International Journal of Research Studies in Biosciences (IJRSB)
Volume 6, Issue 7, 2018, PP 22-30
ISSN No. (Online) 2349-0365
DOI: http://dx.doi.org/10.20431/2349-0365.0607004
www.arcjournals.org

An Alternative Nonparametric Method of Assessing a Difference in the Areas under the Curves (Aucs) for Paired Data

Okeh UM[1], Mbegbu JI[2]

[1]*Department of Industrial Mathematics and Applied Statistics, Ebonyi State University Abakaliki Nigeria*

[2]*Department of Mathematics, University of Benin, Edo State Nigeria*

***Corresponding Author:** **Okeh UM,** Department of Industrial Mathematics and Applied Statistics, Ebonyi State University Abakaliki Nigeria*

Abstract: *The area under the receiver operating characteristic (ROC) curve (AUC) is a popularly used index when comparing two ROC curves. However, this index is less informative when two ROC curves cross while the AUCs are the same. In order to detect differences between ROC curves and to be able to tackle the problem of transformation of the original data and exchangeability of the labels of two diagnostic tests within subject which characterized the methods proposed by Venkatraman and Beggs (1996) as well as Bandos et al(2005), an alternative permutation test based on between-subject permutations of the labels of the subjects is proposed for assessing a change in the AUCs in a matched pair of data from two diagnostic test procedures having both diseased and nondiseased subject in each of the test. Here permutations are made between subjects particularly by shuffling the diseased and nondiseased labels of the subjects within each diagnostic test procedure. The validity of this permutation test is assured even when the scale of measurement of test results differs for each diagnostic test procedure. We demonstrate under the assumption of equality of AUCs that our permutation test is a modified Wilcoxon signed rank test for the symmetry of an underlying discrete distribution with valid sample size. Through extensive data simulation, we show the numerical studies of operating characteristics of our new permutation test and show that our test has equal statistical power to a permutation test proposed by Bandos et al(2005).*

Keywords: *Modified Wilcoxon signed rank test, Diagnostic test procedure, Permutation test, Exchangeability, receiver operating characteristic (ROC) curve, Area under the curve(AUC), Nonparametric test, asymptotic.*

1. INTRODUCTION

Many nonparametric estimators of the variance of a single AUC and the difference between two correlated AUCs have been proposed. The methods proposed by Bamber in 1975 (based on formula from Noether 1967) and Wieand et al (1983) provide unbiased estimators of the variance of a single AUC and the covariance of two correlated AUCs correspondingly. Hence, these estimators are useful for assessing the magnitude of the variability but may provide no advantages in hypothesis testing. The estimator proposed by Hanley & McNeil (1982) explicitly depends only upon the AUC and sample size and thus enables simple estimation of the sample size for a planned study. However, this estimator is known to underestimate or overestimate variance depending on the underlying parameters (Obuchowski 1994; Hanley & Hajian-Tilaki 1997) and thus is not optimal for either variance estimation or hypothesis testing (an improved estimator of the same kind was proposed by Obuchowski in 1994). Perhaps the most widely used estimator which offers both relatively accurate estimator of the variability and leads to acceptable hypothesis testing is the estimator proposed by DeLong et al (1988). Therefore nonparametric inference for a difference in areas under the curve (AUCs) for paired studies was first proposed by DeLong et al (1988), which is based upon asymptotic theory for U-statistics (Hoeffding, 1948) and estimates the covariance of the 2 U-statistics using the jackknife. Other nonparametric inference procedures include those based upon an analysis of variance of jackknife pseudovalues (Dorfman et al, 1992; Song, 1997) and bootstrap-based methods (Campbell, 1994; Moise et al, 1988). However, the validity of each of these methods is founded in large-sample theory and each does not necessarily lead to a valid test of difference in AUC in small samples. A competing approach to the above methods is a permutation test, the size of which will remain nominal in small samples. Permutation based procedures are specific to hypothesis testing. A permutation procedure constructs a

