IOSR Journal of Dental and Medical Sciences (IOSR-JDMS)
e-ISSN: 2279-0853, p-ISSN: 2279-0861.Volume 14, Issue 4 Ver. V (Apr. 2015), PP 64-68
www.iosrjournals.org

Brief review on Sensitivity, Specificity and Predictivities

[1]Arif Habib, [2]Meshiel Alalyani, [3]I Hussain, Musa [4]M. S. Almutheibi
[1,,3,4]College of Applied Medical Sciences, [2]College of Nursing; Khamis Mushait
King Khalid University, Kingdom of Saudi Arabia

Abstract: *This paper is designed as a tool that a researcher could use in planning and conducting quality research. This is a review paper which gives a discussion of various aspects of sensitivity, specificity and predictivities including the sample size estimation. The manuscript intended to discuss and clarify aspects related to key methodological issues in assessing the findings or conclusions based on primary analyses of data in clinical trials. Over the last decades, considerable interest has been focused on medical research and its credibility, thus the results of clinical trials relies on the validity of the methods of analysis or models used and their corresponding assumptions.*

Key words: *sensitivity, specificity, predictivities, disease prevalence, sample size*

I. Introduction:

The words "sensitivity" and "specificity" have their origins in screening tests for diseases. Sensitivity and specificity are the statistical measures of performance of a binary classification tests. The sensitivity measures the proportion of actual positive which are classified as such (e.g. the percentage of sick people who are identified as having the condition); and specificity measures the proportion of negatives who are correctly identified (e.g. the percentage of well people who are identified as not having the condition). In short, sensitivity refers the probability of true showing up true and specificity to the probability of false showing up false. Sensitivity and specificity are usually expressed in percentage. In clinical research, the sensitivity of a medical test is the probability of its giving a 'positive' result when the patient is indeed positive and specificity is the probability of getting a negative result when the patient is indeed negative. A theoretical optimal prediction result can achieve 100% sensitive (i.e. predicts all people from a sick population as sick) and 100% specificity (i.e. not predicts any from the healthy population). Imagine a scenario, where people are tested for a disease. The test outcome can be positive (sick) or negative (healthy), while the actual health status of a person may be different.

Consider the following 2×2 table that gives the number of subjects with positive and negative tests, and with and without disease in any random sample from a target population. It is only for such a sample that all four indices can be calculated from the same data.

		Actual Condition		
		Present (+)	Absent (-)	
Test Result	Disease (+)	**a = True Positive** (Sign and symptoms Present + Positive Result)	**b = False Positive** (Sign and symptoms absent + Positive result)	**a + b**
	No Disease(-)	**c = False negative** (Sign and symptoms Present + Negative Result)	**d = True negative** (Sign and symptoms absent + Negative result)	**c + d**
		a + c	**b + d**	**n**

Concept of Sensitivity and Specificity:

The concept of Sensitivity and specificity is very useful in finding out the utilities of new therapies in medical field as compared to standard therapy. Further these measures assessing new scaling of assessment to improve in patients as compared to old, well established and widely used criteria.

Sensitivity: The probability that the test reports a person has the disease when in fact they do have the disease. i.e the ability of the test to be correctively positive among those who are known to have the disease. So this can be measured by $p\left(T^+ / S^+\right) = \frac{a}{a + c}$. This is a measure of how likely it is for a test to pick up the presence of a disease in a person who has it.

Sensitivity alone does not tell us how well the test predicts the other class (i.e. about negative cases). In the binary classification this is the corresponding specificity test or equivalently the sensitivity for the other classes.

The calculation if sensitivity does not tale into account the intermediate results, the option are either to exclude intermediate samples from analysis (but the number of exclusions should be stated when quoting sensitivity) or alternatively intermediate samples can be treated as false negative.

Specificity: The probability that the test reports a person does not have the disease when in fact they are without disease. i.e The ability of the test to be correctively negative among those who are known to be free of disease. So this can be measured by $p\left(T^- / S^-\right) = \frac{d}{b + d}$.

