Published in final edited form as:

Stat Med. 2013 September 10; 32(20): 3449–3458. doi:10.1002/sim.5777.

On use of partial area under the ROC curve for evaluation of diagnostic performance

Hua Ma¹, Andriy I. Bandos^{1,*}, Howard E. Rockette¹, and David Gur²

¹Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, U.S.A

²Department of Radiology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15216, U.S.A

Abstract

Evaluation of diagnostic performance is a necessary component of new developments in many fields including medical diagnostics and decision making. The methodology for statistical analysis of diagnostic performance continues to develop, offering new analytical tools for conventional inferences and solutions for novel and increasingly more practically relevant questions.

In this paper we focus on the partial area under the Receiver Operating Characteristic (ROC) curve, or pAUC. This summary index is considered to be more practically relevant than the area under the entire ROC curve (AUC), but because of several perceived limitations, it is not used as often. In order to improve interpretation, results for pAUC analysis are frequently reported using a rescaled index such as the standardized partial AUC proposed by McClish (1989).

We derive two important properties of the relationship between the “standardized” pAUC and the defined range of interest, which could facilitate a wider and more appropriate use of this important summary index. First, we mathematically prove that the “standardized” pAUC increases with increasing range of interest for practically common ROC curves. Second, using comprehensive numerical investigations we demonstrate that, contrary to common belief, the uncertainty about the estimated standardized pAUC can either decrease or increase with an increasing range of interest.

Our results indicate that the partial AUC could frequently offer advantages in terms of statistical uncertainty of the estimation. In addition, selection of a wider range of interest will likely lead to an increased estimate even for standardized pAUC.

Keywords

evaluation of diagnostic performance; ROC; partial area under the Receiver Operating Characteristics; standardized pAUC; summary index; variance of standardized pAUC

1. Introduction

Assessment of diagnostic performance is an important problem in many fields. It has been gaining particular relevance in the development of medical diagnostic systems, evaluating biomarkers and building predictive models. A basic problem in assessing diagnostic performance is the accuracy of interpretations of examinations of subjects with a known binary true status (e.g., “normal”/“abnormal”). Typically, test results are either binary (e.g.,

*anb61@pitt.edu.

negative/positive with respect to the perceived abnormality) or have a form of a quantitative “rating” (e.g., likelihood of presence of a specific pre-defined abnormality or the likelihood of an actual finding being abnormal). The most widely used methodology for assessment of performance in this type of diagnostic task is the Receiver Operating Characteristic (ROC) analysis.

The basic quantities in ROC analysis are “sensitivity” (or true positive fraction) and “specificity” (or true negative fraction) which are defined as the probabilities of correct classification of normal and abnormal subjects into “negative” and “positive” groups correspondingly. When results of a diagnostic test are ordinal this classification is performed by comparing the diagnostic ratings for each subject to a fixed threshold. The ROC curve describes all pairs of sensitivity (or TPF) and 1-specificity (or FPF) values, computed for all positive thresholds. It is conventionally plotted with TPF as a vertical and FPF as a horizontal coordinate. The ROC curve is a fundamental tool in ROC analysis and determines most of the indices of diagnostic performance.

One of the most commonly used summary indices derived from the ROC curve is the area under the curve (AUC). AUC has a convenient interpretation and a close relationship to the well-known Wilcoxon statistic [1]; as a result, methods for AUC-based analyses are well developed and widely used [2,3]. One of the major practical drawbacks of the AUC as an index of diagnostic performance is that it summarizes the entire ROC curve, including regions that frequently are not relevant to practical applications (e.g., regions with low levels of specificity). In order to alleviate this deficiency while benefiting from some of the advantageous properties of the area under the ROC curve, one can use a partial area under the ROC curve (pAUC) which summarizes a portion the curve over the pre-specified range of interest (e_1, e_2) [4,5]. A number of approaches have been developed for pAUC-based analysis [4, 6, 7, 8]. However, the same features that increase the practical relevance of the partial AUC introduce some difficult to resolve issues related to arbitrariness of specifying the range of interest.

