



Measure of Association Dependent on Sensitivity and Specificity of Diagnostic Screening Tests in a Study Population

¹Okeh UM, ²Okoro CN

¹Department of Industrial Mathematics and Health Statistics, David Umahi Federal University of Health Sciences Uburu, Ebonyi State, Nigeria.

²Department of Industrial Mathematics and Applied Statistics, Ebonyi State University Abakaliki, Ebonyi State, Nigeria.

Correspondence: Okeh UM. ¹Department of Industrial Mathematics and Health Statistics, David Umahi Federal University of Health Sciences Uburu, Ebonyi State, Nigeria.; okebucheckwumarius@gmail.com; +2348034304278

Article history: Received 16 January 2024, Reviewed 8 March 2024, Accepted for publication 12 March 2024

This is an open access journal and articles are distributed under the terms of the Creative Commons Attribution License (Attribution, Non-Commercial, ShareAlike[®] 4.0) - (CC BY-NC-SA 4.0) that allows others to share the work with an acknowledgement of the work's authorship and initial publication in this journal.

How to cite this article:

Okey UM, Okoro CN. Measure of Association Dependent on Sensitivity and Specificity of Diagnostic Screening Tests in a Study Population. The Nigerian Health Journal 2024; 24(1):1099 – 1108. Doi: <https://www.doi.org/10.60787/tnhj-24-1-776>

Abstract

Background: The traditional odds ratio and relative risk cannot strictly speaking properly and validly be used because in them, the number of subjects testing positive and negative among subjects known or believed not to have a condition in nature usually are not known and hence the total number of subjects testing positive and negative are not also completely known. Paper proposes, develops, and presents a measure of the strength of association between test results and condition in a population, by using sensitivity and specificity of diagnostic screening tests that are independent of the population under study.

Method: A retrospective study was carried out. The proposed measure of association which always lies between -1 and 1 inclusively enables the researcher to determine not only if an association exists between test results and state of nature or condition in a population and if such an association exists, whether it is positive and direct or negative and indirect thereby giving the measure an advantage over and above the traditional odds ratio method.

Result: The proposed method is easier to interpret and understand than those from the traditional odd ratio approach. For comparison and completeness, it also develops modified sample estimate of the traditional odd ratio and its sample variance from observable sample data. The likelihood ratio showed that the test is very informative.

Conclusion: The proposed measure of association is shown to be at least as efficient and hence as powerful as the traditional odds ratio. The modified traditional odds ratio performs better than the traditional odds ratio. The likelihood ratios are at least as efficient as the proposed method but better than the traditional odd ratio.

Keywords: sensitivity, specificity, traditional odds ratio, diagnostic screening tests, state of nature or condition, clinical trial.

Introduction

The purpose of screening is to identify people in an apparently healthy population who are at higher risk of a health problem or a condition, so that an early treatment or intervention can be offered. This, in turn, may lead to better health outcomes for some of the screened individuals.¹ Clinicians practicing evidence-based medicine are familiar with the concepts of sensitivity and

specificity, defined as the probability of a positive test given that the person has the target condition, and the probability of a negative test given that the person does not have the condition, respectively.² These sensitivity and specificity are considered to be characteristics of the test as intrinsic accuracy measures, and independent from the characteristics of the population. The rationale for this assumption relates to the mathematical

calculation of these measures from a classic 2×2 diagnostic table. Some screening programmes choose a highly sensitive threshold for the screening test at the expense of lower specificity. This means that there will be very few false negatives but more false positives. In these circumstances, all the positive results are then investigated with a further test with high specificity to exclude the false positives.¹

Meanwhile, the traditional odds ratio and relative risk are often used as measures of the strength of association between a predisposing or antecedent factor and condition in controlled comparative studies because these two measures are invariant under the three commonly used study methods to generate the data, namely the cross-sectional, prospective and retrospective study design.^{3,4,5} However, in diagnostic screening tests and clinical trials these measures cannot strictly speaking, properly and validly be used. This is because in these tests the number of subjects testing positive among subjects known or believed not to have a condition in nature and the number of subjects testing negative among subjects known or believed to have a condition in nature usually are not known and hence the total number of subjects testing positive and negative respectively are not also completely known.^{6,7}

These values which are contained in the expressions used in the estimation of the traditional odds ratio, relative risk, their associated standard errors and test statistics for significance can therefore not be properly and directly used in calculations.^{8,9} The false rates and proportion of the population expected to test positive should ideally be factored in and reflected in measures used to assess association in diagnostic screening tests.^{3,10,11} Unfortunately, the utility of these three indices is seriously limited by the fact that their formulations contain prevalence rates of the conditions in a population of interest, values that often are not known for many conditions.^{12,13,14}

This paper proposes, develops and presents a statistical measure of the strength of association between diagnostic screening test results and state of nature or condition in a population that depends on only the sensitivity and specificity of screening tests and clinical trials that are estimable using only observable ample values. The proposed measure does not also require knowledge of the prevalence rate of a condition in the population before its estimation.

