Published in final edited form as:

Acad Radiol. 2013 July ; 20(7): 874–882. doi:10.1016/j.acra.2013.03.009.

ROC Analysis in Biomarker Combination with Covariate Adjustment

Danping Liu, PhD and Xiao-Hua Zhou, PhD

Biostatistics and Bioinformatics Branch, Division of Epidemiology, Statistics & Prevention Research, NIH/NICHD, Bethesda, MD 20892 (D.L.); the HSR&D Center for Excellence, VA Puget Sound Health Care System, Seattle, WA 98101 (X.-H.Z.); the Department of Biostatistics, University of Washington, Seattle, WA 98195 (X.-H.Z.).

Abstract

Rational and Objectives—Receiver operating characteristic analysis (ROC) is often used to find the optimal combination of biomarkers. When the subject level covariates affect the magnitude and/or accuracy of the biomarkers, the combination rule should take into account of the covariate adjustment. The authors propose two new biomarker combination methods that make use of the covariate information.

Materials and Methods—The first method is to maximize the area under covariate-adjusted ROC curve (AAUC). To overcome the limitations of the AAUC measure, the authors further proposed the area under covariate standardized ROC curve (SAUC), which is an extension of the covariate-specific ROC curve. With a series of simulation studies, the proposed optimal AAUC and SAUC methods are compared with the optimal AUC method that ignores the covariates. The biomarker combination methods are illustrated by an example from Alzheimer's disease research.

Results—The simulation results indicate that the optimal AAUC combination performs well in the current study population. The optimal SAUC method is flexible to choose any reference populations, and allows the results to be generalized to different populations.

Conclusions—The proposed optimal AAUC and SAUC approaches successfully address the covariate adjustment problem in estimating the optimal marker combination. The optimal SAUC method is preferred for practical use, because the biomarker combination rule can be easily evaluated for different population of interest.

Keywords

Biomarker combination; covariate adjustment; AUC; covariate standardization

In disease screening, a common practice is to obtain various biomarkers from each subject. These biomarkers often reflect diverse aspects of the disease, such as the severity, subtype, etc. Any one of the markers alone may not be a great diagnostic tool, while combining the markers could potentially help achieve better accuracy. The receiver operating characteristic (ROC) curve is a useful tool to evaluate the diagnostic accuracy (1). Many biomarker combination methods rely on the maximization of the area under ROC curve (AUC). The major advantage of optimal AUC combination is that no distributional assumption is needed for the biomarkers (2). In addition, it is flexible to incorporate linear (2, 3) or nonlinear (4-7) combinations.

In practice, individual covariates may be associated with the magnitude and/or accuracy of a biomarker. Covariate adjustment is generally recommended in ROC analysis as long as the decision threshold in diagnosis is dependent on covariates (8, 9). The unadjusted ROC curve pools all the heterogeneous subjects together, and makes an implicit assumption that the decision threshold for test positivity is constant for everybody. This assumption could be violated in practice. For example, patient's age often plays an important role in the disease diagnosis. The irrational decision strategy could lead to a non-concave ROC curve (10), and the comparison of two subjects with different covariates may not have good scientific meanings. The pooling may also result in the Simpson's paradox (11). The idea of covariate adjustment in ROC analysis resembles that in linear regression: the effect of interest is only evaluated with the confounders held fixed. The covariate-specific ROC curve is frequently reported in the diagnostic studies, which examines the accuracy of a marker within a subpopulation stratified by the covariates.

Although covariate adjustment is widely accepted in evaluating a single marker, not much has been done for the biomarker combination with covariate adjustment. Since the unadjusted AUC has its own limitations, the optimal AUC combination could be misleading for practical use. The major aim of this paper is to investigate new biomarker combination methods when individual covariates are available. One option is to maximize the covariate-adjusted AUC (AAUC) measure. Recently proposed by Janes and Pepe (12), AAUC summarizes the “overall” marker accuracy using a weighted average of the covariate-specific ROC curves. Despite its ability for covariate adjustment, AAUC is hard to generalize to different populations, because the calculation of the weights relies on the covariate distribution. It is an undesirable property that the combination rule can only be applied to a population with particular covariate constitutions. Therefore, we propose another overall accuracy measure called “standardized AUC” (SAUC). SAUC can adjust for covariates, while remains generalizable to other populations. We standardize the marker distributions for the cases and controls to a reference covariate population by exploiting the relationship of the marker value and covariates. SAUC is shown to be a generalized form of the covariate-specific AUC (AUC_x).

In this paper, we examine and compare the performance of optimal AUC, AAUC, and SAUC combination rules through simulation studies and an example from the Alzheimer's disease research. We also provide recommendations for the use of the three methods.

MATERIALS AND METHODS

Notations

Let D_i denote the binary disease status with 1 being diseased (case) and 0 being healthy (control). Let T_i^k denote the k th biomarker value for the i th subject with $k = 1, \dots, K$ and $i = 1, \dots, N$. Without loss of generality, we assume that subjects with larger markers are more likely to be diseased. Let X_i be the subject level covariates. We may suppress the scripts i and k sometimes when there is no confusion.

