Research Article

Statistics in Medicine

Received 19 December 2011, Accepted 7 August 2012 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5592

Generalization of Youden index for multiple-class classification problems applied to the assessment of externally validated cognition in Parkinson disease screening

Christos T. Nakas,[a*†] John C. Dalrymple-Alford,[b,c,d] Tim J. Anderson[b,d,e] and Todd A. Alonzo[f]

Routine cognitive screening in Parkinson disease (PD) has become essential for management, to track progression and to assess clinical status in therapeutic trials. Patients with mild cognitive impairment (PD-MCI) are more likely to progress to dementia and therefore need to be distinguished from patients with normal cognition and those with dementia. A three-class Youden index has been recently proposed to select cut-off points in three-class classification problems. In this article, we examine properties of a modification of the three-class Youden index and propose a generalization to k-class classification problems. Geometric and theoretical properties of the modified index J_k are examined. It is shown that J_k is equivalent to the sum of the $k-1$ two-class Youden indices for the adjacent classes of the ordered alternative problem given that the ordering holds. Methods are applied in the assessment of the Montreal Cognitive Assessment test when screening cognition in PD.

Keywords: k-class classification; Kolmogorov–Smirnov statistic; ROC analysis; Youden index

1. Introduction

New evidence of the frequency of cognitive impairment in Parkinson disease (PD) has made it a focus of intense investigation and stressed the need for appropriate therapeutic strategies [1, 2]. Decline to dementia (PD-D) and related nonmotor complications drive PD patients into institutional care and contribute most to caregiver burden [3, 4]. The cumulative prevalence of PD-D is about 80%, but there is wide individual variation in the duration between PD onset and dementia [5]. Routine cognitive screening has therefore become essential for management, to track progression and to assess clinical status in therapeutic trials. It is important to identify patients with mild cognitive impairment (PD-MCI) because they are more likely to progress to dementia [5]. The Montreal Cognitive Assessment (MoCA) has been recommended as a suitable brief screen for this purpose because it samples a broad range of cognitive domains and has better sensitivity for PD-MCI, unlike the commonly used Mini-Mental State Examination [6].

[a] *Laboratory of Biometry, University of Thessaly, Phytokou Street, 38446 N Ionia-Volos, Greece*
[b] *New Zealand Brain Research Institute, 66 Stewart St., Christchurch 8011, New Zealand*
[c] *Department of Psychology, University of Canterbury, Christchurch, New Zealand*
[d] *Department of Medicine, University of Otago, Christchurch, New Zealand*
[e] *Department of Neurology, Christchurch Hospital, Christchurch, New Zealand*
[f] *Division of Biostatistics, University of Southern California Keck School of Medicine, 440 E.Huntington Dr, Suite 400, Arcadia, CA 91066, U.S.A.*
**Correspondence to: Christos T. Nakas, University of Thessaly, Laboratory of Biometry, Phytokou Street, 38446 N Ionia- Volos, Greece.*
† *E-mail: cnakas@uth.gr*

The traditional approach to diagnostic classification, two-class ROC analysis, either combines the PD-D and PD-MCI groups into one impaired group to compare these patients with those showing normal levels of cognition (PD-N) or compares the PD-D group with a combination of all non-dementia patients (e.g., [7, 8]). Combining the PD-MCI group with either PD-D or PD-N patients is less than satisfactory when one wants to target disease modifying therapy to minimize progression to dementia. Thus, the simultaneous comparison of the three PD classes, that is, PD-N, PD-MCI, and PD-D, is ideal because many patients during the early phases of cognitive impairment and even early dementia are not easily discriminated on the basis of clinical judgment alone. ROC surfaces provide a new strategy to evaluate the diagnostic accuracy in such ordered three-class classification problems as a direct generalization of the two-class ROC curve [9]. [7] employed ROC surface methodology to assess the simultaneous discrimination of the three cognitive classes in PD on the basis of the order PD-N > PD-MCI > PD-D in terms of scores expected on a cognitive screening measure. By using externally validated cognitive status, the study by [7] provided clear evidence that the MoCA was significantly superior to the Mini-Mental State Examination when making this simultaneous classification and was nonsignificantly superior to a longer PD-focused test, the Scales for Outcomes in PD-Cognition [10]. While it is important to have a screen that examines all three PD classifications (PD-N, PD-MCI, and PD-D), it is also of interest to know how these groups and especially PD-N compare with healthy controls (i.e., the k-class problem when $k = 4$).

From a practical perspective, the selection of an optimal cut-off point is needed for diagnostic purposes. Use of the Youden index has intuitive appeal because the optimal cut-off point is the one that maximizes the sum of sensitivity and specificity. [7] employed the Youden index approach for pairwise classification. A generalization of the Youden index has been recently proposed for the assessment of accuracy and simultaneous cut-off point selection in three-class classification [11]. In this article, we modify this index and study its properties. This modification results in a unification of the three-class and the two-class approaches and in a natural generalization to k-class classification. We highlight the equivalence of the simultaneous selection of cut-off points by using the three-class, or k-class, approach with the cut-off point selection based on the two-class approach applied twice, or $k - 1$ times, for the adjacent classes, given that the anticipated ordering holds.

In the next section, we describe the Youden index in the framework of two-class classification problems. We provide a generalization in three-class classification problems along with the unification of two-class and three-class approaches in the three-class setting. In Section 3, we extend methods for the general k-class setting. In Section 4, we apply methods in the assessment of the MoCA test as a tool to screen externally validated cognition in PD. We end with a discussion on the utility of the methodological results.

