** hereafter.

Similar to parametric methods, we can use monotone transformation, arcsine-square-root and the logit transformation to construct the confidence region for the transformed $h(\eta) = (h(\text{HUM}_K), h(J_K))^T$ (the details can be found in Appendix A). The bootstrap confidence region with the logit transformation is referred as **BTLT**, and the one with arcsine-square-root transformation as **BTAT**.

In summary, we propose three non-parametric confidence regions in this section: **BTI** (bootstrap confidence region), **BTLT** (bootstrap confidence region with logit transformation), and **BTAT** (bootstrap confidence region with arcsine-square-root transformation).

5 | SIMULATION RESULTS

The proposed methods are evaluated by simulation studies under normality and gamma distribution with three classes ($K = 3$ for HUM_K and J_K). Table 1 presents parameter settings for simulation study under normality and gamma distribution, and density plots can be found in "Supplemental materials." The simulation settings include biomarkers with wide range of values for VUS and J_3. The settings under normality include different means and variances structures across classes, while the settings under Gamma distribution also accommodates different skewness structures. For each parameter setting, 2000 samples are simulated. For generalized confidence regions, we set $B = 2500$. For bootstrap methods, $B = 500$ bootstrap samples are used.

Table 2 presents estimated coverage probabilities at the nominal level of 95% and the estimated areas of the confidence regions under normality from small to large sample sizes. The generalized inference method **GI** generally achieves satisfactory coverage probabilities while it could be slightly conservative when sample sizes are smaller than (50, 50, 50) for some settings, especially for large VUS and J_3 (eg, VUS = 0.843, and $J_3 = 0.683$). The confidence regions by **BTI** are generally liberal, especially when sample sizes are smaller than (50, 50, 50). Both logit transformation (**BTLT**) and arcsine-square-root transformation (**BTAT**) greatly improve coverage probabilities. For some scenarios, **BTLT** tends to be slightly conservative when sample sizes are smaller than (30, 30, 30). In terms of area of confidence regions, the generalized inference methods yield smaller areas than all the bootstrap methods. Among non-parametric bootstrap methods,

TABLE 1 Simulation settings ($K = 3$).

	Parameter settings				
Scenario	**Class 1**	**Class 2**	**Class 3**	**VUS**	J_3
Normal 1	N(0.0, 0.5)	N(1.0, 0.5)	N(2.0, 0.5)	0.843	0.683
Normal 2	N(0.0, 0.5)	N(0.5, 0.5)	N(1.0, 0.5)	0.684	0.533
Normal 3	N(0.0, 1.0)	N(1.0, 1.0)	N(2.0, 1.0)	0.536	0.383
Normal 4	N(0.0, 1.0)	N(1.0, 1.2)	N(2.0, 1.4)	0.471	0.327
Normal 5	N(0.0, 1.0)	N(0.5, 1.0)	N(1.5, 1.0)	0.430	0.290
Normal 6	N(0.0, 1.0)	N(0.5, 1.2)	N(1.5, 1.4)	0.378	0.241
Normal 7	N(0.0, 1.0)	N(0.5, 1.0)	N(1.0, 1.0)	0.337	0.197
Normal 8	N(0.0, 1.0)	N(0.5, 1.2)	N(1.0, 1.4)	0.306	0.167
Normal 9	N(0.0, 1.0)	N(0.5, 1.0)	N(0.8, 1.0)	0.298	0.158
Normal 10	N(0.0, 1.0)	N(0.5, 1.2)	N(0.8, 1.4)	0.277	0.137
Gamma 1	G(3.0, 2.0)	G(3.0, 1.0)	G(3.0, 0.2215)	0.747	0.614
Gamma 2	G(3.0, 2.0)	G(3.0, 1.0)	G(3.0, 0.401)	0.651	0.500
Gamma 3	G(2.0, 1.0)	G(3.0, 1.0)	G(5.0, 0.5)	0.640	0.520
Gamma 4	G(3.0, 2.0)	G(3.0, 1.0)	G(3.0, 0.627)	0.514	0.372
Gamma 5	G(1.0, 2.0)	G(2.5, 1.8)	G(3.0, 1.6)	0.505	0.365
Gamma 6	G(2.0, 2.0)	G(2.0, 1.0)	G(4.0, 1.0)	0.409	0.294
Gamma 7	G(1.0, 3.0)	G(1.5, 2.0)	G(2.0, 1.0)	0.379	0.236
Gamma 8	G(1.0, 3.0)	G(1.0, 2.0)	G(2.0, 2.0)	0.346	0.225
Gamma 9	G(1.5, 1.0)	G(1.5, 0.8)	G(2.0, 0.8)	0.273	0.136
Gamma 10	G(2.0, 4.0)	G(2.0, 2.0)	G(2.0, 1.0)	0.212	0.060

TABLE 2 Coverage probabilities (and average area) of proposed 95% confidence regions for (VUS, J_3) under normality (2000 simulations).

	GI	BTI	BTLT	BTAT
Sample size	**Coverage probability (area of confidence region)**			
	Normal 1: $(VUS, J_3) = (0.843, 0.683)$			
(20,20,20)	0.965 (0.020)	0.905 (0.051)	0.969 (0.058)	0.932 (0.053)
(30,30,30)	0.971 (0.011)	0.922 (0.033)	0.968 (0.038)	0.940 (0.035)
(50,50,50)	0.973 (0.005)	0.929 (0.019)	0.954 (0.020)	0.939 (0.019)
(20,30,50)	0.977 (0.011)	0.921 (0.034)	0.959 (0.037)	0.940 (0.035)
(80,80,80)	0.982 (0.002)	0.934 (0.011)	0.954 (0.012)	0.943 (0.012)
(100,100,100)	0.977 (0.002)	0.926 (0.009)	0.945 (0.009)	0.932 (0.009)
(120,120,120)	0.984 (0.001)	0.943 (0.008)	0.953 (0.008)	0.948 (0.008)
(100,150,100)	0.977 (0.001)	0.940 (0.008)	0.940 (0.008)	0.940 (0.008)
	Normal 2: $(VUS, J_3) = (0.684, 0.533)$			
(20,20,20)	0.967 (0.032)	0.921 (0.077)	0.964 (0.086)	0.942 (0.081)
(30,30,30)	0.960 (0.019)	0.928 (0.050)	0.953 (0.054)	0.941 (0.052)
(50,50,50)	0.965 (0.010)	0.933 (0.030)	0.957 (0.031)	0.946 (0.030)
(20,30,50)	0.967 (0.022)	0.928 (0.055)	0.954 (0.060)	0.939 (0.057)
(80,80,80)	0.963 (0.006)	0.945 (0.018)	0.954 (0.019)	0.949 (0.019)
(100,100,100)	0.947 (0.005)	0.948 (0.015)	0.957 (0.015)	0.951 (0.015)
(120,120,120)	0.957 (0.004)	0.944 (0.012)	0.947 (0.012)	0.945 (0.012)
(100,150,100)	0.956 (0.004)	0.940 (0.012)	0.946 (0.013)	0.944 (0.012)
	Normal 3: $(VUS, J_3) = (0.536, 0.383)$			
(20,20,20)	0.951 (0.030)	0.934 (0.085)	0.960 (0.107)	0.951 (0.090)
(30,30,30)	0.953 (0.017)	0.930 (0.055)	0.949 (0.061)	0.945 (0.057)
(50,50,50)	0.954 (0.009)	0.940 (0.032)	0.951 (0.034)	0.945 (0.033)
(20,30,50)	0.955 (0.018)	0.933 (0.057)	0.954 (0.062)	0.947 (0.059)
(80,80,80)	0.961 (0.005)	0.936 (0.020)	0.947 (0.020)	0.941 (0.020)
(100,100,100)	0.959 (0.003)	0.940 (0.016)	0.942 (0.016)	0.939 (0.016)
(120,120,120)	0.967 (0.003)	0.955 (0.013)	0.960 (0.013)	0.958 (0.013)
(100,150,100)	0.969 (0.002)	0.944 (0.014)	0.950 (0.014)	0.947 (0.014)
	Normal 4: $(VUS, J_3) = (0.471, 0.327)$			
(20,20,20)	0.962 (0.033)	0.926 (0.085)	0.955 (0.124)	0.944 (0.091)
(30,30,30)	0.960 (0.019)	0.942 (0.056)	0.955 (0.066)	0.953 (0.059)
(50,50,50)	0.959 (0.010)	0.946 (0.033)	0.960 (0.035)	0.953 (0.034)
(20,30,50)	0.953 (0.019)	0.928 (0.055)	0.947 (0.062)	0.941 (0.058)
(80,80,80)	0.957 (0.006)	0.949 (0.020)	0.960 (0.021)	0.952 (0.020)
(100,100,100)	0.956 (0.004)	0.946 (0.016)	0.957 (0.016)	0.954 (0.016)
(120,120,120)	0.958 (0.003)	0.953 (0.013)	0.962 (0.014)	0.960 (0.013)
(100,150,100)	0.955 (0.004)	0.944 (0.014)	0.952 (0.014)	0.946 (0.014)

