Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–25

DOI: 10.1177/0962280215611631
smm.sagepub.com
SAGE

A new diagnostic accuracy measure and cut-point selection criterion

Tuochuan Dong,[1] Kristopher Attwood,[2] Alan Hutson,[1] Song Liu[2] and Lili Tian[1]

Abstract

Most diagnostic accuracy measures and criteria for selecting optimal cut-points are only applicable to diseases with binary or three stages. Currently, there exist two diagnostic measures for diseases with general k stages: the hypervolume under the manifold and the generalized Youden index. While hypervolume under the manifold cannot be used for cut-points selection, generalized Youden index is only defined upon correct classification rates. This paper proposes a new measure named maximum absolute determinant for diseases with k stages ($k \geq 2$). This comprehensive new measure utilizes all the available classification information and serves as a cut-points selection criterion as well. Both the geometric and probabilistic interpretations for the new measure are examined. Power and simulation studies are carried out to investigate its performance as a measure of diagnostic accuracy as well as cut-points selection criterion. A real data set from Alzheimer's Disease Neuroimaging Initiative is analyzed using the proposed maximum absolute determinant.

Keywords

Maximum absolute determinant, optimal cut-points, volume for the parallelotope, Alzheimer's Disease

1 Introduction

In diagnostic studies, the most straightforward classification rule is binary, i.e. non-diseased or diseased. The probabilities that a diagnostic biomarker correctly classifies a non-diseased subject or a diseased subject are defined as specificity or sensitivity, respectively. Plotting sensitivity against 1-specificity for all the possible cut-points or decision thresholds from a continuous biomarker leads to what is termed a receiver operating characteristic (ROC) curve. There exist many detailed reviews pertaining to inference in the framework of ROC curve.[1–4] For diseases with binary classification, the most popular diagnostic measure of accuracy is the area under the ROC curve (AUC), which has been extensively studied.[5–13] Another popular diagnostic measure, the Youden index,[14] is defined as

[1]Department of Biostatistics, University at Buffalo, Buffalo, NY, USA
[2]Department of Biostatistics and Bioinformatics, Roswell Park Cancer Institute, Buffalo, NY, USA

Corresponding author:
Lili Tian, Department of Biostatistics, 717 Kimball Tower, 3435 Main St. Bldg. 26 Buffalo, NY 14214, USA.
Email: ltian@buffalo.edu

the maximum of specificity+sensitivity−1. Inference procedures regarding Youden index have also been well studied.[15–18]

In practice, many disease processes can be naturally classified into more than two stages. For example, in Alzheimer's disease there exists a transition stage called mild cognitive impairment (MCI), which is the cognitive change between normal aging (non-diseased) and complete memory loss (fully diseased).[19,20] Therefore, we often have a three-stage classification in Alzheimer's disease: non-diseased, early diseased and fully diseased. Consequently, the ROC curve for binary classification is extended to ROC surface for three-stage classification and the AUC to the volume under the ROC surface (VUS), which is the most widely used quantitative summary index for diseases with three stages.[21] There are several recent papers in the statistical literature focused on VUS research.[22–28] Furthermore, as an extension of the Youden index to diseases with three stages, Nakas et al.[29] introduced the generalized Youden index (GYI), defined as the maximum of the sum of the correct classification rates (CCRs) for three groups minus 1. There are several recent papers in the statistical literature focused on GYI research.[30,31] For general cases of multiple-class diagnoses, the concept of hypervolume under the manifold (HUM) was first proposed as an extension of AUC by Scurfield[32,33] in the field of mathematical psychology. Nakas and Yiannoutsos[26] first proposed the non-parametric estimator for HUM, and Li and Fine[34] proposed the corresponding inference procedures and methods for estimating classification probabilities and calculating HUM. Furthermore, the concept of GYI has also been extended to the diseases with general k stages.[35] A brief review of HUM and GYI is given in Section 2.

The critical step for making diagnoses based on a specific biomarker is to select the optimal cut-points. For diseases with a binary classification, Fluss et al.[16] discussed the performance of Youden index in terms of selecting optimal cut-point under different assumptions. Perkins and Schisterman[36] compared two common approaches for optimal cut-point selection: the point on the ROC curve closest to (0,1) and the Youden index. More recently, Liu[37] developed a method for selecting the optimal cut-point by maximizing the product of specificity and sensitivity. For diseases with three stages, He and Frey[38] discussed the ''optimal'' decision making criteria using likelihoods; Nakas et al.[29] proposed selecting the optimal cut-points using GYI; and Attwood et al.[39] introduced two new criteria as the closest to perfection/minimum distance (MD) method, and the max volume (MV) method which are the extensions of the corresponding criteria for diseases with binary status. For diseases with general k stages, GYI[35] can be used as a criterion for the optimal cut-points selection.

In this paper, we propose a new measure of diagnostic accuracy for diseases with ≥ 2 classes. This new measure directly incorporates all possible true and false classification rates. It has an appealing geometric interpretation, which justifies its use as a measure for classification accuracy. Furthermore, the proposed measure can also be used for cut-point selection. The rest of paper is organized as follows: in Section 2, preliminary background information is introduced and some existing diagnostic accuracy measures are reviewed. In Section 3, a new measure, namely the maximum absolute determinant (MADET), is proposed and specific details are discussed. In Section 4, a power study is carried out to assess the performance of the proposed measure in comparison with HUM and GYI in testing the equality of two biomarkers. In Section 5, issues for optimal cut-points selection are discussed and simulation results comparing criteria MADET with other existing criteria are presented. In Section 6, several important biomarkers from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study are analyzed. In Section 7, summary and discussion are given.

2 Preliminaries

In this section, we present the definitions, notation and a review of existing measures of diagnostic accuracy for diseases with general k stages.

2.1 Notations

For a disease that can be classified into k different classes, let S denote the true disease status for a subject and let T denote the diagnostic result based on a continuous biomarker. A subject with true disease status i $(i = 1, 2, \ldots, k)$ can correctly or incorrectly be classified as having disease state j $(j = 1, 2, \ldots, k)$. Define the conditional probability of a classification given the true disease status as

$$P_{i,j} = P(T = j \,|\, S = i), \quad \text{where} \quad i = 1, \ldots, k, j = 1, \ldots, k \tag{1}$$

Hence, $P_{i,i}$ denotes the probability that a subject from class i is correctly diagnosed and $P_{i,j}$ $(i \neq j)$ is the probability that a subject from class i is incorrectly diagnosed as class j. For a subject in class i, let $\overrightarrow{\mathbf{P}}_i$ denote the vector consisting of k classification rates, i.e.

$$\overrightarrow{\mathbf{P}}_i = \begin{pmatrix} P_{i,1} \\ \vdots \\ P_{i,k} \end{pmatrix} \quad \text{where} \quad i = 1, \ldots, k \tag{2}$$

By its definition, it follows

$$\left\langle \overrightarrow{\mathbf{1}}, \overrightarrow{\mathbf{P}}_i \right\rangle = \sum_{j=1}^{k} P_{i,j} = 1, \quad \text{for} \quad i = 1, \ldots, k \tag{3}$$

Hence, at a given i, out of k components $P_{i,1}, P_{i,2}, \ldots, P_{i,k}$, only $k - 1$ of them are independent.

Now consider k independent subjects whose true disease statuses are from each of the k classes, respectively. Then we have vectors $\overrightarrow{\mathbf{P}}_1, \overrightarrow{\mathbf{P}}_2, \ldots, \overrightarrow{\mathbf{P}}_k$ to fully describe the performance of the biomarkers being considered, where $\overrightarrow{\mathbf{P}}_1, \overrightarrow{\mathbf{P}}_2, \ldots, \overrightarrow{\mathbf{P}}_k$ are mutually independent. Next define a $k \times k$ square probability matrix (SPM) as

$$\mathcal{P} = \begin{pmatrix} \overrightarrow{\mathbf{P}}_1^T \\ \vdots \\ \overrightarrow{\mathbf{P}}_k^T \end{pmatrix} = (P_{i,j}) \tag{4}$$

The definition of SPM captures both the correct and false classification rates for all k classes. Overall, there are k CCRs and $k(k-1)$ false classification rates.

The class probabilities $P_{i,j}$'s for $i, j = 1, \ldots, k$ need to be assessed by some decision rules. To make a diagnosis for a disease with k stages based on a continuous biomarker X, cut-points $c_1, \ldots, c_{k-1}$ are needed. Here $c_1 < \cdots < c_{k-1}$ is assumed. If $c_{j-1} < X \leq c_j$, for $j = 1, \ldots, k$, this subject is classified into stage j. Let X_i denote the measurement for a subject from the ith class. Hence, the corresponding conditional probability $P_{i,j}$ can be written as

$$P_{i,j} = P(c_{j-1} < X_i \leq c_j \,|\, S = i), \quad \text{where} \quad i = 1, \ldots, k, \; j = 1, \ldots, k \tag{5}$$

2.2 Existing diagnostic measures for diseases with k stages

In this subsection, we review two measures of diagnostic accuracy proposed for diseases with general k stages, i.e. the HUM[26,34] and the GYI.[35]

2.2.1 HUM

Let t_i denote the CCR for the ith disease class, i.e. $P_{i,i}$ as defined in equation (1). By plotting the t_i's ($i = 1, \ldots, k$) in the k-dimensional space, a manifold is obtained. The corresponding HUM is utilized as a measure of diagnostic accuracy for diseases with general k stages. A rigorous mathematical definition of HUM is given by Li and Fine[34] as

$$HUM = \int_0^1 \int_0^{f_1(t_1)} \cdots \int_0^{f_{k-2}(t_1,\ldots,t_{k-2})} f_{k-1}(t_1, \ldots, t_{k-1}) dt_{k-1} \cdots dt_2 dt_1 \tag{6}$$

where f_{i-1}'s are the recursive equations defined as $t_i = f_{i-1}(t_1, \ldots_{i-1})$, $i = 2, \ldots, k$. The t_i's can be estimated via multinomial logistic regression or by using cut-points. For details, see the paper by Li and Fine.[34] Both the AUC and VUS are special cases of HUM for diseases with binary and three stages, respectively.

2.2.2 GYI

Nakas et al.[35] generalized the Youden Index to the general k stages classification problem. The k stages GYI is defined as

$$GYI(c_1, c_2, \ldots, c_{k-1}) = \max_{c_1,\ldots,c_{k-1}; c_1 < \cdots < c_{k-1}} \{P_{1,1} + \cdots + P_{k,k} - 1\} \tag{7}$$

Under the condition $c_1 < \cdots < c_{k-1}$, it was shown that GYI is the sum of the Youden indexes for the adjacent classes. As defined, GYI is not only a direct measure of overall CCR, but also a tool for optimal cut-points selection.

2.2.3 Overview of HUM and GYI and justification for the new measure

Both HUM and GYI have certain advantages and disadvantages. HUM is a global measure in nature, i.e. it gauges the diagnostic accuracy of a biomarker across all the possible cut-points. Consequently, HUM cannot be used to determine the optimal cut-points. Hence, Li and Fine[34] proposed to use multinomial logistic regression for estimating class probabilities. On the other hand, GYI can be used to determine the optimal cut-points, which yield the maximum overall correct classification. However, GYI considers only the k CCRs and ignores the $k(k-1)$ false classification rates, which are also the critical and integral components for describing diagnostic accuracy. As the number of classes k increases, less information is utilized by the GYI method. Hence, GYI becomes less optimal in this setting.