We write F_X and G_X to be the cumulative distribution function (CDF) of the biomarker given covariates for the cases and controls respectively, i.e.,

$$\begin{aligned} F_X(t) &= \Pr(T \leq t | D=1, X), \\ G_X(t) &= \Pr(T \leq t | D=0, X). \end{aligned}$$

Let H_d denote the CDF of covariates given the disease status, $X/D = d$ for $d = 1, 2$. Let $F(t) = \int F_x(t) dH_1(x)$ and $G(t) = \int G_x(t) dH_0(x)$ denote the marginal distribution of the biomarker for the cases and controls, respectively.

Description of the NACC data

The National Alzheimer's Coordinating Center (NACC) Uniform Data Set (UDS) comes from a multi-center longitudinal study, which follows the study participants from 34 Alzheimer's disease centers throughout North America since September 2005. The subjects received clinical assessment and neuropsychological tests approximately once each year, and all these information are recorded in standardized forms. The severity of cognitive impairment can often be divided into three categories: normal cognition, mild cognitive impairment (MCI), and dementia. Subjects with MCI could reverse back to normal, but dementia is generally believed to be irreversible. Our primary interest is to predict the progression from MCI to dementia using the baseline markers. We include the individuals who were diagnosed as MCI at their enrollment and made at least one follow-up visit. Among the 3746 subjects in our analysis sample, 672 (17.9%) progressed to dementia ($D = 1$) in the second visit. Our prediction makes use of two baseline markers: clinical dementia rating sum of boxes (CDRSB) and mini-mental state examination (MMSE). The CDRSB (T^1) is the physician's judgment on the severity of a subject's cognitive impairment in six domains: memory, orientation, judgment, community affairs, home and hobby, and personal care. Each domain is scored from 0 to 3 and the summation is the CDRSB. The MMSE (T^2) is a 30-point questionnaire test which is widely used to screen for cognitive impairment. The AUCs for CDRSB and MMSE are 0.728 and 0.686, respectively, suggesting room for improvement. Thus we would like to investigate ways to combine the two markers, which could potentially improve the classification accuracy. We included several additional individual covariates (X) that may affect the magnitude or accuracy of both markers, and these variables are summarized in Table 1. As found in previous studies, many of these covariates are associated with the marker values (13). So it is necessary to incorporate the covariates when coming up with the combination rule.

Existing AUC Measures

We list several existing AUC measures and their formula in Table 2. The traditional unadjusted AUC ignores all the covariate information and only compares the marginal distributions of the marker for the cases and controls. It is interpreted as the probability that a randomly selected case has a larger marker value than a randomly selected control, regardless of the covariates.

The covariate-specific AUC (AUC_x) is also commonly seen in the literature, which is interpreted as the probability that a case has a larger biomarker than a control, given they share the same covariates vector x . One can relax this condition and allow the comparison of any case-control pair (14), namely, the extended AUC_x. It is usually interpreted with $x = y$ because $AUC_{xx} = AUC_x$. We show in the Appendix that the unadjusted AUC can also be written as an "average" of AUC_{xy} .

Recently, Janes and Pepe (12) proposed a new covariate-adjusted AUC (AAUC) measure, which is a weighted average of AUC_x. The weight is taken with respect to the covariate distribution for the cases. The AAUC is also a weighted average of AUC_{xy} , but the covariate values for the case-control pairs are always forced to be equal. AAUC provides a convenient summary of a marker's overall diagnostic accuracy, while adjusting for the covariates in the meantime. However, there are two disadvantages of this measure. Firstly, the AAUC lacks the probability interpretation, so we cannot write it in the form of $P(T_1 > T_2 | D_1 = 1, D_2 = 0)$. Thus the case and control populations under comparison are not defined explicitly.

Secondly, AAUC relies on the covariate distribution $H_1(x)$ in the current population. If we evaluate the same marker in two different populations, AAUC would be different because of different covariate constitutions, which limits the generalizability of AAUC. AAUC can be estimated with a semiparametric method by considering a covariate-specific placement value (12).

Standardized AUC (SAUC)

The key to covariate adjustment in the ROC analysis is making comparisons of cases and controls that are similar in the covariate values. Here we borrow the idea of standardized mortality ratio in epidemiology and try to standardize the biomarker distributions to a reference population. The pseudo case and pseudo control distributions are defined as

$$\begin{aligned} F^*(t) &= \int F_x(t) dH^*(x) \\ G^*(t) &= \int G_x(t) dH^*(x), \end{aligned}$$

where H^* represents the covariates distribution for the reference population. Then SAUC is expressed similarly as the unadjusted AUC, with F and G replaced by F^* and G^* :

$$SAUC = \int G^*(t) dF^*(t) = \iint AUC_{xy} dH^*(x) dH^*(y).$$

Again, we write SAUC to be a weighted average of AUC_{xy} as well.