(Continues)

TABLE 2 (Continued)

Sample size	GI	BTI	BTLT	BTAT
	Coverage probability (area of confidence region)			
	Normal 5: $(VUS, J_3) = (0.430, 0.419)$			
(20,20,20)	0.962 (0.035)	0.929 (0.084)	0.951 (0.134)	0.950 (0.092)
(30,30,30)	0.966 (0.020)	0.932 (0.055)	0.955 (0.070)	0.955 (0.058)
(50,50,50)	0.966 (0.010)	0.934 (0.033)	0.946 (0.036)	0.943 (0.034)
(20,30,50)	0.966 (0.023)	0.930 (0.059)	0.954 (0.069)	0.948 (0.063)
(80,80,80)	0.962 (0.006)	0.938 (0.020)	0.952 (0.021)	0.945 (0.020)
(100,100,100)	0.958 (0.005)	0.935 (0.016)	0.949 (0.016)	0.942 (0.016)
(120,120,120)	0.957 (0.004)	0.944 (0.013)	0.953 (0.014)	0.949 (0.013)
(100,150,100)	0.962 (0.004)	0.939 (0.014)	0.946 (0.014)	0.943 (0.014)
	Normal 6: $(VUS, J_3) = (0.378, 0.241)$			
(20,20,20)	0.956 (0.034)	0.940 (0.080)	0.965 (0.167)	0.960 (0.087)
(30,30,30)	0.959 (0.019)	0.935 (0.054)	0.955 (0.079)	0.952 (0.057)
(50,50,50)	0.963 (0.010)	0.938 (0.032)	0.957 (0.038)	0.957 (0.033)
(20,30,50)	0.960 (0.020)	0.926 (0.054)	0.955 (0.066)	0.949 (0.058)
(80,80,80)	0.969 (0.006)	0.951 (0.020)	0.965 (0.021)	0.960 (0.020)
(100,100,100)	0.958 (0.005)	0.957 (0.016)	0.965 (0.016)	0.963 (0.016)
(120,120,120)	0.964 (0.004)	0.944 (0.013)	0.962 (0.013)	0.953 (0.013)
(100,150,100)	0.963 (0.004)	0.946 (0.013)	0.963 (0.014)	0.955 (0.014)
	Normal 7: $(VUS, J_3) = (0.337, 0.197)$			
(20,20,20)	0.954 (0.036)	0.933 (0.073)	0.955 (0.170)	0.955 (0.084)
(30,30,30)	0.961 (0.020)	0.942 (0.049)	0.950 (0.087)	0.957 (0.055)
(50,50,50)	0.965 (0.009)	0.927 (0.031)	0.936 (0.042)	0.942 (0.031)
(20,30,50)	0.962 (0.021)	0.895 (0.044)	0.963 (0.056)	0.939 (0.048)
(80,80,80)	0.956 (0.004)	0.939 (0.018)	0.945 (0.021)	0.947 (0.019)
(100,100,100)	0.958 (0.003)	0.941 (0.014)	0.959 (0.015)	0.955 (0.014)
(120,120,120)	0.948 (0.003)	0.950 (0.012)	0.955 (0.013)	0.955 (0.012)
(100,150,100)	0.947 (0.003)	0.944 (0.013)	0.948 (0.014)	0.948 (0.013)
	Normal 8: $(VUS, J_3) = (0.306, 0.167)$			
(20,20,20)	0.956 (0.037)	0.944 (0.070)	0.968 (0.177)	0.964 (0.082)
(30,30,30)	0.952 (0.021)	0.930 (0.047)	0.962 (0.092)	0.958 (0.053)
(50,50,50)	0.959 (0.011)	0.938 (0.032)	0.951 (0.034)	0.943 (0.033)
(20,30,50)	0.959 (0.021)	0.933 (0.046)	0.964 (0.064)	0.951 (0.051)
(80,80,80)	0.961 (0.006)	0.932 (0.018)	0.954 (0.022)	0.953 (0.019)
(100,100,100)	0.953 (0.005)	0.941 (0.014)	0.961 (0.017)	0.962 (0.015)
(120,120,120)	0.956 (0.004)	0.938 (0.012)	0.956 (0.013)	0.951 (0.012)
(100,150,100)	0.961 (0.004)	0.930 (0.013)	0.949 (0.014)	0.946 (0.013)

(Continues)

TABLE 2 (Continued)

Sample size	GI	BTI	BTLT	BTAT
	Coverage probability (area of confidence region)			
	Normal 9: $(VUS, J_3) = (0.298, 0.158)$			
(20,20,20)	0.936 (0.038)	0.936 (0.068)	0.953 (0.176)	0.949 (0.079)
(30,30,30)	0.950 (0.022)	0.931 (0.045)	0.955 (0.096)	0.952 (0.052)
(50,50,50)	0.951 (0.011)	0.943 (0.027)	0.949 (0.050)	0.955 (0.030)
(20,30,50)	0.957 (0.022)	0.936 (0.046)	0.955 (0.064)	0.950 (0.050)
(80,80,80)	0.957 (0.005)	0.944 (0.017)	0.944 (0.023)	0.953 (0.018)
(100,100,100)	0.958 (0.003)	0.943 (0.014)	0.955 (0.017)	0.952 (0.014)
(120,120,120)	0.939 (0.003)	0.941 (0.012)	0.939 (0.013)	0.941 (0.012)
(100,150,100)	0.951 (0.003)	0.943 (0.012)	0.947 (0.014)	0.951 (0.013)
	Normal 10: $(VUS, J_3) = (0.277, 0.140)$			
(20,20,20)	0.945 (0.038)	0.960 (0.068)	0.960 (0.201)	0.962 (0.075)
(30,30,30)	0.949 (0.023)	0.961 (0.044)	0.961 (0.109)	0.963 (0.048)
(50,50,50)	0.959 (0.012)	0.956 (0.027)	0.959 (0.059)	0.964 (0.029)
(20,30,50)	0.951 (0.022)	0.952 (0.044)	0.954 (0.068)	0.952 (0.048)
(80,80,80)	0.964 (0.007)	0.954 (0.017)	0.959 (0.026)	0.965 (0.018)
(100,100,100)	0.956 (0.006)	0.948 (0.014)	0.956 (0.018)	0.962 (0.014)
(120,120,120)	0.956 (0.005)	0.937 (0.011)	0.960 (0.014)	0.961 (0.012)
(100,150,100)	0.956 (0.005)	0.952 (0.012)	0.967 (0.015)	0.966 (0.013)

BTI gives the the smallest area, and for most scenarios, it is followed by **BTAT**. **BTAT** yields similar estimated areas as **BTI** when sample sizes are as large as (100,100, 100).

Table 3 presents coverage probabilities at 95% nominal level and estimated confidence regions under gamma assumption. The three approximate generalized inference methods (**GammaGI1**, **GammaGI2**, and **GammaGI3**) maintain satisfactory coverage probabilities under most parameter settings while they can be liberal at small sizes under certain settings. The generalized inference method with Box-Cox transformation (ie, **BCGI**) is liberal for most scenarios, especially the ones with large VUS (eg, $VUS = 0.747$). The **BTI** method is generally liberal, especially when sample sizes are from small to medium. Both logit transformation (**BTLT**) and arcsine-square-root transformation (**BTAT**) greatly improve coverage probabilities, and they are generally satisfactory except that **BTAT** can be slightly liberal for some scenarios at small sample sizes. Overall, the three approximate generalized inference methods (**GammaGI1**, **GammaGI2**, and **GammaGI3**) achieve the smallest average area of confidence region, followed by the generalized inference method based on Box-Cox transformation (**BCGI**). Among non-parametric methods, **BTLT** yields largest areas in comparison to **BTI** and **BTAT**.