The partial AUC has two other limitations that prevent its widespread use. First is the dependence of the scale of possible values on the range of interest. In order to reduce this dependence of partial AUC on the range of interest being considered, several modifications to the partial AUC have been proposed [2,4]. Second, it is generally assumed that due to effective use of less information, its use results in loss of statistical precision, as compared with inferences based on the full AUC, [e.g., 2, 9, 10, 11]. Conjectures about the relative stability of the standardized partial AUC with respect to the range of interest and about the decrease in variance with increasing range are intuitively appealing and can frequently exert an effect on the way statistical analysis is planned and interpreted. However, the issues have not been adequately investigated to date.

In analyzing experimentally ascertained datasets from observer performance studies we frequently encountered scenarios that contradicted the two conjectures. Motivated by these observations we conducted an investigation that is summarized in this paper. In our investigation we focus on the partial area over the high specificity range, i.e., range of interest of the form $(0, e)$, where $e \in (0, 1]$. The range of high specificity is of particular interest in many practical applications including medicine in general and diagnostic imaging in particular. In Section 2 we introduce notations and demonstrate that under some commonly occurring conditions the standardized partial AUC increases with increasing range. In Section 3 we investigate the dependence of the variance of the estimate of standardized partial AUCs on the range of interest. In Section 4 we verify the generality of the findings of Section 3 in a simulation study investigating the width of the distribution of the estimates of the standardized partial AUC. We illustrate the identified patterns with an

analysis of experimentally ascertained data from two observer performance studies in Section 5 followed by a discussion and concluding statements in Section 6.

2. “Standardized” partial AUC and its properties

For assessment of diagnostic performance, a population of subjects is typically considered to consist of normal subjects ($D=0$) and abnormal subjects ($D=1$) (e.g., non-diseased/diseased). The test results for normal and abnormal subjects are assumed to follow probability distributions F_X and F_Y , correspondingly. Then, the True Positive Fraction (TPF, or sensitivity) and False Positive Fraction (FPF, or 1-specificity) are the “survival” functions for the corresponding distributions, i.e. $FPF(\xi)=P(X>\xi)=S_X=1-F_X$, $TPF(\xi)=P(Y>\xi)=S_Y=1-F_Y$. The ROC curve can be described by a pair of functions $(FPF(\xi), TPF(\xi))$, or can be written explicitly as $ROC(e)=S_Y(S_X^{-1}(e))$.

An important characteristic of the ROC curve is the diagnostic likelihood ratio negative, which is defined as $(1-ROC(e))/(1-e)$. The ROC curve with a decreasing negative diagnostic likelihood ratio offer important practical implications. Such ROC curves ensure that starting at any given operating point, a threshold-driven improvement in sensitivity will be better than an improvement achieved by randomly picking subjects that were tested “negative” at the given operating point [12, 13]. Thus, a decreasing negative diagnostic likelihood ratio in the region where operating points are observed is a natural property for many practical diagnostic tests.

The partial area under the ROC curve (pAUC) over the range $(0, e)$ is defined as an integral

of the ROC function over the given range, i.e. $A_e = \int_0^e ROC(f)df$. When $e=1$ the partial area represents the conventional area under the entire ROC curve (AUC). For ROC curve describing better-than-chance performance the partial area A_e can be shown to vary from $e^2/2$ to e , hence, a natural transformation of the partial area aimed to “standardize” the range of its values can be written as follows [4]:

$$\tilde{A}_e = \frac{1}{2} \left(1 + \frac{A_e - e^2/2}{e - e^2/2} \right) = \frac{1}{2} \left(1 + \frac{\int_0^e ROC(f)df - e^2/2}{e - e^2/2} \right) \quad (1)$$

Here, we term this index as the “standardized partial AUC”. For ROC curves describing better-than-chance performance, $\tilde{A}_e$ varies from 0.5 to 1 regardless of e , and for $e=1$ it reduces to the conventional AUC. It can also be shown that the standardized pAUC and the variance of its estimate are always larger than conventional pAUC and the variance of its estimate. Indeed since $1/e/(2e-1)$, is less than 1 for all $e > 1$,

$$\tilde{A}_e \geq \frac{1}{2} \left\{ 1 + 2 \left(A_e - \frac{e^2}{2} \right) \right\} = A_e + \frac{1}{2} - \frac{e^2}{2} \geq A_e$$

and

$$V(\tilde{A}_e) = V(\hat{A}_e) / 4 \left(e - \frac{e^2}{2} \right)^2 \geq V(\hat{A}_e).$$