SAUC is in fact a generalized form of AUCx. If we take H^* to be a point mass at x then SAUC reduces to AUCx. Note that AUCx compares the case and control populations conditional on the covariates, while SAUC compares the two populations with the same covariate constitutions. In a sense, SAUC examines a marginal population-averaged effect. The difference of the marker distribution is only attributed to the disease status, and the standardization removes the confounders. If H^* is taken to be H_1 , then the definitions of SAUC and AAUC look quite similar, except that SAUC do not force the integration along the line of $x = y$. More importantly, SAUC overcomes the two disadvantages of AAUC. It is interpreted as the probability that a case has a greater marker value than a control in the reference population. The arbitrary specification of H^* allows us to generalize the conclusion from one population to another. For example, if two markers are studied in different populations, we could compare their SAUC measures by choosing a common fixed reference population of interest.

To study how the reference population affects the SAUC, we consider two simple cases with a binary covariate X , gender, for example. Assume binormal ROC curves for both male and female. In case A, gender only affects the magnitude of the marker, but not the accuracy of the marker; that is, AUCx is the same for male and female. In case B, gender affects the diagnostic accuracy of the marker, and the male has a higher AUCx than the female. We then choose certain gender decomposition as the reference population and calculate the SAUC. Figure 1 plots the SAUC versus $\Pr(X = 1)$ in the reference population for cases A and B. We observe the quadratic shape in both cases, and that the SAUC is always smaller than the best AUCx. In general, with continuous and/or multiple covariates, we expect the SAUC to get larger if (a) the reference population consists of subjects with higher AUCx, or (b) the reference population is more homogeneous.

The estimation of SAUC follows from estimating the standardized marker distribution F^* and G^* . We propose a semiparametric estimation for SAUC and the detailed procedures are

given in the Appendix. The R codes for computing SAUC and AAUC is not provided along with the paper, but are available upon request.

Biomarker Combination

In this paper we only focus on the linear combination of biomarkers. When no covariate information is available, maximizing the unadjusted AUC is a convenient approach for biomarker combination. It makes no assumptions on the marker distribution, and is efficient when the markers are multivariate normal (2). When the covariate adjustment becomes necessary, maximizing AUC may no longer give optimal combination rules. We propose to maximize the AAUC and SAUC instead. In the case of two markers, the combination rule is given by $\operatorname{argmax}_{\lambda}\{AAUC \text{ of } T^1 + \lambda T^2\}$ or $\operatorname{argmax}_{\lambda}\{SAUC \text{ of } T^1 + \lambda T^2\}$. We could use bootstrap to estimate the 95% confidence interval of $\hat{\lambda}$.

Maximizing AAUC can be regarded as a necessary condition of maximizing every AUCx. In other words, the combined marker will, on average, achieve good accuracy for every single covariate value. In a special case when the markers follow a bivariate normal distribution, and AUCx does not vary with the covariates, we derive the explicit theoretical solution of λ in the Appendix. In general there is no explicit solution for λ , and we need to use numerical methods to maximize AAUC. When the combined marker is employed to make diagnostics, clinicians need to interpret the results using covariate-specific decision threshold.

The optimal SAUC combination is flexible depending on the reference population used for standardization. We should always choose a target population as the reference to which the combination rule is going to be applied. Although the selection of the reference population is quite flexible, there are several issues to keep in mind. Firstly, we should not choose a reference population that is extremely different from the study population. The estimation of SAUC with such reference population requires extrapolation of $F_X(t)$ to new covariate values, which could be rather crude or even meaningless. Secondly, it is helpful to select multiple reference populations in practice, and examine how they affect the combination rule and the SAUC. This could often serve as a sensitivity analysis.

Both AAUC and SAUC incorporate the covariates, but they still have some subtle distinctions in the way of using covariates. AAUC only makes use of the covariates in the current study sample. The covariates for controls are used to model the relationship between T and X , and the covariates for the cases are used to compute the placement value (12). On the other hand, SAUC uses both the covariates in the study sample as well as in the reference population. The covariates in the study sample are all used to model $T|X, D = d$. The covariates in the reference population are used to standardize the marker value.

Simulation Studies

We perform a series of simulation studies to compare the performance of optimal AUC, AAUC and SAUC combination rules under different settings. We consider the combination of two markers with two covariates. In the training sample, the covariates X_1 and X_2 are generated independently from *Bernoulli*(0.4) and *N*(0,1) distributions. The disease status D is generated from a logistic regression:

$$\operatorname{logit}\{P(D=1|X)\} = -1.5 + 0.4X_1 + 0.5X_2.$$

The two markers are generated from multivariate normal distribution given D and X :

$$T^k = \beta_0^k + \beta_1^k D + \beta_2^k X_1 + \beta_3^k X_2 + \beta_4^k D \times X_1 + \beta_5^k D \times X_2 + \epsilon^k$$

for $k = 1, 2$, where

$$\begin{pmatrix} \epsilon^1 \\ \epsilon^2 \end{pmatrix} \sim MVN \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 2 \end{pmatrix} \right].$$

We fix $(\beta_0^1, \dots, \beta_5^1) = (-0.2, 0.6, 0, 0.5, 0.6, 0.4)$ and $(\beta_0^2, \dots, \beta_5^2) = (-0.1, 1, 1.5, 0.3)$, but vary $(\beta_4^2, \beta_5^2) = (0.6, 0.4)$ or $(-0.6, -0.4)$, which we refer to as scenarios A and B, respectively. The positive (negative) β_4^2 and β_5^2 indicate that the accuracy (AUCx) of T^2 is increasing (decreasing) with the values of X_1 and X_2 . We vary the covariance ρ to be -0.8, 0.3, 0, 0.3, or 0.8.