A test with high specificity has a low type I error rate. Specificity is sometimes confused with **precision** or the positive predicted value, both of which refer to the fraction of returned positives that are true positives. A test with high specificity can have a very low precision if there are far more true negatives than true positives and vice versa.

Though these tests are generally quite accurate, they still make errors that we need to account for. For example, we may have a test which has high sensitivity, but this sometimes results in low specificity. Generally we are able to keep both sensitivity and specificity high in screening tests. However, we still get false positives and false negatives.

False Positive: A false positive occurs when the test reports a positive result for a person who is disease free. The false positive rate is given by $p\left(T^+ / S^-\right) = \frac{b}{a + b}$. Ideally we would like the value of b to be zero; however, this is generally impossible to achieve in a screening test involving a large population.

False Negative: A false negative occurs when the test reports a negative result for a person who actually has the disease. The false negative rate is given by: $p\left(T^- / S^+\right) = \frac{c}{c + d}$.

In general one should worry more about false positives in screening tests. We don't want to tell someone that they have a serious disease when they do not really have it. For example if a person did test for a particular disease and the result shows positive, while the person is free of disease.

Concept of Predictivities:

The major limitation of both sensitivity and specificity is that they are of no practical use when it comes to helping the clinician estimate the probability of disease in individual patients. When you see a patient in your clinic who returns a positive result for a particular test, the question that you and your patient would want an answer to is 'what is the chance (probability) of disease given the positive test?' Sensitivity and specificity cannot be used to answer such a question. This is because both sensitivity and specificity are defined on the basis of people with or without a disease. However, because the patient would have presented to you with a set of symptoms rather than a diagnosis, you would not know at the time whether the patient has a disease or not and cannot, therefore, apply these parameters directly to them. What we need to know are predictive values which, in routine clinical practice, are more useful measures of diagnostic accuracy. The whole purpose of a diagnostic test is to use its results to make a diagnosis, so we need to know the probability that the test result will give the correct diagnosis. Positive and NPVs describe a patient's probability of having disease once the results of his or her tests are known.

Positive Predictive Value:

The positive predictive value (PPV) of a test is defined as the proportion of people with a positive test result who actually have the disease. It is an ability of the test to correctly predict the presence of the disease. In

other words it is the probability that the disease is present when the test is positive. This can be measure by $p\left(T^{+}/S^{+}\right)=\frac{a}{a+b}$.

Negative Predictive Value:

The NPV of a test is the proportion of people with a negative test result who do not have disease. It is an ability of the test to correctly predict the presence of the disease. In other words it is the Probability that the disease is not present when the test is negative. This can be measure by $p\left(T^{-}/S^{-}\right)=\frac{d}{d+c}$.

Given the definition of PPV above, it is apparent that this metric is a function of the True Positive and False Positive values. If there were no false positives, the PPV would be TP/TP or 100%. This would occur if the test could reliably distinguish normal persons from persons with disease without any overlap of values. However, this is rarely the situation with laboratory results that fall on a continuous scale. Almost always there is some degree of overlap between results from normal and diseased subjects. For example, low levels of blood markers of myocardial infarction can be detected in healthy subjects. These may overlap the lower range of values found in patients with myocardial infarction. When there is such overlap, some cut-off point must be established to distinguish a positive from a negative result. It is worth noting that even results which are reported as "positive" or "negative" by analytical instruments are generated from continuous scales by using a cut-off point. An example of this is the reporting of HIV-1 antibody screening tests. In a popular assay system, these are detected by ELISA reactions which result in changes in absorbance as measured by a spectrophotometer. The lower end of absorbance readings in HIV-1 positive patients overlap those of normal subjects. The computer in the analytical instrument defines a cut-off point of absorbance readings and then reports the results as positive or negative based on the cut-off value. Given that most assays show some overlap of values between normal and diseased subjects, it is apparent that it is very unusual for assays to have no FP results.