Therefore, there is a need for a comprehensive diagnostic measure, which is able to take into account all the information related to a classification rule, i.e. the information contained in the SPM $\mathcal{P}$, and which can also serve as a tool for optimal cut-point selection. The goal of this paper is to fill this gap.

3 New diagnostic accuracy measure

In this section, we propose a new measure of diagnostic accuracy, which directly utilizes all the classification rates contained in the SPM $\mathcal{P}$ in equation (2).

We term this new measure as the maximum (supremum) absolute determinant (MADET) of the SPM. It is defined as

$$\text{MADET} = \sup_{\mathcal{P}} |det(\mathcal{P})| \tag{8}$$

If the class probabilities are estimated using cut-points as discussed in Section 2.1, MADET can be expressed as

$$\text{MADET} = \max_{c_1,\ldots,c_{k-1}} |det(\mathcal{P})|(c_1,\ldots,c_{k-1}) \tag{9}$$

where $c_1,\ldots,c_{k-1}$ need to be determined by some criteria.

3.1 Geometric interpretation of MADET

In the following section, we investigate the geometric interpretation of MADET and subsequently justify its use as a diagnostic accuracy measure. For this purpose, several definitions from advanced geometry need to be restated.

Definition of parallelotope: Given a set of independent vectors $\vec{a}_1,\ldots,\vec{a}_k$ of dimension k in the Euclidean space, a parallelotope is the set of vectors

$$\mathcal{V}(\vec{a}_1,\ldots,\vec{a}_k) = \left\{ \sum_i x_i \vec{a}_i : 0 \le x_i \le 1 \right\}$$

for $1 \le i \le k$.

Given an orthonormal basis $\vec{e}_1,\ldots,\vec{e}_k$, the vectors $\vec{a}_1,\ldots,\vec{a}_k$ can be expressed as

$$(\vec{a}_1,\ldots,\vec{a}_k)^T = A(\vec{e}_1,\ldots,\vec{e}_k)^T$$

where A is a $k \times k$ matrix. The base of the k-dimensional parallelotope on vectors $\vec{a}_1,\ldots,\vec{a}_k$ is defined as the (k-1)-dimensional parallelotope $\mathcal{V}(\vec{a}_1,\ldots,\vec{a}_{k-1})$ and its height is defined as the orthogonal component of $\vec{a}_k$ with respect to the sub-space spanned by $\vec{a}_1,\ldots,\vec{a}_{k-1}$, denoted as $\mathbf{ort}_{(\vec{a}_1,\ldots,\vec{a}_{k-1})}\vec{a}_k$. The volume of a k-dimensional parallelotope ($k > 1$) is the product of the volume of its base and its height. The volume of a one-dimensional parallelotope $\mathcal{V}(\vec{a}_1)$ is the length of the vector $\overrightarrow{a_1}$. The volume of parallelotope on vectors $\vec{a}_1,\ldots,\vec{a}_k$ can be easily obtained as stated in the following theorem:

Theorem 1: Assume that vectors $\vec{a}_1,\ldots,\vec{a}_k$ are expressed via an orthonormal basis $\vec{e}_1,\ldots,\vec{e}_k$ with matrix A. The volume of the k-dimensional parallelotope $\mathcal{V}(\vec{a}_1,\ldots,\vec{a}_k)$ is equal to the determinant of matrix A, i.e.

$$vol\mathcal{V}(\vec{a}_1,\ldots,\vec{a}_k) = det|A| \tag{10}$$

The detailed proof can be found in many text books in advanced algebra, e.g., Vinberg.[40]

The volume of a k-dimensional parallelotope ($k > 1$), $\mathcal{V}(\vec{a}_1,\ldots,\vec{a}_k)$, is a measure of how divergent the vectors $\vec{a}_1,\ldots,\vec{a}_k$ are in the k-dimensional space. When $k = 2$, the absolute value of a 2×2

matrix determinant is the area of a corresponding parallelogram with the two row vectors as sides. When $k=3$, the absolute value of a 3×3 matrix determinant is the volume of a corresponding parallelepiped with the three row vectors as sides. Generally, the determinant of a matrix is an intuitive way to picture the volume of k-dimensional parallelotope spanned by its row vectors.

Consider the specific settings in medical diagnosis for diseases with k stages. For the ith ($i=1,\ldots,k$) disease class, the classification rate vector $\overrightarrow{\mathbf{P}}_i's$ ($i=1,\ldots,k$) defined in equation (2) corresponds to $\overrightarrow{a_i}$ and the SPM $\mathcal{P}$ in equation (4) corresponds to matrix A. According to the theorem stated above, MADET as defined in equation (9), is equal to the largest volume of the k-dimensional parallelotope spanned by $\overrightarrow{\mathbf{P}}_i's$ ($i=1,\ldots,k$). This directly measures how far apart the $\overrightarrow{\mathbf{P}}_i$'s are in the k-dimensional space. The further apart the $\overrightarrow{\mathbf{P}}_i$'s, the larger the difference between the diagnostic results for the k disease classes and hence the better the diagnostic accuracy of the test. Therefore, a larger MADET indicates a better diagnostic test. When MADET is equal to 1, the classification rate vectors $\overrightarrow{\mathbf{P}}_i's$ ($i=1,\ldots,k$) coincide with the corresponding orthonormal bases $\vec{e}_i's$ ($i=1,\ldots,k$) and A becomes a $k\times k$ identity matrix. For such cases, $P_{i,i}=1$ and $P_{i,j}=0$ for $i\neq j$ indicating a perfect biomarker which can correctly classify subjects in each class. On the other hand, MADET is 0 when the at least one vector $\overrightarrow{\mathbf{P}}_i's$ ($i=1,\ldots,k$) lies in the hyperplane spanned by the others, thus the set of vectors $\overrightarrow{\mathbf{P}}_i's$ ($i=1,\ldots,k$) are dependent. For example, for $k=3$, MADET will be 0 when 1) all three $\overrightarrow{\mathbf{P}}_i's$ ($i=1,\ldots,3$) coincide; 2) any two of the $\overrightarrow{\mathbf{P}}_i's$ ($i=1,\ldots,3$) coincide; 3) all three $\overrightarrow{\mathbf{P}}_i's$ ($i=1,\ldots,3$) fall on the same two-dimensional plane. In other words, from a geometrical perspective, MADET is equal to 0 when $\overrightarrow{\mathbf{P}}_i's$ ($i=1,\ldots,3$) do not form a parallelepiped together. Hence, as the number of stages k is fixed, MADET equal to 0 means the biomarker being considered has no diagnostic ability to distinguish among the k stages.

Note that equation (10) shows the defined volume $volV(\vec{a}_1,\ldots,\vec{a}_k)$, i.e. $det|A|$ can be negative. The sign of the value of the determinant provides some information on the orientation of the parallelotope. The sign of the $det(\mathcal{P})$ becomes opposite when two columns of $\mathcal{P}$ are switched; i.e. when the order of disease classes are switched. Since this paper focuses on diseases with ordinal stages, we use the absolute value of the determinants in the definition of MADET, i.e. $|det(\mathcal{P})|$, in equation (9).

From the definition of MADET, the geometric interpretation to justify its use as a diagnostic accuracy is very straightforward. The probabilistic interpretation of MADET for diseases with binary and three stages classification are also obtainable. However, for $k>3$, the determinant of the matrix $\mathcal{P}$ becomes very complicated, and it becomes a hard task to elucidate the probabilistic interpretation of MADET for diseases with general k stages. In the following, we discuss detailed probabilistic and geometric interpretation for binary and three stages classification.

3.2 Special case 1: binary classification

Consider diseases with binary status (1 for non-diseased, and 2 for diseased). The classification rate vector $\overrightarrow{\mathbf{P}}_i$'s defined in equation (2) equals to $(spe, 1-spe)^T$ for $i=1$ and $(1-sen, sen)^T$ for $i=2$ (where spe stands for the specificity and sen for the sensitivity). Therefore, the 2×2 SPM is

$$\mathcal{P}=\begin{pmatrix} spe & 1-spe \\ 1-sen & sen \end{pmatrix}$$

and the newly proposed diagnostic measure maximum absolute determinant (MADET) is

$$\text{MADET}=\max_{c_1}(spec+sen-1)$$

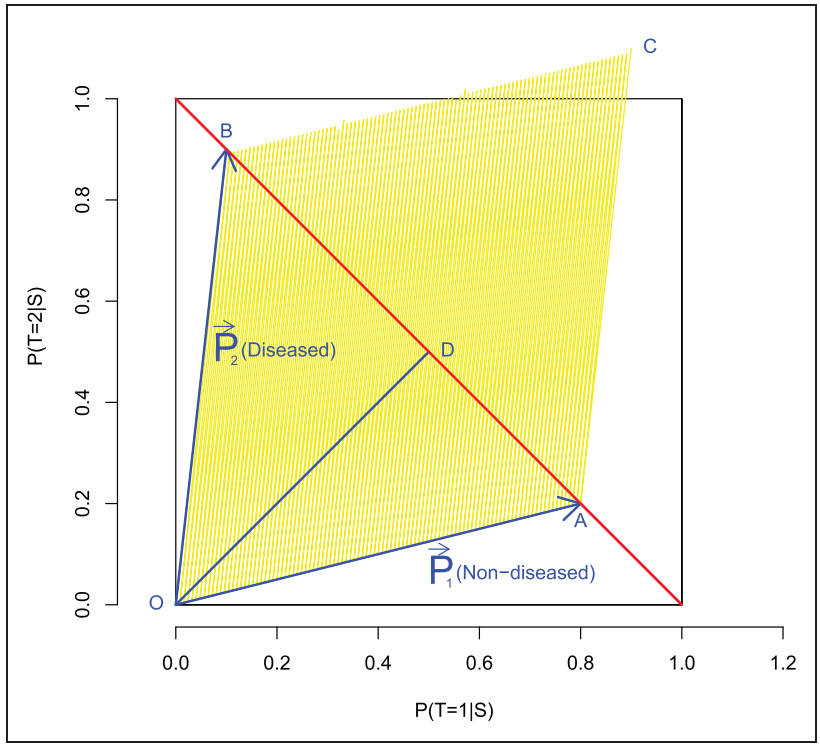

Figure 1. Illustration of MADET for the disease with binary classification. The red line denotes $P(T = 1|S) + P(T = 2|S) = 1$, and the area of the parallelogram OABC represents MADET.

It is obvious that MADET is the same as the Youden index.