We generate two test samples from different populations. Population 1 has the same covariate constitution as the training sample. Population 2 generates X_1 and X_2 independently from *Bernoulli*(0.8) and $N(1.5, 1.2^2)$ distributions. The disease prevalence is about 22% for population 1 and about 40% for population 2. Four methods are used for the biomarker combination:

- M1: Optimal AUC;
- M2: Optimal AAUC;
- M3: Optimal SAUC with covariates standardized to population 1;
- M4: Optimal SAUC with covariates standardized to population 2.

The estimated combination rules are validated in samples from populations 1 and 2, in terms of AUC, AAUC and SAUC. The size of the training sample, validation sample 1, and validation sample 2 are all fixed to be 200. The simulations are repeated for 500 times.

RESULTS

Simulation Results

Tables 3 and 4 show the simulation results. The combination rules from M1-M4 are quite different in all settings, and the difference tends to be larger when ρ increases. We reported the median and interquartile range (IQR) instead of the mean and empirical standard deviation. The reason is that occasionally M1 and M2 may result in quite large estimated coefficients, probably due to the small sample size. The IQR for M1 and M2 are generally comparable. M3 yields slightly smaller IQR than M1 especially for larger ρ . Interestingly, the IQR for M4 is always quite small. This is probably because the case/control ratio for population 2 is more balanced than population 1.

The performance of the combination rules is evaluated by different AUC measures. In population 1, from which the training sample is drawn, it is not strange that M1 is generally the best in terms of AUC; M2 is generally the best in terms of AAUC; M3 is the best in terms of SAUC. This confirms the out-sample optimality of maximizing the respective AUC measures. However, when evaluated in a quite different population, M4 almost always the best or close to the best in terms of AUC, AAUC and SAUC. Therefore, when diagnosis is to be made in a new population other than the study population, it is important to take into account the information of reference population, so that the good diagnostic accuracy can be

achieved. We also observe that the advantage of M4 is more prominent when ρ increases or when $(\beta_4^2, \beta_5^2) = (-0.6, -0.4)$ (Scenario A).

We also performed additional simulation studies where both covariates are continuous that follows gamma distributions (results not presented). We found that the results are quite similar as in the current setting.

In summary, we have two major discoveries from the simulation results. First, the combination rule differs substantially according to different optimizers. This implies that failure to adjust for important covariates may result in a misleading combination of biomarkers. Second, in order that the combined marker can be applied to a new population, it is critical to estimate the combination rule that borrows information from the new reference population. Ignoring the reference population often yields suboptimal diagnostic accuracy, and this a more serious problem as ρ increases.

In practice, we would generally recommend the use of SAUC as the target function for biomarker combination, given its flexibility and generalizability. AAUC can only be used if the combined marker is applied to a similar population. AUC is only useful when covariates are not available or are known to be unrelated to the test results.

NACC data example

We applied the methods for biomarker combination to the NACC UDS data. The reference population can be chosen arbitrarily, according to the target population of interest. In this example, we generated a reference population with average age of 70 years (SD=5 years), balanced gender and years of education distributions, and time between two visits fixed to be 1 year. As higher MMSE is indicative of better cognitive function, we use 30-MMSE instead. The CDRSB and 30-MMSE are both standardized with mean=0 and SD=1, yielding a Z score. The coefficient of MMSE is found to be 0.544 (95% CI: 0.436, 1.115) for the optimal AUC combination, 0.600 (95% CI: 0.415, 0.814) for the optimal AAUC combination, and 0.458 (95% CI: 0.288, 0.681) for the optimal SAUC combination. The three optimal combinations do not differ too much in magnitude. But the AAUC and SAUC combinations have narrower confidence intervals than the AUC combination, because the latter does not make use of any covariate information. Table 5 lists the AUC, AAUC and SAUC measures for each single marker and the combined marker. We can see that the combined marker does seem to improve the diagnostic accuracy.