Results

Method

Study population

A retrospective analysis will be performed on the results of screening tests which was conducted at a named Hospital or clinic. The study period should span for a specified period and a given number of participants which may include both outpatients and inpatients were required to undergo through their initial screening tests. The inclusion criteria depend on the nature of the screening procedures.

Data collection, calculations and test of hypothesis

All the clinical characteristics of the participants and other data will be collected from the clinical laboratory information system (LIS) including the test results, gender, age, reasons for testing, clinical departments, etc. We will calculate the sensitivity, specificity, the sample estimate of the sensitivity and specificity of the screening test, estimated variance, Chi-square test statistic of the number of associations between screening test results and state of nature or condition will be calculated and test of the null hypothesis of the non-significance of association will be carried out. A test statistic expressed in terms of the sensitivity and specificity of the screening test for testing the significance of the proposed measure of association will be provided. Sample data will be used to illustrate the proposed method and also used to compare with the traditional odds ratio measure of association.

Proposed Method

Suppose a research Scientist or Clinician collects a random sample of n_1 subjects known or believed to actually have a certain condition in nature in a population and also collects a second random sample of n_2 subjects from the same population known or believed not to actually have the condition in nature, giving a total random sample of size $n = n_1 + n_2$ subjects to be studied. Research interest is to determine through a diagnostic screening test or clinical trial whether or not each of the sampled subjects actually tests positive or negative to the condition in nature in the population. Let B be the event that a randomly selected subject of the population actually has a condition in nature and \bar{B} be the event that the randomly selected subject, does not actually have the condition in nature. Also let A and \bar{A} be respectively the events that the randomly selected subject tests and does not test positive to the condition in the screening test when screened and tested. The results of such a screening test may be presented in a fourfold table in Table (1).

Table 1: Format for presentation of results in diagnostic screening test

Test Result	State of Nature or condition		Total
	Present (B)	Absent (\bar{B})	
Positive (A)	$n_{11} = F^{++}$	$n_{12} = F^{-+}$	$n_{1.} = F^{+}$
Negative (\bar{A})	$n_{21} = F^{+-}$	$n_{22} = F^{--}$	$n_{2.} = F^{-}$
Total	$n_{.1}$	$n_{.2}$	$n_{..} (= n)$

In Table (1) out of $n_{.1}$ subjects known or believed to actually have a condition in nature, n_{11} subjects test positive and n_{21} test negative. Similarly out of $n_{.2}$ subject known or believed not to actually have a condition in nature, n_{12} test positive and n_{22} test negative. Of the $n_{.1}$ subjects sampled, n_{11} subjects test positive while n_{21} subjects test negative. However, as noted above in diagnostic screening test results only n_{11} and n_{22} subjects which are often of primary interest to the researcher are usually observed and known. The values n_{12} , the number of subjects testing positive among those known or believed not to actually have the condition in nature and n_{21} , the number of subjects testing negative among those known or believed to actually have the condition in nature usually are not known. Hence the marginal frequencies or totals $n_{.1}$ and $n_{.2}$ usually are not completely known. Hence these unknown values may not properly and validly directly be used in calculations. The present proposed measure of association is based on the expectation that if a diagnostic screening test or clinical trial is a good one, then the sum of the proportion of subjects testing positive among the population of subjects known or believed to actually have a condition in nature and the proportion of subjects testing negative among the population of subjects known or believed not to actually have the condition in nature would be much larger than the sum of the proportion of subjects known or believed not to actually have the condition in nature and the proportion of subjects known or believed to actually have the conditions in nature. Similarly if the screening test is a poor one then one would expect the converse result to be obtained, that is, the sum of the proportion of subjects testing positive among the population of

$$\pi = (Se + Sp) = ((1 - Se) + (1 - Sp)) = 2(Se + Sp - 1)$$

To use Equation 1-5 to develop the proposed measure of association we may let

subjects known or believed not to actually have a condition in nature and the proportion of subjects known or believed to actually have the condition in nature would be larger than the sum of the proportion of subjects known or believed to actually have a condition in nature and the proportion of subjects known or believed not to actually have the condition in nature. Now using conditional probabilities of events A and B, the proportion of subjects testing positive among the proportion of subjects known or believed to actually have a condition in nature which is the sensitivity, Se of the screening test [2] is

