International Journal of Clinical Biostatistics and Biometrics
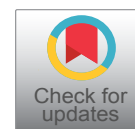
ORIGINAL ARTICLE

The Basic Four Measures and their Derivates in Dichotomous Diagnostic Tests

Tadeusz R Ostrowski, MD[1] and Tadeusz Ostrowski, PhD[2]*

[1]*The Jacob of Paradyż University of Applied Sciences,, Poland*

[2]*WSB University in Gorzow, Poland*

Corresponding author: *Tadeusz Ostrowski, PhD, WSB University in Gorzow, Poland*

Abstract

The paper focuses on four basic statistics of dichotomous diagnostic tests, i.e. sensitivity, specificity, positive and negative predictive value, and some of their derivates, like Youden index and predictive summary index, and on further derivates of these derivates, i.e. Matthews correlation coefficient (or Yule phi), chi squared test and Cramer's V coefficient. The paper contains also the necessary and sufficient conditions for a test to be invalid, to be uninformative and the necessary condition to be possibly valuable. The builder-concept of the paper is the determinant of 2 by 2 matrix.

Keywords

Sensitivity, Specificity, Positive (Negative) predictive value, Youden index, Predictive summary index, Matthews correlation coefficient

Abbreviations

Se: Sensitivity; Sp: Specificity; PPV: Positive Predictive Value; NPV: Negative Predictive Value; DOR: Diagnostic Odds Ratio; γ: Youden Index; ψ: Predictive Summary Index; MCC: Matthews Correlation Coefficient; V: Cramer's Coefficient; detM: Determinant of a Matrix M

Preliminaries

The aim of dichotomous diagnostic tests is to determine or predict the presence or absence of target condition (a disease or an infection) in study subjects. As it is known, clinical developments of new treatments are impossible without them. Different diagnostic measures relate to the different aspects of diagnostic procedure and some of them are used to assess the discriminative property of the test, others to estimate its predictive ability or overall accuracy. Let 2 by 2 matrix M (also called a confusion matrix) representing a contingency table be modeled as

$$M = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix}, \qquad (1)$$

Where according to the commonly used convention:

the upper row, T+ = [TP, FP], stands for a positive test result;

the bottom row, T- = [FN, TN], stands for a negative test result;

the left and right column, D+ = $\begin{bmatrix} TP \\ FN \end{bmatrix}$, D- = $\begin{bmatrix} FP \\ TN \end{bmatrix}$, stands for disease positive, negative (respectively). It should not lead to mix-up in the paper using the convenient convention, namely T+ = TP + FP for the frequency of T+ (the same concerns other concepts).

The elements of the matrix M are also named with the common convention (recall that the intersection of two sets A and B, denoted by A ∩ B, is the set containing all elements of A that also belong to B).

TP = T+ ∩ D+ (true positive; test positive and disease present);

FP = T+ ∩ D- (false positive; test positive and disease absent);

FN = T- ∩ D+ (false negative; test negative and disease present);

TN = T- ∩ D- (true negative; test negative and disease absent).

Therefore the confusion matrix (1) contains all information needed for the quantitative assessment of the

Ostrowski and Ostrowski. Int J Clin Biostat Biom 2020, 6:026

• Page 1 of 5 •

diagnostic test accuracy. Regrettably, only in an ideal world the test can be perfect (a positive patient has the target condition, and a negative patient does not have the condition of interest, so FP = FN = 0). In the real world, i.e. in practice, that kind of test is "a rare bird", so FN, FP cells are not empty, and we have to deal with false results.

Recall that if E, F are two events, then P(E/F) stands for the probability (called the conditional probability of E given F) of the event E occurring given that the event F has occurred. Formally,

$$P(E/F) = \frac{P(E \cap F)}{P(F)}.$$

Closely related to the idea of probability are the odds (familiar to gamblers and in betting), used to describe the chance of an event occurring. Probability and the odds represent a different way of expressing similar concept. The odds of an event E happening, or the odds in favor of E, denoted in this paper by O(E), mean the ratio of the probability that E will occur to the probability that E will not occur. Formally,

$$O(E) = \frac{P(E)}{1 - P(E)} \Leftrightarrow P(E) = \frac{O(E)}{1 + O(E)}.$$

If the probability is low, then the odds and the probability have quite similar values, but when the probability increases to 1, then the odds tend to infinity. Only if E is the impossible event, then P(E) = O(E) = 0. Roughly saying, probability and the odds, even different in numbers, are equivalent in meaning

Denote also by $O_L(K)$ the odds of K in the presence of L:
$$O_L(K) = \frac{P(K/L)}{1 - P(K/L)}.$$

Prevalence (Pr), or *pretest probability* in the context of diagnosis, is a measure of disease defined as the probability of a patient having a disease, or as the ratio of the number of existing conditions over the total sample number

$$Pr = P(D+) = \frac{D+}{(D+) + (D-)}$$

Or equivalently, the prevalence is described by the odds of belonging to the group of people identified as sick in the sample population

$$o(Pr) = \frac{D+}{D-}.$$

Sensitivity (Se) or *true positive rate*, called also *the recall* or *probability of detection*, is a measure quantifying diagnostic test ability to identify subjects with the disease condition, and it gives an answer to the question "If a patient has the disease, how likely is the patient to have a positive test?"

$$Se = P(T+/D+) = \frac{TP}{D+}$$

Or equivalently, it is the measure answering the question "What are the odds of having a positive test in the presence of disease?"

$$O(Se) = O_{D+}(T+) = \frac{TP}{FN}.$$

The relationship above implies that the lower FN, the odds of Se are higher and for this reason Se is referred to as *sensitive to disease*.

