Journal of Biometrics & Biostatistics
ISSN: 2155-6180

Shaikh. J Biomet Biostat 2011, 2:5
DOI: 10.4172/2155-6180.1000128

Review Article Open Access

Measures Derived from a 2 x 2 Table for an Accuracy of a Diagnostic Test

Shaffi Ahamed Shaikh

Department of Family & Community Medicine, College of Medicine, KSU, Riyadh, Kingdom of Saudi Arabia

Abstract

Diagnostic test studies are receiving increasing attention, but are rather challenging to identify efficiently and reliably. Health care professionals seek information regarding the best available evidence on test accuracy, and there is also a growing requirement to understand the measures of diagnostic tests. An outcome of epidemiological studies, diagnostic tests, and comparative therapeutic trails are often presented in the form of 2 x 2 tables. The analysis from these tables and its significance must be interpreted correctly, so as to answer the clinical research questions of the studies. This article discusses about the measures which could be derived from a 2x2 table format, for a diagnostic test situation, where the interest is to observe the relationship between two qualitative (nominal) variables.

Keywords: 2 x 2 contingency table; Sensitivity; Specificity; Likelihood ratio; Predictive values; Diagnostic odds ratio; Youden's index

Introduction

Presenting the results of any research study in a table form is an integral part of statistical analysis. A 2x 2 table (two rows and two columns) is an essential tool to present the data of epidemiological studies, diagnostic test evaluation studies, and studies related to therapeutic comparisons. For a 2 x 2 table, the terms four-field table, contingency table and cross table are also often used. Notation of a base 2 x 2 table is given in (Table 1).

Similar format and considerations of 2x2 tables apply to diagnosis, prognosis and therapy. For more details, reference can be made to Feltcher et al (1), Altman (2) and Campbell et al (3). Apart from the statistical tests (Yates-corrected chi-square, the Mantel Hansel chi-square and the Fisher's exact test), other measures relevant to 2 x 2 table are :(i) the analysis of risk factors (odds ratio, relative risk, absolute and relative risk reduction and number needed to treat) (ii) the analysis of effectiveness of a diagnostic criterion for some condition of interest (sensitivity, specificity, + ve and –ve predictive values, + ve and –ve likelihood ratios, diagnostic accuracy, diagnostic odds ratio and Youden's index) (iii) measures of inter-rater reliability and (iv) other measures of association such as contingency coefficient, Cramer's phi-coefficient and Yule's Q. Bewick et al (4) reviewed the statistical tests related to qualitative data and tests of association. This article illustrates the statistical measures relevant to diagnostic test situation from a 2 x 2 table, which are applied in all disciplines of clinical medicine.

What is a Diagnostic Test?

The aim of a diagnostic test is to confirm the presence or absence of a disease. The clinical performance of a diagnostic test is entirely based on its ability to correctly classify subjects into relevant sub groups. This test helps to find, if a person tests positive, what is the probability that the person really has the disease/condition, and if a person tests negative, what is the probability that the person is free of disease/condition? When new diagnostic tests are introduced, it is necessary to evaluate the comparative diagnostic accuracy and feasibility of this new test in comparison to the existing tests or the gold standard. In other words evaluation of a test gives the answer to the following question:

Exposure	Present	Absent	Total
Yes	a	b	a + b
No	c	d	c + d
Total	a + c	b + d	n

Table 1: Association between Disease and Exposure in a 2x2 table form.

How well does this test discriminate between health and disease? This discriminative ability and measures of diagnostic accuracy can be quantified by calculating [1] the sensitivity and specificity [2] the positive and negative predicative values (PPV, NPV), [3] the positive and negative likelihood ratios [4] the diagnostic odds ratio (DOR) and [5] the diagnostic accuracy & misclassification rate and [6] the Youden's index. Some of these measures are used to assess the discriminative property of the test, and others are used to assess its predictive ability [5,6]. The basic approach in the calculation of above measures is to make a 2x2 table with groups of subjects divided according to a gold standard or (reference method) in columns, and categories according to test in rows as given in (Table 2). Some of the measures are used to assess the discriminative property of the test, and others are used to assess its predictive ability.