$$P(A|B) = \frac{P(AB)}{P(B)} = Se \quad (1)$$

The proportion of subjects testing negative among the population of subjects known or believed not to actually have a condition in nature which is also the specificity, Sp of the screening test is:

$$P(\bar{A}|\bar{B}) = \frac{P(\bar{A}\bar{B})}{P(\bar{B})} = Sp \quad (2)$$

Also, the proportion of subjects testing negative among the population of subjects known or believed to actually have a condition in nature is

$$P(\bar{A}|B) = 1 - P(A|B) = 1 - Se \quad (3)$$

And the proportion of subjects known or believed not to actually have a condition in nature is

$$P(A|\bar{B}) = 1 - P(\bar{A}|\bar{B}) = 1 - Sp \quad (4)$$

Hence the difference between the proportions of subjects testing positive among the subjects known or believed to actually have a condition in nature or testing negative among the subjects known or believed not to actually have the condition in nature and the proportion of subjects testing positive among the subjects known or believed not to actually have a condition in nature or testing negative among the subjects known or believed to actually have a condition in nature is

$$(5)$$

$$v_i = \begin{cases} 1, & \text{if the } i\text{th sampled and screened subject either tests positive and is} \\ & \text{known or believed to actually have the condition in nature or tests} \\ & \text{negative and is known or believed not to actually have the condition in nature} \\ -1, & \text{if the } i\text{th sampled and screened subject either tests positive and is} \\ & \text{known or believed not to actually have the condition in nature or tests} \\ & \text{negative and is known or believed to actually have the condition in nature} \end{cases} \quad (6)$$

for $i = 1, 2, \dots, n$.

Let

$$\pi^+ = P(u_i = 1); \pi^- = P(u_i = -1) \quad (7)$$

where

$$\pi^+ + \pi^- = 1 \quad (8)$$

Define

$$w = \sum_{i=1}^n u_i \quad (9)$$

Now the expected value or mean and the variance of u_i are respectively.

$$E(u_i) = \pi^+ - \pi^-; \text{var}(u_i) = \pi^+ + \pi^- - (\pi^+ - \pi^-)^2 = 1 - (\pi^+ - \pi^-)^2 \quad (10)$$

Also, the expected value or mean of w is

$$E(w) = \sum_{i=1}^n E(u_i) = n(\pi^+ - \pi^-) \quad (11)$$

The corresponding variance of w is from Equation (10)

$$\text{var}(w) = \sum_{i=1}^n \text{var}(u_i) = n(\pi^+ + \pi^- - (\pi^+ - \pi^-)^2) = n(1 - (\pi^+ - \pi^-)^2) \quad (12)$$

Now π^+ is the proportion of subjects or the probability that a randomly selected subject from the population either tests positive and is actually positive in nature or tests negative and is actually negative in nature; while π^- is the proportion of subjects or the probability that a randomly selected subject from the population either tests positive and is negative in nature or test negative and is positive in nature. Their sample estimates using the frequencies of Table(1) are respectively

$$\hat{\pi}^+ = \frac{F^{++} + F^{-+}}{n} = \frac{n_1 \left(\frac{F^{++}}{n_1} \right) + n_2 \left(\frac{F^{-+}}{n_2} \right)}{n} = \frac{n_1 \hat{S}e + n_2 \hat{S}p}{n} \quad (13)$$

And

$$\hat{\pi}^- = \frac{F^{+-} + F^{--}}{n} = \frac{(n_2 - F^{--}) + (n_1 - F^{+-})}{n} = \frac{n_1(1 - \hat{S}e) + n_2(1 - \hat{S}p)}{n} \quad (14)$$

Where as shown in Table (1)

$$F^{++} = n_{11}; F^{+-} = n_{12}; F^{-+} = n_{21}; F^{--} = n_{22} \quad (15)$$

The marginal frequencies of the screening test results are:

$$F^+ = F^{++} + F^{+-} = n_{11} + n_{12} = n_1; F^- = F^{-+} + F^{--} = n_{21} + n_{22} = n_2 \quad (16)$$

Here $\hat{S}e$ and $\hat{S}p$ are respectively the sample estimates of the sensitivity and specificity of the screening test. The sample estimates of the difference between these probabilities or proportions namely sensitivity, S_e and specificity, S_p is

$$\hat{\pi} = \hat{\pi}^+ - \hat{\pi}^- = \frac{(F^{++} + F^{-+}) + (F^{+-} + F^{--})}{n} = \frac{w}{n} \quad (17)$$