Unfortunately, “standardization” of the partial area in (1) is not ideal. Indeed, although the range of $\tilde{A}_e$ is independent of e , the actual value of $\tilde{A}_e$ for a given ROC curve could depend on e . Moreover, as we demonstrate in Proposition 1 below, theoretically it can either increase or decrease with increasing range while remaining constant only for the “straight-line ROC curve” composed of two straight-line segments – one vertical and the other passing through (1,1). Formally, we define the straight line ROC curve passing through point (f, t) as $ROC_{straight,(f,t)}(e) = 1 - (1 - e) \times (1 - t) / (1 - f)$. It is easy to see that partial AUC for the straight-line ROC curve is

$$A_{e, straight, (f,t)} = e^2 \times (1 - t) / 2(1 - f) + e \times \{1 - (1 - t) / (1 - f)\},$$

and the standardized partial AUC does not depend on the range of interest (independent of e):

$$\tilde{A}_{straight, (f,t)} = 1 - (1 - t) / 2(1 - f). \quad (2)$$

Proposition 1

For any $e \in (0, 1)$

- i. $\frac{\partial \tilde{A}_e}{\partial e} > 0 \iff ROC(e) > 2(1 - \tilde{A}_e)e + (2\tilde{A}_e - 1)$
- ii. $\frac{\partial \tilde{A}_e}{\partial e} = 0 \iff ROC(e) = 2(1 - \tilde{A}_e)e + (2\tilde{A}_e - 1)$
- iii. $\frac{\partial \tilde{A}_e}{\partial e} < 0 \iff ROC(e) < 2(1 - \tilde{A}_e)e + (2\tilde{A}_e - 1)$

Proof—By straightforward differentiation of (1) we obtain:

$$\frac{\partial \tilde{A}_e}{\partial e} = \frac{1}{2} \left(e - \frac{e^2}{2} \right)^{-2} \left\{ (ROC(e) - e) \left(e - \frac{e^2}{2} \right) - \left(A_e - \frac{e^2}{2} \right) (1 - e) \right\}.$$

Since $A_e - \frac{e^2}{2} = (2\tilde{A}_e - 1) \left(e - \frac{e^2}{2} \right)$, the derivative of standardized partial AUC can be written as follows:

$$\frac{\partial \tilde{A}_e}{\partial e} = \frac{1}{2} \left(e - \frac{e^2}{2} \right)^{-1} \left\{ (ROC(e) - e) - (2\tilde{A}_e - 1)(1 - e) \right\}.$$

The three claims of this proposition immediately follow.

Proposition 1 implies that given the area over the range $(0, e)$ we can determine whether a small increase in the range would lead to an increase in the standardized pAUC by comparing whether the point on the ROC curve $ROC(e)$ is actually above or below the fixed straight line, which passes through $(1, 1)$ and has a slope of $2(1 - \tilde{A}_e)$. Alternatively, this comparison can be conducted by comparing the negative diagnostic likelihood ratio $(1 - ROC(e))/(1 - e)$ with $2(1 - \tilde{A}_e)$.

While results of proposition 1 are important for judging the dependence of standardized pAUC on small changes in range, they provide little insight into the more global behavior of the standardized pAUC, or the general form of the curves with always increasing/decreasing $\tilde{A}_e$. These questions are addressed by the following proposition and its corollaries.