We also investigate whether the optimal combination rule should vary by age. We fix the covariates gender, education and year, but allow age to vary from 65 to 90. The SAUC is maximized using every point of covariate values as the reference, which is equivalent to maximizing the AUCx. The coefficients of MMSE and their associated 95% CIs are shown in Figure 2. Although there is a very mild decreasing trend for the coefficient of MMSE, it is fair to conclude that the combination rule does not seem to change with age. As for the maximized SAUC (or AUCx in this case), the decreasing trend is obvious (Figure 3). The SAUC can be as high as over 0.8 for the younger age, and gets below 0.75 for the older age. The optimality of the SAUC is with respect to the reference population, so for the older population, the best achievable SAUC for the combined marker is around 0.75. This implies that the practical use of the combined marker in predicting the progression to dementia is more favorable towards younger populations. A possible explanation is that the cognitive decline due to natural aging poses a complication in predicting dementia.

DISCUSSION

In this paper, we proposed a new SAUC measure that allows covariate adjustment while evaluating the overall accuracy of a marker. We then propose to use SAUC and AAUC as the target function in searching for the optimal marker combination. Compared to optimal AUC combination, the optimal AAUC and SAUC methods could make use of the subject level covariate information. The optimal AAUC combination only relies on the covariate distribution in the study sample, so the combination rule cannot be easily extended to other populations. The optimal SAUC combination, on the other hand, has the advantage of flexibility and generalizability. The combination rule can be derived with respect to a reference population, to which we would apply the combined marker in making diagnoses. When covariate information is available, we would always recommend the use of optimal AAUC or SAUC combination. In the situation that the combination rule needs to be applied to other populations, we would recommend the use of the optimal SAUC method.

AUC_x is a special case of SAUC: the former examines case and control pairs with a fixed value of covariates, and the latter looks at case and control pairs drawn from the same population of covariates. In a sense, SAUC is a measure of the population-averaged accuracy. In the context of the marker combination, it is often meaningful to determine whether the optimal combination rule is covariate-dependent. SAUC combination allows us to examine this with a sensitivity analysis: we could choose multiple reference populations and maximize the SAUC. Then the combination rules can be compared across the reference populations.

This article focuses on developing the combination of biomarkers with covariate adjustment, while using the combination to make diagnosis is of much practical interest to the clinicians. A natural question is to find out the optimal decision threshold. Without covariates, the threshold can be estimated by minimizing a user-defined cost function based on the Bayesian decision theory (8). For the combined marker with covariate adjustment, we could employ the similar techniques, but now the decision threshold might be covariate-specific. This motivates our future investigation in the clinical decision making that considers the patient characteristics. We suggest that in practice, simple patient covariates such as demographic information should be collected together with the biomarker values, in order to help with the disease diagnoses.

In this paper, we estimate SAUC and AAUC indirectly, that is, through modeling the marker distribution. Direct estimation might be an alternative approach: we could first estimate AUC_{xy} using a binary regression model, and then take an average to get SAUC or AAUC. As we noted in our previous work (13), the direct estimation approach is often more computationally intensive (with the order of N^2), because the binary regression is constructed using U statistics.

In radiological studies, a “biomarker” or a test is typically a categorical rating of an image by a radiologist. We may wish to combine this rating with other tests, such as computer-aided diagnosis (4), or ratings from multiple images (7). The principle of covariate adjustment for estimating the combination still applies. The scale of the discrete tests may not be linearly corresponding to the disease severity, so it would be interesting to explore nonlinear combination by including quadratic or interaction terms of the tests as well. These parameters can be easily estimated with reasonably large sample size.

We only demonstrated the usefulness of SAUC in biomarker combination, but the use of SAUC can go beyond that. A good example is the comparison of multiple biomarkers when they are studied in different populations. Both AUC and AAUC would be biased in this case, as the diagnostic accuracy is likely to be affected by the covariates compositions.

Acknowledgments

This work was supported in part by NIH/NIA grant U01AG016976.

D.L.'s research is supported by the Intramural Research Program of the National Institute of Health (NIH), Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD).

APPENDIX

Relationship Between AUC and AUC_{xy}

Note that $AUC = \int G(t) dF(t)$, $F(t) = \int F_x(t) dH_1(x)$ and $G(t) = \int G_y(t) dH_0(y)$. So we have

$$\begin{aligned} AUC &= \int \left\{ \int G_y(t) dH_0(y) \right\} d \left\{ \int F_x(t) dH_1(x) \right\} \\ &= \int \int \int G_y(t) dF_x(t) dH_1(x) dH_0(y) \\ &= \int AUC_{xy} dH_1(x) dH_0(y). \end{aligned}$$

Semiparametric Estimation of SAUC

The first step is to estimate a location model (15, 16) for $T/D = d, X$:

$$T = \mu(X, d) + \epsilon,$$

where $\mu(X, d) = E(T/D = d, X)$ and ϵ is the residual whose distribution is left unspecified. A convenient choice of the mean function is a linear model stratified by disease, e.g., $\mu(X, d) = X' \beta_d$. The second step is to estimate the residual distribution empirically:

$$\widehat{R}_d(s) = \frac{\sum_{j=1}^N I \left[T_j - \widehat{\mu}(X_j, d) \leq s, D_j = d \right]}{\sum_{j=1}^n I \left[D_j = d \right]}.$$