Note that the sensitivity, S_e and the specificity, S_p of a screening test can be estimated directly using only the observed sample data of Table (1) obtainable and available from the screening test as respectively.

$$\hat{Se} = \frac{F^{++}}{n_1} = \frac{n_{11}}{n_1}; \hat{Sp} = \frac{F^{--}}{n_2} = \frac{n_{22}}{n_2} \quad (18)$$

The variance of $\hat{\pi}$ is estimated from Equation 12 as

$$\text{var}(\hat{\pi}) = \text{var}(\hat{\pi}^+ - \hat{\pi}^-) = \frac{\text{var}(w)}{n^2} = \frac{1 - (\hat{\pi}^+ - \hat{\pi}^-)^2}{n} \quad (19)$$

When expressed in terms of estimated sensitivity and specificity of the screening test Equation (17) becomes:

$$\hat{\pi} = \hat{\pi}^+ - \hat{\pi}^- = \frac{w}{n} = \frac{n_1(2\hat{Se}-1) + n_2(2\hat{Sp}-1)}{n} \quad (20)$$

With corresponding variance is obtained from Equation (19) as

$$\text{var}(\hat{\pi}) = \text{var}(\hat{\pi}^+ - \hat{\pi}^-) = \frac{1 - (\hat{\pi}^+ - \hat{\pi}^-)^2}{n} = \frac{1 - (n_1(2\hat{Se}-1) + n_2(2\hat{Sp}-1))^2}{n} \quad (19)$$

A null hypothesis that is usually of research interest is whether or not the diagnostic screening test or clinical trial is a good one, which is whether an association exists, between test results and condition of research interest that is actual state of nature in the population. As noted above, if there is no association between screening test results and state of nature or condition in the population,

then the proportion $\pi = \pi^+ - \pi^-$; that is, the difference between the sum of the proportion of subjects who either test positive among the population of subjects

known or believed to actually have a condition in nature or the proportion of subjects testing negative among subjects in the population known or believed not to actually have the condition in nature and the sum of the proportion of subjects who either test positive among the population known or believed not to actually have the condition in nature or the proportion of subjects who test negative among the population of subjects known or believed to actually have the condition in nature would be expected to be zero. Symbolically, a more general null hypothesis would be

$$H_0 : \pi = \pi^+ - \pi^- \leq \theta_0 \text{ versus } H_1 : \pi = \pi^+ - \pi^- > \theta_0 (0 \leq \theta_0 \leq 1). \quad (22)$$

The null hypothesis, H_0 of Equation (22) may be tested using the test statistic:

$$\begin{aligned} \chi^2 &= \frac{(\hat{\pi} - \theta_0)^2}{\text{var}(\hat{\pi})} = \frac{((\hat{\pi}^+ - \hat{\pi}^-) - \theta_0)^2}{\text{var}(\hat{\pi}^+ - \hat{\pi}^-)} = \frac{n((\hat{\pi}^+ - \hat{\pi}^-) - \theta_0)^2}{1 - (\hat{\pi}^+ - \hat{\pi}^-)^2} \\ &= \frac{((n_1(2\hat{Se}-1) + n_2(2\hat{Sp}-1) - \theta_0))^2}{n(1 - (n_1(2\hat{Se}-1) + n_2(2\hat{Sp}-1) - \theta_0))^2} = \frac{((n_1(2\hat{Se}-1) + n_2(2\hat{Sp}-1) - \theta_0))^2}{n - (n_1(2\hat{Se}-1) + n_2(2\hat{Sp}-1))^2} \end{aligned} \quad (23)$$

Which under the null hypothesis H_0 has approximately the Chi-square distribution with 1 degree of freedom for sufficiently large sample size $n = n$. the null hypothesis H_0 of Equation (22) is rejected at the α level of significance if

$$\chi^2 \geq \chi_{1-\alpha;1}^2 \quad (24)$$

Otherwise H_0 is accepted.

Results/Illustrative example

We here use the following data to illustrate the proposed method. A clinician collected a random sample of 98 adult males from a certain population, 12 of whom are suspected to have Prostate cancer and 86 of whom are believed not to have the disease. The Clinicians interest

is to confirm through a diagnostic screening test whether or not each of the sampled adult males is actually Prostate cancer positive or negative. The results of the screening test are presented in Table (2).