2. Two-class and three-class ROC analysis

2.1. Two-class analysis

The goal in a two-class classification problem is to assess the ability of a test to accurately discriminate between two disease classes (1 and 2). Suppose that measurements from class 1, denoted by X_1, follow a distribution function F_1 and measurements from class 2, denoted by X_2, follow a distribution function F_2. If the ordering $X_1 < X_2$ is of interest, using cut-off point c to decide for class 1 when a measurement is less than c and for class 2 otherwise, the ROC curve is constructed by plotting the points $(FCF_{12}(c), TCF_2(c)), c \in (-\infty, \infty)$ in the unit square, where the false-class fraction is $FCF_{12}(c) = P[X_1 > c]$ and the true-class fraction is $TCF_2(c) = P[X_2 > c]$, in the strictly continuous case. If class 1 corresponds to healthy subjects and class 2 corresponds to diseased subjects, then FCF_{12} is equal to 1-specificity (*FPF*) and TCF_2 is equal to the sensitivity (*TPF*) of the test. We refer to false-class and true-class fractions in the sequel instead of false positive/negative and true positive/negative to generalize naturally to three-class and to k-class problems, where the notions of positive and negative are not meaningful.

An equivalent construction of the ROC curve is produced by plotting $(TCF_1(c), TCF_2(c))$, $c \in (-\infty, \infty)$ in the unit square, where $TCF_1(c) = P[X_1 < c]$. Given respective samples from the two classes under study, the empirical estimate of the ROC curve is constructed on the basis of the empirical estimates of $TCF_1(c)$ or $FCF_{12}(c)$, and $TCF_2(c)$. The area under the ROC curve (AUC) is widely used for the assessment of the diagnostic accuracy of the test under study, and it is equal to $P[X_1 < X_2]$ ([12], p.78). The empirical AUC is equivalent to the Wilcoxon–Mann–Whitney statistic. $AUC = 0.5$ when the two distributions completely overlap, and $AUC = 1$ when there is perfect discrimination with X_1 always less than X_2.

Once the diagnostic accuracy of a test is established, the selection of an optimal cut-off point is needed for practical discrimination. Use of the Youden index for optimal cut-off point selection has been supported in a number of articles (e.g., [13]). The Youden index is defined as

$$J_2 = \max_c\{TCF_1(c) + TCF_2(c) - 1\} = \max_c\{F_1(c) - F_2(c)\}. \tag{1}$$

It can be estimated parametrically on the basis of distributional assumptions [13] or nonparametrically by $\hat{J}_2 = \max_c\{F_{1n_1}(c) - F_{2n_2}(c)\}$, where F_{1n_1}, F_{2n_2} are the empirical distribution functions of F_1, F_2 for sample sizes equal to n_1 and n_2, respectively. Specifically, $F_{1n_1}(c) = \frac{1}{n_1}\sum_{i=1}^{n_1} I(X_1 < c)$, and $F_{2n_2}(c) = \frac{1}{n_2}\sum_{j=1}^{n_2} I(X_2 < c)$, where the indicator function $I(\cdot)$ equals one if the expression is true and equals zero otherwise. The relative importance of the TCF proportions for any given problem is reflected in the choice of the optimal cut-off point by introducing weights, ν and μ, in the definition of J_2 as follows: $J_2^+ = \max_c\{\nu \cdot TCF_1(c) + \mu \cdot TCF_2(c) - 1\}$. Properties of the weighted Youden index have been studied in [14]. By its definition, the Youden index is the maximum vertical distance from the ROC curve to the main diagonal ($TCF_2 = 1 - TCF_1$), and in practice, it is equivalent to the two-class Kolmogorov–Smirnov statistic, $\hat{D} = \sup_c|F_{1n_1}(c) - F_{2n_2}(c)|$.

2.2. *Three-class analysis*

ROC surfaces have been proposed for the evaluation of the diagnostic accuracy in ordered three-class classification problems as a direct generalization of the ROC curve (see, e.g., [15, 16]). Generalizations for the nominal three-class case have also been proposed [17, 18]. For the three-class case, suppose that measurements from class 1, denoted by X_1, follow a distribution with d.f. F_1 (i.e., $X_1 \sim F_1$); similarly, $X_2 \sim F_2$, and $X_3 \sim F_3$. A decision rule that classifies subjects in one of these classes can be defined using two ordered threshold points $c_1 < c_2$ (in the strictly continuous case). Specifically, suppose that the ordering of interest is $X_1 < X_2 < X_3$. Decide for class 1 when a measurement is less than c_1, for class 2 when it is between c_1 and c_2, and for class 3 otherwise. This decision rule will result in three TCFs and six FCFs. Then, $TCF_1 = P[X_1 < c_1]$, $TCF_2 = P[c_1 < X_2 < c_2]$, and $TCF_3 = P[X_3 > c_2]$. Also, $FCF_{12} = P[c_1 < X_1 < c_2]$, and the remaining five possible FCF_{ij}, $i, j = 1, 2, 3$, $i \neq j$, are defined accordingly.