Proposition 2

If the ROC curve has a decreasing negative diagnostic likelihood ratio in $(0, e_0)$, namely

$$\frac{\partial}{\partial e} \left\{ \frac{1 - ROC(e)}{1 - e} \right\} < 0, \text{ then } \frac{\partial \tilde{A}_e}{\partial e} > 0 \text{ in the same range.}$$

Proof—Let us consider e from $(0, e_0)$. Since for any $e' \in (0, e)$ $\frac{\partial}{\partial f} \left\{ \frac{1 - ROC(f)}{1 - f} \right\} \Big|_{f=e'} < 0$ we then immediately obtain the following inequality :

$$\frac{1 - ROC(e)}{1 - e} < \frac{1 - ROC(e')}{1 - e'} \quad \text{or} \quad ROC(e') < 1 - (1 - e') \times \frac{ROC(e) - 1}{e - 1}.$$

Hence over the range $(0, e]$, the partial area (A_e) and the standardized partial area under the ROC curve ($\tilde{A}_e$) are smaller than the corresponding areas under the straight line ROC curve passing through $(e, ROC(e))$. Indeed:

$$A_e = \int_0^e ROC(f) df < \int_0^e \left\{ 1 + (f - 1) \times \frac{ROC(e) - 1}{e - 1} \right\} df = A_{e, \text{straight}, (e, ROC(e))} \Rightarrow \tilde{A}_e < \tilde{A}_{\text{straight}, (e, ROC(e))}.$$

On the other hand, from (2) we obtain the following equality:

$$ROC(e) = 2 \left(1 - \tilde{A}_{\text{straight}, (e, ROC(e))} \right) e + \left(2 \tilde{A}_{\text{straight}, (e, ROC(e))} - 1 \right) = 2(1 - e) \tilde{A}_{\text{straight}, (e, ROC(e))} + 1.$$

Also, since $\tilde{A}_e < \tilde{A}_{\text{straight}, (e, ROC(e))}$, from above we obtain:

$$ROC(e) > 2(1 - e) \tilde{A}_e + 1 = 2(1 - \tilde{A}_e)e + (2\tilde{A}_e - 1)$$

Finally, applying the result (i) of proposition 1 we obtain $\frac{\partial \tilde{A}_e}{\partial e} > 0$.

As we discussed in the beginning of this section a decreasing negative diagnostic likelihood ratio is a natural property for many practical diagnostic tests. We also note that the result of proposition 2 is directly applicable to concave ROC curves as they automatically have decreasing diagnostic likelihood ratios. Figure 1 illustrates the increase of the standardized partial AUC with increasing range for five concave binomial ROC curves.

We note that the result of proposition 2 is directly extendable to the partial area index [4,5] as well as to the “non-standardized” partial area. Results summarized in this section indicate that in practical scenarios current approaches to standardization of the partial AUC do not eliminate the effect of the range of interest on values of the standardized pAUC. Moreover, increasing range of can frequently increase the apparent level of diagnostic performance. In the next two sections we examine the statistical uncertainty of the estimated standardized partial AUC.

3. Variance of the estimate of standardized partial AUC and range of interest

The partial AUC and other ROC related characteristics are typically estimated from a sample of n_0 normal and n_1 abnormal subjects with observed diagnostic test results of $\{x_i\}_{i=1}^{n_0}$ and $\{y_i\}_{i=1}^{n_1}$ correspondingly. Using these data the partial AUC can be estimated using non-parametric [6,7], semi-parametric [14], or parametric [2] approaches. For correctly specified models and large enough sample sizes all approaches provide similar results. We focus here on the relationship between the variance of the standardized partial AUC and the size of the range of interest. In particular we examine the common conjecture that in regular scenarios the variance would decrease with increasing range, since a larger range incorporates more available information on operating characteristics.

We begin by considering a simple variance estimate for the partial area under the binormal ROC curve [4]. In Section 5 we present simulation results that demonstrate the generality of the derived conclusions. The ROC curve for normally distributed test results depends on two parameters a and b , i.e. $ROC(e) = \Phi\{a + b\Phi^{-1}(e)\}$. The maximum likelihood estimates of these parameters, $\hat{a}$, $\hat{b}$, can be obtained by using the sample mean and the MLE of the variance for the diagnostic results for normal and abnormal subjects [2]. The estimate of the partial area can be computed by numerical integration of the estimated binormal ROC curve and its variance can be estimated using $\hat{a}$, $\hat{b}$ and the range of interest $(0, e)$ in the following manner [2]:

$$\widehat{V}(\widehat{A}_e) = f^2 V(\widehat{a}) + g^2 V(\widehat{b}) + 2fgC(\widehat{a}, \widehat{b}) \quad (3)$$

where:

$$\begin{aligned} \widehat{V}(\widehat{a}) &= \frac{n_0(a^2+2)+2n_1b^2}{2n_0n_1}, \quad \widehat{V}(\widehat{b}) = \frac{(n_1+n_0)b^2}{2n_0n_1}, \quad \widehat{C}(\widehat{a}, \widehat{b}) = \frac{ab}{2n_1}, \\ &\text{and} \\ f &= \frac{\exp\left\{-\frac{a^2}{2(1+b^2)}\right\}}{\sqrt{2\pi(1+b^2)}} \{\Phi(h)\}, \quad g = \frac{\exp\left\{-\frac{a^2}{2(1+b^2)}\right\}}{2\pi(1+b^2)} \left\{-\exp\left(-\frac{h^2}{2}\right)\right\} \\ &\quad - \frac{ab\exp\left\{-\frac{a^2}{2(1+b^2)}\right\}}{\sqrt{2\pi(1+b^2)^3}} \{\Phi(h)\}, \end{aligned}$$

and,

$$h = \left\{ \Phi^{-1}(e) + \frac{ab}{1+b^2} \right\} \sqrt{1+b^2}.$$

We can compute the variance of the estimated standardized partial AUC as:

$$V(\hat{A}_e) = \frac{V(\hat{A}_e)}{4\left(e - \frac{e^2}{2}\right)^2},$$

where $V(\hat{A}_e)$ is computed according to (3).

Figure 2 demonstrates the variance of the estimated standardized pAUC as a function of the length of the range e , for two different binormal as well as straight-line ROC scenarios. These scenarios are based on a sample size of 100, ($n_0 = n_1 = 50$) and describe different shapes of ROC curves, including concave curves ($b=1$) and typical improper curves ($b=0.5$) [10]. Each figure shows variance functions for five ROC curves with AUCs of 0.55, 0.65, 0.75, 0.85, and 0.95. We note that here, and in the investigations that follow, we consider binormal ROC curves with $b \leq 1$ since the corresponding shapes of these ROC curves are more common in practical applications including, but not limited to, medical imaging. Indeed, a binormal ROC curve with $b > 1$ implies a worse-than-chance performance in evaluations of highly suspicious subjects which rarely happens in practice.

As shown in Figure 2b, for an improper binormal ROC curve, the variance frequently increases with increasing range, in particular the variance of the full AUC ($e=1$) tends to be larger than the variances for standardized partial AUCs over most ranges considered. The anticipated decrease in the variance when switching to full AUC is evident only for the ROC curve with the largest AUC (0.95) considered here. For concave ROC curves (Figure 2a) the variance of the full AUC can exhibit both patterns, namely, it can be either smaller or larger than variance of the standardized partial AUCs on $(0, e)$. The decrease in variance with increasing range is observed only for ROC curves with AUC values greater than 0.75. In all straight-line ROC scenarios – where all standardized partial AUCs are exactly the same as the full AUC—the variance of the standardized partial AUC increases.

If they hold in general, these results have the important implications that in a number of practical scenarios the estimated partial AUC may be no less precise than the estimated variance for the full AUC. The variance is an important characteristic of the statistical uncertainty of the estimated standardized pAUC. However, its usefulness for non-symmetric distribution is limited (e.g., sampling distribution of estimates of high pAUC). Furthermore, the generality of the trends in Figure 2 could be questioned based on the assumption of normality used in computation of the variance. In order to verify these trends we conducted a simulation study as described in the following section.

4. Simulation study

We conducted a simulation study to assess the length of the equal-tail 95% range (97.5th – 2.5th percentile) of the sampling distribution of the standardized pAUC. In the simulation study the test results for normal and abnormal subjects were generated from normal distributions with parameters selected to generate binormal ROC curves with specific values of AUC (AUCs (ranging from 0.55 to 0.95) and of the three values for shape parameter b (1, 0.5 and 0.33). Values for the parameters of binormal ROC curves were selected to reflect shapes typically encountered in performance assessment studies in diagnostic medicine. We also investigated scenarios when the test results obey exponential distributions corresponding to power-law ROC curves [15], and uniform distributions corresponding to the straight-line ROC curves. For each scenario we generated 10,000 datasets of with $n_0=50$ and $n_1=50$ subjects.