permutation sample space, which consists of the equally likely permutation samples. The permutation samples are created by interchanging the units of the data that are assumed to be "exchangeable" under the null hypothesis. The permutation sample space is the exact probability space of the possible arrangements of the data under the null hypothesis given the original sample. This natural permutation test is characterized by exchanging the paired units when two diagnostic systems are to be compared with paired data. Two permutation tests for paired receiver operating characteristic (ROC) studies currently exist: one proposed by Venkatraman and Begg (1996) and the other one from Bandos et al (2005). The test of Bandos et al(2005) directly tests for an equality of AUCs, while the test of Venkatraman and Begg(1996) is more general and tests for equality of the underlying ROC curves. As a result, the test of Venkatraman and Begg is less powerful for testing equality of AUCs. In other words, this permutation method is for detecting any differences between two ROC curves so that the authors used a measure specifically designed to detect the differences at every operating point. Both permutation tests are executed by permuting the labels of the two diagnostic tests within each diseased and non-diseased subject. Such an approach implicitly assumes that both diagnostic tests are exchangeable within subject and requires an appropriate transformation, such as ranks, for diagnostic tests differing in scale measurement. This means that both of these tests assume the same condition of exchangeability of the diagnostic results under the null hypothesis, but differ with respect to their sensitivity to specific alternatives and the availability of an asymptotic version. Namely our permutation test better detects different ROC curves if they differ with respect to the AUC, and it has an easy-to-implement and precise approximation which is unavailable for the test of Venkatraman & Begg. The availability of the asymptotic approximation to the permutation test can be an important issue since the exact permutation tests are practically impossible to implement with even moderate sample sizes and the Monte Carlo approximation to the permutation test is associated with a sampling error. Fortunately, in some cases the asymptotic approximation can be constructed by appealing to the asymptotic normality of the summary statistic and using the estimator of its variance, if the latter is derivable. For the nonparametric estimator of the difference in the AUC we demonstrated (Bandos et al, 2005) that the exact permutation variance can be calculated directly without actually permuting the data previously mention estimation methods which provide estimators of the variance regardless of the magnitude of the difference. However, the properties of the statistical tests can be compared directly with Monte Carlo and the availability of the closed-form solution for the permutation variance greatly alleviates the computational burden of this task. The comparison of the asymptotic permutation test with the widely used procedure of DeLong et al. indicate the advantages of the former for the range of parameters common in diagnostic imaging, i.e. AUC greater than 0.8 and correlation between scores greater than 0.4 (Bandos et al., 2005).

Meanwhile, the estimator proposed by DeLong et al(1988) possesses an upward bias which on the one hand results in an improved (compared to the unbiased estimator) type I error of the statistical test for equality of the AUCs when AUCs are small, but on the other hand results in loss of statistical power when AUCs are large (Bandos 2005; Bandos et al, 2005). Bandos et al (2005) compared the performance of their test to that of DeLong et al (1988) via simulation and found that the permutation test had greater power than the nonparametric test developed by DeLong et al(1988) when there was moderate correlation between diagnostic tests, large AUCs, and small sample sizes.

We propose an alternative permutation test based on between-subject permutations of the diseased non-diseased labels of the subjects. These permutations do not require the exchangeability of the two diagnostic tests and do not require transformations of the original data. In Section 2, we derive our permutation test and a corresponding asymptotic normal approximation. In Section 3, we discuss simulation results regarding the validity and power of our test in relation to that of Bandos and others (2005). In Section 4, we make concluding remarks.

2. MATERIALS AND METHOD

2.1. Proposed Permuation Test (Modified Wilcoxon Signed Rank Test)

This study is aimed at comparing two diagnostic tests in terms of their AUCs, namely, AUC_1 *and* AUC_2 with a view to identifying a change where both diagnostic tests are each having diseased and nondiseased subject. A total of N subjects passed through the two diagnostic tests out of which altogether a total of m subjects are nondiseased while n represents the total number of subjects who are diseased. For the

nondiseased subject(i) in both diagnostic tests, let X_{i1} *and* X_{i2} respectively represent the tests results from diagnostic test 1 and 2 where $i=1,2,...,m.$ Also for the diseased subject (j),let Y_{j1} *and* Y_{j2} respectively denote the test results from diagnostic test 1 and 2 where $j=1,2,...,n.$ Given the two diagnostic tests,the vector of paired test results for the nondiseased subjects are denoted by $X=\{(X_{11},X_{12}),(X_{21},X_{22}),....,(X_{m1},X_{m2})\}$ while the vector of paired test result for the diseased subjects in diagnostic test 1 and 2 is denoted as $Y=\{(Y_{11},Y_{12}),(Y_{21},Y_{22}),....,(Y_{n1},Y_{n2})\}.$ Based on these definitions,the nonparametric estimate of the difference in AUC between the two diagnostic tests denoted as $AUC_{\Delta}=AUC_2-AUC_1$ is given by $\hat{AUC}_{\Delta}=\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}S_{ij}$ *where* S_{ij} is an indicator variable for the two diagnostic tests denoted as $S_{ij}=U_{ij2}-U_{ij1}$ *and* $U_{ijk}=I\left(X_{ik}<Y_{jk}\right)+\frac{1}{2}I(X_{ik}=Y_{jk})$, *for* $k=1,2.$ Note that U_{ij1} *and* U_{ij2} are indicator variables for diagnostic tests 1 and 2 respectively. Given the nonparametric estimate of AUC_{Δ} above, the null hypothesis suitable for testing it is given as $AUC_{\Delta}=0.$ The test this hypothesis formally in terms of permutation test here, we combine all the subjects into 1 group of N subjects. Let the N test results from diagnostic test 1 be $Z_1=\{Z_{11},Z_{12},...,Z_{1m},Z_{1,m+1},Z_{1,m+2},...,Z_{1N}\}$ in which the subscripts $l=1,2,...,m$ denotes values of subjects having non-diseased test results and $l=m+1,m+2,m+3,...,N$ representing values of subjects having diseased test results. Based on these information, we compare every subject's value to every other subject's value. This means comparing every diseased subject to all nondiseased subjects and all (n-1) other diseased subjects. These comparison yields $V_{ll'1}=I(Z_{l1}<Z_{l'1})+\frac{1}{2}I(Z_{l1}=Z_{l'1}),$ *for* $l\neq l'.$ In a similar way, we are comparing every nondiseased subject to all diseased subjects and all (m-1) other nondiseased subjects. Following the same procedure, from diagnostic test 1 and 2,we have that $Z_2=\{Z_{21},Z_{22},...,Z_{2m},Z_{2,m+1},Z_{2,m+2},...,Z_{2N}\}$ denoting N test results from diagnostic test 2 in which the subscripts have already been defined in diagnostic test 1 above. Also for diagnostic test 2 $V_{ll'2}=I(Z_{l2}<Z_{l'2})+\frac{1}{2}I(Z_{l2}=Z_{l'2}),$ *for* $l\neq l'.$ By these definition $V_{ll'}=1-V_{l'lk},$ *for* $k=1,2.$ Following these definitions and in other to adjust for the possible presence of tied observations, we generalize the definition of the estimate of AUC_{Δ} as