In summary, generalized inference method (**GI**) has the most accurate coverage probabilities and the smallest average area among all proposed methods under normal assumption. With gamma assumption, **GammaGI2**, and **GammaGI3** are good options because they have satisfactory coverage probabilities and smallest average area among the proposed methods. Without any distribution assumptions, both **BTLT** and **BTAT** are reasonable options and **BTAT** generally have slightly smaller area than **BTLT**.

6 | DATA EXAMPLE

ADNI is a global collaborative research project established to develop clinical, imaging, genetic, and biochemical biomarkers for early detection and tracking of Alzheimer's disease (AD). ADNI includes more than 800 participants aging from 55 to 90, recruited from over 50 sites across United States and Canada (publicly accessible

TABLE 3 Coverage probabilities (and average area) of proposed 95% confidence regions for (VUS, J_3) under gamma distributions (2000 simulations).

	GammaGI1	GammaGI2	GammaGI3	BCGI	BTI	BTLT	BTAT
Sample size	**Coverage probability (area of confidence region)**						
	Gamma 1: $(VUS, J_3) = (0.747, 0.614)$						
(20,20,20)	0.965 (0.027)	0.951 (0.025)	0.948 (0.026)	0.905 (0.031)	0.902 (0.072)	0.961 (0.080)	0.929 (0.075)
(30,30,30)	0.962 (0.017)	0.949 (0.017)	0.956 (0.020)	0.901 (0.017)	0.927 (0.047)	0.955 (0.051)	0.939 (0.049)
(50,50,50)	0.960 (0.009)	0.958 (0.009)	0.952 (0.009)	0.898 (0.009)	0.932 (0.028)	0.953 (0.029)	0.945 (0.029)
(20,30,50)	0.968 (0.018)	0.952 (0.016)	0.946 (0.017)	0.927 (0.020)	0.908 (0.055)	0.933 (0.060)	0.916 (0.056)
(80,80,80)	0.966 (0.005)	0.955 (0.005)	0.959 (0.006)	0.925 (0.004)	0.941 (0.017)	0.949 (0.018)	0.945 (0.018)
(100,100,100)	0.947 (0.004)	0.955 (0.004)	0.957 (0.004)	0.912 (0.004)	0.937 (0.014)	0.945 (0.014)	0.940 (0.014)
(120,120,120)	0.957 (0.004)	0.956 (0.003)	0.956 (0.004)	0.925 (0.003)	0.938 (0.012)	0.947 (0.012)	0.940 (0.012)
(100,150,100)	0.958 (0.004)	0.956 (0.003)	0.963 (0.004)	0.933 (0.003)	0.931 (0.012)	0.939 (0.012)	0.937 (0.012)
	Gamma 2: $(VUS, J_3) = (0.651, 0.500)$						
(20,20,20)	0.963 (0.030)	0.958 (0.027)	0.946 (0.028)	0.922 (0.040)	0.931 (0.085)	0.966 (0.096)	0.948 (0.090)
(30,30,30)	0.963 (0.017)	0.949 (0.016)	0.942 (0.029)	0.906 (0.018)	0.938 (0.056)	0.963 (0.060)	0.951 (0.057)
(50,50,50)	0.961 (0.009)	0.957 (0.008)	0.960 (0.008)	0.930 (0.008)	0.945 (0.033)	0.964 (0.034)	0.952 (0.033)
(20,30,50)	0.968 (0.017)	0.941 (0.016)	0.951 (0.017)	0.929 (0.019)	0.931 (0.058)	0.961 (0.062)	0.946 (0.060)
(80,80,80)	0.969 (0.005)	0.951 (0.004)	0.954 (0.010)	0.930 (0.004)	0.945 (0.020)	0.957 (0.021)	0.951 (0.020)
(100,100,100)	0.956 (0.004)	0.943 (0.003)	0.945 (0.006)	0.950 (0.003)	0.949 (0.016)	0.958 (0.016)	0.954 (0.016)
(120,120,120)	0.961 (0.003)	0.967 (0.003)	0.947 (0.006)	0.925 (0.003)	0.953 (0.013)	0.958 (0.014)	0.954 (0.014)
(100,150,100)	0.966 (0.003)	0.958 (0.003)	0.959 (0.006)	0.940 (0.003)	0.945 (0.013)	0.951 (0.014)	0.949 (0.014)
	Gamma 3: $(VUS, J_3) = (0.640, 0.520)$						
(20,20,20)	0.962 (0.040)	0.948 (0.038)	0.955 (0.039)	0.963 (0.077)	0.913 (0.083)	0.954 (0.092)	0.931 (0.086)
(30,30,30)	0.959 (0.025)	0.950 (0.027)	0.904 (0.019)	0.962 (0.039)	0.928 (0.055)	0.958 (0.058)	0.943 (0.056)
(50,50,50)	0.961 (0.015)	0.955 (0.014)	0.959 (0.014)	0.956 (0.017)	0.930 (0.033)	0.950 (0.034)	0.942 (0.033)
(20,30,50)	0.961 (0.029)	0.950 (0.024)	0.913 (0.019)	0.959 (0.055)	0.912 (0.063)	0.948 (0.067)	0.931 (0.066)
(80,80,80)	0.957 (0.009)	0.950 (0.009)	0.926 (0.007)	0.942 (0.009)	0.944 (0.020)	0.954 (0.021)	0.948 (0.02 0)
(100,100,100)	0.953 (0.007)	0.949 (0.007)	0.950 (0.007)	0.950 (0.007)	0.932 (0.016)	0.942 (0.016)	0.935 (0.016)
(120,120,120)	0.956 (0.006)	0.958 (0.006)	0.930 (0.005)	0.960 (0.006)	0.935 (0.013)	0.945 (0.014)	0.943 (0.014)
(100,150,100)	0.957 (0.006)	0.958 (0.006)	0.936 (0.005)	0.950 (0.006)	0.945 (0.014)	0.940 (0.014)	0.935 (0.014)
	Gamma 4: $(VUS, J_3) = (0.514, 0.372)$						
(20,20,20)	0.949 (0.033)	0.931 (0.031)	0.927 (0.032)	0.938 (0.087)	0.934 (0.089)	0.968 (0.110)	0.955 (0.094)
(30,30,30)	0.952 (0.020)	0.938 (0.019)	0.943 (0.019)	0.941 (0.047)	0.948 (0.058)	0.963 (0.064)	0.958 (0.061)
(50,50,50)	0.959 (0.011)	0.943 (0.010)	0.959 (0.010)	0.936 (0.017)	0.941 (0.034)	0.952 (0.036)	0.945 (0.035)
(20,30,50)	0.956 (0.020)	0.935 (0.019)	0.943 (0.019)	0.950 (0.044)	0.947 (0.057)	0.963 (0.062)	0.957 (0.059)
(80,80,80)	0.958 (0.006)	0.942 (0.006)	0.952 (0.013)	0.931 (0.007)	0.943 (0.021)	0.953 (0.022)	0.948 (0.022)
(100,100,100)	0.949 (0.005)	0.947 (0.004)	0.955 (0.004)	0.928 (0.005)	0.944 (0.017)	0.952 (0.017)	0.947 (0.017)
(120,120,120)	0.959 (0.004)	0.961 (0.004)	0.946 (0.009)	0.925 (0.004)	0.949 (0.014)	0.958 (0.014)	0.955 (0.014)
(100,150,100)	0.962 (0.004)	0.957 (0.004)	0.958 (0.010)	0.926 (0.004)	0.936 (0.014)	0.944 (0.015)	0.941 (0.014)

(Continues)

TABLE 3 (Continued)