For each dataset we estimated the ROC curve (MLE for binormal and exponential and empirical for uniform scenario) and using numerical integration, we computed the standardized partial AUC over different ranges starting from 0 and ending at 0.2, 0.4, 0.6, 0.8, and 1. The difference between the 9750th largest and 250th smallest estimate of AUC for a given scenario was used to estimate the length of the equal-tail 95% range of the sampling distribution. We note that transformations (e.g., logit) are often used to improve on Wald-type confidence intervals. In the simulation study, however, we have the ability to assess the width of distribution more precisely by using quantiles of the simulated distribution.

The results for the parametric estimator of the standardized pAUC are summarized in Table 1 (results for the empirical estimator are similar and were omitted for brevity). The results summarized in Table 1 closely agree with results from the previous section (Figure 1). In particular, the lengths of the equal-tail 95% ranges of the sampling distributions of the estimated standardized pAUCs increase with increasing ranges for the ROC curves with lower AUCs (e.g., AUC for concave ROC curves is less than 0.75). With increasing “improperness” of the ROC curves (i.e., decreasing b in Table 1) decreasing trends, even for ROC curve with large AUCs, are gradually diminishing. For example, for a binormal ROC curve with $b=0.33$, the length of the equal-tail 95% interval of sampling distribution of standardized pAUC decreases with increasing range of interest for all considered ROC curves. However, for the straight-line ROC curves, which guarantee constancy of the standardized partial AUC, the width of the sampling distribution always increases.

5. Examples

In this section we illustrate the patterns described in the previous sections with an example obtained from two datasets from observer performance studies we previously conducted [16]. One dataset (307 cases, 103 abnormal and 204 normal) includes observer’s ratings for classification of images as depicting/non-depicting lung nodules. The second dataset (307 cases, 84 abnormal and 223 normal) includes observer’s ratings for classification of images in regard to presence/absence of subtle interstitial disease. For both datasets the diagnostic ratings were provided by a group of radiologists using a pseudo-continuous scale from 0 to 100.

For each dataset we estimated empirical ROC curves by connecting empirical points with straight lines [2,3]. The estimates of the standardized partial AUC were computed by integration for ranges starting at 0 and ending at 0.2, 0.4, 0.6, 0.8, and 1. Variance of the empirical estimator of the standardized partial AUC was estimated using a nonparametric bootstrap approach [17]. Alternatively, and virtually equivalently for the given datasets (the results are omitted for brevity), the variance can be estimated using the two-sample jackknife approach [18], which can be viewed as an adaptation of the method described by He and Escobar [7] to linearly interpolated empirical ROC curves. The bootstrap percentile confidence intervals were computed using 10,000 random bootstrap samples.

Figure 4 illustrates the empirical ROC curves for the two datasets. Table 2 summarizes the standardized partial area, its bootstrap variance, and the length of the 95% bootstrap confidence interval. In agreement with our findings in Section 2 for both empirical ROC curves the standardized partial areas were increasing with increasing range. In agreement with our findings in Section 3, for the ROC curve with AUC=0.84, the variance estimator of the standardized partial area first decreased and then remained virtually unchanged. Since data for interstitial disease included very subtle cases, the ROC curve had a relatively low AUC of 0.64, and the bootstrap variance of standardized partial area for the ROC curve increases over the considered ranges. The same trend was observed for the length of the 95% bootstrap confidence interval.

6. Discussion

We developed two important properties of the partial AUC which should facilitate a wider and more appropriate use of this important summary index. First, for ROC curves typically encountered in practical applications the “standardized” partial AUC actually increases with increasing range of interest. For example, the standardized partial AUC is always increasing for concave ROC curves and also, when considering short ranges for improper ($b < 1$) binormal ROC curves. Second, the statistical uncertainty of the estimated standardized partial AUC in general, and its variance in particular, could frequently be smaller than those of the full AUC. In particular, a decrease of the variance with increasing range (as often conjectured) can be observed only in the case of concave binormal ROC curve ($b = 1$) with AUC of at least 0.75, or in the case of improper binormal ROC curves with increasingly larger AUCs. This decrease in the width of distribution with increasing range for large AUCs, is likely to be the result of the true value of standardized pAUC approaching its upper boundary. As demonstrated in Figure 3 for the straight-line ROC curves, where standardized pAUC is a constant, the variance increases regardless of how high the ROC curve is.