$$\hat{AUC}_{\Delta}=\frac{1}{mn}\sum_{l}\sum_{l'<l}\omega_{ll'}r\left(|T_{ll'}|\right) \tag{2.1}$$

Where

$$\omega_{ll'}=\begin{cases}1, if\ subject\ l\ is\ nondiseased\ and\ subject\ l'\ is\ diseased,\\ 0, if\ subjects\ l\ and\ l'\ are\ both\ diseased\ or\ both\ nondiseased,\\ -1, if\ subject\ l\ is\ diseased\ and\ subject\ l'\ is\ nondiseased,\end{cases}$$

And $T_{ll'}=(V_{ll'2}-V_{ll'1}).$

Here $\omega_{ll'}$ is the sign rank of $T_{ll'}$, $|T_{ll'}|$ is absolute value of the difference between two diagnostic tests resulting from when every subject's value is compared to every other subject's value, $r\left(|T_{ll'}|\right)$ is the signed rank of $|T_{ll'}|$, while $V_{ll'1}$ *and* $V_{ll'2}$ are figures obtained when every subject's value is compared to every other subject's value from diagnostic test 1 and 2 respectively. The second summation sign in (2.1) is restricted to $l'<l$ because of the fact that $T_{ll'}=-T_{l'l}.$ This simply means that both values correspond to a comparison of the same diseased and non-diseased subject, but in reverse orders. In other to validate the permutation test based on the condition that the two diagnostic tests have continuous distributions, we prove here that testing the hypothesis $AUC_{\Delta}=0$ is equal to testing the hypothesis that $T_{ll'}$ has a distribution symmetric about zero.

Suppose $V_{ll'1}$ *and* $V_{ll'2}$ have distribution

$$\Pr ob(V_{ll'1}=v)=\begin{cases} p_{10}, & if\ v=0,\\ 1-p_{10}-p_{11}, & if\ v=0.5,\\ p_{11}, & if\ v=1,\end{cases} \text{ and } \Pr ob(V_{ll'2}=v)=\begin{cases} p_{20}, & if\ v=0\\ 1-p_{20}-p_{21}, & if\ v=0.5\\ p_{21}, & if\ v=1,\end{cases} \tag{2.2}$$

So that $AUC_1=p_{10}+\frac{1}{2}(1-p_{10}+p_{11})$ *and* $AUC_2=p_{20}+\frac{1}{2}(1-p_{20}+p_{21})$. But $P_d=\frac{2mn}{[N(N-1)]}$ is the probability that only one of the subjects l *and* l' is diseased, while $(1-P_d)=1-\frac{2mn}{[N(N-1)]}$ is the probability that subjects l *and* l' are both diseased or both nondiseased. Because of that, $T_{ll'}=V_{ll'1}-V_{ll'2}$ has distribution

$$\Pr ob(T_{ll'}=t)=\begin{cases} p_d\,p_{10}\,p_{21}, & if\ t=-1,\\ p_d[p_{10}(1-p_{20}-p_{21})+ & \\ p_{21}(1-p_{10}-p_{11})], & if\ t=-0.5,\\ (1-p_d)+p_d[p_{10}p_{20}+p_{11}p_{21}+ & \\ (1-p_{10}-p_{11})(1-p_{20}-p_{21})], & if\ t=0,\\ p_d[p_{11}(1-p_{20}-p_{21})+ & \\ p_{20}(1-p_{10}-p_{11})], & if\ t=0.5,\\ p_d\,p_{20}\,p_{11}, & if\ t=1.\end{cases} \tag{2.3}$$