By varying c_1, c_2 in the union of the supports of F_1, F_2, and F_3, all the TCFs can be plotted in three dimensions to produce the ROC surface in the unit cube. The true-class fractions take values in $[0, 1]$ with corner coordinates $\{(1,0,0),\ (0,1,0),\ (0,0,1)\}$. Thus, the ROC surface is the three-dimensional plot in the unit cube depicting $(F_1(c_1), F_2(c_2) - F_2(c_1), 1 - F_3(c_2))$, for all cut-off points (c_1, c_2), with $c_1 < c_2$. The functional form of the ROC surface [9] is $ROC_s(TCF_1, TCF_3) = F_2(F_3^{-1}(1 - TCF_3)) - F_2(F_1^{-1}(TCF_1))$. It can be seen that this is a generalization of the ROC curve in three dimensions because projecting the ROC surface to the plane defined by TCF_2 versus TCF_1, that is, setting $TCF_3 = 0$, the ROC curve between classes 1 and 2 is produced, that is, $ROC(TCF_1) = 1 - F_2(F_1^{-1}(TCF_1))$. The latter is the equivalent construction of the ROC curve depicting $(TCF_1(c_1), TCF_2(c_1))$ instead of $(FCF_{12}(c_1), TCF_2(c_1))$. Similarly, the projection of the ROC surface to the plane defined by the axes TCF_2, TCF_3, yields the ROC curve between classes 2 and 3, that is, $ROC(TCF_3) = F_2(F_3^{-1}(1-TCF_3))$, the latter being the functional form of TCF_2 versus TCF_3 analogous to specificity versus sensitivity rather than the other way around.

The Volume Under the ROC Surface (VUS) is equal to $P[X_1 < X_2 < X_3]$. An unbiased nonparametric estimator of VUS is given by $\hat{VUS} = \frac{1}{n_1 n_2 n_3}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} I(X_{1i}, X_{2j}, X_{3k})$, where $I(X_1, X_2, X_3)$ equals one if X_1, X_2, X_3 are in the correct order and zero otherwise [9]. The definition of $I(X_1, X_2, X_3)$ can be adapted to adjust for the presence of ties. Parametric approaches for the estimation of VUS have been discussed in [19]. VUS takes the value $\frac{1}{3!} = \frac{1}{6}$ when the three distributions completely overlap and the value one when the three classes are perfectly discriminated in the correct order.

A three-class Youden index has been recently proposed for the assessment of accuracy and cut-off point selection in the three-class setting [11]. Given that $X_1 < X_2 < X_3$, a modification of this index is defined here as follows:

$$\begin{aligned} J_{3;(1,2,3)} &= \max_{c_1, c_2; c_1 < c_2}\{TCF_1 + TCF_2 + TCF_3 - 1\} \\ &= \max_{c_1, c_2; c_1 < c_2}\{F_1(c_1) + F_2(c_2) - F_2(c_1) - F_3(c_2)\}. \end{aligned} \tag{2}$$

This is a constrained optimization problem with $c_1 < c_2$. J_3 can be estimated nonparametrically by using the empirical distribution functions in the definition earlier, that is, $\hat{J}_3 = \max_{c_1,c_2;c_1<c_2}\{F_{1n_1}(c_1) + F_{2n_2}(c_2) - F_{2n_2}(c_1) - F_{3n_3}(c_2)\}$, or parametrically on the basis of distributional assumptions for the data as in [11]. The pair of cut-off points c_1, c_2 that corresponds to J_3 is considered optimal and can be used in practice for screening purposes. As in the two-class setting, weights can be added to the definition of J_3 to reflect the relative importance of the three TCFs. Then, $J_3^+ = \max_{c_1,c_2;c_1<c_2}\{\nu \cdot TCF_1 + \mu \cdot TCF_2 + \lambda \cdot TCF_3 - 1\}$. Weights ν, μ, and λ will reflect prior probabilities of class assignments and/or classification costs. Variance estimates for J_3 and CIs can be constructed using resampling techniques, such as the bootstrap or permutation methods as described in [11] because in the three-class case, J_3 is a linear transformation of the index proposed in [11].

2.3. *Unification of two-class and three-class analysis approaches in a three-class setting*

The modified index lends itself to a natural unification of the two-class and three-class analysis approaches. Denote by $J_{3;(1,2,3)}$ the J_3 index corresponding to the ordering $X_1 < X_2 < X_3$ and by $J_{2;(i,j)}$ the J_2 index corresponding to the ordering $X_i < X_j$, $i, j = 1, 2, 3$. Then, by the definitions of J_2 and J_3 earlier, it follows that

$$\begin{aligned} J_{3;(1,2,3)} &= \max_{c_1,c_2;c_1<c_2}\{F_1(c_1) - F_2(c_1) + F_2(c_2) - F_3(c_2)\} \\ \overset{[c_1<c_2]}{} &= \max_{c_1}\{F_1(c_1) - F_2(c_1)\} + \max_{c_2}\{F_2(c_2) - F_3(c_2)\} \\ &= J_{2;(1,2)} + J_{2;(2,3)}. \end{aligned}$$

Thus, J_3 is the sum of the Youden index for the two-class analysis of classes 1 and 2 and the Youden index for the two-class analysis of classes 2 and 3, when $c_1 < c_2$. This result holds if weights are introduced because λ can be set to one and $\nu^* = \frac{\nu}{\lambda}$, $\mu^* = \frac{\mu}{\lambda}$ can be used instead of ν, μ in the definition of J_3^+. Then, $J^+_{3;(1,2,3)} = \max_{c_1,c_2;c_1<c_2}\{\nu^* \cdot TCF_1 + \mu^* \cdot TCF_2 + TCF_3 - 1\} = J^+_{2;(1,2)} + J^+_{2;(2,3)}$. This result holds whenever the ordering $X_1 < X_2 < X_3$ is true and $c_1 < c_2$. Counterexamples can easily be constructed.