Table 2: Results of Prostate Cancer Screening test adult males in a population

		Histologic Diagnosis
--	--	----------------------

Clinical Diagnosis	Present(B)	Absent (\bar{B})	Total (n_i)
Prostate Cancer positive (+)	$4(n_{11} = F^{++})$	$2(n_{12} = F^{+-})$	$6(n_{1.} = F^+)$
Prostate Cancer negative (-)	$8(n_{21} = F^{-+})$	$84(n_{22} = F^{--})$	$92(n_{2.} = F^-)$
Total ($n_{.j}$)	$12(n_{.1})$	$86(n_{.2})$	$98(n_{..} = n)$

Now from Table (2) we have that the sample estimate of the sensitivity and specificity of the screening test are respectively from Equation (18)

$$\hat{S}_e = \frac{F^{++}}{n_{.1}} = \frac{n_{11}}{n_{.1}} = \frac{4}{12} = 0.333; \hat{S}_p = \frac{F^{--}}{n_{.2}} = \frac{n_{22}}{n_{.2}} = \frac{84}{86} = 0.977.$$

These results show that the screening test is low in sensitivity but high in specificity. Now from Equations 13 and 14 the sample estimates of π^+ and π^- are respectively.

$$\hat{\pi}^+ = \frac{F^{++} + F^{-+}}{n} = \frac{n_{11} + n_{22}}{n} = \frac{4 + 84}{98} = 0.898.$$

And

$$\hat{\pi}^- = \frac{F^{+-} + F^{--}}{n} = \frac{(n_{12} - n_{22}) + (n_{11} - n_{12})}{n} = \frac{(86 - 84) + (12 - 4)}{98} = \frac{(2) + (8)}{98} = 0.102.$$

Hence from Equations 17 and 20, we have that $\hat{\pi} = \hat{\pi}^+ - \hat{\pi}^- = 0.898 - 0.102 = 0.796$ with estimated variance obtained from

$$\text{Equation 21 as } \text{var}(\hat{\pi}) = \text{var}(\hat{\pi}^+ - \hat{\pi}^-) = \frac{1 - (0.796)^2}{98} = 0.004.$$

Hence the test statistic of no association between screening test results and state of nature or condition (Prostate cancer) H_0 of Equation 22 is from Equation 23 with $\theta_0 = 0$ as

$$\chi^2 = \frac{(\hat{\pi})^2}{\text{var}(\hat{\pi})} = \frac{(0.796)^2}{0.004} = 158.500 (P\text{-value} = 0.0000)$$

which with 1 degree of freedom is highly statistically significant at the 5 percent significance level ($\chi_{0.95;1}^2 = 3.841$) leading to the rejection of the null hypothesis H_0 of Equation (22). We may therefore conclude that there is a strong degree of association between screening test results and

state of nature or condition (presence of Prostate cancer in the population). Also since $\hat{\pi} = \hat{\pi}^+ - \hat{\pi}^- = 0.796$ is positive, the association between the screening test results and state of nature namely, presence of Prostate cancer in the population is positive and direct. It would be instructive to compare the present result with what would have been obtained if we had used the traditional odds ratio to analyze the data of Table (2) in spite of odds ratio's short comings as already pointed out above when used in the analysis of diagnostic screening test results. The sample estimate of the traditional odds ratio for the data of Table 2 is

$$o = \frac{n_{11} \cdot n_{22}}{n_{21} \cdot n_{12}} = \frac{(4)(84)}{(2)(8)} = 21.00$$

This means that for everyone adult male who is found to have Prostate cancer among those tested and erroneously informed that they are free of the disease, 21 adult males from the population among those tested and found to have Prostate cancer would be expected to be correctly so informed that they are Prostate cancer positive. This is clearly more difficult to interpret and understand than the simple information conveyed by the simple difference in rates, $\hat{\pi} = \hat{\pi}^+ - \hat{\pi}^- = 0.898 - 0.102 = 0.796$, namely that the

proportion of adult males in the population testing positive among adult males who have Prostate cancer or testing negative among adult males who do not have Prostate cancer is about 79.68 higher than the proportion of adult males testing positive among those who do not have Prostate cancer or testing negative among adult males who have the disease in the population. In other words, the number of adult males who test positive among those who actually have Prostate cancer or negative among those who do not have the disease is about 79.6 percent more than the

number of adult males who test positive among the adult males known not to have the disease or who test negative among those known or believed to actually have Prostate cancer in the population. Thus, the rate π

$$Se(o) = O \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} = 21.00 \sqrt{\frac{1}{4} + \frac{1}{2} + \frac{1}{8} + \frac{1}{84}} = ($$