Under the null hypothesis $AUC_1=AUC_2=AUC^*$, we find that $\omega_{ll'}$ is the sign of $(T_{ll'})$ and its meanis given by

$$E(\omega_{ll'})=p_d[(p_{11}-p_{21})AUC^*+p_{20}(1-p_{11})-p_{10}(1-p_{21})]. \tag{2.4}$$

If continuous distributions are assumed for both diagnostic tests, then $AUC^*=p_{10}=(1-p_{11})=p_{20}=(1-p_{21})$ and we find that $E(\omega_{ll'})=0$,thereby proving that $T_{ll'}$ is equally likely to be positive or negative. Diagnostic tests producing discrete values will lead to a distribution for $T_{ll'}$ that is skewed positively or negatively from zero. However, the effect of the skewness will vanish asymptotically as $N\to\infty$, assuming that the ratio m/N remains constant (Romano,1990).As a result, our permutation test will be (asymptotically) valid when $AUC_\Delta=0$, irrespective of the value of AUC_1 *and* AUC_2.

A standard and most appropriate nonparametric test for matched continuous test results and for symmetry about zero is the Wilcoxon signed rank test. This means that the null distribution of AUC_Δ is obtained by calculating AUC_Δ for every permutation of $\omega_{ll'}$, the sign of the $T_{ll'}$. This is done when we permute $\omega_{ll'}$, by switching the labels of nondiseased subject l and diseased subject l'. This corresponds to permuting the vector of diseased/nondiseased labels among the subjects. Based on this style of permutation, the connection between a subject's disease status and the values from the two diagnostic tests are broken. This permutation scheme validates any test whose $AUC_1=AUC_2=c$ where c is a real value between 0.5 and 1.0 inclusive. This does not only mean that our permutation test is a valid test of $AUC_\Delta=0$, that is, the differences in AUC of two diagnostic tests are equally not useful for detecting disease. Similar to all Wilcoxon signed rank tests,the validity of our test is observed when $T_{ll'}$ has a distribution symmetric about zero and has power impacted by $p_0=\Pr ob(T_{ll'}=0)$, that is, the quantity of mass the discrete distribution of $T_{ll'}$ has at zero. Here as either AUC_1 or AUC_2 increases toward 1.0, p_0 increases also. In particular as the overlap in the distributions of diseased and non-diseased subjects for both diagnostic tests decreases, then the likelihood that $V_{ll'}=1$ for the two diagnostic tests increases, which will lead to increased probability that $T_{ll'}=0$ and decrease in the power of our permutation test. Based on this scenario, we adopt the traditional approach of improving power in Wilcoxon signed rank tests by proposing a modified statistic

$$D^*=S_+ + f\ S_0, \tag{2.5}$$

Where $S_0=\sum_i\sum_j I(S_{ij}=0), S_+=\sum_i\sum_j I(S_{ij}>0),$ and f is a proportion describing the degree to which we use the number of zeros as evidence against $AUC_\Delta=0.$ Surprisingly, the optimal value of f in (2.5) is difficult

to obtain because of the correlation of the $\omega_{ll'}$. If our $\omega_{ll'}$ were independent, the methods of Irle and Klosener (1980), based upon the Neyman–Pearson lemma, would show that the optimal value of f is a function of $p_- = \Pr ob(S_{ll'} < 0)\ and\ p_+ = \Pr ob(S_{ll'} > 0),$ both of which in our setting will vary depending on the actual values of AUC_1 and AUC_2. Although Putter (1955) suggested f = 1/2 and Coakley and Heise (1996) proposed f = 2/3 for use with the standard Wilcoxon signed rank test. Nonetheless,we therefore suggest a formula for f denoted as $\hat{f}$ that appears to perform well in most settings. It is given by

$$\hat{f} = \frac{\log(p+) - \log([1 - p_0]/2)}{\log(p+) - lop(p-)}, \tag{2.6}$$

Here $\hat{f}$ is based upon the optimal value proposed by Irle and Klosener (1980) in the setting of independent S_{ij}. However, we do not claim this estimate to be necessarily optimal in our setting, but rather one that appears to offer our test nominal size and excellent power across a variety of settings.