Another useful result from the definition of J_3 in Section 2.2 is that J_3 is the maximum vertical distance from the ROC surface to the plane defined by the equation $TCF_1 + TCF_2 + TCF_3 = 1$. Figure 1 illustrates this fact. It appears that J_3 generalizes the Kolmogorov–Smirnov statistic in three-class discrimination problems. J_3 is defined in $[0, 2]$, being zero in the uninformative case when $F_1 = F_2 = F_3$ and two when there is no overlap in the three distributions and the ROC surface passes through the point $(1, 1, 1)$ of the unit cube. In the latter case, the maximum vertical distance to the plane $TCF_1 + TCF_2 + TCF_3 = 1$ equals 2. As a consequence, the cut-off points derived using the three-class approach are the same as those derived by using the two-class approach twice for adjacent classes (1 vs 2) and (2 vs 3), given that $c_1 < c_2$ for the separate analyses of the adjacent classes.

3. The k-class classification problem

Generalizing in k-class classification, suppose that $X_1 \sim F_1, \ldots, X_k \sim F_k$ and that the ordering of interest is $X_1 < \cdots < X_k$. A k-dimensional ROC manifold can be defined, but cannot be visualized, using $k-1$ ordered cut-off points $c_1 < \cdots < c_{k-1}$ and a decision rule analogous to the three-class case. By varying the $k-1$ cut-off points in the union of the supports of $F_1, \ldots, F_k$, points $(TCF_1(c_1, \ldots, c_{k-1}), \ldots, TCF_k(c_1, \ldots, c_{k-1}))$ are defined that produce the manifold. The hypervolume under the k-dimensional ROC manifold (HUM) is equal to $P[X_1 < \cdots < X_k]$, and it varies from $\frac{1}{k!}$, when $F_1 = \cdots = F_k$, to 1 when the class distributions are perfectly discriminated in the anticipated ordering [9].

Statistics in Medicine

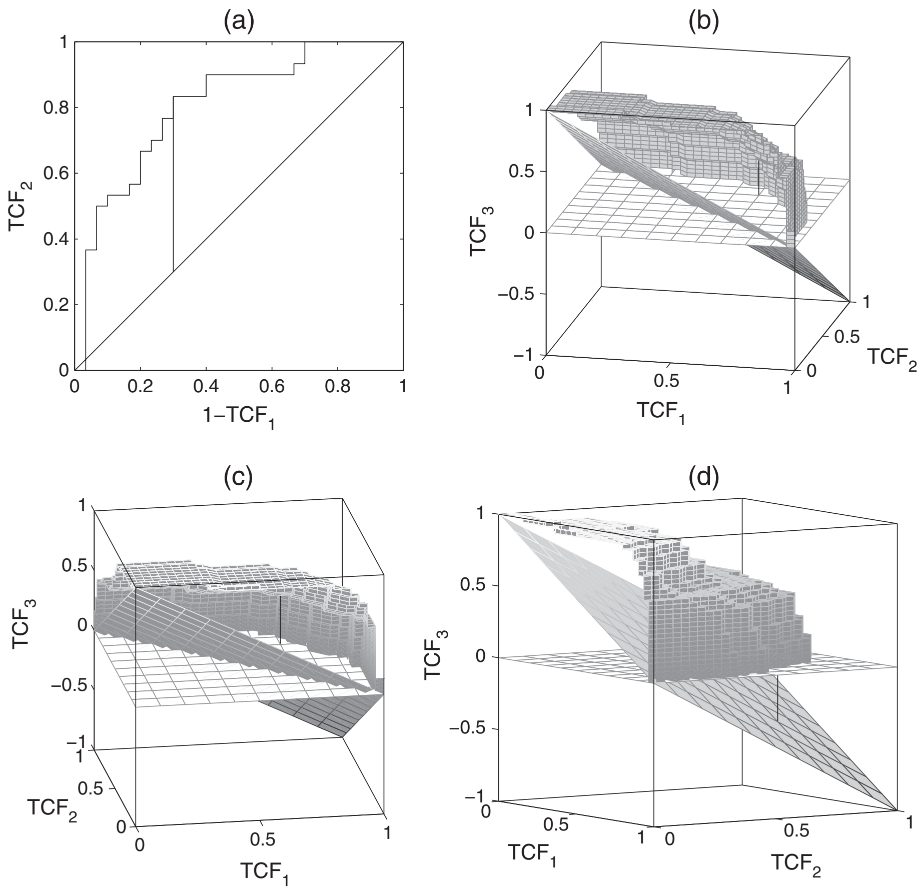

Figure 1. Panel (a) illustrates J_2, which is the maximum vertical distance from the ROC curve to the line $TCF_1 + TCF_2 = 1$. Panels (b), (c), and (d) illustrate J_3 from different viewpoints. J_3 is the maximum vertical distance from the ROC surface to the plane $TCF_1 + TCF_2 + TCF_3 = 1$. A cube of dimensions $1 \times 1 \times 2$ divided in two by the plane $z = 0$ is shown. The upper half includes the ROC surface. The lower half is shown to fully illustrate J_3.