This measure of the error of O, the sample estimate of traditional odds ratio namely 19.782 is clearly much larger than the error of only $Se(\hat{\pi}) = \sqrt{0.004} = 0.063$ of the estimated value of $\hat{\pi}$ for our sample data of Table (2). These results show that the proposed measure of association π is relatively more efficient than the traditional odds ratio measure of association, w. The Chi-square test statistic for the significance of the traditional odds ratio is:

$$\chi^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_1 n_2 n_{.1} n_{.2}} = \frac{98((4)(84) - (2)(8))^2}{(6)(92)(12)(86)} = 17.616(P$$

which with 1 degree of freedom is also statistically significant at the 5 percent significance level ($\chi^2_{0.95;1} = 3.841$),

again leading to a rejection of the null hypothesis, H_0 of no association between screening test results and presence of Prostate cancer in the population. However, although the proposed method and the traditional odds ratio method here both lead to a rejection of the null hypothesis, nevertheless the relative sizes of the calculated Chi-square values suggest that the traditional odds ratio method is less efficient and likely to lead to an acceptance of a false null hypothesis (Type II Error) more frequently and hence is likely to be less powerful than the proposed method. In any case, note that the expression for the traditional population odds ratio expressed in terms of conditional probabilities of events A and B using the notations of Table (1) is

obtained using the proposed method is relatively easier to interpret and understand than the traditional odds ratio, w the standard error of the estimated odds ratio O is

$$w = \frac{P(A|B) \cdot P(A|\bar{B})}{P(\bar{A}|B) \cdot P(\bar{A}|\bar{B})}$$

Its sample estimate is $\hat{w} = o$

$$o = \frac{F^{++} \cdot F^{--}}{F^{+-} \cdot F^{-+}} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

Its sample variance is

$$\text{var}(o) = o^2 \left(\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}} \right)$$

And the accompanied test statistic

$$\chi^2 = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_1 n_2 n_{.1} n_{.2}}$$

Notwithstanding the above rather heuristic calculation cannot strictly speaking be really directly and properly evaluated without modifications from the results of a diagnostic screening test or clinical trial and used to analyze the data of Table (2). This is because as already

noted above, $P(A|\bar{B})$ and $P(\bar{A}|B)$, the sample values

n_{12}, n_{21} and their derivatives $n_{.1}$ and $n_{.2}$ are not obtainable from the results of a diagnostic screening test or clinical trial. However, algebraic manipulation would easily enable one reformulate Equation (25) to obtain an expression for the traditional population odds ratio, w that could enable its application and use with sample observation obtained from diagnostic screening tests or clinical trials and calculate the sample estimates of the sensitivity and specificity of the screening test as well as an appropriate test statistic. Thus Equation (25) is actually equivalent to

$$w = \frac{P(A|B) \cdot P(\bar{A}|\bar{B})}{(1 - P(A|\bar{B})) \cdot (1 - P(\bar{A}|B))} = \frac{Se \cdot Sp}{(1 - Se)(1 - Sp)} \quad (29)$$

Whose sample estimate using the cell frequencies of Table (1) is:

$$\hat{w} = o = \frac{\hat{Se} \cdot \hat{Sp}}{(1 - \hat{Se}) \cdot (1 - \hat{Sp})} = \frac{(n_{11}/n_{.1}) \cdot (n_{22}/n_{.2})}{(1 - n_{11}/n_{.1})(1 - n_{22}/n_{.2})} = \frac{n_{11} \cdot n_{22}}{(n_{.1} - n_{11})(n_{.2} - n_{22})}$$

The corresponding modified sample estimate of the variance of the traditional sample odds ratio, O is

$$\text{var}(o) = o^2 \left(\frac{1}{n_{11}} + \frac{1}{(n_{.2} - n_{22})} + \frac{1}{(n_{.1} - n_{11})} + \frac{1}{n_{22}} \right) = o^2 \left(\frac{1}{n_{.1} \cdot \hat{Se}} + \frac{1}{n_{.1} \cdot (1 - \hat{Se})} \right)$$

The modified test statistic expressed in terms of sample data or values that are obtainable in a diagnostic

screening test or clinical trial for testing the statistical significance of the modified odds ratio is

$$\chi^2 = \frac{n(n_{11}n_{22} - (n_{.1} - n_{11})(n_{.2} - n_{22}))^2}{n_{.1}n_{.2} \cdot (n_{.1} - n_{11} + n_{22})(n_{.2} - n_{22} + n_{11})} = \frac{1}{\hat{S}e +}$$

Which under the null hypothesis, H_0 has approximately the Chi-square distribution with 1 degree of freedom for sufficiently large sample size $n = n_{\dots}$. The null hypothesis, H_0 is rejected at the α level of significance if Equation 24 is satisfied, otherwise the null hypothesis, H_0 is accepted.