Furthermore, because the exact correlation structure of the $\omega_{ll'}$ is quite complicated, permutation theory cannot be used to derive the asymptotic variance of D^*. As an alternative, we collect the values of $S_+ \text{ and } S_0$ from each permutation to give us a joint permutation distribution for S_+ and S_0. From this distribution, we compute $\hat{\mu}_+ \ and \ \hat{\mu}_0$, the respective sample means of S_+ and S_0, as well as $\hat{\sigma}_+^2, \hat{\sigma}_0^2 \ and \ \hat{\sigma}_{+0}$, the respective sample variances and covariance of $S_+ \text{ and } S_0$. Assuming a value for f, we estimate the null mean and variance of D^* to be $\mu_{D^*} = f\hat{\mu}_0 + \hat{\mu}_+ \ and \ \hat{\sigma}_{D^*}^2 = f^2\hat{\sigma}_0^2 + \hat{\sigma}_+^2 + 2f\hat{\sigma}_{+0}$, respectively. An asymptotic version of our permutation test would therefore compare the value of $(D^* - \hat{\mu}_{D^*})/\hat{\sigma}_{D^*}$ to the appropriate critical value in a standard normal distribution.

3. RESULTS

3.1. Description of Extensive Simulation and Results

We have simulated measurements for both diagnostic tests as follows. We drew the 2 continuous measurements for each nondiseased subject from a bivariate normal distribution centered at $\mu_X = 0,$ with both measurements having a marginal variance of 1.0 and correlation ρ. We drew the 2 continuous measurements for each diseased subject from a bivariate normal distribution centered at $\mu_Y,$ also with both measurements having a marginal variance of 1.0 and correlation ρ; the values in μ_Y are directly determined from AUC_1 and AUC_2. To generate discrete outcomes, we first generated continuous outcomes as described above. Within each modality, we then assigned a value of 1 to outcomes in the lowest quintile of outcomes, a value of 2 outcomes between the first and the second quintiles, etc., with a value of 5 to outcomes above the fourth quintile.

For each simulation, we examined 3 fixed values f = {1/2, 1/3, 1/4} in our D^* statistic, as well as the value $\hat{f}$ which varied among simulations depending upon the values $\hat{p}_+, \hat{p}_0 \ and \ \hat{p}_-$ the proportions of the S_{ij} that were greater than, equal to, and less than zero, respectively. This value $\hat{f}$ is based upon the optimal value proposed by Irle and Klosener (1980) in the setting of independent S_{ij}. However, we do not claim this estimate to be necessarily optimal in our setting, but rather one that appears to offer our test nominal size and excellent power across a variety of settings. Tables 3.1 and 3.2 examine the size of the 2 competing permutation tests (both exact and asymptotic versions) for assessing a difference in AUC for 2 continuous or discrete diagnostic tests, while Tables 3.3 and 3.4 are the corresponding comparisons of the power for the 2 tests. Each setting in Tables 3.1–3.4 is defined by m, the number of non-diseased and number of diseased subjects, ρ, the within-subject correlation of the 2 diagnostic tests, and the values of AUC_1 and AUC_2. The size and power of each test were computed as the percentage of 5000 simulations in which the null hypothesis $AUC_\Delta = 0$ was rejected at a level of α = 0.05.We generated the permutation distribution of D^* in each simulation by generating 5000 random permutations of the diseased/nondiseased labels.

4. DISCUSSION

4.1. Numerical Studies of Operating Characteristics

Our extensive computer simulation results are presented as a series of 4 tables, which also contains specific computational details for the simulations. Based upon 5000 simulations, an approximate 95% confidence interval around a nominal size of 0.05 is (0.036, 0.064). Thus, we see in Table 3.1 that our permutation test is valid with any value of f when both AUC_1 and AUC_2 are less than 0.7, regardless of sample size or within-subject correlation. However, for higher AUC values, we see that the appropriate value of f will vary and if f is set too high, our test can actually have size above the desired level of 0.05. When $AUC_1 = AUC_2 = 0.80$, the results suggest that $f = 1/3$ may be appropriate, while when $AUC_1=AUC_2 = 0.90$, f should be set greater than 1/3, but less than 1/2. In contrast to using a fixed value of f, our proposed value $\hat{f}$ appears to work well (whether exact or asymptotic) among all settings, although it may produce a slightly conservative test at extreme AUC values. These findings continue to hold for discrete modalities, as demonstrated in Table 3.2. Furthermore, although there are slight variations in size between our test using $\hat{f}$ and that of Bandos et al (2005), none of the differences are significant, except with discrete diagnostic tests with AUC of 0.80 or higher. Therefore, we conclude overall that both approaches have similar size in most reasonable settings.

In Tables 3.3 and 3.4, we see that the power of the proposed permutation test using an appropriate value of f is comparable with the power of the test of Bandos et al (2005). In fact, the power of the proposed permutation test can be increased marginally above that of Bandos et al (2005) with some values of f. For example, when testing for a difference in $AUC_1 = 0.7$ and $AUC_2 = 0.8$ with continuous modalities using 40 diseased and 40 non-diseased subjects with intra-subject correlation $\rho = 0.5$, we see in Table 3.3 that the proposed test has power 0.471 when $f = 1/4$, as compared to power 0.420 for the test of Bandos et al (2005). Nonetheless, choosing a fixed value of f to increase power will be difficult in practice as the true AUC values of both diagnostic tests will be unknown.