It follows immediately that

$$\widehat{G}_x(t) = \widehat{R}_1(t - \widehat{\mu}(x, 0)) = \frac{\sum_{j=1}^N I \left[T_j - \widehat{\mu}(X_j, d) \leq t - \widehat{\mu}(x, 0), D_j = d \right]}{\sum_{j=1}^n I \left[D_j = d \right]}.$$

Since $G^*(t)$ is an integral of $G_x(t)$ with respect to $H^*(x)$, the covariate distribution in the reference population, we need to have a random sample from $H^*(x)$ to evaluate the integral, which is often available in practice. We write this sample as $X_1^*, \dots, X_M^*$. Now we have

$$\widehat{G}^*(t) = \frac{\sum_{m=1}^M \sum_{j=1}^N I \left[T_j - \widehat{\mu}(X_j, d) \leq t - \widehat{\mu}(X_m^*, 0), D_j = d \right]}{M \times \sum_{j=1}^N I \left[D_j = d \right]}.$$

Note that $\widehat{G}^*(t)$ is merely a step function with at most $M \times \sum_{j=1}^N I \left[D_j = d \right]$ jumps. Similarly, $\widehat{F}^*(t)$ could also be estimated empirically. The last step is to form the SAUC estimate, which

depends on the rankings of the jump points in $\hat{G}^*(t)$ and $\hat{F}^*(t)$. This would be exactly the same as we compute the Mann-Whitney statistic as the unadjusted AUC.

A Special Case of Maximizing AAUC

Suppose that the markers T^1 and T^2 follow multivariate normal distribution. Their mean and variance matrix are

$$\mu_T = \begin{pmatrix} X' \alpha_1 + \gamma_1 D \\ X' \alpha_2 + \gamma_2 D \end{pmatrix}, \Sigma_T = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Here μ_T does not include any interaction terms of X and D , indicating that the covariates only affect the magnitude of the marker, but not the accuracy of the marker. Consider the linear combination $T^1 + \lambda T^2$, which follows a univariate normal distribution with

$$\begin{aligned} \mu_c &= X' (\alpha_1 + \lambda \alpha_2) + (\gamma_1 + \lambda \gamma_2) D, \\ \sigma_c^2 &= \sigma_1^2 + 2\rho \sigma_1 \sigma_2 \lambda + \sigma_2^2 \lambda^2. \end{aligned}$$

The AUC_x of $T^1 + \lambda T^2$ is common for all the covariates, which is

$$AAUC = AUC_x = \Phi \left(\frac{\gamma_1 + \lambda \gamma_2}{\sqrt{2\sigma_1^2 + 4\rho \sigma_1 \sigma_2 \lambda + 2\sigma_2^2 \lambda^2}} \right).$$

Maximizing AAUC is then equivalent to maximizing AUC_x . According to Su and Liu (3), the linear discriminant function maximize the AUC_x ; that is,

$$(1, \lambda) \propto (\gamma_1, \gamma_2) \Sigma_T^{-1}.$$

After some simple algebra, we can show that $\lambda = \frac{\gamma_2 \sigma_1^2 - \gamma_1 \rho \sigma_1 \sigma_2}{\gamma_1 \sigma_2^2 - \gamma_2 \rho \sigma_1 \sigma_2}$.

We should note that this special case no longer holds if μ_T includes the interaction between X and D . When the covariates affect the diagnostic accuracy, AAUC is a weighted average of AUC_x , and no close form solution for λ is generally available.

Reference List

1. Metz CE. Basic principles of ROC analysis. Seminars in Nuclear Medicine. 1978; 8(4):283–98. [PubMed: 112681]
2. Pepe MS, Thompson ML. Combining diagnostic test results to increase accuracy. Biostatistics. 2000; 1(2):123–40. [PubMed: 12933515]
3. Su JQ, Liu JS. Linear Combinations of Multiple Diagnostic Markers. Journal of the American Statistical Association. 1993; 88(424):1350–5.
4. Jiang Y, Metz CE. A Quadratic Model for Combining Quantitative Diagnostic Assessments from Radiologist and Computer in Computer-Aided Diagnosis. Academic Radiology. 2006; 13(2):140–51. [PubMed: 16428048]
5. Liu B, Metz CE, Jiang Y. Effect of correlation on combining diagnostic information from two images of the same patient. Med Phys. 2005; 32(11):3329–38. [PubMed: 16372412]