Given that the ordering of interest is $X_1 < \cdots < X_k$, we introduce the k-class Youden index as follows:

$$\begin{aligned}
J_{k;(1,2,\ldots,k-1)} &= \max_{c_1,\ldots,c_{k-1};c_1<\cdots<c_{k-1}}\{TCF_1 + \cdots + TCF_k - 1\} \\
&= \max_{c_1,\ldots,c_{k-1};c_1<\cdots<c_{k-1}}\{F_1(c_1) - F_2(c_1) + F_2(c_2) - F_3(c_2) \\
&\quad + F_3(c_3) - F_4(c_3) + \cdots + F_{k-1}(c_{k-1}) - F_k(c_{k-1})\} \\
&\overset{[c_1<\cdots<c_{k-1}]}{=} \max_{c_1}\{F_1(c_1) - F_2(c_1)\} + \max_{c_2}\{F_2(c_2) - F_3(c_2)\} \\
&\quad + \cdots + \max_{c_{k-1}}\{F_{k-1}(c_{k-1}) - F_k(c_{k-1})\} \\
&= J_{2;(1,2)} + J_{2;(2,3)} + \cdots + J_{2;(k-1,k)}.
\end{aligned}$$

Thus,

$$J_{k;(1,2,\ldots,k-1)} = \sum_{i=1}^{k-1} J_{2;(i,i+1)} = J_{k-1;(1,2,\ldots,k-2)} + J_{2;(k-1,k)}. \tag{3}$$

That is, J_k is the sum of the Youden indices for the adjacent classes, when $c_1 < \cdots < c_{k-1}$. It is also the maximum vertical distance from the k-dimensional ROC manifold to the hyperplane defined by $TCF_1 + \cdots + TCF_k = 1$. J_k varies from zero when $F_1 = \cdots = F_k$ because all pairwise Youden indices are zero in that case to $k-1$ when the class distributions are perfectly discriminated in the correct order because all pairwise Youden indices will then be equal to one. As in the three-class case, J_k can be estimated nonparametrically by using the empirical distribution functions or parametrically on the basis of distributional assumptions for the data.

The bootstrap or permutation methods are an easy approach, however computationally intensive, to calculate the variance of J_k or construct CIs. Weights can be introduced in the definition of J_k to reflect the relative importance of the different true-class fractions. Although not described in detail here, weightings are a crucial consideration when classes have very different prevalences.

4. Parkinson disease screening

Our consecutive convenience sample of volunteers at the central Movement disorders clinic for a region of 400,000 people (Christchurch, NZ) comprised of 24 PD-D patients, 36 PD-MCI patients, and 80 PD-N patients. The MoCA scores of a control group of 50 age-matched healthy volunteers were also available. A subset of these data has been examined in [7]. External validation of cognitive status was made using 20 measures across four domains (executive function, attention and working memory, learning and memory, and visuoperception) and assessment of everyday cognitive function [2,20]. Descriptive statistics for the MoCA scores of the four classes are given in Table I. Figure 2 shows boxplots of the MoCA scores for each class. Nonparametric estimates of all quantities of interest were based on the corresponding empirical distribution functions [9, 11, 18]. Standard percentile bootstrap methodology was used for the construction of 95% CIs by using $b = 1000$ replicates, following the suggestions in [21]. The percentile bootstrap uses the percentiles of the bootstrap distribution of an estimate for the construction of CIs (see, e.g., [22], p.202). As the first step, diagnostic accuracy for the discrimination of the four classes using the MoCA was assessed via four class ROC methodology [9]. A nonparametric approach was adopted for the application. The hypervolume under the ROC manifold was estimated to be equal to 0.299 (95% CI: 0.220, 0.389) showing significant discrimination of the four classes under study compared with the uninformative case where $HUM = \frac{1}{24}$. The respective J_4 index was also calculated ($J_4 = 1.204$; 95% CI: 1.106, 1.477). However, the healthy controls and the PD-N patients were virtually indistinguishable ($AUC = 0.535$; 95% CI: 0.438,0.631, whereas $J_{2;(\text{PD-N, Controls})} = 0.093$; 95% CI: 0.025, 0.273). Illustration of the k-class classification problem was therefore restricted to the three PD groups, which is the primary question in a clinical setting.

ROC surface methodology was applied to the discrimination between PD-D, PD-MCI, PD-N groups by using the MoCA. Figure 3 depicts the ROC curves that are used to assess the discrimination of adjacent classes and the ROC surface corresponding to the PD-D $<$ PD-MCI $<$ PD-N ordering.

Panels (c) and (d) of Figure 3 provide the ROC surface illustrating the ability of MoCA to accurately discriminate the three groups of PD patients in the ordering PD-D $<$ PD-MCI $<$ PD-N. The corresponding VUS is 0.717 (95% CI: 0.625, 0.807). By using Equation 2, $J_3 = 1.111$ (95% CI: 0.993, 1.367), with screening cut-off points $c_1 = 21$ (95% CI: 17, 23) and $c_2 = 25$ (95% CI: 24, 25). That is, patients whose MoCA score is less than or equal to 21 will be considered as PD-D, whereas patients with scores greater than 21 but less than or equal to 25 will be considered as PD-MCI, and the remaining will be classified as PD-N. These choices result in $TCF_{\text{PD-D}} = 0.833$, $TCF_{\text{PD-MCI}} = 0.528$, $TCF_{\text{PD-N}} = 0.750$. Estimated cut-off points do not differ from those proposed by [7], where the Youden index of pairwise analyses was used for a subset of the current dataset.