Table3.1. *Comparison of size for proposed permutation test and that of Bandos et al (2005) for assessing a difference in AUC of 2 paired continuous diagnostic tests with within-subject correlation ρ in samples of n diseased and m non-diseased subjects. The proposed permutation test is applied with f = {1/4, 1/3, 1/2} as well as the value $\hat{f}$ presented in (2.6). The asymptotic size of the proposed permutation test is based upon $f = \hat{f}$*

ρ	**M**	AUC_1	AUC_2	**Proposed**					**Bandos et al(2005)**	
				Exact				**Asymptotic**	**Exact**	**Asymptotic**
				$f = 1/4$	$f = 1/3$	$f = 1/2$	$\hat{f}$			
0.0	40	0.6	0.6	0.053	0.053	0.054	0.054	0.053	0.057	0.055
		0.7	0.7	0.056	0.052	0.055	0.055	0.055	0.062	0.060
		0.8	0.8	0.055	0.042	0.040	0.045	0.046	0.061	0.063
		0.9	0.9	0.111*	0.044	0.013	0.039	0.039	0.050	0.051
	80	0.6	0.6	0.051	0.050	0.048	0.040	0.040	0.041	0.041
		0.7	0.7	0.047	0.042	0.030	0.036	0.036	0.038	0.039
		0.8	0.8	0.094*	0.055	0.021	0.035	0.036	0.047	0.045
		0.9	0.9	0.297	0.098*	0.009	0.041	0.043	0.046	0.047
0.5	40	0.6	0.6	0.062	0.062	0.061	0.055	0.056	0.060	0.060
		0.7	0.7	0.051	0.048	0.046	0.043	0.044	0.052	0.053
		0.8	0.8	0.057	0.044	0.033	0.037	0.036	0.046	0.041
		0.9	0.9	0.111*	0.047	0.013	0.040	0.040	0.042	0.041
	80	0.6	0.6	0.043	0.043	0.044	0.051	0.049	0.051	0.052
		0.7	0.7	0.047	0.047	0.049	0.043	0.039	0.058	0.058
		0.8	0.8	0.077*	0.051	0.032	0.041	0.045	0.048	0.050
		0.9	0.9	0.229*	0.089*	0.011	0.036	0.038	0.050	0.048

**Significantly above desired level of 0.05*

Table3.2. *Comparison of size for proposed permutation test and that of Bandos et al(2005) for assessing a difference in AUC of 2 paired discrete diagnostic tests with within-subject correlation ρ in samples of m diseased and m non-diseased subjects. The proposed permutation test is applied with f = {1/4, 1/3, 1/2} as well as the value $\hat{f}$ presented in (2.6). The asymptotic size of the proposed permutation test is based upon $f = \hat{f}$*

ρ	**M**	AUC_1	AUC_2	**Proposed**					**Bandos et al(2005)**	
				Exact				**Asymptotic**	**Exact**	**Asymptotic**
				$f = 1/4$	$f = 1/3$	$f = 1/2$	$\hat{f}$			
0.0	40	0.6	0.6	0.053	0.054	0.056	0.056	0.057	0.057	0.055
		0.7	0.7	0.056	0.052	0.054	0.056	0.057	0.064	0.060
		0.8	0.8	0.055	0.042	0.040	0.048	0.045	0.059	0.061
		0.9	0.9	0.111*	0.044	0.013	0.032	0.030	0.047	0.045
0.0	80	0.6	0.6	0.051	0.050	0.048	0.050	0.050	0.044	0.045
		0.7	0.7	0.047	0.042	0.030	0.034	0.039	0.037	0.037
		0.8	0.8	0.094*	0.055	0.021	0.038	0.038	0.042	0.044
		0.9	0.9	0.297	0.098*	0.009	0.028	0.030	0.036	0.034
0.5	40	0.6	0.6	0.062	0.062	0.061	0.063	0.065	0.061	0.061
		0.7	0.7	0.051	0.048	0.046	0.051	0.050	0.056	0.056
		0.8	0.8	0.057	0.044	0.033	0.045	0.046	0.046	0.048
		0.9	0.9	0.111*	0.047	0.013	0.035	0.035	0.042	0.040
0.5	80	0.6	0.6	0.043	0.043	0.044	0.045	0.044	0.051	0.044
		0.7	0.7	0.047	0.047	0.049	0.052	0.051	0.058	0.059
		0.8	0.8	0.077*	0.051	0.032	0.041	0.043	0.048	0.056
		0.9	0.9	0.229*	0.089*	0.011	0.028	0.030	0.050	0.036