Sample size	GammaGI1	GammaGI2	GammaGI3	BCGI	BTI	BTLT	BTAT
	Coverage probability (area of confidence region)						
	Gamma 5: $(VUS, J_3) = (0.505, 0.365)$						
(20,20,20)	0.924 (0.035)	0.934 (0.033)	0.944 (0.035)	0.963 (0.103)	0.920 (0.086)	0.950 (0.110)	0.943 (0.093)
(30,30,30)	0.953 (0.022)	0.936 (0.021)	0.941 (0.022)	0.957 (0.065)	0.937 (0.057)	0.956 (0.065)	0.951 (0.059)
(50,50,50)	0.961 (0.012)	0.957 (0.011)	0.962 (0.012)	0.953 (0.031)	0.946 (0.034)	0.958 (0.036)	0.956 (0.035)
(20,30,50)	0.949 (0.021)	0.941 (0.020)	0.936 (0.021)	0.960 (0.058)	0.942 (0.055)	0.964 (0.061)	0.955 (0.058)
(80,80,80)	0.970 (0.007)	0.956 (0.007)	0.962 (0.007)	0.955 (0.013)	0.939 (0.021)	0.947 (0.021)	0.943 (0.021)
(100,100,100)	0.956 (0.005)	0.960 (0.006)	0.957 (0.005)	0.949 (0.008)	0.937 (0.017)	0.947 (0.017)	0.941 (0.017)
(120,120,120)	0.962 (0.004)	0.963 (0.004)	0.960 (0.004)	0.940 (0.006)	0.941 (0.014)	0.945 (0.014)	0.942 (0.014)
(100,150,100)	0.963 (0.004)	0.958 (0.004)	0.968 (0.004)	0.935 (0.006)	0.945 (0.014)	0.953 (0.015)	0.947 (0.014)
	Gamma 6: $(VUS, J_3) = (0.409, 0.294)$						
(20,20,20)	0.954 (0.037)	0.943 (0.033)	0.950 (0.033)	0.943 (0.087)	0.932 (0.089)	0.969 (0.103)	0.958 (0.094)
(30,30,30)	0.962 (0.022)	0.950 (0.020)	0.961 (0.020)	0.949 (0.045)	0.942 (0.058)	0.963 (0.063)	0.954 (0.060)
(50,50,50)	0.966 (0.012)	0.967 (0.011)	0.970 (0.011)	0.953 (0.017)	0.938 (0.034)	0.956 (0.035)	0.948 (0.035)
(20,30,50)	0.958 (0.023)	0.957 (0.021)	0.956 (0.022)	0.942 (0.053)	0.916 (0.061)	0.960 (0.066)	0.937 (0.063)
(80,80,80)	0.968 (0.007)	0.956 (0.007)	0.968 (0.007)	0.938 (0.007)	0.936 (0.021)	0.951 (0.022)	0.943 (0.021)
(100,100,100)	0.965 (0.006)	0.952 (0.005)	0.953 (0.005)	0.940 (0.005)	0.938 (0.017)	0.946 (0.017)	0.942 (0.017)
(120,120,120)	0.956 (0.005)	0.959 (0.004)	0.952 (0.004)	0.936 (0.004)	0.943 (0.014)	0.951 (0.014)	0.947 (0.014)
(100,150,100)	0.965 (0.005)	0.943 (0.004)	0.958 (0.004)	0.940 (0.004)	0.941 (0.014)	0.946 (0.014)	0.943 (0.014)
	Gamma 7: $(VUS, J_3) = (0.379, 0.236)$						
(20,20,20)	0.948 (0.034)	0.942 (0.030)	0.956 (0.031)	0.921 (0.031)	0.930 (0.083)	0.961 (0.101)	0.953 (0.089)
(30,30,30)	0.966 (0.020)	0.959 (0.017)	0.956 (0.018)	0.919 (0.017)	0.942 (0.054)	0.964 (0.059)	0.956 (0.056)
(50,50,50)	0.967 (0.010)	0.960 (0.008)	0.964 (0.009)	0.910 (0.009)	0.947 (0.032)	0.961 (0.033)	0.953 (0.032)
(20,30,50)	0.964 (0.021)	0.952 (0.018)	0.957 (0.019)	0.932 (0.019)	0.943 (0.056)	0.962 (0.061)	0.953 (0.058)
(80,80,80)	0.965 (0.005)	0.968 (0.004)	0.963 (0.004)	0.930 (0.004)	0.955 (0.020)	0.959 (0.020)	0.957 (0.020)
(100,100,100)	0.976 (0.004)	0.966 (0.003)	0.954 (0.003)	0.930 (0.003)	0.940 (0.016)	0.945 (0.016)	0.942 (0.016)
(120,120,120)	0.970 (0.003)	0.967 (0.003)	0.956 (0.003)	0.920 (0.003)	0.940 (0.013)	0.947 (0.013)	0.944 (0.013)
(100,150,100)	0.966 (0.003)	0.956 (0.003)	0.959 (0.003)	0.910 (0.003)	0.944 (0.013)	0.949 (0.014)	0.946 (0.013)
	Gamma 8: $(VUS, J_3) = (0.346, 0.225)$						
(20,20,20)	0.954 (0.044)	0.926 (0.039)	0.942 (0.042)	0.964 (0.117)	0.933 (0.084)	0.963 (0.151)	0.955 (0.091)
(30,30,30)	0.956 (0.027)	0.937 (0.025)	0.954 (0.026)	0.960 (0.088)	0.923 (0.056)	0.959 (0.081)	0.947 (0.060)
(50,50,50)	0.959 (0.016)	0.956 (0.014)	0.957 (0.014)	0.961 (0.063)	0.937 (0.033)	0.958 (0.038)	0.952 (0.035)
(20,30,50)	0.963 (0.030)	0.944 (0.027)	0.952 (0.028)	0.953 (0.096)	0.924 (0.059)	0.965 (0.076)	0.951 (0.064)
(80,80,80)	0.962 (0.010)	0.942 (0.008)	0.962 (0.009)	0.970 (0.048)	0.934 (0.021)	0.960 (0.022)	0.950 (0.021)
(100,100,100)	0.960 (0.008)	0.948 (0.007)	0.959 (0.007)	0.958 (0.042)	0.934 (0.016)	0.954 (0.017)	0.944 (0.017)
(120,120,120)	0.963 (0.006)	0.946 (0.006)	0.950 (0.006)	0.965 (0.037)	0.934 (0.014)	0.952 (0.014)	0.943 (0.014)
(100,150,100)	0.961 (0.007)	0.943 (0.006)	0.958 (0.006)	0.968 (0.039)	0.942 (0.014)	0.951 (0.015)	0.946 (0.014)

(Continues)

TABLE 3 (Continued)

Sample size	GammaGI1	GammaGI2	GammaGI3	BCGI	BTI	BTLT	BTAT
	Coverage probability (area of confidence region)						
	Gamma 9: $(VUS, J_3) = (0.273, 0.136)$						
(20,20,20)	0.920 (0.038)	0.925 (0.035)	0.931 (0.039)	0.904 (0.048)	0.958 (0.063)	0.963 (0.166)	0.957 (0.068)
(30,30,30)	0.949 (0.024)	0.943 (0.023)	0.945 (0.024)	0.921 (0.030)	0.943 (0.042)	0.957 (0.088)	0.949 (0.045)
(50,50,50)	0.966 (0.013)	0.947 (0.012)	0.956 (0.013)	0.932 (0.017)	0.942 (0.026)	0.953 (0.050)	0.953 (0.028)
(20,30,50)	0.953 (0.026)	0.932 (0.024)	0.942 (0.026)	0.925 (0.031)	0.939 (0.045)	0.957 (0.064)	0.945 (0.048)
(80,80,80)	0.958 (0.008)	0.951 (0.007)	0.960 (0.007)	0.944 (0.010)	0.932 (0.017)	0.948 (0.024)	0.947 (0.018)
(100,100,100)	0.957 (0.006)	0.960 (0.006)	0.970 (0.006)	0.952 (0.008)	0.947 (0.013)	0.957 (0.018)	0.960 (0.014)
(120,120,120)	0.968 (0.005)	0.964 (0.004)	0.963 (0.005)	0.965 (0.006)	0.934 (0.011)	0.952 (0.014)	0.951 (0.012)
(100,150,100)	0.965 (0.005)	0.957 (0.005)	0.957 (0.005)	0.956 (0.006)	0.931 (0.012)	0.950 (0.014)	0.948 (0.013)
	Gamma 10: $(VUS, J_3) = (0.212, 0.060)$						
(20,20,20)	0.935 (0.034)	0.932 (0.031)	0.941 (0.032)	0.938 (0.087)	0.923 (0.089)	0.951 (0.113)	0.943 (0.095)
(30,30,30)	0.950 (0.020)	0.941 (0.018)	0.945 (0.019)	0.940 (0.045)	0.931 (0.059)	0.956 (0.065)	0.947 (0.061)
(50,50,50)	0.955 (0.011)	0.949 (0.010)	0.954 (0.010)	0.939 (0.015)	0.940 (0.035)	0.955 (0.036)	0.949 (0.035)
(20,30,50)	0.957 (0.021)	0.942 (0.019)	0.949 (0.019)	0.940 (0.046)	0.931 (0.059)	0.962 (0.064)	0.950 (0.061)
(80,80,80)	0.958 (0.006)	0.954 (0.006)	0.955 (0.006)	0.932 (0.006)	0.942 (0.021)	0.948 (0.022)	0.946 (0.022)
(100,100,100)	0.960 (0.005)	0.949 (0.004)	0.955 (0.004)	0.928 (0.004)	0.953 (0.017)	0.962 (0.017)	0.958 (0.017)
(120,120,120)	0.968 (0.004)	0.958 (0.003)	0.957 (0.003)	0.940 (0.004)	0.943 (0.014)	0.947 (0.014)	0.944 (0.014)
(100,150,100)	0.964 (0.004)	0.946 (0.004)	0.957 (0.004)	0.942 (0.004)	0.948 (0.014)	0.956 (0.015)	0.952 (0.015)