A perfect diagnostic procedure has the ability to completely discriminate subjects with and without disease. Values of a perfect test which are above the specific cut-off (which could be labeled as "abnormal and normal" or " positive and negative") are always indicating the disease which are known as true positive values (TP), while below the specific cut-off values are always excluding the disease which are known as true negative values (TN). Values above the cut-off are not always indicative of a disease since subjects without disease can also sometimes have higher values. Such high values of certain parameter of interest are called false positive values (FP). On the other

New Test	Subjects with the disease	Subjects without the disease	Total
Positive	a (TP)	b (FP)	a +b (TP + FP)
Negative	c (FN)	d (TN)	c +d(FN + TN)
Total	a +c (TP + FN)	b +d(FP + TN)	a+ b+ c+ d (TP+TN+FP+FN)

Table 2: Assessment of new diagnostic test accuracy in relation to the gold standard

***Corresponding author:** Shaffi Ahamed Shaikh, Dept. of Family & Community Medicine, College of Medicine, King Saud University, Tel: +966-1-4671544 ; Fax: +966-1-4671967; E-mail: shaffi786@yahoo.com

Received November 03, 2011; **Accepted** November 21, 2011; **Published** December 25, 2011

Citation: Shaikh SA (2011) Measures Derived from a 2 x 2 Table for an Accuracy of a Diagnostic Test. J Biomet Biostat 2:128. doi:10.4172/2155-6180.1000128

hand, values below the cut-off are mainly found in subjects without disease. However, few subjects with the disease may have such values. Those values are known as false negative values (FN). Therefore, the cut-off value divides the population of examined subjects with and without disease in four subgroups considering parameter values of interest (Table 2).

Sensitivity and Specificity

Sensitivity is expressed as the proportion of correctly classified as true positives among the total disease (a/ a+ c or TP/TP+FN). In other words, sensitivity of a test is its ability to correctly identify the proportion of patients with the disease. It rarely misses a patient with the disease. A highly sensitive test is useful, when someone do not want to miss a disease, in the early stage of diagnostic work up and in screening the population for the target disorder. Moreover a sensitive test is most useful if it is negative. Whereas specificity is the ability to identify the proportion of population who do not have the disease that is the true negatives and is expressed as the proportion of correctly classified as true negatives among the total non disease (health)(b/ b+ d or TN/TN+FP). A specific test will rarely misclassify individual without the disease as diseased. A highly specific test is useful, if false positives are nil or rare and to confirm a diagnosis (rule-in). A specific test is most useful if it is positive [7].

Predictive Values

Predictive values (positive and negative) reflect the characteristics of a test. The positive predictive value of test is the probability of a study subject who has the disease when restricted to those subjects who had a positive test. It can be calculated as (a/a+ b or TP/ (TP+FP). It can be observed that denominator of positive predictive value is the number of subjects who test positive.

The negative predictive value of a test is the probability of a study subject who will not have the disease when restricted to those subjects who test negative. It can calculate as (d/c+ d or TN/ FN+ TN). Here the denominator is the number of subjects who test negative. It is meaningless to calculate the positive and negative predictive values on a sample where the prevalence of disease was artificially controlled that is by recruiting healthy and diseased patients in a one to one ratio. Unlike sensitivity and specificity, predictive values are largely dependent on disease prevalence in the population. Therefore, predictive values from one study should not be used (or referred) to some other setting with a different prevalence of the disease in the population. Prevalence affects PPV and NPV where these two values move in opposite direction. As the prevalence of disease in population increases, the PPV will be increasing, while NPV decreases [8].