However, the result obtained using the proposed method can be compared using another method of measuring the accuracy of diagnostic test called likelihood ratio. Here, we can compute the sample data and interpret our result in terms of the likelihood ratio. Recall that likelihood ratios can be used to update the pre-test probability of disease using Bayes' theorem, once the test result is known. The updated probability is referred to as the post-test probability. For a test that is informative, the post-test probability should be higher than the pre-test probability if the test result is positive, whereas the post-test probability should be lower than the pre-test probability if the test result is negative. The positive likelihood ratio describes how many times more likely positive test results were in the diseased group compared to the non-diseased group. The positive likelihood ratio is given as

$$LR+ = \frac{sen}{(1-spec)} = \frac{a/a+c}{b/b+d} = \frac{F^{++}}{n_{.1}} \div \left[1 - \frac{F}{n} \right]$$

The positive likelihood ratio is here greater than 1 showing that the test is informative. The negative likelihood ratio describes how many times less likely negative test results were in the diseased group compared to the non-diseased group. The negative likelihood ratio is given as

$$LR- = \frac{1-sen}{spec} = \frac{c/a+c}{d/b+d} = 1 - \frac{4}{12} \times \frac{1}{\frac{84}{86}} = 0.683$$

The negative likelihood ratio is here less than 1 showing that the test is informative. These results show that the screening test is low in negative likelihood ratio but high in positive likelihood ratio meaning that there is association between screening test results and presence of Prostate cancer in the population in which case significant relationship exists hence the test is informative.

Discussion

The traditional odds ratio (TOR) as a measure of test performance combines the strengths of sensitivity and specificity, as prevalence independent indicators, with the advantage of accuracy as a single indicator. These characteristics lend the TOR particularly useful for comparing tests whenever the balance between false negative and false positive rates is not of immediate importance. These features are also highly convenient in systematic reviews and meta-analyses. In decisions on the introduction of a test in clinical practice, we are aware that the actual balance between the true positive rate and false positive rate often matters.¹⁵ Whenever false positives and false negatives are weighted differentially, both the prevalence and the conditional error rates of the test have to be taken into consideration to make a balanced decision. In these cases, the TOR is less useful, as it does not distinguish between the two types of diagnostic mistake. Also, the relative sizes of the calculated Chi-square values as seen suggest that the traditional odds ratio method is less efficient and likely to lead to an acceptance of a false null hypothesis (Type II Error) more frequently and hence is likely to be less powerful than the proposed method. Meanwhile, since ruling-out or ruling-in of the target condition is the primary intended use of a test, conditional indicators, or accuracy measures such as sensitivity and specificity was used in the proposed method.

As all available measures of test performance, the TOR of a test is unlikely to be a test-specific constant. Its magnitude likely depends on the spectrum of disease as well as on pre-selection using other tests.^{16,17} Despite this universal caveat for indicators of diagnostic tests, we feel that a more systematic use of the odds ratio in diagnostic research can contribute to more consistent applications of diagnostic knowledge. This called for the proposed method in this paper. Some may object that there are already too many indicators of test performance. With such an abundance of choices, there is little need for yet another statistic. This may be true, but it is hard to see how the selection can or should be produced. Each of the indicators serves a different purpose. Sensitivity and specificity are expressions of the conditional hit rates of the test. Predictive values or posterior probabilities are the numbers that are most salient for clinical practice. The so-called likelihood ratios come in handy for comparing the diagnostic content of multiple possible test results and for transforming those into posttest probabilities. Among those helpful indicators, the proposed method is comparatively more efficient and powerful than other indicators of diagnostic test accuracies.

Conclusion

We have in this paper proposed, developed and presented a statistical method for measuring the strength of association between test results and state of nature or condition in a population exposed to a diagnostic screening test or clinical trial. The proposed measure is based on only the sensitivity and specificity of the screening test which are independent of the population of interest and estimated using only observed sample values obtainable in a diagnostic screening test or clinical trial. The proposed measure which always lies between -1 and 1 inclusively could be used to establish whether an association is strong and direct, strong and indirect or zero and whether or not the association is statistically significant. Estimates of the variance of the proposed measure of association, π expressed in terms of obtainable sample data in a diagnostic screening test or clinical trial was provided. A test statistic expressed in terms of the sensitivity and specificity of the screening test for testing the significance of the proposed measure of association was also provided. Sample data were used to illustrate the proposed method which was shown to be at least as efficient and powerful as the traditional odds ratio measure of association which strictly speaking was shown not to be appropriate to use for the analysis of results obtained in a diagnostic screening test or clinical trial in which some of the required sample data are not obtainable.