via http://adni.loni.usc.edu/data-samples/access-data). The ADNI study has four phases: ADNI 1, ADNI GO, ADNI 2, and ADNI 3. The baseline data of ADNI 1 include 379 subjects including 107 of healthy controls (HC), 182 of mild cognitive impairment (MCI) subjects, and 90 of diseased (AD) subjects.

Early diagnosis of Alzheimer's disease is critical for selecting optimal patient care and targeting important AD neuropathological features in clinical trails. Three core cerebrospinal fluid (CSF) biomarkers have been incorporated into research diagnostic criteria for AD, namely the "42 amino acid long amyloid-beta peptide" ($A\beta_{1-42}$), "total tau protein" (T-tau), and "tau phosphorylated at threonine 181" (P-tau$_{181}$). Researchers have observed that $A\beta_{1\text{-}42}$ decreases while T-tau and P-tau$_{181}$ increase in the CSF of individuals with AD.[39] Additionally, blood-based biomarkers, such as "plasma amyloid β peptides," have been studied as a screening method to identify individuals at risk of dementia, since they are less invasive and cost-effective.[40] Studies have shown that a lower $A\beta 42/A\beta 40$ ratio ($A\beta 42/40$) in plasma indicates higher risk of dementia.[41] Furthermore, Plasma isoprostanes, such as "8-iso-PGF$_{2\alpha}$," has been shown to have increased level in patients with AD, indicating the development and propagation of AD.[42] We will study the classification accuracy of $A\beta_{1-42}$, T-tau, P-tau$_{181}$, $A\beta 42/40$, and 8-iso-PGF$_{2\alpha}$ by presenting their joint confidence regions of VUS and J_3.

Table 4 presents the descriptive statistics of $A\beta_{1-42}$, T-tau, P-tau$_{181}$, $A\beta 42/40$, and 8-iso-PGF$_{2\alpha}$. Note that some records are not included due to missingness and all four biomarkers are measured on a continuous positive scale. Figure 2 presents density plots of HC, MCI, and AD for all five biomarkers. Note that for $A\beta_{1-42}$ and $A\beta 42/40$, lower marker value indicate worse disease status. By eyeballing Figure 2, we observe the substantial differences among these biomarkers. To quantitatively evaluate their accuracy, we estimate their confidence region of (VUS, J_3) using the proposed methods. For $A\beta_{1-42}$, T-tau, P-tau$_{181}$, and $A\beta 42/40$, normality cannot be achieved for either original data or Box-Cox transformed data based on Shapiro-Wilk test, while for 8-iso-PGF$_{2\alpha}$, normality can be achieved via Box-Cox transformation. Furthermore, gamma assumption is not satisfied for these five biomarkers based on gamma goodness-of-fit test.[43] Therefore, we use the non-parametric method **BTAT** to evaluate the joint confidence region of VUS and J_3 of these biomarkers, and we also use the parametric method **BCGI**, that is, the generalized inference method for Box-Cox transformed data, for 8-iso-PGF$_{2\alpha}$.

TABLE 4 Descriptive statistics of Alzheimer's markers ($A\beta_{1\text{-}42}$, T-tau, P-tau_{181}, $A\beta 42/40$, and 8-iso-$PGF_{2\alpha}$).

	HC			MCI			AD		
Biomarker	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**
$A\beta_{1\text{-}42}$	107	207.075	53.979	181	163.398	54.632	90	143.378	42.153
T-tau	107	69.112	28.306	178	100.382	51.729	88	121.943	57.737
P-tau_{181}	107	24.748	13.756	182	35.555	16.823	90	41.600	19.864
$A\beta 42/40$	106	0.272	0.086	176	0.253	0.086	89	0.252	0.066
8-iso-$PGF_{2\alpha}$	106	4.493	2.281	179	4.476	2.119	90	4.767	2.285

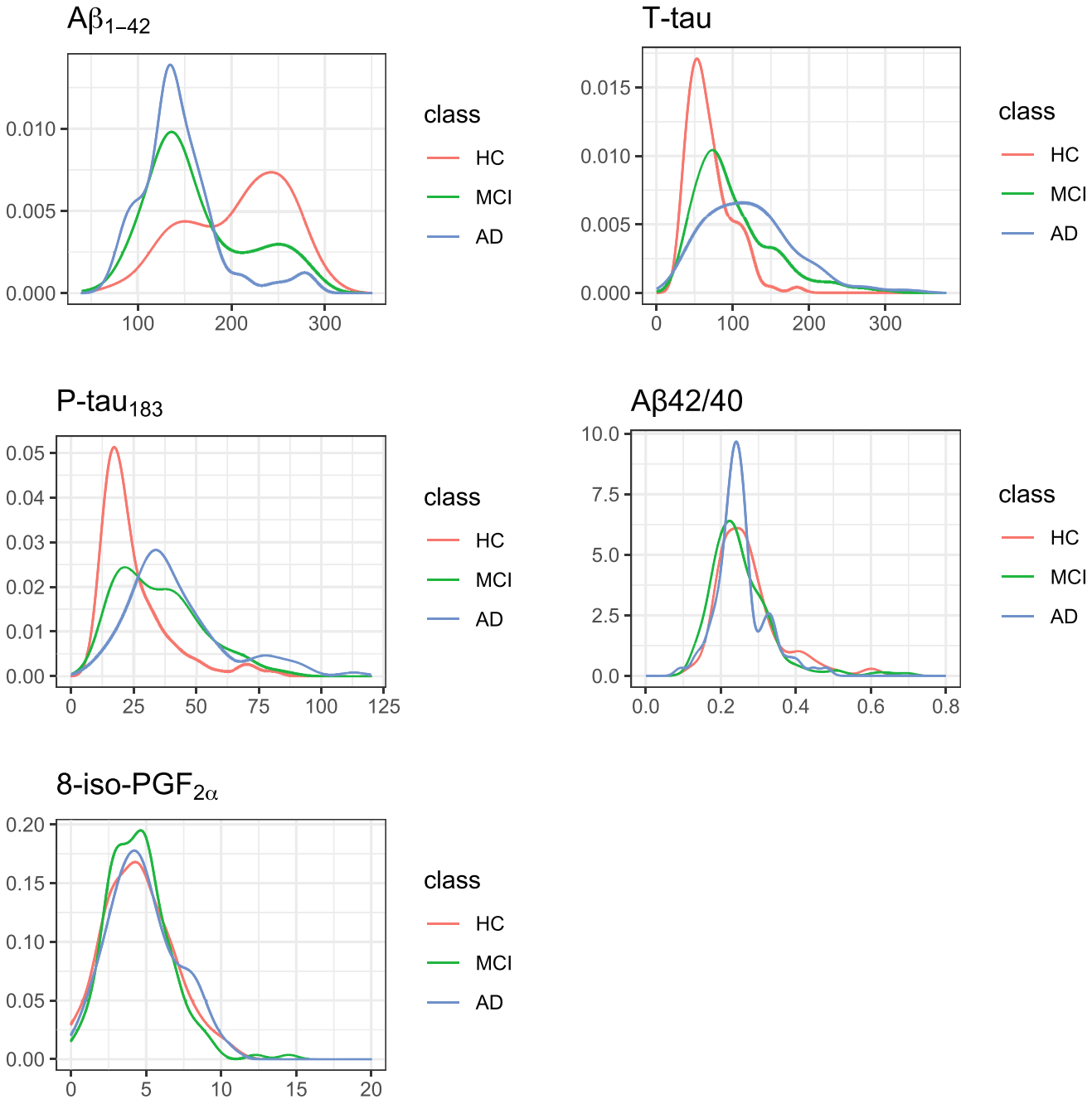

FIGURE 2 Density plots of $A\beta_{1\text{-}42}$, T-tau, and P-tau_{181}, $A\beta 42/40$, and 8-iso-$PGF_{2\alpha}$.