Likelihood Ratio (LR)

Likelihood ratio is a very useful and mostly widely applied measure of diagnostic accuracy. It can summarize information about the diagnostic test, where it combines the values of sensitivity and specificity. It indicates how much a positive or negative test result changes the likelihood that a patient would have the disease. This measure incorporates both the sensitivity and specificity of the test and provides a direct estimate of how much a test result will change the odds of having a disease. The likelihood ratio for a positive result (LR+) shows how much the odds of the disease increases when a test is positive. This can be calculated as LR+ = Sensitivity/ (1-Specificity). Whereas, the likelihood ratio for a negative result (LR-) shows how much the odds of disease decreases when a test is negative. This can be calculated as LR- = (1-Sensitivity)/ Specificity. Like predictive values, LR's does not depend on prevalence of disease of population, as only sensitivity and specificity values were used to calculate both the positive and negative likelihood ratios. As a result the LR's of one study could be used in another setting with the condition that the definition of disease is not changed.

LR's can be directly related the pre-test and post-test probability of a disease in a specific patient, where the effect of diagnostic test will be quantified. By specifying the information about the patient, pre-test odds (the likelihood that the patient would have a specific disease prior to testing), the post-test odds of disease could be determined. The pre-test odds are related to the prevalence of the disease and it is important to specify as the diagnostic test will be adapted to patient rather than patient to the diagnostic test [9].

Diagnostic Odds Ratio (DOR)

The diagnostic odds ratio (DOR) is an overall measure to summarize test performance. It is the positive likelihood ratio divided by the negative likelihood ratio [(LR+) ÷ (LR-)] or [TP* TN/ FN * TN]. The DOR is used to estimate the discriminative ability of diagnostic test procedures and also to compare the diagnostic accuracies of between two more diagnostic tests. In Meta analysis to combine the results of multiple, DOR is being increasingly used as a clinical parameter. From the above formula, it can be observed that the DOR depends on sensitivity and specificity of a test. The value of DOR will be high with higher values of sensitivity and specificity and with low false positive and negative values. DOR depends on the criteria used to define the disease but not on the prevalence of disease [10].

Diagnostic Effectiveness (Accuracy)

The diagnostic accuracy of a test is expressed as the proportion of those individuals correctly categorized by the test (those with disease who had a positive test plus those without disease who had a negative test result). It can be calculated as: (a+d) ÷ (a+ b+ c+ d) or (TP+ TN/ TP+ FP+FN+TN). This measure is affected by the prevalence of disease. The accuracy of a test increases as the prevalence of disease decreases, by keeping the sensitivity and specificity same. This measure of classification of subjects as true positives and true negatives should be preferred and used after taking into account the other measures of accuracy, particularly predictive values.

Misclassification Rate

The misclassification rate is the proportion of those individuals incorrectly categorized by the test (those with disease who had a negative test plus those without disease who had a positive test result).It can be calculated as: (b + c) ÷ (a+ b+ c+ d) or (FP + FN/TP + FP + FN+ TN). The misclassification rate is the complement of the diagnostic accuracy of the test, i.e., misclassification rate = 1 - diagnostic accuracy.

Youden's Index

Youden's index is one of the well known measures of diagnostic measure of accuracy [11]. It is a global measure of a test performance, used in the evaluation of overall discriminative power of a diagnostic procedure and comparison of one test with other tests. It is an index which summarizes the sensitivity and specificity of a test. It can be calculated as (sensitivity + specificity)-1 or [(a/a + c) + (b/b +d) -1]. It ranges from 0 for a poor diagnostic accuracy and to a "perfect" value of 1.0 for a perfect diagnostic test. This index in not affected by the disease prevalence, but it is affected by the spectrum of the disease. The prime

disadvantage of this index is, it does not change for the differences in the sensitivity and specificity of the test. That is a test with sensitivity 0.7 and specificity of 0.8 has the same You den's index (0.5) as a test with sensitivity 0.9 and specificity 0.6.

Examples

The following 3 examples from literature for a cohort study design, case control study design & randomized controlled trail and a hypothetical example for diagnostic test evaluation study, indicates the application of 2 x 2 table:

Cohort study

The British Regional Heart Study (12) was a cohort of 7735 men aged 40-59 years randomly selected from genera practices in 24 British towns, with the aim of identifying risk factors for ischemic heart disease. Of the 7718 men who provided information on smoking status, 5899(76.4%) had smoked at some stage during their lives. Over the subsequent 10 years, 650 of these 7718 men (8.4%) had a myocardial infraction (MI). The following were the results displayed in 2x2 table show the number and percentage of smokers and non-smokers who developed and did not develop and MI over the 10 year period.