As stated in the Introduction, the Youden index is a diagnostic measure as well as a criterion for optimal cut-point selection. The fact that Youden index is the same as MADET offers a new geometric interpretation for this well-known index, as shown in Figure 1. The classification rate vector for a randomly chosen non-diseased subject, $\vec{\mathbf{P}}_1 = (P_{1,1}, P_{1,2})$, is represented by the vector $\overrightarrow{OA}$, and the one for a diseased subject, $\vec{\mathbf{P}}_2 = (P_{2,1}, P_{2,2})$, is denoted by the vector $\overrightarrow{OB}$. Due to the fact $P_{i,1} + P_{i,2} = 1$, for $i = 1, 2$ as stated in equation (3), the ends of two vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$ always fall on the line $P(T = 1|S) + P(T = 2|S) = 1$. As stated in Section 3, MADET, i.e. the Youden index for diseases with binary classification, is equal to the area of the corresponding parallelogram with the two row vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$ as sides, indicated by the yellow shaded area: $OABC$. Meanwhile, one half of the MADET, is the area of the triangular ΔOAB. Since OD, the height of the triangular, is fixed as $\frac{\sqrt{2}}{2}$, MADET simply measures the length of the line segment AB, i.e. how far apart the two vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$ fall. Hence, for $k = 2$, MADET or Youden index serves as a direct measure of diagnostic accuracy; i.e. a measure of the difference between the classification rate vectors $\vec{\mathbf{P}}_1$ for a random non-diseased subject and $\vec{\mathbf{P}}_2$ for a random diseased subject. If $\overrightarrow{OA}$ rests at (1, 0) and $\overrightarrow{OB}$ at (0, 1), the corresponding MADET reaches the maximum 1. This means the biomarker has the perfect sensitivity and specificity. On the other side, if the diagnostic biomarker has no power for diagnosis, the classification rate vectors for a non-diseased subjects and a diseased subject will completely overlap, i.e. $\overrightarrow{OA}$ and $\overrightarrow{OB}$ overlap, yielding MADET as 0.

3.3 Special case 2: Diseases with three stages

Consider diseases with three ordinal stages. Let $i = 1,\ 2,\ 3$ denote the non-diseased stage, the early diseased stage and the fully diseased stage respectively. The corresponding SPM is

$$\mathcal{P} = \begin{pmatrix} P_{1,1} & P_{1,2} & P_{1,3} \\ P_{2,1} & P_{2,2} & P_{2,3} \\ P_{3,1} & P_{3,2} & P_{3,3} \end{pmatrix}$$

where $P_{i,j}$ stands for the probability of a subject from ith class being classified as jth class. We know that

$$P_{i,1} + P_{i,2} + P_{i,3} = 1, \quad \text{for} \quad i = 1, 2, 3 \tag{11}$$

Simple manipulations lead to

$$\begin{aligned} \text{MADET} &= \max_{c_1,c_2} \left| det \begin{pmatrix} P_{1,1} & P_{1,2} & P_{1,3} \\ P_{2,1} & P_{2,2} & P_{2,3} \\ P_{3,1} & P_{3,2} & P_{3,3} \end{pmatrix} \right| \\ &= \max_{c_1,c_2} |P_{1,1}P_{2.2}P_{3,3} + P_{1,2}P_{2,3}P_{3,1} + P_{1,3}P_{3,2}P_{2,1} \\ &\quad - P_{1,1}P_{2,3}P_{3,2} - P_{1,2}P_{2,1}P_{3,3} - P_{1,3}P_{3,1}P_{2,2}| \\ &= \max_{c_1,c_2} |1 - \left[P_{1,1}(1 - P_{2,2}) + P_{2,2}(1 - P_{3,3}) + P_{3,3}(1 - P_{1,1})\right] \\ &\quad - (P_{1,2}P_{2,1} + P_{1,3}P_{3,1} + P_{2,3}P_{3,2})|. \end{aligned} \tag{12}$$

Unlike the binary case, MADET for diseases with three stages is quite different from GYI, defined as $\max_{c_0,c_1,c_2,c_3}(P_{1,1} + P_{2,2} + P_{3,3} - 1)$.[29] Instead, MADET is a rather complicated function of the correct and false classification rates, due to the fact that it simultaneously takes all the correct and false classification rates into consideration.

From equation (12), MADET has the following probabilistic interpretation. Consider two subjects 1 and 2 from two different disease groups $(i,j;\ i,j \in \{1,2,3\},\ i \neq j)$ respectively, let S_l denote the true disease status for subject l, and T_l denote the corresponding diagnostic result, for $l = 1,\ 2$. MADET can be further written as

$$\begin{aligned} \text{MADET} = \max_{c_1,c_2} |1 &- P(T_1 = i, T_2 \neq j | S_1 = i, S_2 = j;\ i,j \in \{1,2,3\}, i \neq j) \\ &- P(T_1 \neq i, T_2 \neq j | S_1 = i, S_2 = j;\ T_1, T_2 \in \{i,j\};\ i,j \in \{1,2,3\}, i \neq j)| \end{aligned}$$

where $P(T_1 = i, T_2 \neq j | S_1 = i, S_2 = j;\ i,j \in \{1,2,3\}, i \neq j)$ stands for the probability that the first subject is correctly classified but the second is not, and $P(T_1 \neq i,\ T_2 \neq j | S_1 = i, S_2 = j;\ T_1, T_2 \in \{i,j\};\ i,j \in \{1,2,3\}, i \neq j)$ stands for the conditional probability that two subjects are classified into the wrong classes given the fact that the two disease classes for the two subjects are known, while which subject belongs to which class is unknown. In other words, MADET measures the total probability 1 with the penalization of the probability that the diagnostic biomarker correctly diagnoses the first subject but fails for the second, and the conditional probability that the diagnostic biomarker falsely classifies both subjects given the two groups $(i,\ j)$.

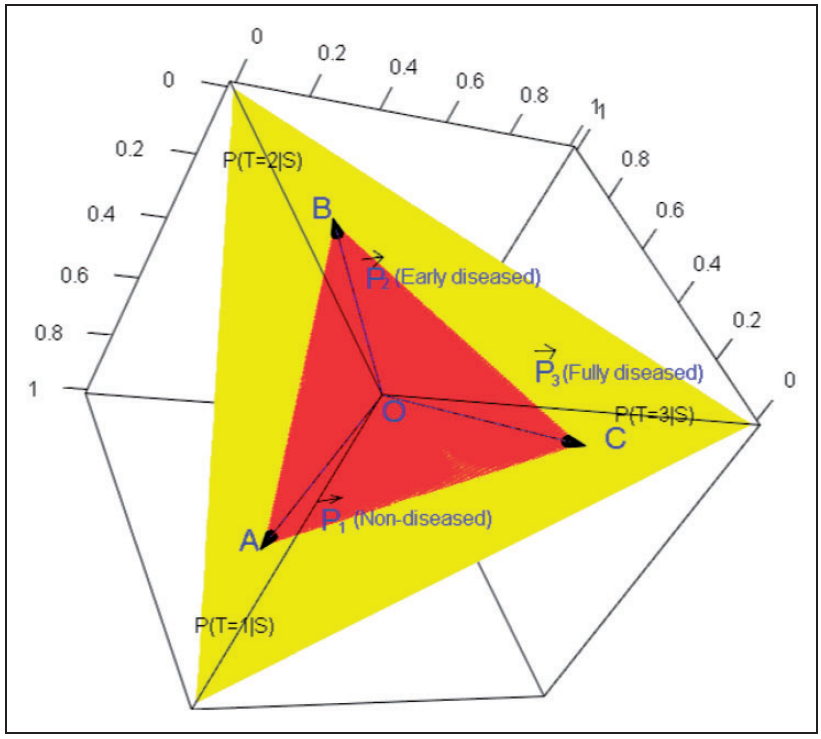

Figure 2. Illustration of MADET for diseases with three stages. The yellow plane denotes $P(T = 1|S) + P(T = 2|S) + P(T = 3|S) = 1$, and MADET is equal to the volume of the parallelepiped spanned by the three vectors, and one sixth of the MADET is the volume of the red tetrahedron *OABC*.

The geometric interpretation of MADET for the disease with three stages is as follows. In Figure 2, the classification rate vectors for three disease classes, i.e. $\vec{\mathbf{P}}_1$, $\vec{\mathbf{P}}_2$ and $\vec{\mathbf{P}}_3$ for randomly chosen non-diseased, early stage and diseased subjects are represented by vectors $\overrightarrow{OA}$, $\overrightarrow{OB}$ and $\overrightarrow{OC}$ respectively. Based on equation (11), the ends of $\overrightarrow{OA}$, $\overrightarrow{OB}$ and $\overrightarrow{OC}$ fall on the yellow plane defined as $P(T = 1|S) + P(T = 2|S) + P(T = 3|S) = 1$. According to the aforementioned theorem in Section 3, MADET is equal to the volume of the parallelepiped spanned by $\overrightarrow{OA}$, $\overrightarrow{OB}$ and $\overrightarrow{OC}$, and one sixth of the MADET denotes the volume of the tetrahedron $OABC$, i.e. the red shaded polyhedron. Similar to the two-dimensional case, the height of the tetrahedron, from point O to the plane $P(T = 1|S) + P(T = 2|S) + P(T = 3|S) = 1$, is fixed as $\frac{\sqrt{3}}{6}$. Therefore MADET measures the area of the triangular ΔABC on the yellow plane $P(T = 1|S) + P(T = 2|S) + P(T = 3|S) = 1$. The larger the MADET, the more diverse $\vec{\mathbf{P}}_1$, $\vec{\mathbf{P}}_2$ and $\vec{\mathbf{P}}_3$ are. Hence, larger value of MADET indicates larger difference among the classification rates vectors $\vec{\mathbf{P}}_1$, $\vec{\mathbf{P}}_2$ and $\vec{\mathbf{P}}_3$, and therefore better diagnostic ability of the biomarker. For a perfect diagnostic biomarker, $\vec{\mathbf{P}}_1$, $\vec{\mathbf{P}}_2$ and $\vec{\mathbf{P}}_3$ will be (1, 0, 0), (0, 1, 0) and (0, 0, 1), and the MADET will reach the maximum value 1. For a diagnostic biomarker with no diagnostic ability for distinguishing among the three disease classes, the three vectors $\overrightarrow{OA}$, $\overrightarrow{OB}$ and $\overrightarrow{OC}$ fall on the same plane, including scenarios such as all three vectors or any two of them overlap, yielding MADET as 0.

Table 1. Distributional scenarios for the power study.

	Distributions under H_0			Distributions under H_a		
Scenario	Non-diseased	Early diseased	Fully diseased	Non diseased	Early diseased	Fully diseased
Normal 1	N(0, 1)	N(0, 1)	N(0, 1)	N(0, 1)	N(0.5, 1)	N(1, 1)
Normal 2	N(0, 1)	N(0, 1)	N(0, 1)	N(0, 1)	N(0,2)	N(0,3)
Normal 5	N(0, 1)	N(5.3, 1)	N(5.4, 1)	N(0, 1)	N(5.3, 1)	N(5.4, 1)
Normal 3	N(0, 1)	N(5.3, 1)	N(5.4, 1)	N(0, 1)	N(4.8, 2)	N(5.5, 4)
Normal 4	N(0, 1)	N(5.3, 1)	N(5.4, 1)	N(0, 1)	N(2.0, 1)	N(3.7, 4)
Gamma 1	G(2, 6)	G(2, 6)	G(2, 6)	G(2, 6)	G(3, 6)	G(6, 6)
Gamma 2	G(2, 6)	G(2, 6)	G(2, 6)	G(2, 6)	G(2, 4)	G(2, 2)
Gamma 5	G(2, 1)	G(3.2, 4)	G(5.5, 2)	G(2, 1)	G(3.2, 4.0)	G(5.5, 2)
Gamma 3	G(2, 1)	G(3.2, 4)	G(5.5, 2)	G(2, 1)	G(5.0, 0.5)	G(6.1, 1)
Gamma 4	G(2, 1)	G(3.2, 4)	G(5.5, 2)	G(2, 1)	G(3.4, 0.2)	G(9.4, 1)
Exponential 1	E(1)	E(1)	E(1)	E(1)	E(0.5)	E(0.2)
Exponential 2	E(1)	E(1)	E(1)	E(1)	E(0.5)	E(0.1)
Exponential 5	E(1)	E(0.96)	E(0.002)	E(1)	E(0.96)	E(0.002)
Exponential 3	E(1)	E(0.96)	E(0.002)	E(1)	E(0.43)	E(0.029)
Exponential 4	E(1)	E(0.96)	E(0.002)	E(1)	E(0.56)	E(0.029)

4 Power study

In order to assess the performance of the newly proposed diagnostic measure MADET, a power study ***was*** conducted comparing our new methodology with two existing measures: GYI and VUS, in testing for differences between biomarker levels across disease stages.