**Significantly above desired level of 0.05*

Table3.3. *Comparison of power for proposed permutation test and that of Bandos et al (2005) for assessing a difference in AUC of 2 paired continuous diagnostic tests with within-subject correlation ρ in samples of m diseased and m non-diseased subjects. The proposed permutation test is applied with f = {1/4, 1/3, 1/2} as well as the value $\hat{f}$ presented in (2.6). The asymptotic size of the proposed permutation test is based upon $f = \hat{f}$*

ρ	**m**	AUC_1	AUC_2	**Proposed**					**Bandos et al(2005)**	
				Exact				**Asymptotic**	**Exact**	**Asymptotic**
				$f = 1/4$	$f = 1/3$	$f = 1/2$	$\hat{f}$			
0.0	40	0.6	0.7	0.197	0.184	0.166	0.190	0.190	0.201	0.195
		0.6	0.8	0.703	0.682	0.622	0.685	0.683	0.691	0.692
		0.7	0.8	0.295	0.257	0.179	0.235	0.237	0.239	0.237
		0.7	0.9	0.111*	0.861	0.731	0.842	0.837	0.834	0.834
		0.8	0.9	0.571*	0.448	0.208	0.374	0.378	0.363	0.361
	80	0.6	0.7	0.413	0.391	0.341	0.379	0.379	0.380	0.377
		0.7	0.8	0.615*	0.556	0.407	0.491	0.489	0.487	0.483
		0.8	0.9	0.913*	0.821	0.506	0.706	0.705	0.696	0.690
0.5	40	0.6	0.7	0.349	0.338	0.320	0.347	0.350	0.363	0.349
		0.6	0.8	0.911	0.900	0.875	0.919	0.917	0.921	0.923
		0.7	0.8	0.471	0.422	0.337	0.409	0.409	0.420	0.413
		0.7	0.9	0.987*	0.980	0.948	0.979	0.978	0.977	0.978
		0.8	0.9	0.758*	0.636	0.392	0.626	0.622	0.610	0.598
	80	0.6	0.7	0.611	0.593	0.555	0.607	0.609	0.619	0.612
		0.7	0.8	0.796*	0.743	0.630	0.718	0.717	0.716	0.712
		0.8	0.9	0.975*	0.935*	0.751	0.905	0.907	0.890	0.889

**Based upon test with supra-nominal size.*

Table3.4. *Comparison of power for proposed permutation test and that of Bandos et al(2005) for assessing a difference in AUC of 2 paired discrete diagnostic tests with within-subject correlation ρ in samples of m diseased and m non-diseased subjects. The proposed permutation test is applied with f = {1/4, 1/3, 1/2} as well as the value $\hat{f}$ presented in (2.6). The asymptotic size of the proposed permutation test is based upon $f = \hat{f}$*

ρ	**M**	AUC_1	AUC_2	**Proposed**					**Bandos et al(2005)**	
				Exact				**Asymptotic**	**Exact**	**Asymptotic**
				$f = 1/4$	$f = 1/3$	$f = 1/2$	$\hat{f}$			
0.0	40	0.6	0.7	0.197	0.184	0.166	0.176	0.180	0.177	0.178
		0.6	0.8	0.703	0.682	0.622	0.659	0.660	0.659	0.665
		0.7	0.8	0.295	0.257	0.179	0.217	0.213	0.225	0.218
		0.7	0.9	0.903*	0.861	0.731	0.807	0.807	0.817	0.813
		0.8	0.9	0.571*	0.448	0.208	0.358	0.356	0.365	0.363
0.0	80	0.6	0.7	0.413	0.391	0.341	0.358	0.363	0.370	0.368
		0.7	0.8	0.615*	0.556	0.407	0.462	0.460	0.454	0.453
		0.8	0.9	0.913*	0.821	0.506	0.669	0.669	0.665	0.669
0.5	40	0.6	0.7	0.349	0.338	0.320	0.328	0.331	0.349	0.339
		0.6	0.8	0.911	0.900	0.875	0.888	0.889	0.893	0.890
		0.7	0.8	0.471	0.422	0.337	0.389	0.388	0.390	0.385
		0.7	0.9	0.987*	0.980	0.948	0.970	0.971	0.967	0.968
		0.8	0.9	0.758*	0.636	0.392	0.562	0.558	0.553	0.543
0.5	80	0.6	0.7	0.611	0.593	0.555	0.568	0.569	0.569	0.570
		0.7	0.8	0.796*	0.743	0.630	0.664	0.668	0.671	0.665
		0.8	0.9	0.975*	0.935*	0.751	0.861	0.859	0.854	0.853

**Based upon test with supra-nominal size.*

5. Suggestion for Future Research

Due to the complicated correlation structure of the elements used in the statistic, we have not yet derived a theoretically optimal value of the value *f* necessary for the statistic. Although we have developed one possible value that appears in simulations to work well across many settings, we are continuing research into deriving a formula for *f* that will maximize power in all settings. We are also seeking to use our permutation test to generate a confidence interval for AUC_Δ as a complement to the hypothesis test. Furthermore, unlike the test of Bandos et al(2005), our proposed test does not require diagnostic tests that are measured on identical scales and thus may prove to be more powerful in settings in which the diagnostic test values are skewed; we are pursuing this conjecture in current research.