6. Liu C, Liu A, Halabi S. A min–max combination of biomarkers to improve diagnostic accuracy. *Statist Med.* 2011; 30(16):2005–14.
7. Zou KH, Bhagwat JG, Carrino JA. Statistical Combination Schemes of Repeated Diagnostic Test Data. *Academic Radiology.* 2006; 13(5):566–72. [PubMed: 16627197]
8. Pepe, MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction.* Oxford University Press; Oxford: 2003.
9. Zhou, XH.; Obuchowski, NA.; McClish, DK. *Statistical Methods in Diagnostic Medicine.* Wiley; New York: 2011.
10. Pesce LL, Metz CE, Berbaum KS. On the Convexity of ROC Curves Estimated from Radiological Test Results. *Academic Radiology.* 2010; 17(8):960–8. [PubMed: 20599155]
11. Janes H, Pepe MS. Adjusting for Covariates in Studies of Diagnostic, Screening, or Prognostic Markers: An Old Concept in a New Setting. *American Journal of Epidemiology.* 2008; 168(1):89–97. [PubMed: 18477651]
12. Janes H, Pepe MS. Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve. *Biometrika.* 2009; 96(2):371–82. [PubMed: 22822245]
13. Liu D, Zhou XH. Covariate adjustment in estimating the area under ROC curve with partially missing gold standard. *Biometrics.* 2013
14. Dodd LE, Pepe MS. Semiparametric Regression for the Area under the Receiver Operating Characteristic Curve. *Journal of the American Statistical Association.* 2003; 98(462):409–17.
15. Pepe MS. Three Approaches to Regression Analysis of Receiver Operating Characteristic Curves for Continuous Test Results. *Biometrics.* 1998; 54(1):124–35. [PubMed: 9544511]
16. Liu D, Zhou XH. Semiparametric Estimation of the Covariate-Specific ROC Curve in Presence of Ignorable Verification Bias. *Biometrics.* 2011; 67(3):906–16. [PubMed: 21361890]

Figure 1.

The plot of SAUC versus the mixture probability of a binary covariate in the reference population. (a) same AUC_x for $X=1$ and $X=0$; (b) AUC_x is larger for $X=1$.

Figure 2.

The coefficient of MMSE versus age. (The vertical lines are the 95% CI. The dashed lines represent women with 13-16 years of education; the time lag between the first two visits is fixed to be 15 months. The solid lines represent men with 13-16 years of education; the time lag between the first two visits is fixed to be 12 months.)

Figure 3.

The maximized SAUC versus age. (The vertical lines are the 95% CI. The dashed lines represent women with 13-16 years of education; the time lag between the first two visits is fixed to be 15 months. The solid lines represent men with 13-16 years of education; the time lag between the first two visits is fixed to be 12 months.)

Table 1

Summary statistics of the study subjects in their initial visit. Means $\pm$ standard deviations (SD) and ranges are reported for the continuous variables, and percentages are reported for the categorical variables.

	Not Progressed to Dementia ($n_0 = 3074$)	Progressed to Dementia ($n_1 = 672$)
CDRSB (Range: 0-18)	1.1 $\pm$ 1.0	2.0 $\pm$ 1.4
MMSE (Range: 0-30)	27.6 $\pm$ 2.3	26.0 $\pm$ 2.6
Age (Range: 28.2-104.4)	73.9 $\pm$ 9.3	76.1 $\pm$ 9.5
Sex (Male)	47.1%	50.6%
Education 12 years	29.7%	28.9%
13-16 years	40.8%	42.3%
17 years	29.5%	28.9%
Year^I	1.21 $\pm$ 0.40	1.25 $\pm$ 0.49

^IThe time lag between the first and second visit.

Table 2

Existing AUC measures and their mathematical formula.

AUC Measures	Formula
Unadjusted AUC	$AUC = \int G(t) dF(t) = \iint AUC_{xy} dH_1(x) dH_0(y)$
Covariate-specific AUC (AUC _x)	$AUC_x = \int G_x(t) dF_x(t)$
Extended AUC _x	$AUC_{xy} = \int G_x(t) dF_y(t)$
Covariate-adjusted AUC (AAUC)	$AAUC = \int AUC_x dH_1(x) = \iint_{x=y} AUC_{xy} dH_1(x) dH_1(y)$

Estimated combination rules by maximizing AUC (M1), AAUC (M2), SAUC with reference population 1 (M3), and SAUC with reference population 2 (M4), and their out-sample performance in populations 1 and 2. Scenario A: $(\beta_4^2, \beta_5^2) = (-0.6, -0.4)$.

Table 3

ρ	$COEF^2 \times 100$	Population 1				Population 2			
		$(IQR^3 \times 100)$							
		$AUC \times 100$	$AAUC \times 100$	$SAUC \times 100$		$AUC \times 100$	$AAUC \times 100$	$SAUC \times 100$	
-0.8	M1	55.4 (43.6)	82.5⁴	85.3	74.0	87.9	91.9	79.9	
	M2	62.1 (36.6)	82.4	85.6	73.8	86.7	91.0	78.0	
	M3	45.6 (47.3)	82.2	84.5	73.9	88.8	92.3	81.6	
	M4	6.2 (30.9)	76.1	76.0	69.7	90.6	91.2	87.7	
-0.3	M1	49.2 (49.9)	79.2	78.9	71.3	85.6	86.4	78.1	
	M2	49.7 (49.3)	79.2	78.9	71.3	85.8	86.8	78.5	
	M3	38.3 (50.5)	78.9	78.6	71.2	86.9	87.7	80.4	
	M4	-5.8 (31.6)	72.3	72.8	67.0	89.5	90.0	87.7	
0.0	M1	44.5 (55.3)	77.7	76.5	70.1	84.7	84.5	77.7	
	M2	37.8 (53.2)	77.6	76.6	70.1	85.9	85.9	79.6	
	M3	33.2 (53.3)	77.4	76.4	70.1	86.3	86.3	80.3	
	M4	-12.9 (31.8)	70.5	71.8	65.8	89.3	90.1	88.2	
0.3	M1	39.2 (57.3)	76.4	74.8	69.2	84.0	83.2	77.5	
	M2	27.1 (55.1)	76.1	74.9	69.1	86.1	85.9	80.8	
	M3	26.3 (53.5)	76.1	74.9	69.2	86.1	85.7	80.7	
	M4	-19.8 (30.6)	69.1	71.2	64.9	89.6	90.7	89.2	
0.8	M1	26.7 (56.3)	74.8	73.1	68.1	84.2	83.0	78.7	
	M2	6.6 (59.3)	73.5	73.5	67.7	88.0	87.9	84.7	
	M3	12.2 (51.7)	74.5	73.7	68.2	87.1	86.8	83.2	
	M4	-30.9 (28.1)	67.5	71.6	64.1	91.3	92.9	91.7	