MI in Subsequent 10 years			
Smoking status at baseline	Yes (%)	No (%)	Total
Ever smoked	563(9.5)	5336(90.5)	5899
Never smoked	87(4.8)	1732(95.2)	1819
Total	650(8.4)	7068(71.6)	7718

The estimated relative risk = (563/5899)/ (87/1819) = 2.0

The relative risk 2.0 mean that a middle-aged man who has ever smoked is twice as likely to suffer an MI over the next 10 year period as a man who has never smoked. That is the risk of suffering an MI for a man who has ever smoked is 100% greater than that of a man who had never smoked.

Case Control Study

A total of 1327 women aged 50-81 years with hip fractures, who lived in a largely urban area in Sweden, were investigated in an unmatched case-control study (13). They were compared with 3262 controls with the same age range randomly selected from the national registry. The objective was to determine whether women currently taking postmenopausal hormone replacement therapy (HRT) were less likely to have hip fractures than those not taking it. The results were given in a 2x2 table format show the number of women who were current users of HRT and those who had never used or formerly used HRT.

	Current user of HRT	Never used HRT/ former user of HRT	Total
With hip fracture (cases)	40	1287	1327
Without hip fracture (controls)	239	3023	3262
Total	279	4310	4589

The observed odds ratio= (40 x 3023)/ (239 x 1287) = 0.39. Thus the odds of a hip fracture in a postmenopausal women in the age range 50 -81 in Sweden who was a current user of HRT was 39% of that of a woman who had never used or formerly used HRT, i.e. being a current user of HRT reduced the odds of hip fracture by 61%.

Randomized clinical trail

A randomized, placebo-controlled, multicenter trial was conducted in South Africa by recruiting 4939 infants, to evaluate the efficacy of a live, oral rotavirus vaccine in preventing severe rotavirus gastroenteritis (14). Of these infants, 1647 received two doses of the vaccine, 1651 infants received three doses of the vaccine and 1641 received placebo. One of the outcome "decreasing severe diarrhea from all cause gastroenteritis" is given in following 2 x 2 table:

	Outcome		
Group	Yes	No	Total
Vaccine	2718	256	2974
Placebo	1265	178	1443
Total	3983	434	4417

Experimental event rate (EER) = 2718/2974 = 0.914
Control event rate (CER) = 1265/1443 = 0.876
Absolute risk reduction (ARR) =0.914-0.876 = 0.038
Number needed to treat (NNT) = 1/0.038 = 26.31

That is 27 infants need to be vaccinated to have 1 infant with decreased risk to develop severe diarrhea from any cause of gastroenteritis (including Rotavirus).

Diagnostic Test Evaluation

The illustration and interpretation of diagnostic test parameters which are discussed above, data are generated from a hypothetical study of diabetic eye tests. Assume that diabetic patients were screened for eye problems using direct opthalmoscopy (the new test) and slit lamp biomicroscopy (the reference or gold standard test). The test was studied and the gold standard was applied to 378 subjects and the new test's diagnostic accuracy was determined.

The following 2 x 2 table shows the relationship between the new test and gold standard:

	Gold Standard		
New test	Positive	Negative	Total
Positive	105	171	276
Negative	15	87	102
Total	120	258	378

Sensitivity (true positive rate) = 105/120 = 0.88 or 88%
Specificity (true negative rate) = 87/258 = 0.34 or 34%
Positive predictive value = 105/276 = 0.38 or 38%
Negative predictive value = 87/102 = 0.85 0r 85%
Likelihood ratio for positive test = 88% / (100-34%) =1.3
Likelihood ratio for negative test = (100%-88%)/ 34% =0.4
Diagnostic odds ratio = 1.3/0.4 = 3.25
Diagnostic effectiveness (accuracy) = (105 + 87)/378 = 50.8%
Misclassification rate = (171 + 15)/378 = 49.2%
Younden's index = (0.88 + 0.34)-1 = 0.22

