Measures of Diagnostic Accuracy: Sensitivity, Specificity, PPV and NPV

Hwee Bee Wong[1], *MSc(Statistics),* Gek Hsiang Lim[1,2], *MSc(Biostatistics)*

[1]Health Services Research and Evaluation Division, Ministry of Health, Singapore

[2]National Registry of Diseases Office, Health Promotion Board, Singapore

INTRODUCTION

In biomedical studies, diagnostic tests are used to determine the presence or absence of diseases in study subjects. Examples include testing for the presence or absence of Alzheimer's disease and invasive carcinoma. A diagnostic test is validated by comparing test results against a gold standard that establishes the true status of the subject. Test validation is an evaluation method used to determine the fitness of a test for a particular use and through it, one can assess how good the test is at identifying subjects with and without a disease or condition. Validation involves calculating four objective measures of test performance, namely, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). The ideal diagnostic test would correctly identify subjects with and without the disease with 100% accuracy. Details of the four measures are provided below.

Table 1 shows the results of a diagnostic test presented in a 2x2 table: columns summarise gold standard results and the rows summarise test results. The labels positive and negative refer to the presence or absence, respectively, of the condition of interest. The number of subjects with the condition testing positive and negative are denoted by a and c. The number of subjects without the condition testing positive and negative are denoted by b and d. The total number of study subjects should be a + b + c + d.

The basic measures to quantify the diagnostic accuracy of a test include sensitivity and specificity[1]. The sensitivity of a diagnostic test quantifies its ability to correctly identify subjects with the disease condition. It is the proportion of true positives that are correctly identified by the test, given by:

$$Sensitivity = \frac{True\ positives}{True\ positives + False\ negatives} = \frac{a}{a+c}$$

The specificity is the ability of a test to correctly identify subjects without the condition. It is the proportion of true negatives that are correctly identified by the test:

$$Specificity = \frac{True\ negatives}{False\ positives + True\ negatives} = \frac{d}{b+d}$$

As both sensitivity and specificity are proportions, their confidence intervals can be computed using the standard methods for proportions[2].

The PPV and NPV are the other two basic measures of diagnostic accuracy. They are related to sensitivity and specificity through disease prevalence (∏). The PPV is the probability that the disease is present given a positive test result, and is defined as:

$$PPV = \frac{Sensitivity * \pi}{Sensitivity * \pi + (1 - specificity) * (1 - \pi)}$$

Similarly, the NPV is the probability that the disease is absent given a negative test result, and is defined as:

$$NPV = \frac{Specificity * (1 - \pi)}{Specificity * (1 - \pi) + (1 - sensitivity) * \pi}$$

Table 1. Results of a diagnostic test presented as a 2x2 table.

Result of diagnostic tests	**Results of Gold Standard Test**	
	Disease present	**Disease absent**
Test positive	True positive (***a***)	False positive (***b***)
Test negative	False negative (***c***)	True negative (***d***)

Fig 1. Effect of disease prevalence on PPV and NPV

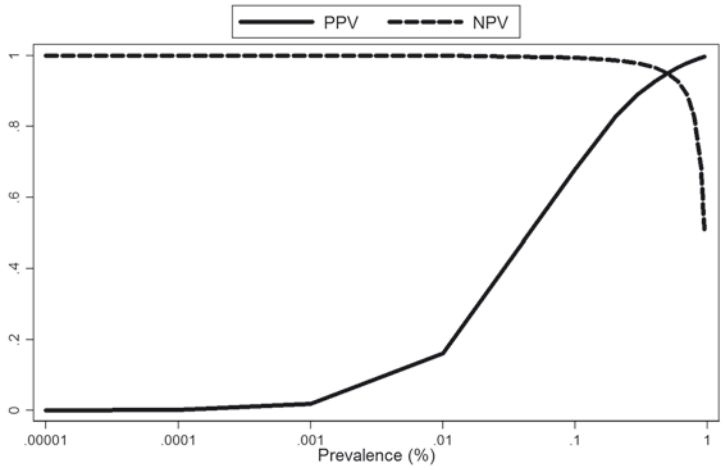

Table 2. Results of a mammography and pathology for NHMN study.