Our findings have direct practical implications for the design and analysis of diagnostic performance studies in which it is common to disregard partial area indices in favor of inferences based on full area under the ROC curve. Specifically, our results on the statistical uncertainty of estimation indicate that in many practical scenarios inferences based on the partial AUC could be no less statistically advantageous than inferences based on the full AUC. Our results on the values of the standardized pAUC indicate that the estimates should be interpreted in the context of the range of interest, even if the standardization is employed. In particular, using a wider range of interest than that which is of interest clinically, could lead to overoptimistic estimates of performance in many practically relevant scenarios.

Results presented in this paper resolve several issues related to use of partial AUC for evaluating performance of an individual diagnostic system. This should encourage investigators to use this practically relevant index more often. Our work also provides a foundation for development of better understanding of properties of comparisons of partial AUCs for several diagnostic tests.

Acknowledgments

Research reported in this publication was supported by the National Institute of General Medical Science of the National Institutes of Health under award number R01GM098253.

References

1. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982; 143:29–36. [PubMed: 7063747]
2. Zhou, XH.; Obuchowski, NA.; McClish, DK. *Statistical methods in diagnostic medicine*. Wiley & Sons Inc; New York: 2002.
3. Pepe, MS. *Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press; Oxford: 2003.
4. McClish DK. Analyzing a portion of the ROC curve. *Medical Decision Making*. 1989; 9:190–195. [PubMed: 2668680]
5. Jiang Y, Metz CE, Nishikawa RM. A receiver operating characteristic partial area index for highly sensitive diagnostic tests. *Radiology*. 1996; 201:745–750. [PubMed: 8939225]
6. Dodd LE, Pepe MS. Partial AUC estimation and regression. *Biometrics*. 2003; 59:614–623. [PubMed: 14601762]

7. He Y, Escobar M. Nonparametric statistical inference method for paired areas under receiver operating characteristics curves, with application to genomic studies. *Statistics in Medicine*. 2008; 27:5991–5308. [PubMed: 18792086]
8. Zhang DD, Zhou XH, Freeman DH Jr, Freeman JL. A non-parametric method for the comparison of partial areas under ROC curves and its application to large health care data sets. *Statistics in Medicine*. 2002; 21:701–715. [PubMed: 11870811]
9. Hanley JA. Receiver operating characteristic (ROC) methodology: state of the art. *Critical Reviews in Diagnostic Imaging*. 1989; 29:307–335. [PubMed: 2667567]
10. Obuchowski NA, McClish DK. Sample size determination for diagnostic accuracy studies involving binormal ROC curve indices. *Statistics in Medicine*. 1997; 16:1529–1542. [PubMed: 9249923]
11. Wieand S, Gail MH, James BR, James KL. A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Bometrika*. 1989; 76(3):585–592.
12. Norman DA. A comparison of data obtained with different false-alarm rates. *Psychological Review*. 1964; 71(3):243–246. [PubMed: 14146348]
13. Bandos AI, Rockette HE, Gur D. Use of likelihood ratios for comparisons of binary diagnostic tests: Underlying ROC curves. *Medical Physics*. 2010; 37(11):5821–5830. [PubMed: 21158294]
14. Zou K, Hall WJ. Semiparametric and parametric transformation models for comparing diagnostic markers with paired design. *Journal of Applied Statistics*. 2000; 29:803–816.
15. Egan, JP. *Signal Detection Theory and ROC Analysis*. 1. Academic Press; New York: 1975.
16. Gur D, Abrams GS, Chough DM, Ganott MA, Hakim CM, Perrin RL, Rathfon GY, Sumkin JH, Zuley ML, Bandos AI. Digital breast tomosynthesis: observer performance study. *American Journal of Roentgenology*. 2009; 193(2):586–591. [PubMed: 19620460]
17. Efron, B.; Tibshirani, RJ. *An Introduction to the Bootstrap*. 3. Chapman & Hall; New York: 1993.
18. Arveson JN. Jackknifing U-statistics. *Annals of Mathematical Statistics*. 1969; 40(6):2076–2100.