Prevalence of Disease Influences Predictive Values:

A crucial point is that prevalence affects the predictive value of any test. The prevalence of the disease can be interpreted as the probability that a randomly chosen member of the population being screened has the disease. The predictive value of a test is determined by the test's sensitivity and specificity and by the prevalence of the condition for which the test is used. Both PPV and NPV vary with changing prevalence of disease. Let us suppose that there are no false positive results, namely, if everyone in the population had the disease. In this situation, every positive result would be a true positive. There could be no false positive results, and the PPV would be 100%. Conversely, if no one in the population had the disease, every positive result would be a false positive. There could be no true positives, and the PPV would be 0%. This leads us to conclude that the disease prevalence influences the PPV by influencing the true positive and false positive rates. It will therefore be wrong for researchers/clinicians to directly apply published predictive values of a test to their own populations, when the prevalence of disease in their population is different from the prevalence of disease in the population in which the published study was carried out.

The following table illustrates this phenomenon. It holds sensitivity and specificity constant, at 99% and 95%. You can just look at the second and last two rows, As prevalence rises from 5% to 9%, PPV increases from 51.47% to 66.32%: a huge difference in the clinical interpretation of the same test result and same can seen in the NPV as prevalence increases NPN decreases.

Population	700	700	700
Disease Prevalence	5	7	9
Diseased	35	49	63
Not Diseased	665	651	637
True Positive [*No.of diseased*×0.9	34.65 =35	48.5 =49	62.37=62
False Positive [*No.of notdiseased*×	33.25=33	32.55 =33	31.85=32
Total positive	67.9=68	81.05=81	94.22=94
Positive Predictive Value	**51.47%**	**59.76%**	**66.32%**
Negative Predictive Value	**95%**	**93%**	**91.13%**

Significance of Bayes rule:

If prevalence of the disease in the target group is known, predictivities can be obtained by using sensitivity and specificity. Herein lays the importance of these two concepts. Based on confirmed cases, sensitivity and specificity are easy to obtain, and these then help to calculate diagnostically important predictivities with the help of Bayes' rule.

High fever, rigors, spleenomegaly, and presence of parasite in the blood are the stages that progressively confirm malaria. As the information increases, the diagnosis becomes focused, and the probability of absence or presence of the disease firms up. The probability depends on what information is already available. The chance part is the uncovered information. The probability of any event without availability of any information is called **prior probability** and the probability after some information is available is called **posterior probability**. The latter obviously depends on the kind of information available to alter the probability. The function of Bayes' Rule is to convert one posterior probability to its directional inverse. It converts probability of A given B to probability of B given A. This is useful to convert sensitivity of a test to its productivity. The PPV is $p\left(T^+ / S^+\right)$ and can be related to the preceding by an application of Bayes theorem, i.e. P(A|B)=P(A & B)/Pr(B), so

$$p\left(T^+ / S^+\right) = \frac{sensitivity \times prevalence}{sensitivity \times prevalence + (1-sensitivity)\times(1-prevalence)}.$$

and for negative predictive value it can be calculated in a same manner and is equal to

$$p\left(T^- / S^-\right) = \frac{specificity \times (1-prevalence)}{(1-sensitivity)\times(prevalence + specificity)\times(1-prevalence)}.$$

Sample size for a sensitivity of a Test:

Based on the literature review identify the sensitivity and specificity of the diagnostic test being study-

For sensitivity

$T_+ + F_- = (Z_{\alpha/2})^2 \frac{\{S_n\ (1-S_n)\}}{E^2}$, where T_+ is true positive, F_- is false negative S_n is sensitivity and E accuracy which is usually taken 0.05.

$nS_n = \frac{T_+ + F_-}{p}$, where p is the prevalence of the disease in test population.

For Specificity

$F_+ + T_- = (Z_{\alpha/2})^2 \frac{\{S_p\ (1-S_p)\}}{E^2}$, where F_+ is false positive, T_- is true negative S_p is specificity.

$nS_p = \frac{F_+ + T_-}{(1-p)}$.