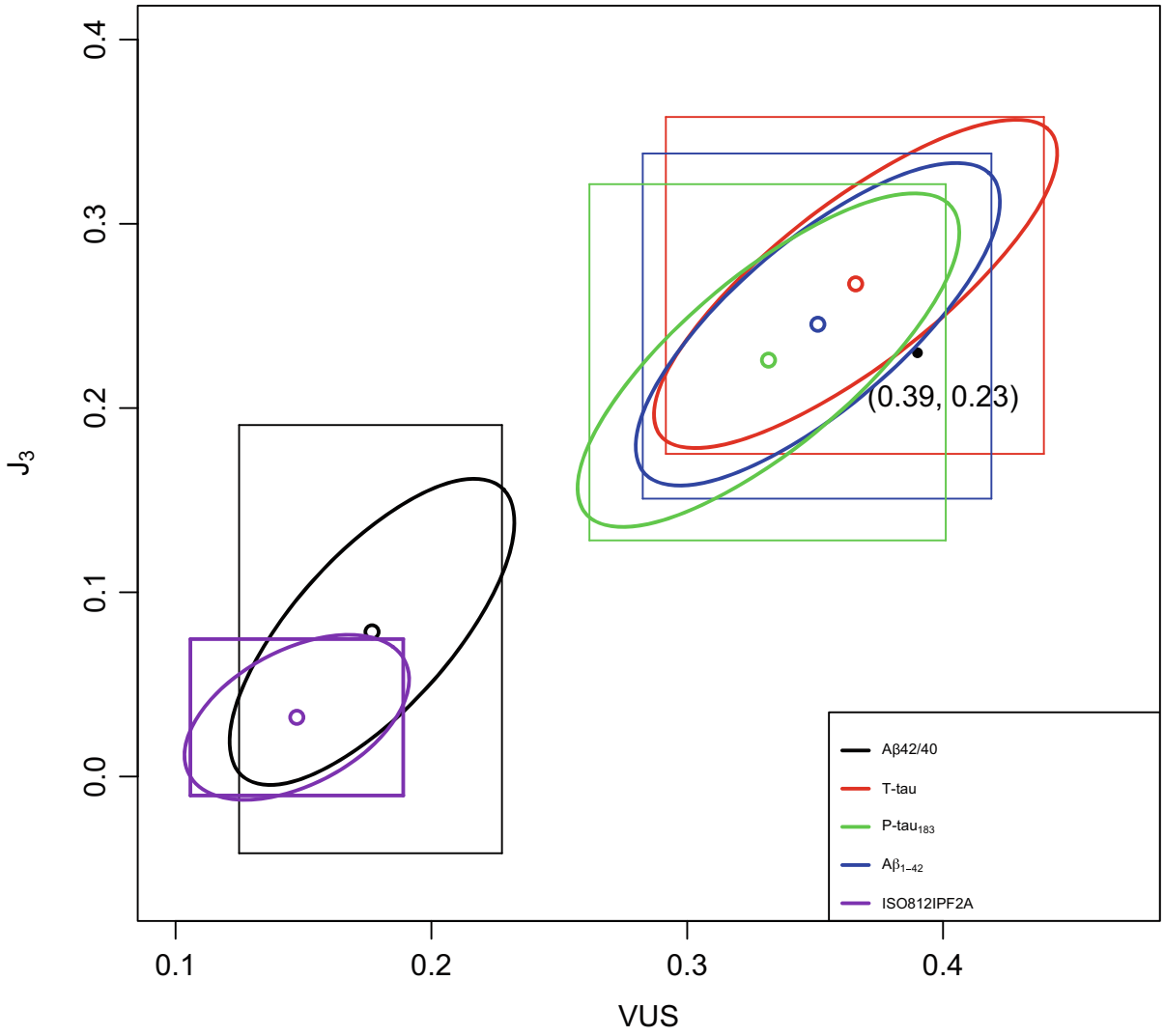

FIGURE 3 Estimated 95.0% joint confidence regions by **BTAT** method and Bonferroni correction for Aβ_{1-42}, T-tau, P-tau$_{181}$, and Aβ42/40 ratio, and estimated 95.0% joint confidence regions by **BCGI** method and Bonferroni correction for 8-iso-PGF$_{2\alpha}$. The **BTAT** confidence region for Aβ_{1-42} is given by the elliptical equation $\frac{(x-0.351)^2}{0.107}+\frac{(y-0.245)^2}{0.038}=1$ with major axis $(1, 1.131)^T$ and origin $(0.351, 0.245)$. Similarly, the elliptical equation for T-tau is $\frac{(x-0.365)^2}{0.114}+\frac{(y-0.267)^2}{0.038}=1$ with major axis $(1, 1.115)^T$ and origin $(0.365, 0.267)$; $\frac{(x-0.331)^2}{0.112}+\frac{(y-0.226)^2}{0.039}=1$ for P-tau$_{181}$ with major axis $(1, 1.300)^T$ and origin $(0.331, 0.226)$; $\frac{(x-0.176)^2}{0.095}+\frac{(y-0.079)^2}{0.034}=1$ for Aβ42/40 with major axis $(1, 1.697)^T$ and origin $(0.176, 0.079)$. The **BCGI** confidence region for 8-iso-PGF$_{2\alpha}$ is given by elliptical equation $\frac{(x-0.147)^2}{0.005}+\frac{(y-0.032)^2}{0.003}=1$ with major axis $(1, 1.106)^T$ and origin $(0.147, 0.032)$. The rectangular regions are formed by corresponding Bonferroni-adjusted 97.5% confidence intervals in Table 5.

Figure 3 presents the elliptical confidence regions of VUS and J_3 by the **BTAT** method for Aβ_{1-42}, T-tau, P-tau$_{181}$, and Aβ42/40, and by **BCGI** method for 8-iso-PGF$_{2\alpha}$, along with rectangular regions formed by Bonferroni-adjusted confidence intervals. While widely used in multiple testing in practice, it is well-known that Bonferroni-adjusted method leads to conservative results due to the fact that Bonferroni-adjusted method ignores the correlation between VUS and J_3. For example, for all three CSF biomarkers, the point with VUS as 0.39 and J_3 as 0.23 is out of all three elliptical confidence regions; however, this point is contained in all three rectangular regions by Bonferroni-adjusted method, indicating that the proposed confidence regions could yield different results from the conservative Bonferroni-adjusted confidence regions.

Table 5 presents the estimated areas of the joint confidence regions and Bonferroni-adjusted regions, the point estimates and the confidence intervals with and without Bonferroni adjustment for both VUS and J_3, respectively. It is easy to see that the areas by proposed joint confidence region are substantially smaller than those by Bonferroni-adjusted confidence regions, clearly indicating the benefit of making proper joint inference in biomarker evaluation.

TABLE 5 Summary of simultaneous confidence region and interval estimations of VUS and J_3 for Alzheimer's markers A$\beta_{1\text{-}42}$, T-tau, P-tau$_{181}$, Aβ42/40, and 8-iso-PGF$_{2\alpha}$.