We ***considered*** diseases with three stages; i.e. $k = 3$ under a variety of underlying distributions. The settings are presented in the Table 1. The true values for each measure under H_0 and H_a are shown in Table 2. Under normal, Gamma and exponential respectively, the first two scenarios have null hypothesis for which the three distributions are identical. Such scenarios are for testing the hypothesis if the biomarkers have any diagnostic accuracy at all. The last 3 scenarios are for testing if the biomarkers have some diagnostic ability. For scenario 3, H_0 and H_a are the same. Hence, scenario 3 is used for some level of checking type-I error control.

For each scenario, the sample sizes for three classes (n_1, n_2, n_3) are assumed as equal and set as 10, 20, 40. Random samples are simulated under H_0 for 2000 runs and 95% percentiles for the values of the statistics (VUS in equation (6), GYI in equation (7) and MADET in equation (8)) are estimated from the smoothed kernel distribution functions.[41] Under H_a, H_0 is rejected if the estimated statistics are greater than their corresponding 95% percentiles from H_0. The power for each measure is calculated as the proportion that H_0 is rejected out of the 1000 replications. The results are presented in Table 3.

As Table 3 shows, for scenario 3, all the three measures generally maintain the nominal level. For all other scenarios, as the sample sizes go up from 10 to 40, the powers increase. There is no clear winner in term of power. For the scenarios 1 and 2, the tests based on MADET have lower or comparable power than those based on GYI or VUS. Explanations can also be obtained based on Table 2. From null hypothesis consisting of three identical distributions to alternative hypothesis in the neighborhood, MADET increase more slowly than GYI or VUS, yielding a lower power for

Table 2. MADET, GYI and VUS under H_0 and H_a in Table 1.

	Normal 1			Normal 2			Normal 3			Normal 4			Normal 5		
Hypothesis	MADET	GYI	VUS	MADET	GYI	VUS	MADET	GYI	VUS	MADET	GYI	VUS	MADET	GYI	VUS
H_0	0.0000	0.0000	0.1667	0.0000	0.0000	0.1667	0.0392	0.5159	0.5281	0.0392	0.5159	0.5281	0.0392	0.5159	0.5281
H_a	0.0159	0.3948	0.3372	0.0103	0.2581	0.1674	0.0392	0.5159	0.5281	0.2109	0.5593	0.5480	0.3159	0.5691	0.5993
	Gamma 1			Gamma 2			Gamma 3			Gamma 4			Gamma 5		
	MADET	GYI	VUS	MADET	GYI	VUS	MADET	GYI	VUS	*MADET*	*GYI*	*VUS*	*MADET*	*GYI*	*VUS*
H_0	0.0000	0.0000	0.1667	0.0000	0.0000	0.1667	0.0516	0.3996	0.1811	0.0516	0.3996	0.1811	0.0516	0.3996	0.1811
H_a	0.1134	0.8181	0.5534	0.0480	0.5773	0.4259	0.0516	0.3996	0.1811	0.2980	0.4306	0.2065	0.4471	0.4556	0.2118
	Exp 1			Exp 2			Exp 3			Exp 4			Exp 5		
	MADET	GYI	VUS	MADET	GYI	VUS	MADET	GYI	VUS	*MADET*	*GYI*	*VUS*	*MADET*	*GYI*	*VUS*
H_0	0.0000	0.0000	0.1667	0.0000	0.0000	0.1667	0.0147	0.5001	0.5102	0.0147	0.5001	0.5102	0.0147	0.5001	0.5102
H_a	0.0529	0.5757	0.4202	0.1077	0.7850	0.5208	0.0147	0.5001	0.5102	0.2172	0.5344	0.6421	0.3005	0.5732	0.7054

Table 3. Simulated power for MADET, GYI and VUS.

	Normal 1			Normal 2			Normal 3			Normal 4			Normal 5		
Sample size	MADET	GYI	VUS	MADET	GYI	VUS	MADET	GYI	VUS	MADET	GYI	VUS	MADET	GYI	VUS
10	0.245	0.507	0.601	0.282	0.102	0.043	0.042	0.047	0.054	0.068	0.067	0.049	0.224	0.137	0.140
20	0.448	0.793	0.864	0.465	0.340	0.067	0.059	0.063	0.056	0.163	0.075	0.058	0.488	0.137	0.161
40	0.701	0.972	0.993	0.688	0.725	0.054	0.052	0.056	0.055	0.463	0.140	0.085	0.947	0.213	0.303
	Gamma 1			Gamma 2			Gamma 3			Gamma 4			Gamma 5		
	MADET	GYI	VUS	MADET	GYI	VUS	MADET	GYI	VUS	MADET	GYI	VUS	MADET	GYI	VUS
10	0.779	0.970	0.981	0.501	0.818	0.851	0.040	0.049	0.041	0.620	0.011	0.083	0.927	0.061	0.100
20	0.967	1.000	1.000	0.794	0.992	0.995	0.040	0.049	0.041	0.883	0.03	0.118	0.989	0.241	0.136
40	0.993	1.000	1.000	0.962	1.000	1.000	0.040	0.049	0.041	0.992	0.049	0.130	1.000	0.515	0.176
	Exp 1			Exp 2			Exp 3			Exp 4			Exp 5		
	MADET	GYI	VUS	MADET	GYI	VUS	MADET	GYI	VUS	MADET	GYI	VUS	MADET	GYI	VUS
10	0.779	0.970	0.981	0.501	0.818	0.851	0.047	0.058	0.043	0.211	0.226	0.228	0.412	0.392	0.407
20	0.967	1.000	1.000	0.794	0.992	0.995	0.043	0.048	0.048	0.448	0.322	0.420	0.758	0.591	0.724
40	0.993	1.000	1.000	0.962	1.000	1.000	0.044	0.063	0.059	0.726	0.435	0.649	0.991	0.840	0.952

distinguishing a scenario from such null hypothesis. For scenarios 4 and 5, the tests based on MADET have higher power than those based on GYI or VUS. For example, for the scenario "Gamma 4", when the sample size is 40, the test based on MADET achieves a power of 0.992, while the one based on GYI is merely at 0.05 nominal level and the one for VUS is only 0.13. Again, some explanations can be obtained using the results presented in Table 2. Under the scenario

Gamma 4, MADET is 0.2960 under H_a, a big increase from 0.0516 under H_0, while GYI and VUS only increase very slightly. This shows that MADET can capture some differences between null and alternative hypotheses in classification rates that GYI and VUS cannot.

5 Selection of optimal cut-point(s)

In diagnostic studies, it is critical to determine the optimal cut-points of a continuous biomarker for the purpose of making diagnosis. Not all the diagnostic measures of accuracy that we have discussed can be utilized as criteria for optimal cut-point selection. For example, HUM cannot be used to select cut-points, since it is defined over all the possible cut-points. The proposed new measure, MADET in equation (8), not only serves as an overall diagnostic measure, as explained in Section 3, but also a criterion for the optimal cut-points selection. For cases with general k stages, the set of cut-points $\overrightarrow{\mathbf{c}} = (c_1, c_2, \ldots, c_{k-1})$ that maximizes the absolute value of the determinant $det(\mathcal{P})$, is the optimal cut-points determined by MADET.

In the following, we address the issues regarding cut-point selections using the proposed criterion MADET in comparison with several existing methods. A simulation study was performed to assess their relative performance.

5.1 Selection criteria

There are three popular methods for obtaining the optimal cut-off point in the binary outcome disease case: Youden index (YI),[16,17] the closest-to-(0, 1) criterion (MD),[36] and the maximum area method (MV).[37] Note that the proposed MADET equals to Youden index for the binary case. Their corresponding objective functions, $YI(c_1)$ for the Youden index, $MD(c_1)$ for the closest-to-(0, 1) criterion, and $MV(c_1)$ for the maximum area approach, are as follows:

$$\begin{aligned} YI(c_1) &= P_{1,1} + P_{2,2} - 1, \\ MV(c_1) &= P_{1,1} * P_{2,2}, \\ MD(c_1) &= \sqrt{(1 - P_{1,1})^2 + (1 - P_{2,2})^2}, \text{respectively} \end{aligned}$$

The optimal cut-off point is obtained by maximizing $YI(c_1)$, $MV(c_1)$ or minimizing $MD(c_1)$ for c_1 respectively. Liu[37] also discussed the comparisons of optimal cut points selected by the three criteria.

For the 3-class disease case, we have as follows

$$\begin{aligned} \text{MADET}(c_1, c_2) &= \max_{c_1, c_2} |1 - [P_{1,1}(1 - P_{2,2}) + P_{2,2}(1 - P_{3,3}) + P_{3,3}(1 - P_{1,1})] \\ &\quad - (P_{1,2}P_{2,1} + P_{1,3}P_{3,1} + P_{2,3}P_{3,2}), \\ GYI(c_1, c_2) &= P_{1,1} + P_{2,2} + P_{3,3} - 1, \\ MV(c_1, c_2) &= P_{1,1} * P_{2,2} * P_{3,3}, \\ MD(c_1, c_2) &= \sqrt{(1 - P_{1,1})^2 + (1 - P_{2,2})^2 + (1 - P_{3,3})^2} \end{aligned}$$

The $MV(c_1, c_2)$, now called the max volume criterion, and $MD(c_1, c_2)$, the closest to perfection method, have been proposed by Attwood et al.[39] Similarly, the optimal cut-off points are obtained by maximizing $GYI(c_1, c_2)$, $MV(c_1, c_2)$ or minimizing $MD(c_1, c_2)$ for c_1, c_2 respectively.

For general case with k stages, the Youden index, the closest-to-(0, 1) criterion, and the maximum volume (MV) approach can be easily extended to any k stages diseases as follows

$$\begin{aligned}
\text{MADET}(c_1, c_2, \ldots, c_{k-1}) &= |det(\mathcal{P})|(c_1, \ldots, c_{k-1}),\\
GYI(c_1, c_2, \ldots, c_{k-1}) &= P_{1,1} + P_{2,2} + \cdots + P_{k,k} - 1,\\
MV(c_1, c_2, \ldots, c_{k-1}) &= P_{1,1} * P_{2,2} * \ldots * P_{k,k},\\
MD(c_1, c_2, \ldots, c_{k-1}) &= \sqrt{(1 - P_{1,1})^2 + (1 - P_{2,2})^2 + \cdots + (1 - P_{k,k})^2}
\end{aligned}$$

5.2 Indexes to evaluate the selection criteria

Before we start to compare the different selection criteria, it is critical to establish a few judging rules or indexes. The performance indexes concerning the accuracy of the estimation of the cut-points include relative bias and root mean square error (RMSE). The relative bias for a statistic T based on N rounds of iterations is

$$\text{Relative bias } (\hat{T}) = \left[\left(\frac{1}{N}\sum_{i=1}^{N}\widehat{T}_i - T\right)/T\right]$$

and its variance is

$$\text{Variance } (\hat{T}) = \sum_{i=1}^{N}(\widehat{T}_i - T)^2/(N-1)$$

The RMSE is defined as $\sqrt{Bias^2 + Variance\ (\hat{T})}$.