Specificity (Sp), or *true negative rate*, called sometimes *selectivity*, is a measure giving an answer to the question "If a patient does not have the disease, how likely is the patient to have a negative test?"

$$Sp = P(T-/D-) = \frac{TN}{D-}$$

Or equivalently, it is a measure answering the question "What are the odds of having a negative test in the absence of disease?"

$$O(Sp) = O_{D-}(T-) = \frac{TN}{FP}.$$

A low proportion of FP implies high the odds of Sp and for this reason Sp is referred to as *specific to health*.

Positive predictive value (PPV), called also *precision* or *purity*, is a measure of exactness which answers the question "If a patient has a positive test, how likely is the patient to have the disease?"

$$PPV = P(D+/T+) = \frac{TP}{T+}$$

or equivalently, it is a measure answering the question "What are the odds of having a positive test in the presence of disease?"

$$O(PPV) = O_{T+}(D+) = \frac{TP}{FP}.$$

A low proportion of FP impies that the odds of PPV are high. A practical clinical value of that is starting treatment.

Negative predictive value (NPV) is a measure which answers the question "If a patient has a negative test, how likely is the patient to not have the disease?"

$$NPV = P(D-/T-) = \frac{TN}{T-}$$

or equivalently, it is a measure answering the question "What are the odds of having a negative test in the absence of disease?"

$$O(NPV) = O_{T-}(D-) = \frac{TN}{FN}.$$

A low proportion of FN impies high the odds of NPV, and practical clinical value of that is discontinued or no treatment.

The derivate of sensitivity and specificity is so called Youden index, and the derivate of predictive values is known as predictive summary index.

Youden index (γ) is a measure of the goodness of the detectability, formally defined by the following formula:

$$\gamma = \text{Se} + \text{Sp} - 1 \in\ <\ -1,\ 1\ >.$$

Predictive Summary Index (ψ), **in construction similar to γ**, and introduced by Linn and Grunau [1], is a measure of the goodness of the predictability in a diagnostic test, and defined as follows:

$$\psi = \text{PPV} + \text{NPV} - 1 \in\ <\ -1,\ 1\ >.$$

Note that, as a derivate of sensitivity and specificity, Youden index γ can be interpreted as four "excess coins", namely:

- γ reflects the excess in the proportion of a *positive* result among patients with the disease versus patients without the disease:

γ = Se - (1 - Sp);

- γ reflects the excess in the proportion of a *negative* result among patients *without* the disease versus patients with the disease:

γ = Sp - (1 - Se);

- γ reflects the excess in the balanced detective accuracy versus the balanced detective error:

$$\gamma = \frac{\text{Se} + \text{Sp}}{2} - \frac{(1 - \text{Se}) + (1 - \text{Sp})}{2};$$

- γ reflects the excess in the joint probabilities of correct detection (positive and negative) versus the joint probabilities of incorrect detection (positive and negative). So it is the gained probability of correct detection information:

γ = Se * Sp - (1 - Se) * (1 - Sp).

Similarly, as a derivate of PPV and NPV, summary predictive index ψ also can be interpreted as four excess coins, namely:

- ψ reflects the *excess* in the proportion of a *positive predictability* goodness among patients with the disease versus patients without the disease:

ψ = PPV - (1 - NPV);

- ψ reflects the excess in the proportion of a *negative predictability* goodness among patients *without* the disease versus patients with the disease:

ψ = NPV - (1 - PPV);

- ψ reflects the excess in the balanced predictive accuracy versus the balanced predictive error:

$$\psi = \frac{\text{PPV} + \text{NPV}}{2} - \frac{(1 - \text{PPV}) + (1 - \text{NPV})}{2};.$$

- ψ reflects the excess in the joint probabilities of correct diagnosis (positive or negative) versus the joint probabilities of incorrect diagnosis. So it is the gained probability of correct diagnosis information:

ψ = PPV * NPV - (1 - PPV) * (1 - NPV).

Diagnostic Odds Ratio (DOR) is the ratio of the odds of the test being positive if the subject has a disease relative to the odds of the test being positive if the subject does not have the disease:

$$\text{DOR} = \frac{O_{D(+)}(T+)}{O_{D(-)}(T+)}.$$

From above we obtain well known relationships

$$\text{DOR} = \frac{\text{Se} * \text{Sp}}{(1 - \text{Se})\,(1 - \text{Sp})} = \frac{\text{PPV} * \text{NPV}}{(1 - \text{PPV})\,(1 - \text{NPV})} = \frac{\text{TP} * \text{TN}}{\text{FN} * \text{FP}}$$

or equivalently, diagnostic odds ratio as the product of the odds of sensitivity and specificity or positive and negative predictive values, i.e.