		A$\beta_{1\text{-}42}$	**T-tau**	**P-tau$_{181}$**	**Aβ42/40**	**8-iso-PGF$_{2\alpha}$ $*$**	
		BTAT	**BTAT**	**BTAT**	**BTAT**	**BTAT**	**BCGI**
Area	Joint CR	0.012	0.013	0.013	0.010	0.008	0.004
	Bonferroni	0.026	0.029	0.028	0.024	0.025	0.012
VUS	Point est.	0.349	0.365	0.331	0.176	0.196	0.194
	97.5% Bonfer. CI	(0.283, 0.418)	(0.292, 0.439)	(0.262, 0.401)	(0.130, 0.219)	(0.152, 0.246)	(0.143, 0.246)
	95.0% ind. CI	(0.292, 0.409)	(0.301, 0.430)	(0.271, 0.392)	(0.130, 0.200)	(0.159, 0.240)	(0.149, 0.240)
J_3	Point est.	0.245	0.264	0.227	0.076	0.042	0.026
	97.5% Bonfer. CI	(0.152, 0.334)	(0.175, 0.358)	(0.128, 0.321)	(0.000, 0.186)	(0.000, 0.170)	(0.000, 0.108)
	95.0% Ind. CI	(0.163, 0.322)	(0.187, 0.347)	(0.140, 0.309)	(0.000, 0.172)	(0.000, 0.154)	(0.005, 0.101)

Abbreviations: Area, area of confidence region of (VUS, J_3); Bonfer. CI, Bonferroni-adjusted simultaneous confidence interval; Ind. CI, individual confidence interval; $*$, normality is satisfied via Box-Cox transformation.

7 | SUMMARY AND DISCUSSION

This article fills the gap of statistical methods for evaluating the accuracy of biomarkers (or tests) with multi-class setting under ROC framework. In biomedical research, both hypervolume under ROC manifold (HUM$_K$) and generalized Youden index (J_K) are popular measures for biomarker evaluation for multi-class classification. While HUM$_K$ summarizes discriminating ability of a biomarker across all possible cut-points, it lacks the direct link to classification accuracy and fails to provide the optimal cut-points. This can be compensated by J_K as it not only provides the cut-points, but also measures discriminatory accuracy with the maximum sum of correct classification rates a biomarker can possibly achieve. Evaluating HUM$_K$ and J_K simultaneously provides a comprehensive picture about the discriminating ability of a biomarker for multi-class classification. This article explores parametric and non-parametric approaches on constructing joint inference for (HUM$_K$, J_K) when there are three or more ordinal classes. Among all confidence region estimation methods investigated in this research, the generalized inference method (**GI**) has the most accurate coverage probabilities and the smallest average area under normal assumption. With gamma assumption, **GammaGI2**, and **GammaGI3** are more accurate thus more preferred. The bootstrap method with arcsine-square-root **BTAT** is recommended when estimating without any distribution assumption.

In addition to HUM$_K$ and J_K, there are other measures to evaluate diagnostic accuracy for ordered multi-class setting. For example, maximum absolute determinant (MADET)[44] serves not only as a comprehensive measure which utilizes all the information involved in a classification problem, but also as a cut-points selection method. The adjusted Youden index (AYI) updates the generalized Youden index by introducing the weighted sum of misclassification rates as a penalty term into generalized Youden index.[45] Furthermore, weighted aggregated AUC and weighted aggregated *YI* using multi-step procedure[46] provide flexibility in examining performance of biomarkers under ordered classes. The proposed methods in this article can be easily extended to any pair of accuracy measures under multi-class classification.

Certainly, other estimating methods such as the non-parametric estimation based on kernel smoothing[33,47,48] and parametric bootstrap methods under normal/gamma distribution assumptions can also be utilized for the proposed problem.[19,33] Furthermore, there exist some research on joint hypothesis testing under binary disease classification.[22,49] We aim to explore joint testing of HUM$_K$ and J_K under multi-class classification in future research.

Regarding computing, an R program is available upon request from ltian@buffalo.edu if needed. Furthermore, it is noted that the use of many existing R packages (eg, trinROC, ThresholdROC) for accuracy measures in multiple class classification should greatly facilitate the programming process.[50,51]

ACKNOWLEDGEMENTS

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI

investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

DATA AVAILABILITY STATEMENT

http://adni.loni.usc.edu/data-samples/access-data.

ORCID

Jingjing Yin https://orcid.org/0000-0003-4843-613X
Lili Tian https://orcid.org/0000-0003-4935-7484

REFERENCES

1. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford, UK: Oxford University Press; 2003.
2. Zhou XH. *Statistical Methods in Diagnostic Medicine*. 2nd ed. Hoboken, NJ: Wiley; 2011.
3. Zou K, Liu A, Bandos A, Ohno-Machado L, Rockette H. *Statistical Evaluation of Diagnostic Performance: Topics in ROC Analysis*. Boca Raton, FL: CRC Press; 2016.
4. Nakas C, Bantis L, Gatsonis C. *ROC Analysis for Classification and Prediction in Practice*. Boca Raton, FL: CRC Press; 2023.
5. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32-35.
6. Fluss R, Faraggi D, Reiser B. Estimation of the Youden index and its associated cutoff point. *Biom J*. 2005;47(4):458-472.
7. Morris JC, Cummings J. Mild cognitive impairment (MCI) represents early-stage Alzheimer's disease. *J Alzheimers Dis*. 2005;7(3):235-239.
8. Scinto LFM, Daffner KR. *Early Diagnosis of Alzheimer's Disease*. Totowa, NJ: Humana Press; 2000.
9. Partheen K, Kristjansdottir B, Sundfeldt K. Evaluation of ovarian cancer biomarkers HE4 and CA-125 in women presenting with a suspicious cystic ovarian mass. *J Gynecol Oncol*. 2011;22(4):244-252.
10. Scurfield BK. Multiple-event forced-choice tasks in the theory of signal detectability. *J Math Psychol*. 1996;40(3):253-269.
11. Mossman D. Three-way ROCs. *Med Decis Making*. 1999;19(1):78-89.
12. Nakas CT, Yiannoutsos CT. Ordered multiple-class ROC analysis with continuous measurements. *Stat Med*. 2004;23(22):3437-3449.
13. Li J, Fine JP. ROC analysis with multiple classes and multiple tests: methodology and its application in microarray studies. *Biostatistics*. 2008;9(3):566-576.
14. Xiong C, Belle vG, Miller JP, Morris JC. Measuring and estimating diagnostic accuracy when there are three ordinal diagnostic groups. *Stat Med*. 2006;25(7):1251-1273.
15. Nakas CT, Alonzo TA, Yiannoutsos CT. Accuracy and cut-off point selection in three-class classification problems using a generalization of the Youden index. *Stat Med*. 2010;29(28):2946-2955.
16. Luo J, Xiong C. Youden index and associated cut-points for three ordinal diagnostic groups. *Commun Stat Simul Comput*. 2012;42(6):1213-1234.
17. Karakaya J, Akin S, Karagaoglu E, Gurlek A. The performance of hemoglobin A1c against fasting plasma glucose and oral glucose tolerance test in detecting prediabetes and diabetes. *J Res Med Sci*. 2014;19(11):1051-1057.
18. Wellens H, BeGole E, Kuijpers-Jagtman A. ROC surface assessment of the ANB angle and Wits appraisal's diagnostic performance with a statistically derived 'gold standard': does normalizing measurements have any merit? *Eur J Orthod*. 2017;39(4):358-364.
19. Yin J, Tian L. Joint confidence region estimation for area under ROC curve and Youden index. *Stat Med*. 2014;33(6):985-1000.
20. Bantis LE, Nakas CT, Reiser B. Construction of confidence regions in the ROC space after the estimation of the optimal Youden index-based cut-off point. *Biometrics*. 2014;70(1):212-223.
21. Bantis LE, Nakas CT, Reiser B, Myall D, Dalrymple-Alford JC. Construction of joint confidence regions for the optimal true class fractions of receiver operating characteristic (ROC) surfaces and manifolds. *Stat Methods Med Res*. 2017;26(3):1429-1442.
22. Yin J, Samawi H, Tian L. Joint inference about the AUC and Youden index for paired biomarkers. *Stat Med*. 2022;41(1):37-64.
23. Li J, Zhou XH. Nonparametric and semiparametric estimation of the three way receiver operating characteristic surface. *J Stat Plan Inference*. 2009;139(12):4133-4142.
24. Hong CS, Cho MH. VUS and HUM represented with Mann-Whitney statistic. *Commun Stat Appl Methods*. 2015;22(3):223-232.
25. Schisterman EF, Perkins NJ. Confidence intervals for the Youden index and corresponding optimal cut-point. *Commun Stat Simul Comput*. 2007;36(3):549-563.
26. Tsui KW, Weerahandi S. Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. *J Am Stat Assoc*. 1989;84(406):602-607.
27. Weerahandi S. Generalized confidence intervals. *J Am Stat Assoc*. 1993;88(423):899-905.
28. Weerahandi S. *Exact Statistical Methods for Data Analysis*. New York: Springer-Verlag; 1995.
29. Krishnamoorthy K, Lu Y. Inferences on the common mean of several normal populations based on the generalized variable method. *Biometrics*. 2003;59(2):237-247.
30. Lin S, Lee JC, Wang R. Generalized inferences on the common mean vector of several multivariate normal populations. *J Stat Plan Inference*. 2007;137(7):2240-2249.
31. Gao Y, Tian L. Confidence interval estimation for sensitivity and difference between two sensitivities at a given specificity under tree ordering. *Stat Med*. 2021;40(16):3695-3723.