Another index for evaluating the selection criterion Q (MADET/GYI/MV/MD) is the total correct classification rate (TCCR) at the optimal cut-points $\vec{c}_Q$, which is defined as

$$TCCR_Q = \sum_{i=1}^{k} P_{i,i}(\vec{c}_Q).$$

The maximum of TCCR, i.e. $\text{TCCR}_{GY\ I}$, is achieved at the optimal cut-points $\vec{c}_{GYI}$. Therefore, the loss of TCCR of MADET is defined as

$$Loss = [(TCCR_{GYI} - TCCR_Q)/TCCR_{GYI}] \times 100\ \%$$

The final consideration for cut-points selection is the balance of CCRs among classes. For example, for $k = 2$, Youden index 0.4 corresponds to many scenarios, e.g. 1) 0.7 sensitivity and 0.7 specificity, 2) 0.5 sensitivity and 0.9 specificity, 3) 0.9 sensitivity and 0.5 specificity. While the preference depends on the specific diseases, generally speaking, the balance between sensitivity and specificity is more desirable and one may favor scenario 1) instead of the other two. To measure the balance, we define a measure called the maximum minimum difference (MMDIF) as

$$\text{MMDIF} = [max(P_{1,1}, \ldots, P_{k,k}) - min(P_{1,1}, \ldots, P_{k,k})]/min(P_{1,1}, \ldots, P_{k,k})$$

The smaller the MMDIF, the better the balance.

5.3 Simulations

Simulations were performed for the three-stage and four-stage diseases to evaluate the four selection criteria: MADET, GYI, MV and MD.

5.3.1 Three-stage diseases

Simulation was performed to compare MADET, GYI, MV and MD in selecting the optimal cut-off points for the three-class diseases. Two distributional settings from[29] and four new ones are employed and listed in Table 4. The sample sizes were set as (20, 20, 20), (50, 50, 50), (100, 100, 100) and separately. For a total of $R = 10{,}000$ rounds of replications, the $P_{i,j}$'s are estimated with the smoothed kernel distribution functions proposed by.[41] The empirical approach for the estimation was also attempted, but its performance ***was*** much worse than the kernel one. Hence, it was removed from consideration.

Tables 5 and 6 contain the simulation results for c_1 and c_2 given by MADET, GYI, MV and MD under the six different distributional settings. Generally estimates of c_1 and c_2 from MV and MD have the smallest RMSE, implying the estimates given by these two criteria are more accurate to the true cut-off points given by each criterion respectively. MADET has larger RMSE but is comparable or slightly better than GYI in terms of estimation accuracy.

Table 7 presents the loss of the TCCRs and the balance of the estimated CCRs. MADET has a smaller loss of total CCR for scenario 2, 4 and 6 than MD or MV. Moreover, MADET achieves better balance in scenario 1 and 3, especially for the larger or unbalanced sample sizes.

5.3.2 Four-stage diseases

The performance of MADET, GYI, MV and MD in the four-class disease scenario for determining optimal cut-points was also compared via a simulation study. Three scenarios from normal, gamma and exponential distributions respectively were selected and presented in Table 8. The sample sizes were set as (20, 20, 20, 20), (50, 50, 50, 50), and (100, 100, 100, 100) separately. The $P_{i,j}$'s are estimated with the smoothed kernel distribution functions, and c_1, c_2 and c_3 are selected under the MADET, GYI, MV and MD criteria for $R = 10{,}000$ rounds of replications. Relative Bias, RMSE, the Loss of total CCR and MMDIF are used as the key indexes to judge the performance of each criterion. The results are presented in Tables 9 to 12.

From Table 9 to 11, we can see that MD and MV provide more accurate estimate of cut-points than MADET and GYI. Table 12 shows that MADET can provide more balanced CCRs, i.e. small MMDIF with relatively higher percentage of loss of total CCR.

Table 4. Simulation distributional scenarios for diseases with three stages.

Scenario	Non-diseased	Early diseased	Fully diseased
1	N(0, 1.0)	N(0.5, 1.0)	N(1, 1.0)
2	N(0, 1.2)	N(0.5, 0.8)	N(1, 1.4)
3	Gamma(2, 1)	Gamma(3, 1.0)	Gamma(4, 1.0)
4	Gamma(2, 1)	Gamma(7, 0.3)	Gamma(12, 0.3)
5	Exp(1)	Exp(0.9)	Exp(0.035)
6	Exp(1)	Exp(0.5)	Exp(0.035)

Table 5. Relative bias and root mean squared error of the c_1 estimates for the three-class disease.

				n = m = k 20		n = m = k 50		n = m = k 100		n = 50,m = 30, k = 20	
Distribution	Scenario	Method	True c_1	R Bias	RMSE	R Bias	RMSE	R Bias	RMSE	R Bias	RMSE
Norm	1	MADET:	−0.229	−0.3419	0.5369	−0.3492	0.5179	−0.3469	0.5044	−0.1877	0.5286
		GYI:	0.250	−0.9106	0.8889	−0.4587	0.5027	−0.2353	0.3642	−0.5851	0.6087
		MD:	−0.035	0.8135	0.3250	0.6146	0.2100	0.4146	0.1545	0.3213	0.2358
		MV:	−0.046	0.3354	0.2800	0.3525	0.1929	0.2778	0.1442	0.2137	0.2180
	2	MADET:	−0.258	−0.0249	0.4517	0.0362	0.3166	0.0466	0.2402	0.0759	0.3693
		GYI:	−0.238	0.0136	0.5243	0.0722	0.3227	0.0852	0.2427	0.0655	0.3711
		MD:	−0.060	0.2558	0.2740	0.2659	0.1800	0.2391	0.1350	0.2103	0.2122
		MV:	−0.066	0.1318	0.2612	0.2116	0.1746	0.2034	0.1318	0.2217	0.2034
Gamma	3	MADET:	1.382	0.4038	0.8079	0.3561	0.7193	0.2855	0.6196	0.3304	0.6990
		GYI:	2.000	0.0891	0.7083	0.0817	0.4734	0.0680	0.3881	0.0843	0.5295
		MD:	1.727	0.0678	0.3285	0.0454	0.2183	0.0321	0.1615	0.0558	0.2529
		MV:	1.721	0.0699	0.3299	0.0462	0.2220	0.0322	0.1636	0.0543	0.2586
	4	MADET:	7.048	−0.1004	1.5364	−0.0543	1.0757	−0.0380	0.8156	−0.0586	1.1133
		GYI:	7.269	−0.1169	1.6291	−0.0554	1.1499	−0.0344	0.8770	−0.0602	1.1675
		MD:	5.734	−0.0698	1.1294	−0.0548	0.7763	−0.0445	0.5848	−0.0506	0.8355
		MV:	7.098	−0.1081	1.5444	−0.0551	1.0854	−0.0377	0.8255	−0.0609	1.1147
Exp	5	MADET:	1.030	0.1397	0.6499	0.1564	0.6410	0.1570	0.6398	−0.0216	0.6490
		GYI:	1.054	0.3155	1.3432	0.2821	1.2368	0.2536	1.0966	0.1985	1.4692
		MD:	0.735	0.1281	0.2695	0.0867	0.1613	0.0633	0.1141	0.0976	0.1886
		MV:	0.727	0.1486	0.2669	0.1003	0.1652	0.0728	0.1187	0.1137	0.1945
	6	MAD:	1.341	0.1454	0.6779	0.1266	0.5328	0.0845	0.4076	0.0951	0.5927
		GYI:	1.386	0.1748	0.7847	0.1179	0.5436	0.0729	0.4065	0.1456	0.6501
		MD:	1.026	0.1166	0.2800	0.0872	0.1802	0.0636	0.1333	0.0936	0.2026
		MV:	1.054	0.1446	0.3400	0.1021	0.2191	0.0732	0.1625	0.1122	0.2493

R Bias: relative bias; RMSE: root mean square error.

6 ADNI example

Alzheimer's disease (AD) is an irreversible neurodegenerative disease that results in a loss of mental function due to the deterioration of brain tissue. It is the most common cause of dementia among people over the age of 65, affecting an estimated 5.3 million Americans, but no prevention methods or cures have been discovered yet. The ADNI is a global research effort that aims to track the progression of the disease using biomarkers to assess the brain structure and function. It is a longitudinal study that assesses clinical, imaging, genetic and biospecimen biomarkers at 58 sites in the United States and Canada. The study had three phases: ADNI1, ADNIGO and ADNI2 and participants were followed and reassessed over time to track the pathology of the disease as it progresses.

The severity of dementia is staged by the clinical dementia rating (CDR) and a global CDR is derived from individual ratings in multiple domains by an experienced clinician. Here CDR 0 indicates no dementia and CDR 0.5, 1, 2, and 3 represent very mild, mild, moderate, and severe dementia, respectively. As patients with large CDR such as 2 or 3 are rarely available, we combined patients with CDR greater than or equal to 1 as the fully diseased group, while CDR 0 and 0.5 refer to the non-diseased and early diseased groups respectively. Data such as MRI and PET images,

Table 6. Relative bias and root mean squared error of the c_2 estimates for the three-class disease.

				n = m = k 20		n = m = k 50		n = m = k 100		n = 50,m = 30, k = 20	
Distribution	Scenario	Method	True c_1	R Bias	RMSE	R Bias	RMSE	R Bias	RMSE	R Bias	RMSE
Norm	1	MADET:	1.229	−0.0134	0.5385	−0.0395	0.5221	−0.0410	0.5001	−0.0591	0.5386
		GYI:	0.750	0.4384	1.0547	0.1902	0.5286	0.0960	0.3647	0.3358	0.8498
		MD:	1.035	0.0302	0.3180	0.0165	0.2082	0.0115	0.1561	0.0364	0.2807
		MV:	1.046	0.0191	0.2793	0.0127	0.1925	0.0094	0.1448	0.0239	0.2532
	2	MADET:	1.356	0.0240	0.4248	0.0256	0.2921	0.0189	0.2222	0.0175	0.3838
		GYI:	1.373	0.0383	0.4546	0.0319	0.2913	0.0246	0.2229	0.0392	0.4368
		MD:	1.171	0.0246	0.2992	0.0214	0.1922	0.0171	0.1442	0.0265	0.2594
		MV:	1.171	0.0174	0.2729	0.0188	0.1836	0.0156	0.1379	0.0166	0.2441
Gamma	3	MADET:	3.618	0.1469	1.1534	0.1223	1.0332	0.0987	0.9253	0.1069	1.0677
		GYI:	3.000	0.1934	1.3685	0.1200	0.8397	0.0840	0.6470	0.1537	1.1245
		MD:	3.551	0.0553	0.5976	0.0375	0.3960	0.0282	0.2929	0.0514	0.5395
		MV:	3.559	0.0532	0.5590	0.0369	0.3809	0.0282	0.2854	0.0481	0.5056
	4	MADET:	29.660	0.0211	2.9509	0.0143	2.0614	0.0107	1.5491	0.0187	2.7351
		GYI:	29.624	0.0190	2.9355	0.0128	2.0470	0.0095	1.5405	0.0167	2.7188
		MD:	29.989	0.0077	2.0260	0.0052	1.3158	0.0037	0.9389	0.0082	1.8675
		MV:	29.820	0.0137	2.3685	0.0095	1.6081	0.0070	1.1914	0.0137	2.1962
Exp	5	MADET:	5.950	−0.3002	2.2638	−0.2763	2.1732	−0.2552	2.1394	−0.3309	2.5147
		GYI:	3.754	0.1196	1.2012	0.1551	0.9935	0.1409	0.8194	0.1609	1.1921
		MD:	5.045	−0.1109	1.3829	−0.0010	1.0497	0.0359	0.8656	−0.0415	1.2469
		MV:	4.378	−0.0110	1.2128	0.0727	0.9868	0.0832	0.8134	0.0536	1.1676
	6	MADET:	7.857	−0.1201	2.3486	−0.0178	1.7272	0.0108	1.3681	−0.0512	2.1954
		GYI:	5.719	0.1106	1.7490	0.1211	1.3364	0.1140	1.1073	0.1435	1.6870
		MD:	7.145	−0.0082	1.8531	0.0455	1.3349	0.0469	1.0545	0.0436	1.7379
		MV:	6.508	0.0444	1.8112	0.0761	1.3311	0.0730	1.0757	0.0936	1.7376