Figure 1.
Values of the standardized partial AUC for concave binormal ROC curves.

Figure 2. Variance of standardized pAUC estimates for binormal ROC curves over $(0, e)$ as a function of the size of the range of interest e .

Figure 3. Variance of standardized pAUC estimates for straight-line ROC curves over $(0, e)$ as a function of the size of the range of interest e .

Figure 4.
Empirical ROC curves for the two datasets

Table 1

Differences of 2.5% and 97.5% estimated percentiles of sampling distributions of standardized pAUC for binormal, power-law and straight-line ROC curves.

Parameters of the ROC curves	Ranges of False Positive Fractions				
	0–0.2	0–0.4	0–0.6	0–0.8	0–1
Binormal Distribution					
<i>b=0.33</i>					
<i>auc=0.55</i>	0.1294	0.1443	0.1641	0.1877	0.2181
<i>auc=0.65</i>	0.1341	0.1454	0.1627	0.1848	0.2119
<i>auc=0.75</i>	0.1320	0.1410	0.1542	0.1709	0.1918
<i>auc=0.85</i>	0.1170	0.1205	0.1278	0.1383	0.1511
<i>auc=0.95</i>	0.0770	0.0739	0.0745	0.0773	0.0818
<i>b=0.50</i>					
<i>auc=0.55</i>	0.1317	0.1521	0.1714	0.1953	0.2209
<i>auc=0.65</i>	0.1416	0.1546	0.1706	0.1894	0.2096
<i>auc=0.75</i>	0.1463	0.1510	0.1603	0.1732	0.1870
<i>auc=0.85</i>	0.1373	0.1347	0.1374	0.1426	0.1496
<i>auc=0.95</i>	0.0922	0.0828	0.0789	0.0782	0.0792
<i>b=1.0</i>					
<i>auc=0.55</i>	0.1107	0.1528	0.1826	0.2053	0.2189
<i>auc=0.65</i>	0.1460	0.1773	0.1941	0.2058	0.2100
<i>auc=0.75</i>	0.1779	0.1881	0.1888	0.1863	0.1849
<i>auc=0.85</i>	0.1875	0.1697	0.1560	0.1466	0.1430
<i>auc=0.95</i>	0.1390	0.1028	0.0855	0.0765	0.0737
Exponential Distribution					
<i>auc=0.55</i>	0.1124	0.1508	0.1745	0.1885	0.1936
<i>auc=0.65</i>	0.1413	0.1610	0.1705	0.1754	0.1769
<i>auc=0.75</i>	0.1779	0.1881	0.1888	0.1863	0.1849
<i>auc=0.85</i>	0.1486	0.1500	0.1492	0.1483	0.1479
<i>auc=0.95</i>	0.0522	0.0445	0.0407	0.0388	0.0381
Uniform Distribution					
<i>auc=0.55</i>	0.128	0.161	0.190	0.214	0.224

Parameters of the ROC curves	Ranges of False Positive Fractions				
	0-0.2	0-0.4	0-0.6	0-0.8	0-1
<i>auc=0.65</i>	0.151	0.173	0.190	0.206	0.215
<i>auc=0.75</i>	0.150	0.164	0.178	0.190	0.196
<i>auc=0.85</i>	0.133	0.143	0.150	0.156	0.160
<i>auc=0.95</i>	0.088	0.088	0.093	0.096	0.098

Table 2

Standardized partial areas and variances of standardized partial areas

	0–0.2	0–0.4	0–0.6	0–1
<u>Nodule</u>				
Stand pAUC	0.796	0.819	0.835	0.843
Standard deviation	0.0270	0.0261	0.0257 *	0.0257 *
Length of 95% bootstrap CI	0.1058	0.1020	0.0993	0.0998
<u>Interstitial</u>				
Stand pAUC	0.534	0.579	0.613	0.644
Standard deviation	0.0206	0.0298	0.0329	0.0334
Length of 95% bootstrap CI	0.0799	0.1160	0.1271	0.1304

* Further increase of the range does not increase the number of included empirical operating points.