Two-class methodology, that is, pairwise analyses, produced equivalent results: for classes PD-D and PD-MCI, denote by TCF' the resulting true-class fractions. From Equation 1, we estimate $J_{2;(\text{PD-D, PD-MCI})} = 0.583$ (95% CI: 0.486,0.792), $c_1 = 21$ (95% CI: 17, 23), $AUC = 0.900$ (95% CI: 0.827,0.959), resulting in $TCF'_{\text{PD-D}} = 0.833$, $TCF'_{\text{PD-MCI}} = 0.750$. The corresponding ROC curve is shown in Panel (a) of Figure 3. Similarly for classes PD-MCI and PD-N, denote by TCF^* the resulting true-class fractions. Then, $J_{2;(\text{PD-MCI, PD-N})} = 0.528$ (95% CI: 0.379,0.694), $c_2 = 25$ (95% CI: 24,

Table I. Descriptive statistics for the MoCA scores of the PD-D, PD-MCI, and PD-N classes.

	PD-D	PD-MCI	PD-N	Controls
Mean	17.33	23.58	26.79	27.08
SD	4.20	2.80	2.07	1.98
Median	18	24	27	27
Min	10	18	22	23
Max	23	29	30	30
N	24	36	80	50

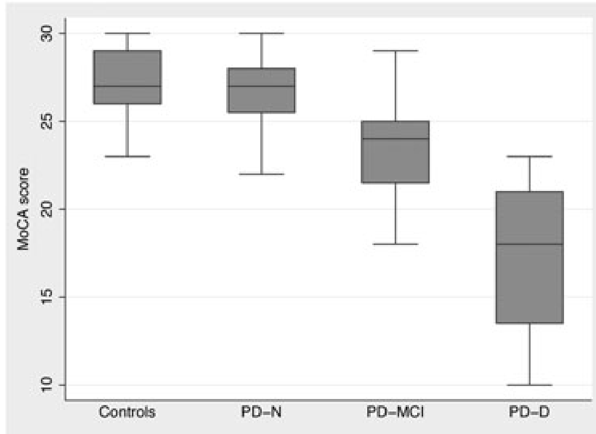

Figure 2. Boxplots of MoCA scores for the four classes.

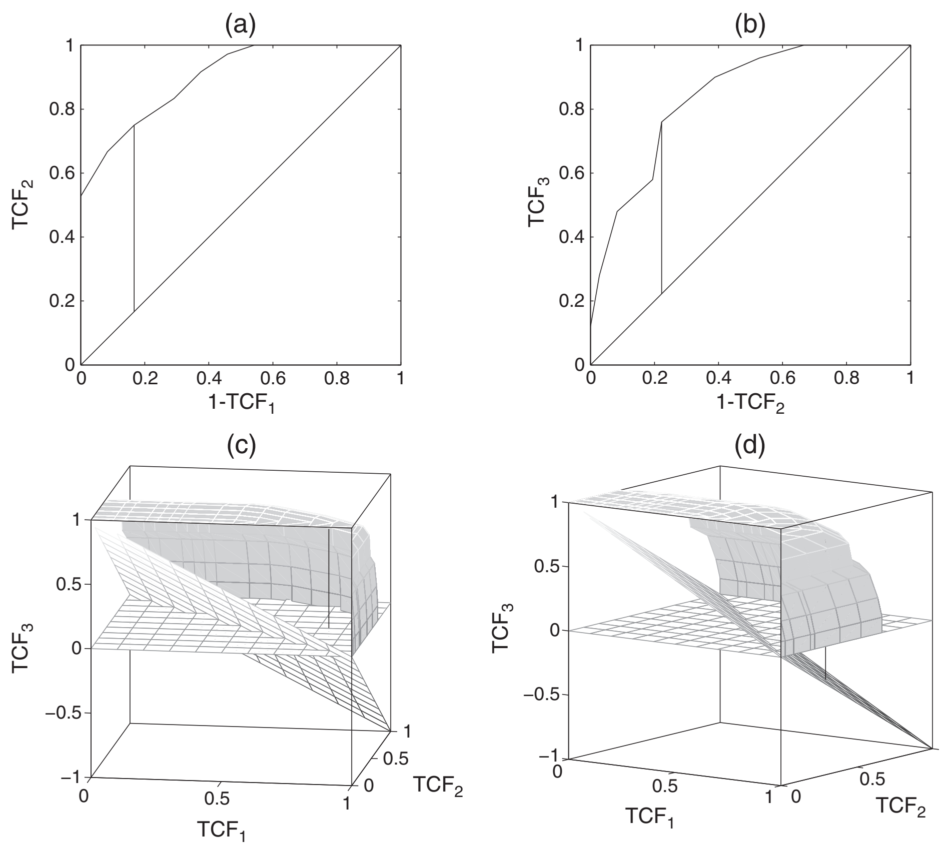

Figure 3. (a) ROC curve for PD-D, PD-MCI discrimination. (b) ROC curve for PD-MCI, PD-N discrimination. (c) and (d) Two viewpoints of the ROC surface depicting the 1: PD-D < 2: PD-MCI < 3: PD-N ordering for the MoCA test. J_2 is shown as the vertical lines on Panels (a) and (b); J_3 is shown across Panels (c) and (d).