R Bias: relative bias; RMSE: root mean square error.

genetics, cognitive tests, cerebrospinal fluid (CSF) and blood biomarkers are collected. Neuroimaging measures include brain volume, ventricular volume, and bilateral hippocampal volumes. Cognitive measures represent five domains respectively: memory, language, executive function, spatial ability, and attention. The CSF variables include T-tau, $A\beta_{42}$, p-tau181, the ratio of the first two variables, and the ratio of the last two variables. Genetic and other lab tests are not considered in our case. According to hypotheses proposed by Heneka et al.[42] and Mudher and Lovestone,[43] the biomarkers of interest include hippocampus volume (Hippocampus), brain volume (WholeBrain), T-tau (TAU), $A\beta_{42}$ (ABETA142), and p-tau181 (PTAU181P).

For the purpose of this paper, only the data at the 24th month visit is included and the disease status defined at that time point is adopted. There are 194, 290 and 183 subjects for the non-diseased, the early diseased, and the fully diseased groups respectively. However, due to missing values, the actual sample sizes for each variable are smaller. We use three diagnostic accuracy measures MADET, GYI and VUS, as well as the cut-off points selection criteria MD and MV discussed in the paper to evaluate the five biomarkers of interests: hippocampus volume (Hippocampus), brain volume (WholeBrain), T-tau (TAU), $A\beta_{42}$ (ABETA142), and p-tau181 (PTAU181P). The corresponding cut-points are determined by MADET, GYI, MD and MV, ***respectively***.

Table 7. Correct classification rate of each disease group, total CCR and MMDIF for the three-class disease.

Sample size	Distribution	Scenario	Method	$P_{1,1}$	$P_{2,2}$	$P_{3,3}$	Total CCR	Loss of CCR(%)	MMDIF
n = m = k = 20	Norm	1	MADET	0.4615	0.4616	0.4412	1.3643	6.1400	0.0462
			GYI	0.5622	0.3537	0.5377	1.4536	0.0000	0.5895
			MD	0.4945	0.4265	0.4929	1.4138	2.7300	0.1594
			MV	0.4914	0.4251	0.4894	1.4059	3.2800	0.1560
		2	MADET	0.4429	0.6646	0.4150	1.5224	1.1100	0.6014
			GYI	0.4541	0.6748	0.4106	1.5395	0.0000	0.6434
			MD	0.4884	0.5658	0.4584	1.5126	1.7400	0.2343
			MV	0.4883	0.5629	0.4595	1.5107	1.8700	0.2250
	Gamma	3	MADET	0.5588	0.4598	0.4462	1.4648	4.7600	0.2524
			GYI	0.6367	0.3252	0.5761	1.5381	0.0000	0.9579
			MD	0.5425	0.4410	0.5149	1.4984	2.5800	0.2302
			MV	0.5415	0.4415	0.5124	1.4954	2.7700	0.2265
		4	MADET	0.9968	0.7845	0.7901	2.5715	0.0100	0.2706
			GYI	0.9980	0.7808	0.7928	2.5716	0.0000	0.2782
			MD	0.9766	0.7853	0.7896	2.5514	0.7900	0.2436
			MV	0.9970	0.7821	0.7909	2.5700	0.0600	0.2748
	Exp	5	MADET	0.6303	0.3898	0.8316	1.8517	4.5500	1.1334
			GYI	0.6666	0.4421	0.8313	1.9399	0.0000	0.8803
			MD	0.5433	0.5202	0.8259	1.8894	2.6000	0.5877
			MV	0.5464	0.5145	0.8288	1.8897	2.5900	0.6109
		6	MADET	0.7652	0.4970	0.7794	2.0416	0.8000	0.5682
			GYI	0.7874	0.4792	0.7916	2.0581	0.0000	0.6519
			MD	0.6699	0.5832	0.7771	2.0302	1.3600	0.3325
			MV	0.6908	0.5657	0.7826	2.0391	0.9200	0.3834
n = m = k = 50	Norm	1	MADET	0.4598	0.4528	0.4501	1.3628	4.0900	0.0216
			GYI	0.5709	0.2870	0.5629	1.4208	0.0000	0.9892
			MD	0.4870	0.4140	0.4894	1.3904	2.1400	0.1821
			MV	0.4834	0.4177	0.4853	1.3865	2.4200	0.1618
		2	MADET	0.4275	0.6782	0.4041	1.5098	0.2600	0.6783
			GYI	0.4329	0.6827	0.3982	1.5138	0.0000	0.7145
			MD	0.4813	0.5607	0.4518	1.4939	1.3200	0.2410
			MV	0.4802	0.5609	0.4523	1.4934	1.3500	0.2401
	Gamma	3	MADET	0.5431	0.4640	0.4579	1.4650	2.9300	0.1861
			GYI	0.6284	0.2877	0.5931	1.5092	0.0000	1.1842
			MD	0.5309	0.4354	0.5138	1.4801	1.9300	0.2193
			MV	0.5293	0.4377	0.5118	1.4788	2.0200	0.2093
		4	MADET	0.9950	0.7789	0.7941	2.5680	0.0000	0.2774
			GYI	0.9967	0.7751	0.7963	2.5681	0.0000	0.2859
			MD	0.9740	0.7851	0.7913	2.5503	0.6900	0.2406
			MV	0.9954	0.7787	0.7932	2.5673	0.0300	0.2783
	Exp	5	MADET	0.6387	0.3765	0.8328	1.8480	2.9200	1.2120
			GYI	0.6539	0.4166	0.8331	1.9036	0.0000	0.9998
			MD	0.5325	0.5125	0.8194	1.8644	2.0600	0.5988
			MV	0.5331	0.5083	0.8262	1.8676	1.8900	0.6254

(continued)

Table 7. Continued

Sample size	Distribution	Scenario	Method	$P_{1,1}$	$P_{2,2}$	$P_{3,3}$	Total CCR	Loss of CCR(%)	MMDIF
		6	MAD	0.7638	0.4949	0.7651	2.0238	0.4900	0.5460
			GYI	0.7729	0.4686	0.7923	2.0338	0.0000	0.6908
			MD	0.6599	0.5807	0.7708	2.0115	1.1000	0.3274
			MV	0.6760	0.5631	0.7803	2.0194	0.7100	0.3857
n = m = k = 100	Norm	1	MADET	0.4573	0.4556	0.4479	1.3607	3.3500	0.0210
			GYI	0.5828	0.2465	0.5785	1.4079	0.0000	1.3643
			MD	0.4853	0.4093	0.4869	1.3816	1.8700	0.1896
			MV	0.4810	0.4152	0.4827	1.3789	2.0600	0.1626
		2	MADET	0.4218	0.6810	0.4022	1.5050	0.1200	0.6932
			GYI	0.4263	0.6845	0.3959	1.5067	0.0000	0.7290
			MD	0.4797	0.5593	0.4508	1.4898	1.1200	0.2407
			MV	0.4784	0.5603	0.4511	1.4897	1.1300	0.2421
	Gamma	3	MADET	0.5168	0.4767	0.4681	1.4616	2.4800	0.1040
			GYI	0.6220	0.2708	0.6060	1.4988	0.0000	1.2969
			MD	0.5251	0.4346	0.5147	1.4744	1.6200	0.2082
			MV	0.5232	0.4375	0.5127	1.4735	1.6900	0.1959
		4	MADET	0.9936	0.7773	0.7973	2.5683	0.0100	0.2783
			GYI	0.9954	0.7737	0.7993	2.5685	0.0000	0.2865
			MD	0.9737	0.7856	0.7932	2.5524	0.6300	0.2394
			MV	0.9940	0.7782	0.7956	2.5679	0.0200	0.2773
	Exp	5	MADET	0.6414	0.3729	0.8351	1.8494	2.2000	1.2395
			GYI	0.6548	0.3973	0.8390	1.8911	0.0000	1.1118
			MD	0.5284	0.5102	0.8201	1.8588	1.7100	0.6074
			MV	0.5276	0.5064	0.8299	1.8639	1.4400	0.6388
		6	MADET	0.7548	0.4970	0.7627	2.0145	0.5700	0.5346
			GYI	0.7629	0.4668	0.7964	2.0261	0.0000	0.7061
			MD	0.6538	0.5794	0.7729	2.0061	0.9900	0.3340
			MV	0.6678	0.5621	0.7835	2.0134	0.6300	0.3939
n = 50, m = 30, k = 20	Norm	1	MADET	0.4464	0.4544	0.4633	1.3641	5.1800	0.0379
			GYI	0.5634	0.3209	0.5543	1.4386	0.0000	0.7557
			MD	0.4922	0.4193	0.4910	1.4026	2.5000	0.1739
			MV	0.4871	0.4211	0.4888	1.3970	2.8900	0.1608
		2	MADET	0.4250	0.6740	0.4164	1.5155	0.6300	0.6186
			GYI	0.4342	0.6815	0.4094	1.5251	0.0000	0.6646
			MD	0.4823	0.5629	0.4562	1.5014	1.5600	0.2339
			MV	0.4798	0.5619	0.4583	1.5000	1.6500	0.2261
	Gamma	3	MADET	0.5312	0.4616	0.4761	1.4688	3.7200	0.1508
			GYI	0.6280	0.3039	0.5936	1.5255	0.0000	1.0665
			MD	0.5352	0.4396	0.5168	1.4916	2.2200	0.2175
			MV	0.5323	0.4414	0.5156	1.4894	2.3700	0.2059
		4	MADET	0.9942	0.7804	0.7929	2.5675	0.0100	0.2740
			GYI	0.9961	0.7761	0.7954	2.5677	0.0000	0.2835
			MD	0.9734	0.7851	0.7896	2.5480	0.7700	0.2398
			MV	0.9946	0.7802	0.7916	2.5664	0.0500	0.2748

(continued)

Table 7. Continued

Sample size	Distribution	Scenario	Method	$P_{1,1}$	$P_{2,2}$	$P_{3,3}$	Total CCR	Loss of CCR(%)	MMDIF
	Exp	5	MADET	0.5681	0.4261	0.8346	1.8287	4.2600	0.9587
			GYI	0.6279	0.4531	0.8291	1.9101	0.0000	0.8298
			MD	0.5339	0.5166	0.8201	1.8706	2.0700	0.5875
			MV	0.5352	0.5128	0.8243	1.8723	1.9800	0.6074
		6	MADET	0.7439	0.5074	0.7698	2.0210	0.8900	0.5171
			GYI	0.7750	0.4751	0.7891	2.0392	0.0000	0.6609
			MD	0.6619	0.5842	0.7712	2.0172	1.0800	0.3201
			MV	0.6787	0.5682	0.7777	2.0246	0.7200	0.3687

MMDIF is defined as : $[max(P_{1,1},P_{22},P_{33}) - min(P_{1,1},P_{22},P_{33})]/min(P_{1,1},P_{22},P_{33})$.