E.g. Let sensitivity be 85% and prevalence be 28% , using above formula the sample size is as fallows

$T_+ + F_- = (1.96)^2 \frac{\{0.85\ (1-0.85)\}}{0.05^2} = 195.9$

$nS_n = \frac{195.9}{0.28} = 699.79$ can be taken as **700**.

Now take specificity as 70% and prevalence 28%

$F_+ + T_- = (1.96)^2 \frac{\{0.70\ (1-0.70)\}}{0.05^2} = 322.72$ can be taken as **323**.

Note: if the researcher is interested in both sensitivity and specificity, than take the higher number. In the above example we can take 700 sample size.

II. Discussion:

From the above review, the concept of sensitivity, specificity, predictivities and its other related concepts are very useful for medical science and applied medical research. They are important in interpreting or establishing the credibility of the findings. If, the results remain robust under different assumptions, methods or

scenarios, these methods can strengthen their credibility. From statistician point of view they are closely related to type I and Type II errors in testing hypothesis. In statistical hypothesis testing type I error is usually denoted by α and $1-\alpha$ is defined as specificity. Increasing the specificity of the test lower the probability of type I error. Thus, specificity is a statistical measure how well a binary classification test correctly identifies the negative cases.

Similarly Type II error is denoted by β. In classical language of statistical hypothesis testing, the sensitivity of a test is called the statistical power of the test, although the word power in that context has a more general usage that is not applicable the present context. A sensitive test have a fewer Type II error. If the sample sizes in the positive (Disease present) and the negative (Disease absent) groups do not reflect the real prevalence of the disease, then the Positive and Negative predicted values cannot be estimated and one should ignore those values. Alternatively, when the disease prevalence is known then the positive and negative predictive values can be calculated using the following formula's based on Bayes' theorem:

III. Conclusions:

The sensitivity and specificity of a test is useful for cross sectional studies where as predictivities are useful for case control studies. Both PPV and NPV vary according to disease prevalence, and published predictive values should not be applied to populations whose prevalence of disease is different from the population in the published study. It is also recommended that if the researcher is interested in both sensitivity and specificity, than take the higher number of sample among the sensitivity and specificity.

Disclosure: The authors report no conflicts of interest in this work.

Acknowledgements:

We sincerely thank the editor and the anonymous referees and reviewers, careful reading of the brief report and some thought provoking comments. The valuable comments have forced us to give clarification to quite a few questions and to correct some mistakes improving the quality of the paper substantially.

References:

[1]. Altman, D.G. and Bland, J.M. (1994). Statistics Notes: Diagnostic tests 1: sensitivity and specificity, Br. Med. Jour., 308, p. 1152.
[2]. Arif Habib, Ayman Johargy et al (2014). Design and Determination of Sample Size in Medical Research. IOSR Journal of Dental and Medical Sciences (IOSR-JDMS) e-ISSN: 2279-0853, p-ISSN: 2279-0861.Volume 13, Issue 5 Ver. VI. (May. 2014), PP 21-31.
[3]. Askin, S.R. and Bryan, R.S. (1976). Femoral neck fractures in young adults, Clin. Orthop., 114, p. 259-264.
[4]. Gaddis, G.M. and Gaddis, M.L. (1990). Introduction to biostatistics: Part 3, sensitivity, specificity, predictive values and hypothesis testing. Ann. Emerg. Med., 19(5), p. 591-597.
[5]. Loong, T.W. (2003). Understanding sensitivity and specificity with the right side of brain. Br. Med. Jour., 327, p. 716-719.
[6]. Chatfield C. Confession of a pragmatic statistician. The Statistician 2002; 51 (part 1):1-20.
[7]. Gardner IA, Greiner M (2006) Receiver-operating characteristic curves and likelihood ratios: improvements over traditional methods for the evaluation and application of veterinary clinical pathology tests. Veterinary Clinical Pathology 35:8-17.
[8]. Parshall MB (2013) Unpacking the 2 x 2 table. Heart & Lung 42:221-226.
[9]. Zweig MH, Campbell G (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry 39:561-577.