Furthermore, the proposed measure of association is shown to be relatively easier to interpret and understand than the traditional odds ratio measure of association. Although the proposed method and the traditional population odds ratio method here both lead to a rejection of the null hypothesis, nevertheless the relative sizes of the calculated Chi-square values suggest that the traditional population odds ratio method is less efficient and likely to lead to an acceptance of a false null hypothesis (Type II Error) more frequently and hence is likely to be less powerful than the proposed method. The implication of this finding is that the chances of making wrong conclusion that there is no association between screening test results and state of nature or condition (presence of Prostate cancer in the population). Another setback of the traditional population odds ratio is that if it is expressed in terms of the conditional probabilities of events A and B, its sample estimates, variance, as well as test statistic, their calculations cannot strictly speaking be really directly and properly evaluated using sample data without modifications from the results of a diagnostic screening test or clinical trial. This is because the conditional

probabilities, the sample values and their derivatives are not obtainable from the results of a diagnostic screening test or clinical trial. This means that it is only by algebraic manipulation that will enable one to reformulate an expression for the traditional population odds ratio, which will enable its application and use with sample observation obtained from diagnostic screening tests or clinical trials and then its subsequent use to calculate the sample estimates of the sensitivity and specificity of the screening test as well as an appropriate test statistic. In clear terms, the proposed method provides information concerning the discriminatory power of a diagnostic test. Also results obtained using the proposed method can be used to choose the best clinical strategy in clinical practice.

Declarations

Ethical consideration: No ethical approval.

Authors' contribution: Dr. Okeh, U.M.: Conceived and proposed the method, analyzed, and interpreted the data, wrote the paper. Dr. C.N. Okoro: Proofread the manuscript and made useful inputs to data collected. Also, supplied useful materials that helped in research.

Conflict of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement: Author wishes to appreciate the laboratory staff of the Nnamdi Azikiwe University Teaching Hospital Nnewi Anambra State Nigeria who took pain to collect the data used in this research work.

References

1. Raffles A, Mackie A, Muir Gray JA. Screening: evidence and practice. 2nd ed. Oxford: Oxford University Press. 2019
2. Rosner B. Fundamentals of biostatistics. 8th ed. Boston: Brooks Cole, Cengage Learning. 2015.
3. Fleiss JL, Bruce Levin and Myunghee Cho Paik.. Statistical Method for Rates and proportions (3rd ed), Wiley Interscience, New Jersey. 2003.
4. Li J, Fine JP. Assessing the dependence of sensitivity and specificity on prevalence in meta-analysis. *Biostatistics*. 2011; 12:710-22.
5. Agresti AA. An Introduction to Categorical Data Analysis (Wiley Series in Probability and Statistics) 3rd Edition. New York: Wiley. 2019
6. Pepe MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford



- statistical series 28, Oxford: University Press, U.K; 2003
7. Naeger DM, Kohi MP, Webb EM, Phelps A, Ordovas KG, Newman TB. Correctly using sensitivity, specificity, and predictive values in clinical practice: how to avoid three common pitfalls. *AJR Am J Roentgenol.* 2013; ;200(6): W566-70.
 8. Bartol T. Thoughtful use of diagnostic testing: Making practical sense of sensitivity, specificity, and predictive value. *Nurse Pract.* 2015; ;40(8):10-2.
 9. Baron JA. Clinical epidemiology, in *Teaching Epidemiology* eds. Olsen J; Saracci R; and Trichopoulos D; Oxford: Oxford University Press, 2001; 237-249.
 10. Leeftang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. *Cmaj.* 2013;185(11): E537-44.
 11. Perry GP, Roderer NK and Asnar SA. Current perspective of medical informatics and health sciences librarianship. *J Med Libr Assoc.* 2005; 33:199–206.
 12. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of clinical epidemiology.* 2005;58(10):982-90.
 13. Ma X, Nie L, Cole SR, Chu H. Statistical methods for multivariate meta-analysis of diagnostic tests: an overview and tutorial. *Statistical methods in medical research.* 2016 Aug;25(4):1596-619.
 14. Oyeka IC, Umeh EU. Measuring strength of association in repeated samples. *African Journal of Mathematics and Computer Science Research.* 2012;5(14):274-7.
 15. Glasziou P, Hilden J. Test selection measures. *Med Decis Making* 1989;9(2):133–41.
 16. Kraemer HC. Risk ratios, odds ratio, and the test QROC. In: *Evaluating medical tests.* Newbury Park, CA: SAGE Publications, Inc.; 1992; 103–13.
 17. Hlatky MA, Lee KL, Botvinick EH, Brundage BH. Diagnostic test use in different practice settings. A controlled comparison. *Arch Intern Med* 1983;143(10):1886–9.