Note that the semi-parametric regression model of Dodd and Pepe (2003) uses the U_{ijk} defined in Section 2 in a generalized estimating equation (GEE) and yields a standardized value of the nonparametric estimate of AUC_Δ when using an independence working covariance structure of the U_{ijk}. Although, inference for AUC_Δ could be based upon the sandwich (robust) variance estimator of GEE methods, Braun and Feng (2001) showed that score and Wald tests using this approach are known to have liberal sizes in smaller sample sizes and developed a permutation test as an alternative to large-sample theory. As a result, we are also pursuing use of our permutation approach to the methods of Dodd and Pepe (2003) that would lead to exact inference for semi-parametric estimates of difference in AUCs.

Acknowledgement

I wish to appreciate my departmental typist in the person of Miss Nwogbu Divine for type setting this work before it was set for submission.

References

Bamber D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology* 12, 387-415.

BANDOS, A. I., ROCKETTE, H. E. AND GUR, D. (2005). A permutation test sensitive to differences in areas for comparing ROC curves from a paired design. *Statistics in Medicine* 24, 2873–2893.

BRAUN, T. M. AND FENG, Z. (2001). Optimal permutation tests for the analysis of group randomized trials. *Journal of the American Statistical Association* 96, 1424–1432.

CAMPBELL, G. (1994). General methodology I: advances in statistical methodology for the evaluation of diagnostic and laboratory tests. *Statistics in Medicine* 13, 499–508.

COAKLEY, C. W. AND HEISE, M. A. (1996). Versions of the sign test in the presence of ties. *Biometrics* 52, 1242–1251.

DELONG, E. R., DELONG, D. M. AND CLARKE-PEARSON, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44, 837–845.

DODD, L. E. AND PEPE, M. S. (2003). Semiparametric regression for the area under the receiver operating characteristic curve. *Journal of the American Statistical Association* 98, 409–417.

DORFMAN, D. D., BERBAUM, K. S. AND METZ, C. E. (1992). Receiver operating characteristic rating analysis: generalization to the population of readers and patients with the jackknife method. *Investigative Radiology* 27,723–731.

Hanley, J.A., McNeil, B.J. (1982). The meaning and use of the Area under Receiver Operating Characteristic (ROC) Curve. *Radiology* 143, 29-36.

Hanley, J.A., Hajian-Tilaki, K.O. (1997). Sampling variability of nonparametric estimates of the areas under receiver operating characteristic curves: an update. *Academic Radiology* 4, 49-58.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics* 19, 293–325.

IRLE, A. AND KLOSENER, K.-H. (1980). Note on the sign test in the presence of ties. *Annals of Statistics* 8, 1168–1170.

MOISE, A., CLEMENT, B. AND RAISSIS, M. (1988). A test for crossing receiver operating characteristic (ROC)curves. *Communications in Statistics—Theory and Methods* 17, 1985–2003.

Noether GE. Elements on Nonparametric Statistics. *Wiley & Sons Inc.:* New York 1967.

Obuchowski, N.A. (1994). Computing sample size for receiver operating characteristics studies. *Investigative Radiology* 29, 238-243.

PUTTER, J. (1955). The treatment of ties in some nonparametric tests. *Annals of Mathematical Statistics* 26, 368–386.

ROMANO, J. P. (1990). On the behavior of randomization tests without a group invariance assumption. *Journal of the American Statistical Association* 85, 686–692.

SONG, H. H. (1997). Analysis of correlated ROC areas in diagnostic testing. *Biometrics* 53, 370–382.

VENKATRAMAN, E. AND BEGG, C. B. (1996). A distribution-free procedure for comparing receiver operating characteristic curves from a paired experiment. *Biometrika* 83, 835–848.

Wieand, H.S., Gail, M.M., Hanley, J.A. (1983). A nonparametric procedure for comparing diagnostic tests with paired or unpaired data. *I.M.S. Bulletin* 12, 213-214.

Citation: *U. Okeh & Mbegbu JI, "An Alternative Nonparametric Method of Assessing a Difference in the Areas under the Curves (Aucs) for Paired Data", International Journal of Research Studies in Biosciences (IJRSB), vol. 6, no. 7, pp. 22-30, 2018. http://dx.doi.org/10.20431/2349-0365.0607004*