25), $AUC = 0.817$ (95% CI: 0.723,0.898), resulting in $TCF^*_{\text{PD-MCI}} = 0.778$, $TCF^*_{\text{PD-D}} = 0.750$. The corresponding ROC curve is shown in Panel (b) of Figure 3.

An important contribution of this analysis is, comparing two-class and three-class problems, to verify that

$$J_3 = J_{2;(\text{PD-D, PD-MCI})} + J_{2;(\text{PD-MCI, PD-N})} = 0.583 + 0.528 = 1.111,$$

and the same cut-off points are produced. Also, $TCF_{\text{PD-D}} = TCF'_{\text{PD-D}}$, $TCF_{\text{PD-N}} = TCF^*_{\text{PD-N}}$, and $TCF_{\text{PD-MCI}} = 1 - (1 - TCF'_{\text{PD-MCI}}) - (1 - TCF^*_{\text{PD-MCI}})$. In addition, $J_3 = J_4 - J_{2;(\text{PD-N, Controls})}$.

Interestingly, by discarding the PD-N group and analyzing the data for the ordering PD-D < PD-MCI < Controls, equivalent results are produced with $c_1 = 21$ and $c_2 = 25$, whereas $TCF_{\text{PD-D}} = 0.833$, $TCF_{\text{PD-MCI}} = 0.528$, $TCF_{\text{Controls}} = 0.760$, and $VUS = 0.741$, $J_3 = 1.121$. Also, c_3 in the four-class analysis is the same as the optimal cut-off point based on $J_{2;(\text{PD-N, Controls})}$, that is, $c_3 = 29$.

We conclude that all three classes are significantly discriminated, and estimated screening cut-off points can be used when routinely assessing cognitive performance by using the MoCA. Timely intervention for the PD-MCI group could result in an optimal management of PD complications.

5. Discussion

Our work extends the Youden index to k-class classification problems. The index J_k has intuitive theoretical and geometric properties and can be used in practice for the assessment of screening tests in k-class problems and for the detection of optimal cut-off points in such problems when a monotone ordering exists between the classes under study. Our new analysis has shown that in the common k-class classification problem with monotone ordering, one can arrive at the same results as with the k-class approach by following two-class methodology for the adjacent classes. In this context, however, the k-sample approach directly addresses the problem of k-class classification to show explicitly the value of a given measure in discriminating k-groups simultaneously. The two-class approach is therefore recommended as a standard post hoc approach to target discrimination between specific adjacent classes. A rather restrictive limitation of this approach, which was underlined earlier, is that the $k-1$ cut-off points for the adjacent classes need to be ordered, that is, $c_1 < c_2 < \cdots < c_{k-1}$. Pairwise analyses may well result in cut-off points that violate this assumption. In that case, pooling adjacent classes is an option. This strategy is used often in practice for constrained statistical inference applications [23].

In practice, J_k also constitutes a generalization of the Kolmogorov–Smirnov statistic in k-class discrimination problems. Such extensions have been studied in the literature (e.g., [11, 24–26]) but lack the intuitive generalization and geometric properties of J_k. In this work, we have studied properties of the theoretical index and have used empirical estimates for the application. Various estimation methods can be used for J_k such as empirical, other nonparametric methods such as kernel estimation, or parametric approaches based on distributional assumptions. Assessing accuracy and precision of different estimation approaches for J_k is an issue of future research. However, the simulation results for bias and mean squared error (MSE) in [11] are relevant to the three-class setting. Formal comparison with alternative decision rules for multiple-class classification, for example, based on likelihood ratios, also warrants further research. Bayesian decision methods are primarily based on a minimum misclassification error approach using the expected utility index. Such approaches are presented in Chapter 2 of [27]. The weighted Youden index is in fact a degenerate case of the expected utility index.

A simultaneous k-class classification approach has shown the value of the MoCA as a screen for different stages of cognition status in PD and comparison with a fourth group, such as healthy controls. The combination of an intermediate group, such as PD-MCI, within either a non-dementia group or a cognitively impaired group that includes dementia fails to specify the direct usefulness of a single measure to discriminate three classes simultaneously. The current work has shown that under the ordering restriction, one can determine the relevant $k-1$ Youden index cut-offs required for any given ordered k-class classification by subsequently examining the ordered pairwise two-class cut-off points. Such an approach is thus helpful also for the broader context of many disease states when the identification of more than two classes is of practical interest. From the clinical perspective, it is therefore pertinent to know that pairwise two-class comparisons when there are three (or k) relevant ordered comparisons is an appropriate strategy. For example, specific neurobehavioural tests or biomarkers could be used to address $k = 3$ degrees of impairment within a dementia condition (mild, moderate, or severe), including PD-D or other dementias such as Alzheimer's disease or dementia with Lewy Bodies. Similarly, the identification and characterization of multiple states, such as comparison between healthy nonrisk participants, those at risk with prodromal markers or showing different MCI states, and those with existing dementia, are now a more tractable problem.

Acknowledgements

The authors would like to thank three anonymous reviewers whose comments greatly improved the final output.