Table 8. Simulation distributional scenarios for the four-class disease.

Scenario	Non-diseased	Early diseased	Fully diseased	Severely diseased
1	N(0, 1)	N(0.5, 1)	N(1, 1)	N(1.5, 1)
2	Gamma(2, 1)	Gamma(3, 1)	Gamma(4, 1)	Gamma(5, 1)
3	Exp(1)	Exp(0.9)	Exp(0.035)	Exp(0.02)

Table 9. Relative bias and root mean squared error of the c_1 estimates for the four-class disease.

				n = m = k = 1 20		n = m = k = 1 50		n = m = k = 1 100	
Distribution	Scenario	Method	True c_1	R Bias	RMSE	R Bias	RMSE	R Bias	RMSE
Norm	1	MADET:	−0.527	−0.1970	0.4948	−0.2518	0.4726	−0.3078	0.4569
		GYI:	0.250	−1.2528	0.9475	−0.5683	0.5067	−0.3090	0.3711
		MD:	−0.129	0.3852	0.3845	0.2579	0.2339	0.2062	0.1776
		MV:	−0.174	0.1590	0.2876	0.1417	0.1987	0.1286	0.1511
Gamma	2	MADET:	1.071	0.5591	0.7773	0.5310	0.6993	0.5016	0.6487
		GYI:	2.000	0.0802	0.6702	0.0770	0.4652	0.0683	0.3829
		MD:	1.616	0.0740	0.3324	0.0484	0.2174	0.0367	0.1655
		MV:	1.576	0.0698	0.3079	0.0437	0.2044	0.0328	0.1554
Exp	3	MADET:	1.017	0.0555	0.7202	0.1099	0.6624	0.1185	0.5942
		GYI:	1.054	0.2985	1.4941	0.2854	1.4199	0.2593	1.2601
		MD:	0.716	0.1500	0.2762	0.1059	0.1674	0.0808	0.1205
		MV:	0.716	0.1582	0.2685	0.1119	0.1694	0.0847	0.1236

R Bias: relative bias; RMSE: root mean square error.

Table 10. Relative bias and root mean squared error of the c_2 estimates for the four-class disease.

				n = m = k = l 20		n = m = k = l 50		n = m = k = l 100	
Distribution	Scenario	Method	True c_1	R Bias	RMSE	R Bias	RMSE	R Bias	RMSE
Norm	1	MADET:	0.750	0.0116	0.4052	0.0066	0.3943	0.0125	0.3898
		GYI:	0.750	0.0250	0.4527	0.0190	0.3491	0.0096	0.2936
		MD:	0.750	0.0059	0.3288	−0.0007	0.2370	−0.0020	0.1851
		MV:	0.750	0.0041	0.2698	−0.0006	0.1924	−0.0025	0.1479
Gamma	2	MADET:	2.658	0.2839	1.0222	0.2705	0.9573	0.2548	0.9134
		GYI:	3.000	0.0940	0.7913	0.0644	0.6084	0.0494	0.5038
		MD:	3.004	0.0728	0.5621	0.0490	0.3906	0.0367	0.3027
		MV:	3.008	0.0646	0.4776	0.0440	0.3329	0.0329	0.2564
Exp	3	MADET:	5.422	−0.2774	2.0902	−0.2519	1.9733	−0.2403	1.9430
		GYI:	3.754	0.1335	1.2305	0.1649	1.0274	0.1413	0.8199
		MD:	3.896	0.0776	1.1724	0.1228	0.9569	0.1094	0.7587
		MV:	3.888	0.0789	1.1698	0.1232	0.9527	0.1105	0.7590

R Bias: relative bias; RMSE: root mean square error.

Table 11. Relative bias and root mean squared error of the c_3 estimates for the four-class disease.

				n = m = k = l 20		n = m = k = l 50		n = m = k = l 100	
Distribution	Scenario	Method	True c_1	R Bias	RMSE	R Bias	RMSE	R Bias	RMSE
Norm	1	MADET:	2.027	−0.0407	0.4971	−0.0601	0.4718	−0.0725	0.4558
		GYI:	1.250	0.3084	1.1176	0.1421	0.5648	0.0750	0.3721
		MD:	1.629	0.0380	0.4080	0.0196	0.2316	0.0142	0.1757
		MV:	1.674	0.0211	0.2858	0.0143	0.1958	0.0114	0.1503
Gamma	2	MADET:	5.272	0.0673	1.1563	0.0504	1.0453	0.0325	0.9674
		GYI:	4.000	0.2049	1.8667	0.1293	1.1143	0.0850	0.8185
		MD:	4.839	0.0577	0.8798	0.0387	0.5618	0.0278	0.4115
		MV:	4.922	0.0482	0.7011	0.0333	0.4786	0.0252	0.3586
Exp	3	MADET:	42.996	−0.0940	16.3399	−0.0723	13.0439	−0.0601	11.1184
		GYI:	37.308	0.0383	17.7952	0.0350	13.1900	0.0308	11.0858
		MD:	32.623	0.1319	9.7402	0.1013	6.5223	0.0817	4.8652
		MV:	32.791	0.1330	9.7954	0.0998	6.6383	0.0794	4.9482

R Bias: relative bias; RMSE: root mean square error.

The summary statistics of the five biomarkers are presented in Table 13, and optimality statistics, optimal cut-points, CCRs are presented in Table 14. Note that for Hippocampus, WholeBrain and ABETA142, smaller values indicate a more diseased status.

Hippocampus has the highest estimated MADET and VUS among the five biomarkers, while ABETA142 has the highest estimated GYI. Furthermore, the intervals between the two cut-points

Table 12. Correct classification rate of each disease group, total CCR and MMDIF for the four-class disease.

Sample size	Distribution	Scenario	Method	$P_{1,1}$	$P_{2,2}$	$P_{3,3}$	$P_{4,4}$	Total CCR	Loss of CCR(%)	MMDIF
n = m = k = l = 20	Norm	1	MADET	0.3658	0.3917	0.3900	0.3612	1.5087	9.6800	0.0844
			GYI	0.5370	0.3047	0.3049	0.5238	1.6704	0.0000	0.7624
			MD	0.4557	0.3539	0.3543	0.4535	1.6174	3.1700	0.2877
			MV	0.4390	0.3566	0.3571	0.4376	1.5903	4.8000	0.2311
	Gamma	2	MADET	0.4862	0.3993	0.3553	0.3779	1.6188	7.9300	0.3684
			GYI	0.6334	0.2872	0.2942	0.5434	1.7582	0.0000	1.2054
			MD	0.5124	0.3690	0.3606	0.4600	1.7019	3.2000	0.4210
			MV	0.4961	0.3726	0.3665	0.4455	1.6806	4.4100	0.3536
	Exp	3	MADET	0.5977	0.4028	0.5576	0.5131	2.0712	5.2100	0.4839
			GYI	0.6643	0.4444	0.5496	0.5266	2.1850	0.0000	0.4948
			MD	0.5406	0.5170	0.5470	0.5167	2.1213	2.9100	0.0586
			MV	0.5440	0.5128	0.5499	0.5142	2.1208	2.9400	0.0723
n = m = k = l = 50	Norm	1	MADET	0.3728	0.3850	0.3847	0.3696	1.5121	7.0400	0.0417
			GYI	0.5614	0.2573	0.2570	0.5509	1.6267	0.0000	1.1844
			MD	0.4484	0.3428	0.3431	0.4488	1.5832	2.6700	0.3092
			MV	0.4320	0.3515	0.3520	0.4321	1.5676	3.6300	0.2293
	Gamma	2	MADET	0.4791	0.4027	0.3525	0.3843	1.6186	5.5800	0.3591
			GYI	0.6270	0.2605	0.2674	0.5593	1.7143	0.0000	1.4069
			MD	0.4986	0.3594	0.3560	0.4563	1.6703	2.5700	0.4006
			MV	0.4828	0.3673	0.3644	0.4432	1.6577	3.3000	0.3249
	Exp	3	MADET	0.6178	0.3860	0.5737	0.4897	2.0673	3.0300	0.6005
			GYI	0.6593	0.4122	0.5561	0.5043	2.1319	0.0000	0.5995
			MD	0.5287	0.5077	0.5388	0.5135	2.0886	2.0300	0.0613
			MV	0.5308	0.5056	0.5404	0.5124	2.0892	2.0000	0.0688
n = m = k = l = 100	Norm	1	MADET	0.3801	0.3815	0.3786	0.3745	1.5147	5.9200	0.0187
			GYI	0.5770	0.2310	0.2321	0.5699	1.6099	0.0000	1.4978
			MD	0.4467	0.3387	0.3387	0.4470	1.5710	2.4200	0.3198
			MV	0.4297	0.3497	0.3501	0.4299	1.5594	3.1400	0.2293
	Gamma	2	MADET	0.4683	0.4043	0.3506	0.3944	1.6177	4.7800	0.3357
			GYI	0.6217	0.2516	0.2448	0.5806	1.6988	0.0000	1.5396
			MD	0.4921	0.3566	0.3531	0.4586	1.6604	2.2600	0.3937
			MV	0.4769	0.3658	0.3635	0.4447	1.6508	2.8300	0.3120
	Exp	3	MADET	0.6273	0.3772	0.5847	0.4763	2.0655	2.1700	0.6630
			GYI	0.6598	0.3923	0.5635	0.4958	2.1114	0.0000	0.6819
			MD	0.5237	0.5032	0.5387	0.5140	2.0797	1.5000	0.0705
			MV	0.5251	0.5020	0.5400	0.5131	2.0802	1.4800	0.0757

by GYI are shorter than those by MADET, implying that the classification criterion based on GYI tends to diagnose fewer subjects into the early diseased group than that based on MADET. Especially, for biomarkers Tau and PTAU181P, the cut-points c_1 and c_2 estimated by GYI coincide and the corresponding $P_{2,2}$'s are 0, which means none of the subjects will be diagnosed into the early diseased stage by GYI criterion. Such cut-points obviously lead to an inappropriate diagnosis since the results contradict with the three-stage setting for Alzheimer's disease. The estimates given by MD and MV are also reasonable. Hence, for Tau and PTAU181P, MADET/MD/MV might be better criteria to use for diagnosis of the three-stage Alzheimer's Disease.

Table 13. Summary statistics for ADNI data.

	CDR 0			CDR 0.5			CDR 1		
Biomarker	N	Mean	Std	N	Mean	Std	N	Mean	Std
Hippocampus	139	7037	914	175	6232	1181	107	5264	1086
WholeBrain	139	991771	95138	175	1006766	108749	107	928522	109449
Tau	24	72.12	28.09	50	103.90	57.63	26	120	66.27
ABETA142	24	212.04	52.83	48	163.15	47.54	26	137.85	26.16
PTAU181P	24	23.67	8.95	49	34.65	15.81	26	37.04	13.77

Table 14. Optimality statistics and cut-points estimates for ADNI data.