32. Hannig J. On generalized fiducial inference. *Stat Sin*. 2009;19(2):491-544.
33. Yin J, Nakas C, Tian L, Reiser B. Confidence intervals for differences between volumes under receiver operating characteristic surfaces (VUS) and generalized Youden indices (GYIs). *Stat Methods Med Res*. 2017;27:675-688.
34. Zou KH, Hall WJ. Two transformation models for estimating an ROC curve derived from continuous data. *J Appl Stat*. 2000;27(5):621-631.
35. Chen P, Ye ZS. Approximate statistical limits for a gamma distribution. *J Qual Technol*. 2017;49(1):64-77.
36. Wang BX, Wu F. Inference on the gamma distribution. *Dent Tech*. 2018;60(2):235-244.
37. Krishnamoorthy K, Mathew T, Mukherjee S. Normal-based methods for a gamma distribution: prediction and tolerance intervals and stress-strength reliability. *Dent Tech*. 2008;50(1):69-78.
38. Krishnamoorthy K, Wang XG. Fiducial confidence limits and prediction limits for a gamma distribution: censored and uncensored cases. *Environmetrics*. 2016;27:479-493.
39. Gireadă B, Dobrin R, Onofrei S, Dobre CE, Sîrghie I, Roşu I. Cerebrospinal fluid biomarkers for early and differential diagnosis of Alzheimer's disease. *Bull Integr Psychiatry*. 2022;93(2):121-128.
40. Fandos N, Pérez-Grijalba V, Pesini P, et al. Plasma amyloid β42/40 ratios as biomarkers for amyloid β cerebral deposition in cognitively normal individuals. *Alzheimers Dement*. 2017;8(1):179-187.
41. Graff-Radford NR, Crook JE, Lucas J, et al. Association of low plasma Abeta42/Abeta40 ratios with increased imminent risk for mild cognitive impairment and Alzheimer disease. *Arch Neurol*. 2007;64(3):354-362.
42. Casadesus G, Smith MA, Basu S, et al. Increased isoprostane and prostaglandin are prominent in neurons in Alzheimer disease. *Mol Neurodegener*. 2007;2(1):2.
43. Villasenor JA, Gonzalez-Estrada E. A variance ratio test of fit for gamma distributions. *Stat Probab Lett*. 2015;96:281-286.
44. Dong T, Attwood K, Hutson A, Liu S, Tian L. A new diagnostic accuracy measure and cut-point selection criterion. *Stat Methods Med Res*. 2015;26:2832-2852.
45. Hua J, Tian L. A comprehensive and comparative review of optimal cut-points selection methods for diseases with multiple ordinal stages. *J Biopharm Stat*. 2020;30(1):46-68.
46. Feng Y, Tian L. Flexible diagnostic measures and new cut-point selection methods under multiple ordered classes. *Pharm Stat*. 2022;21(1):220-240.
47. Yin J, Samawi H, Linder D. Improved nonparametric estimation of the optimal diagnostic cut-off point associated with the Youden index under different sampling schemes. *Biom J*. 2016;58(4):915-934.
48. Yin J, Hao Y, Samawi H, Rochani H. Rank-based kernel estimation of the area under the ROC curve. *Stat Methodol*. 2016;32:91-106.
49. Yin J, Mutiso F, Tian L. Joint hypothesis testing of the area under the receiver operating characteristic curve and the Youden index. *Pharm Stat*. 2021;20(3):657-674.
50. Noll S, Furrer R, Reiser B, Nakas CT. Inference in receiver operating characteristic surface analysis via a trinormal model-based testing approach. *Stat*. 2019;8(1):e249.
51. Perez-Jaume S, Pallares N, Skaltsa K, Carrasco JL. ThresholdROC: optimum threshold estimation tools for continuous diagnostic tests in R. *J Stat Softw*. 2017;82(4):1-21.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wang J, Yin J, Tian L. Evaluating joint confidence region of hypervolume under ROC manifold and generalized Youden index. *Statistics in Medicine*. 2024;43(5):869-889. doi: 10.1002/sim.9998

APPENDIX A. MONOTONIC TRANSFORMATIONS

To improve the performance of proposed confidence region, we can use some monotonic transformations. Since the arcsine-square-root transformation is variance stabilized for binomial probabilities, and the logit function is commonly used for values between 0 to 1, these two transformations can be used to transform HUM_K and J_K. In the following, we present the steps of constructing the generalized confidence region with such transformation.

Let $h(.)$ stand for a transformation function. To obtain confidence region of transformed $\eta = (\mathrm{HUM}_K, J_K)^T$; that is, $\eta^h = (h(\mathrm{HUM}_K), h(J_K))^T$, we follow the same steps of computing the parametric confidence regions presented in Section 3.1, whereas both $\widehat{\mathrm{HUM}}_K$ and $\widehat{J}_K$ are transformed using either the logit or arcsine-square-root transformation. The $100(1-\alpha)\%$ confidence region of $\eta^h = (h(\mathrm{HUM}_K), h(J_K))^T$ is

$$\left\{\eta^h : (\eta^h - \hat{\eta}^h)^T \left(\hat{\Sigma}^h\right)^{-1} \left(\eta^h - \hat{\eta}^h\right) \leq q^2_{1-\alpha}\right\},$$

where $\hat{\eta}^h$ and $\hat{\Sigma}^h$ stand for sample mean and the sample covariance matrix of the transformed simulated samples. For the generalized inference method **GI** under normality, the transformed sample is $\left(h\left(R^b_{\mathrm{HUM}_K}\right), h\left(R^b_{J_K}\right)\right)^T$ for $b = 1, 2, \ldots, B$ where $R^b_{\mathrm{HUM}_K}$ and $R^b_{J_K}$ are the generalized pivotal quantities for HUM_K and J_K as presented in Section 3.1.

Furthermore, in order to obtain the confidence region of $\eta = (\mathrm{HUM}_K, J_K)^T$, the confidence region $(h(\mathrm{HUM}_K), h(J_K))^T$ needs to be transformed back. That is, the $100(1-\alpha)\%$ confidence region of $\eta = \left[h^{-1}(h(\mathrm{HUM}_K)), h^{-1}(h(J_K))\right]^T = (\mathrm{HUM}_K, J_K)^T$ is

$$\left\{\eta : \left[\eta - h^{-1}(\hat{\eta}^h)\right]^T \left(\hat{\Sigma}^{\mathrm{inv}}\right)^{-1} \left[\eta - h^{-1}(\hat{\eta}^h)\right] \leq q^2_{1-\alpha}\right\},$$

where $h^{-1}(\hat{\eta}^h) = \left(h^{-1}\left(\widehat{\mathrm{HUM}}^h_K\right), h^{-1}\left(\hat{J}^h_K\right)\right)^T$, the inverse transformed sample mean $\hat{\eta}^h$, and $\hat{\Sigma}^{\mathrm{inv}} = \mathbf{J}^T_{\mathrm{inv}} \hat{\Sigma}^h \mathbf{J}_{\mathrm{inv}}$ is the inverse transformed sample covariance matrix $\hat{\Sigma}^h$, where $\mathbf{J}_{\mathrm{inv}}$ is the Jacobian matrix of η with the respect to η^h, calculated by taking the first derivative of the inverse function $h^{-1}(.)$ evaluated at $\hat{\eta}^h$.

Similarly, these two transformations can be used in combination with non parametric methods **BTI** yielding **BTAT** and **BTLT**.