Results of mammogram	**Pathological Results**		
	Positive	**Negative**	**Total**
Abnormal	132	639	771
Normal	37	5,344	5,381
Total	169	5,983	6,152

As both PPV and NPV are related to sensitivity, specificity and ∏ :

(a) PPV is greater when ∏ is high and PPV is always greater than ∏. Hence, when a disease is rare (i.e. ∏ is low), a greater specificity is needed to achieve a higher PPV.

(b) NPV is greater when ∏ is low and NPV is always greater than 1 — ∏. Hence, when a disease is common (i.e. ∏ is high), a greater sensitivity is needed to achieve a higher NPV.

Figure 1 illustrates the effect of disease prevalence on PPV and NPV when both sensitivity and specificity

are fixed at 0.95. The results show that increasing disease prevalence, ∏, will lead to an increasing PPV but a decreasing NPV of a diagnostic test.

To illustrate the above-mentioned calculations for sensitivity, specificity, PPV and NPV, consider an example of a diagnostic mammography study of 6,152 women with signs or symptoms of breast cancer (Table 2). This data was retrieved from a one-year study of women from the New Hampshire Mammography Network (NHMN) which was designed to describe key performance measures of screening and diagnostic breast radiography[3]. A positive cancer status was defined as any tissue specimen, including malignant cytologic findings, revealing invasive carcinoma or ductal carcinoma in-situ reported within a year.

Based on the NHMN study, the sensitivity and specificity of the diagnostic mammography is 132 / 169 = 0.78 and 5,344 / 5,983 = 0.89, respectively. We would expect 78% of women with positive pathology results to have abnormal mammograms, while 89% of those with negative pathology would have normal mammograms. Supposing that the prevalence of breast cancer is only 0.8%, the PPV is then 0.78*0.008 / (0.78*0.008 + [1 – 0.89]*[1 – 0.008]) = 0.05. Correspondingly, the NPV is 0.89*(1 - 0.008) / (0.89*[1 – 0.008] + [1 – 0.78]*0.008) = 0.99. Based on the breast cancer prevalence of 0.8%, we would expect only 5% of women with abnormal mammograms to have positive pathology results, while 99% of those with normal mammograms would have negative pathology results.

We have shown mathematically that the prevalence of a condition affects the PPV and NPV of the test; however, it is worthy to note that, prevalence does not affect the sensitivity and specificity of a diagnostic or screening test. This is because sensitivity and specificity are calculated from subgroups of subjects with and without the disease condition, respectively. Given this mathematical property, the implication is that measures of sensitivity and specificity can be computed and compared across study populations with different prevalence rates, although this assumes populations do not differ in important characteristics (e.g. disease severity) that would affect sensitivity and specificity of the diagnostic test.

Although the sensitivity and specificity of diagnostic tests are not affected by the prevalence of the condition, they can be influenced by differences in disease characteristics (such as clinical severity or anatomic extent of a disease) and characteristics of subjects such as age and breast density in the diagnostic mammography study[4]. For example, large, palpable breast cancer tumours or lesions in fatty, less dense breasts of older women are easier to detect with mammography. Hence, mammography has greater sensitivity when it is used in patients with advanced breast cancer or in older women. It is thus critical to evaluate the group of subjects to whom the diagnostic test has been applied when the sensitivity and specificity of two diagnostics tests are compared.

REFERENCES

1. Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. BMJ. 1994 Jun 11;308(6943):1552.
2. Altman DG, Machin D, Bryant TN, Gardner MJ, editors. Statistics with confidence: confidence intervals and statistical guidelines. 2nd ed. BMJ Books, 2000.
3. Poplack SP, Tosteson AN, Grove MR, Wells WA, Carney PA. Mammography in 53,803 Women from the New Hampshire Mammography Network. Radiology. 2000 Dec;217(3):832–40.
4. Zhou XH, Obuchowski NA, McClish DK. Statistical methods in diagnostic medicine. New York: Wiley-Interscience; 2002.