²The average of estimated coefficients of T_2 .

³The interquartile range of the estimated coefficient values.

⁴The largest AUC/AAUC/SAUC of the combined marker across methods M1-M4 is in **bold**.

Estimated combination rules by maximizing AUC (M1), AAUC (M2), SAUC with reference population 1 (M3), and SAUC with reference population 2 (M4), and their out-sample performance in populations 1 and 2. Scenario B: $(\beta_4^2, \beta_5^2) = (0.6, 0.4)$.

Table 4

ρ		$COEF^5 \times 100$	Population 1			Population 2		
			$AUC \times 100$	$AAUC \times 100$	$SAUC \times 100$	$AUC \times 100$	$AAUC \times 100$	$SAUC \times 100$
			$(IQR^6 \times 100)$					
-0.8	M1	54.4 (51.9)	<u>82.3</u> ⁷	86.9	72.4	93.6	97.9	88.3
	M2	75.4 (50.6)	82.2	<u>88.0</u>	71.7	92.9	<u>98.2</u>	86.6
	M3	36.8 (47.3)	81.9	85.1	<u>72.5</u>	<u>94.0</u>	97.2	89.4
	M4	23.1 (32.5)	80.9	82.3	72.3	<u>94.0</u>	96.0	<u>90.3</u>
-0.3	M1	54.4 (60.8)	<u>80.2</u>	82.5	70.7	92.0	95.7	86.3
	M2	77.7 (65.5)	80.1	<u>83.1</u>	70.1	91.3	<u>95.9</u>	84.7
	M3	32.8 (50.9)	79.6	81.0	<u>70.8</u>	<u>92.5</u>	95.0	87.8
	M4	16.8 (33.7)	78.4	78.7	70.6	<u>92.5</u>	93.8	<u>88.99</u>
0.0	M1	56.8 (65.4)	<u>79.0</u>	80.4	69.7	91.1	94.3	85.2
	M2	78.9 (72.1)	78.9	<u>80.8</u>	69.1	90.4	<u>94.4</u>	83.6
	M3	29.7 (52.3)	78.4	79.1	<u>70.0</u>	<u>91.7</u>	93.8	87.0
	M4	12.6 (33.7)	77.0	77.0	69.8	91.6	92.5	<u>88.2</u>
0.3	M1	56.5 (73.7)	<u>77.9</u>	78.7	68.8	90.2	<u>93.0</u>	84.1
	M2	80.2 (78.0)	<u>77.9</u>	<u>79.0</u>	68.3	89.5	<u>93.0</u>	82.6
	M3	25.4 (53.5)	77.4	77.5	<u>69.3</u>	<u>91.0</u>	92.6	86.5
	M4	7.7 (33.2)	75.8	75.6	69.0	90.9	91.5	<u>87.8</u>
0.8	M1	58.8 (89.8)	<u>76.5</u>	76.5	67.7	88.8	<u>91.0</u>	82.6
	M2	91.5 (115.4)	76.4	<u>76.6</u>	67.2	88.0	90.8	81.0
	M3	16.1 (51.6)	75.7	75.5	<u>68.5</u>	<u>90.1</u>	<u>91.0</u>	86.1
	M4	-1.7 (30.7)	73.8	74.0	68.2	89.8	90.2	<u>87.6</u>

⁵The median of estimated coefficients of η_2 .

⁶The interquartile range of the estimated coefficient values.

⁷The largest AUC/AAUC/SAUC of the combined marker across methods M1-M4 is in **bold**.

Table 5

The AUC, AAUC and SAUC for CDRSB, MMSE and the combination.

	CDRSB	MMSE	Combined (95% CI)
Optimal AUC	0.728	0.686	0.758 (0.739, 0.778)
Optimal AAUC	0.733	0.680	0.757 (0.738, 0.778)
Optimal SAUC	0.764	0.686	0.781 (0.755, 0.808)