References

1. Caviness JN, Lue L, Adler CH, Walker DG. Parkinson's disease dementia and potential therapeutic strategies. *CNS Neuroscience & Therapeutics* 2010; **17**:32–44. DOI: 10.1111/j.1755-5949.2010.00216.x.
2. Dalrymple-Alford JC, Livingston L, MacAskill MR, Graham CF, Melzer TR, Porter RJ, Watts R, Anderson TJ. Characterizing mild cognitive impairment in Parkinson's disease. *Movement Disorders* 2011; **26**:629–636. DOI: 10.1002/mds.23592.
3. Aarsland D, Larsen JP, Tandberg E, Laake K. Predictors of nursing home placement in Parkinson's disease: a population-based, prospective study. *Journal of the American Geriatrics Society* 2000; **48**:938–942.
4. Martinez-Martin P, Benito-León J, Alonso F, Catalán MJ, Pondal M, Zamarbide I, Tobias A, de Pedro J. Quality of life of caregivers in Parkinson's disease. *Quality of Life Research* 2005; **14**:463–472.
5. Aarsland D, Kurz MW. The epidemiology of dementia associated with Parkinson's disease. *Brain Pathology* 2010; **20**:633–639. DOI: 10.1111/j.1750-3639.2009.00369.x.
6. Chou KL, Amick MM, Brandt J, Camicioli R, Frei K, Gitelman D, Goldman J, Growdon J, Hurtig HI, Levin B, Litvan I, Marsh L, Simuni T, Tröster AI, Uc EY. A recommended scale for cognitive screening in clinical trials of parkinsons disease. *Movement Disorders* 2010; **25**:2501–2507. DOI: 10.1002/mds.23362.
7. Dalrymple-Alford JC, MacAskill MR, Nakas CT, Livingston L, Graham C, Crucian GP, Melzer TR, Kirwan J, Keenan R, Wells S, Porter RJ, Watts R, Anderson TJ. The MoCA: well suited screen for cognitive impairment in Parkinson disease. *Neurology* 2010; **75**:1717–1725.
8. Hoops S, Nazem S, Siderowf AD, Duda JE, Xie SX, Stern MB, Weintraub D. Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology* 2009; **73**:1738–1745.
9. Nakas CT, Yiannoutsos CT. Ordered multiple-class ROC analysis with continuous measurements. *Statistics in Medicine* 2004; **23**:3437–3449. DOI: 10.1002/sim.1917.
10. Marinus J, Visser M, Verwey NA, Verhey FRJ, Middelkoop HAM, Stiggelbout AM, van Hilten JJ. Assessment of cognition in Parkinson's disease. *Neurology* 2003; **61**:1222–1228.
11. Nakas CT, Alonzo TA, Yiannoutsos CT. Accuracy and cut-off point selection in three-class classification problems using a generalization of the Youden index. *Statistics in Medicine* 2010; **29**:2946–2955. DOI: 10.1002/sim.4044.
12. Pepe MS. *The Statistical Evaluation of Medical Diagnostic Tests for Classification and Prediction*. Oxford University Press: Oxford, 2003.
13. Fluss R, Faraggi D, Reiser B. Estimation of the Youden index and its associated cutoff point. *Biometrical Journal* 2005; **47**:458–472. DOI: 10.1002/bimj.200410135.
14. Gail MH, Green SB. A generalization of the one-sided two-sample Kolmogorov-Smirnov statistic for evaluating diagnostic tests. *Biometrics* 1976; **32**:561–570.
15. Krzanowski WJ, Hand DJ. *ROC Curves for Continuous Data*. CRC Press: Boca Raton, 2009.
16. Dreiseitl S, Ohno-Machado L, Binder M. Comparing three-class diagnostic tests by three-way ROC analysis. *Medical Decision Making* 2000; **20**:323–331. DOI: 10.1177/0272989X0002000309.
17. Mossman D. Three-way ROCs. *Medical Decision Making* 1999; **19**:78–89. DOI: 10.1177/0272989X9901900110.
18. Li J, Fine JP. ROC analysis with multiple tests and multiple classes: methodology and its application in microarray studies. *Biostatistics* 2008; **9**:566–576. DOI: 10.1093/biostatistics/kxm050.
19. Xiong C, van Belle G, Miller JP, Morris JC. Measuring and estimating diagnostic accuracy when there are three ordinal diagnostic groups. *Statistics in Medicine* 2006; **25**:1251–1273. DOI: 10.1002/sim.2433.
20. Emre M, Aarsland D, Brown R, Burn DJ, Duyckaerts C, Mizuno Y, Broe GA, Cummings J, Dickson DW, Gauthier S, Goldman J, Goetz C, Korczyn A, Lees A, Levy R, Litvan I, McKeith I, Olanow W, Poewe W, Quinn N, Sampaio C, Tolosa E, Dubois B. Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Movement Disorders* 2007; **22**:1689–1707. DOI: 10.1002/mds.21507.
21. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Statistics in Medicine* 2006; **25**:4279–4292. DOI: 10.1002/sim.2673.
22. Davison AC, Hinkley DV. *Bootstrap Methods and Their Application*. Cambridge University Press: Cambridge, 1997.
23. Sivapulle MJ, Sen PK. *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*. Wiley: New Jersey, 2005.
24. Zhang J, Wu Y. k-sample tests based on the likelihood ratio. *Computational Statistics and Data Analysis* 2007; **51**:4682–4691. DOI: 10.1016/j.csda.2006.08.029.
25. Conover WJ. Several k-sample Kolmogorov-Smirnov tests. *Annals of Mathematical Statistics* 1965; **36**:1019–1026.
26. David HT. A three-sample Kolmogorov-Smirnov test. *Annals of Mathematical Statistics* 1958; **29**:842–851.
27. Duda RO, Hart PE, Stork DG. *Pattern Classification, 2nd Edition*. Wiley: New York, 2001.