Biomarkers	Method	Optimal Statistic	Estimated C_1 / C_2	$P_{1,1}$	P_{22}	P_{33}
Hippocampus	MADET	0.0673	6593 / 4909	0.6937	0.4907	0.4536
	GYI	0.6615	6436 / 5192	0.7412	0.3813	0.5390
	MD	–	6737 / 5137	0.6406	0.4786	0.5243
	MV	–	6701 / 5130	0.6549	0.4704	0.5223
	VUS	0.4645	–	–	–	–
WholeBrain	MADET	0.0156	1,124,301 / 901,318	0.0819	0.6884	0.4291
	GYI	0.2444	1,018,843 / 9,22,049	0.4057	0.3516	0.4871
	MD	–	1,022,376 / 911,824	0.3937	0.3908	0.4583
	MV	–	1,021,869 / 909,893	0.3954	0.3938	0.4529
	VUS	0.2514	–	–	–	–
Tau	MADET	0.0336	68 / 112	0.5150	0.3285	0.3897
	GYI	0.4088	88 / 88	0.7803	0	0.6285
	MD	–	71 / 100	0.5651	0.2288	0.4936
	MV	–	69 / 109	0.5404	0.2972	0.4133
	VUS	0.3068	–	–	–	–
ABETA142	MADET	0.0131	161 / 142	0.8099	0.1918	0.6508
	GYI	0.6726	164/ 153	0.7906	0.0943	0.7877
	MD	–	176/ 137	0.7137	0.3360	0.5503
	MV	–	174/ 136	0.7249	0.3497	0.5223
	VUS	0.4416	–	–	–	–
PTAU181P	MADET	0.0372	27/ 40	0.7318	0.2916	0.3385
	GYI	0.5284	30/ 30	0.8280	0	0.7004
	MD	–	25/ 34	0.6143	0.2295	0.5446
	MV	–	24/ 36	0.5568	0.2852	0.4986
	VUS	0.3592	–	–	–	–

7 Discussion

For diseases with multiple stages, there are a few measures of diagnostic accuracy and cut-point selection criteria available. In this paper, we propose a new measure of diagnostic accuracy for the k stage setting. The new MADET method directly utilizes all the classification information and has an appealing geometric justification of its use as a diagnostic measure for disease with k stages ($k > 2$).

Additionally, the new measure can be utilized as a criterion to select optimal cut-points. Simulation results show that MADET can detect the differences in diagnostic accuracy that GYI and HUM fail to observe. Furthermore, simulation results about cut-point selection show that the proposed MADET method has performance comparable to that of GYI, and is capable of achieving better balanced rates while maintaining minimal loss of TCCR.

The MADET will be zero when any two (or more) classes are indistinguishable regardless of whether or not there is separation in other classes. When two classes or more overlap, there is an indication that class-merging should occur before MADET is calculated. We suggest the users of MADET to check the SPM in equation (4) when $\text{MADET} = 0$. When two or more rows of SPM overlap, a good practice is to merge these classes so that only one row in SPM is kept. Subsequently, MADET can be calculated based on the reduced SPM.

In this paper, we focus on the scenarios for diseases with ordinal stages. In many applications, e.g. genomic studies, we need to deal with k nominal classes. For such cases, there actually exist $k!$ possible classification decision rules; i.e. one for each possible permutation. The proposed measure MADET can be easily adapted to such cases due to the fact that switching columns (or rows) of SPM does not change the absolute value of the determinant of SPM (MADET), via two steps: (1) assume a hypothetical order for the k classes and calculate the corresponding MADET and the resulting SPM; (2) Permute the columns of SPM to maximize the CCRs or to minimize the false classification ones or both, depending on the interest. Consequently, the final optimal SPM as well as the corresponding optimal classification decision rule can be obtained.

It is worth noting that Skaltsa et al.[44] and Batterton and Schubert[45] have proposed two interesting metrics based on different cost functions in the context of incorporating the classification cost to evaluate a biomarker for the multiple-class diagnostic problems. For future work, we are also interested in incorporating classification costs to MADET.

An R-program is available on request from the corresponding author (ltian@buffalo.edu).

Acknowledgments

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.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.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimers Association; Alzheimers Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics,

N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. The ADNI research was also supported by NIH grants P30 AG010129 and K01 AG030514.

References

1. Shapiro D. The interpretation of diagnostic tests. *Stat Methods Med Res* 1999; **8**: 113–134.
2. Zhou XH, Obuchowski N and McClish D. *Statistical methods in diagnostic medicine*. Wiley: New York, 2002.
3. Pepe MS. *The statistical evaluation of medical tests for classification and prediction*. New York: Oxford University Press, Oxford Statistical Science Series Vol. 28, 2003.
4. Zou KH, Liu A, Bandos AI, et al. *Statistical evaluation of diagnostic performance topics in ROC analysis*. Chapman & Hall/CRC Biostatistics Series, 2011.
5. Bamber D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J Math Psychol* 1975; **12**: 387–415.
6. Hanley JA and McNeil BJ. The meaning and use of the area under the receiver operating characteristic (ROC) curve. *Radiology* 1982; **143**: 29–36.
7. DeLong E, DeLong D and Clarke-Pearson C. Comparing the area under two or more correlated ROC curves: a non-parametric approach. *Biometrics* 1988; **44**: 837–845.
8. Wieand S, Gail MH, James BR, et al. A family of nonparametric statistics for comparing diagnostic markers with paired and unpaired data. *Biometrika* 1989; **76**: 585–592.
9. Faraggi D and Reiser B. Estimation of the area under the ROC curve. *Stat Med* 2002; **21**: 3093–3106.
10. Obuchowski N. Fundamentals of clinical research for radiologists: ROC analysis. *Am J Roentgenol* 2005; **184**: 364–372.
11. Molodianovitch K, Faraggi D and Reiser D. Comparing the areas under two correlated ROC curves: parametric and non-parametric approaches. *Biometr J* 2006; **48**: 745–757.
12. Zou KH, O'Malley AJ and Mauri L. Receiver operating characteristic analysis for evaluating diagnostic tests and predictive models. *Stat Primer Cardiovasc Res* 2007; **115**: 654–657.
13. Mallett S, Halligan S, Thompson M, et al. Interpreting diagnostic accuracy studies for patient care. *BMJ* 2012; **345**: e3999.
14. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950; **3**: 32–35.
15. Schisterman EF, Perkins NJ, Liu A, et al. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology* 2005; **16**: 73–81.
16. Fluss R, Faraggi D and Reiser B. Estimation of the Youden index and its associated cutoff point. *Biometr J* 2005; **47**: 458–441.
17. Schisterman EF and Perkins NJ. Confidence intervals for the Youden Index and corresponding optimal cut-point. *Commun Stat: Simulation Comput* 2007; **36**: 549–563.
18. Lai C, Tian L and Schisterman EF. Exact confidence interval estimation for the Youden index and its corresponding optimal cut-point. *Comput Stat Data Anal* 2010; **56**: 1103–1114.
19. Scinto LFM and Daffner KR. *Early diagnosis of Alzheimer's disease*. Totowa, NJ: Humana Press, 2000.
20. Morris JC, Storandt M, Miller JP, et al. Mild cognitive impairment represents early-stage Alzheimer disease. *Arch Neurol* 2001; **58**: 397–405.
21. Xiong CJ, Belle GV, Miller JP, et al. Measuring and estimating diagnostic accuracy when there are three ordinal diagnostic groups. *Stat Med* 2006; **25**: 1251–1273.
22. Alonzo TA, Nakas TC, Yiannoutsos CT, et al. A comparison of tests for restricted orderings in the three-class case. *Stat Med* 2009; **28**: 1144–1158.
23. Mossman D. Three-way ROCs. *Med Decision Making* 1999; **19**: 78–89.
24. Dreiseitl S, Ohno-Machado L and Binder M. Comparing three-class diagnostic tests by three-way ROC analysis. *Med Decision Making* 2000; **20**: 323–331.
25. Heckerling PS. Parametric three-way receiver operating characteristic surface analysis using Mathematica. *Med Decision Making* 2001; **21**: 409–417.
26. Nakas CT and Yiannoutsos CT. Ordered multiple-class ROC analysis with continuous measurement. *Stat Med* 2004; **23**: 3437–3449.
27. Xiong CJ, Belle GV, Miller JP, et al. A parametric comparison of diagnostic accuracy with three ordinal diagnostic groups. *Biometr J* 2007; **49**: 682–693.
28. Tian L, Xiong C, Lai C, et al. Exact confidence interval estimation for the difference in diagnostic accuracy with three ordinal diagnostic groups. *J Stat Planning Inf* 2010; **141**: 549–558.
29. Nakas CT, Alonzo TA and Yiannoutsos CT. Accuracy and cut-off point selection in three-class classification problems using a generalization of the Youden index. *Stat Med* 2010; **29**: 2946–2955.
30. Maturi TC, Elkhafifi FF and Coolen FPA. Nonparametric predictive inference for three-group ROC analysis. Technical Report 2013.
31. Luo J and Xiong C. Youden index and associated cut-off points for three ordinal diagnostic groups. *Commun Statist: Simul Comput* 2013; **42**: 1213–1234.
32. Scurfield BK. Multiple-event forced-choice tasks in the theory of signal detectability. *J Math Psychol* 1996; **40**: 253–269.

33. Scurfield BK. Generalization of the theory of signal detectability to n-event m-dimensional forced-choice tasks. *J Math Psychol* 1998; **42**: 5–31.
34. Li J and Fine JP. ROC analysis with multiple classes and multiple tests: methodology and its application in microarray studies. *Biostatistics* 2008; **9**: 566–576.
35. Nakas CT, Dalrymple-Alford JC, Anderson TJ, et al. Generalization of Youden index for multiple-class classification problems applied to the assessment of externally validated cognition in Parkinson disease screening. *Stat Med* 2013; **32**: 995–1003.
36. Perkins NJ and Schisterman EF. The inconsistency of 'optimal' cut-off points using two roc based criteria. *Am J Epidemiol* 2006; **163**: 670–675.
37. Liu X. Classification accuracy and cut point selection. *Stat Med* 2012; **31**: 2676–2686.
38. He X and Frey EC. Three-class ROC analysis – the equal error utility assumption and the optimality of three class ROC surface using the ideal observer. *IEEE Trans Med Imaging* 2006; **25**: 979–986.
39. Attwood K, Tian L and Xiong C. Diagnostic thresholds with three ordinal groups. *J Biopharm Sta* 2014; **24**: 608–633.
40. Vinberg EB. *A course in algebra*. Rhode Island: American Mathematical Society, 2003.
41. Lloyd CJ and Yong Z. Kernel estimators of the ROC curve are better than empirical. *Stat Probab Lett* 1999; **44**: 221–228.
42. Heneka MT, Nadrigny F, Regen T, et al. Locus ceruleus controls Alzheimer's disease pathology by modulating microglial functions through norepinephrine. *Proc Natl Acad Sci USA* 2010; **107**: 6058–6063.
43. Mudher A and Lovestone S. Alzheimer's disease-do tauists and baptists finally shake hands? *Trends Neurosci* 2002; **25**: 22–26.
44. Skaltsa K, Jover L, Fuster D, et al. Optimum threshold estimation based on cost function in a multistate diagnostic setting. *Stat Med* 2011; **31**: 1098–1109.
45. Batterton KA and Schubert CM. Confidence intervals around Bayes Cost in multi-state diagnostic settings to estimate optimal performance. *Stat Med* 2014; **33**: 3080–3099.