From the above calculated values, sensitivity of 88% shows that new test correctly identified 105 subjects out of 120 who have the disease (eye problems). Specificity of 34% refers to the ability of new test to correctly identify subjects who do not have the disease (eye problems). Positive predictive value of 38% refers to the proportion that a positive test result indicates the presence of the disease condition. Negative predictive value of 85% refers to the proportion that a negative test result indicates the absence of the disease condition. The positive

likelihood ratio of a test 1.3 indicates that with a positive result a subject is 1.3 times more likely to be truly positive than negative, as determined by the gold standard. The negative likelihood ratio of 0.4 indicates that with a negative result a subject is 0.6 times as likely to be positive than negative, as determined by the gold standard. Diagnostic odds ratio of 3.25 for the new test indicates that, the odds for positivity among subjects with disorder (eye problems) is 3.25 times higher than the odds of positivity among the subjects without disorder (eye problems). The diagnostic effectiveness (accuracy) 50.8% of new test is expressed as the proportion of subjects correctly categorized by a new test, in relation to gold standard. The compliment of diagnostic accuracy (misclassification rate) 49.2% shows a proportion of subjects, who were incorrectly classified by new test, in relation to the gold standard. The Youndex's index of 0.22 is another measure of the diagnostic accuracy of new test.

Conclusions

Studies which are designed to measure the performance of diagnostic tests are often presented in the form of a 2 x 2 Table These studies are important not only for patient care but also for effective management of health care cost. Proper understanding and interpretation of measures related to diagnostic test accuracy are necessary for clinicians so as to make valid conclusions. This article has listed out the basic definitions of statistical measures derived from a 2 x 2 table for a diagnostic test accuracy assessment, its method of calculation with examples, application and limitations.

References

1. Fletcher RH, Fletcher SW (1996) Clinical Epidemiology –the essentials. (3rd edn), Williams & Wilkins, USA.
2. Altman DG (1991) Practical statistics for medical research, Chapman and Hall, London.
3. Campbell MJ, Machin D, Walters SJ (2007) Medical statistics: a textbook for the health sciences. (4th edn), John Willey and Sons, Chi Chester, England.
4. Bewick V, Cheek L, Ball J (2004) Statistics review 8: Qualitative data-tests of association. Crit Care 8: 46-53.
5. Irwig L, Bossuyt P, Glasziou P, Gatsonis C, Lijmer J (2002) Designing studies to ensure that estimates of test accuracy are transferable. BMJ 324: 669-671.
6. Raslich MA, Markert RJ, Stutes SA (2007) Selecting and interpreting diagnostic tests. Biochemia Medica 17: 151-161.
7. Altman DG, Bland JM (1994) Statistics Notes: Diagnostic tests 1: sensitivity and specificity. BMJ 308: 1552.
8. Altman DG, Bland JM (1994) Statistics Notes: Diagnostic tests 2: predictive values. BMJ 309: 102.
9. Deeks JJ, Altman DG (2004) Diagnostic tests 4: likelihood ratios. BMJ 329: 168-169.
10. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM (2003) The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 56: 1129-1135.
11. Youden WJ (1950) Index for rating diagnostic tests. Cancer 3: 32-35.
12. Shaper AG, Pocock SJ, Walker M, Cohen NM, Wale CJ, et al. (1981) British Regional Heart Study: cardiovascular risk factors in middle-aged men in 24 towns. Br Med J, 283: 179-186.
13. Michaelsson K, Baron JA, Farahmand BY, Johnell O, Magnusson C, et al. (1998) Hormone replacement therapy and risk of hip fracture: population based case-control study. British Medical Journal 316: 1858-1863.
14. Madhi SA, Cunliffe NA, Steele D, Desiree W, Mari K, et al. (2010) Effect of Human Rotavirus Vaccine on Severe Diarrhea in African Infants. The New England Journal of Medicine 362: 289-298.