DOR = O(Se) * O(Sp) = O(PPV) * O(NPV).

By the way, there is another way [2,3], to determine this measure widely used by clinicians. Namely, the ratio $\frac{\text{PPV}}{1 - \text{NPV}}$, i.e relative risk of the disease for an exposure to that for non-exposure, called *positive predictive ratio* (PPR). As it is known, the clinicians commonly prefer to use it, but from a patient point of view usually more preferred seems to be *negative predictive ratio* (NPR), i.e. relative risk of non-disease for the exposure to that for non-exposure, referred as $\frac{1 - \text{PPV}}{\text{NPV}}$. Therefore diagnostic odds ratio is the ratio of these two ratios, PPR relative to NPR.

Results

Recall the determinant of two by two matrix M (detM) is determined as

detM = TP * TN - FN * FP   (2)

Dividing the first and the second column of (1) by D(+), D(-), respectively, we obtain the matrix M_D, based on the concepts of sensitivity and specificity, where

$$M_D = \begin{bmatrix} \text{Se} & 1 - \text{Sp} \\ 1 - \text{Se} & \text{Sp} \end{bmatrix}, \det M_D = \text{Se} + \text{Sp} - 1 = \quad .$$

On the other side, $\det M_D = \dfrac{\det M}{(D+)*(D-)}$, and we have the following relationship

$$\frac{\det M}{(D+)*(D-)} = \gamma.$$

Note that applying notation Se = 1 - β, Sp = 1 - α, equivalent matrix to M_D has the following form

$$\begin{bmatrix} 1 - \beta & \alpha \\ \beta & 1 - \alpha \end{bmatrix},$$

where α is related with I type error (the error probability of falsely classifying a healthy person as diseased), β is related with II type error (the error probability of

falsely classifying a diseased person as healthy). Obviously, the lower α, β, the higher specificity, sensitivity (respectively).

Dividing the first and the second row of (1) by T(+), T(-) respectively, we obtain the matrix M_T, based on the concepts of predictive values, where

$$M_T = \begin{bmatrix} PPV & 1 - PPV \\ 1 - NPV & NPV \end{bmatrix}, \det M_T = PPV + NPV - 1 = \psi.$$

On the other side, $\det M_T = \dfrac{\det M}{(T+)*(T-)}$, and we have the relationship

$$\frac{\det M}{(T+)*(T-)} = \psi.$$

Dividing the first and the second column of (1) by FN, FP respectively, we obtain the matrix M_O, based on the odds of sensitivity and specificity, where

$$M_O = \begin{bmatrix} O(Se) & 1 \\ 1 & O(Sp) \end{bmatrix}, \det M_O = DOR - 1.$$

On the other side, $\det M_O = \dfrac{\det M}{FN*FP}$, and we have the relationship

$$\frac{\det M}{FP*FN} = DOR - 1.$$

Similarly, dividing the first and the second row of the matrix (1) but now by FP, FN respectively, we also obtain the matrix based on the odds of predictive values and with the same its determinant equal to DOR - 1.

Corollary 1: Summarizing all above, we have the relationships which connect the determinant of any dichotomous test matrix with three important indexes of the test

$$DOR = \frac{\det M}{FP*FN} + 1, \tag{3}$$

$$\gamma = \frac{\det M}{(D+)*(D-)}, \tag{4}$$

$$\psi = \frac{\det M}{(T+)*(T-)}. \tag{5}$$

It easy to check that (3) is equivalent to the cross-product of the 2 × 2 diagnostic contingency table, i.e. $DOR = \dfrac{TP+TN}{FN+FP}$.

Corollary 2: Widely used in many fields Matthews correlation coefficient (MCC), also known as the Yule φ, is determined by a little sophisticated relationship [4] as

$$MCC = \frac{TP*TN - FN*FP}{\sqrt{(TP+FN)(FP+TN)(TP+PF)(FN+TN)}} \in <-1, 1>$$

And is closely related to Youden Index γ and Predictive Summary Index ψ via a very simple formula, namely

$$\left|MCC\right| = \sqrt{\gamma*\psi}, \tag{6}$$

Where the sign of MCC is determined by the sign of

detM, and the product γ * ψ on the right side is always nonnegative because γ and ψ are both negative, zero or positive.

Really, by (2) and (4), (5) we immediately obtain

$$MCC^2 = \frac{(\det M)^2}{(D+)*(D-)*(T+)*(T-)} = \frac{\det M}{(D+)*(D-)} * \frac{\det M}{(T+)*(T-)} = \gamma*\psi,$$

and it proves (6).

Generally, Cramer V coefficient equals

$$V = \sqrt{\frac{}{n \cdot m}} \in < 0, 1 >,$$

where $m = \min(p - 1, q - 1)$; n is the total number of observations, p is the number of rows, q is the number of columns in M. Note that V reduces to φ for 2 by 2 matrix.

So for the contingency matrix M = [2 × 2] is fulfilled the relationship

MCC² = φ² = V²,

and $\chi^2 = n * \gamma * \psi$, where $n = (D+) + (D-) = (T+) + (T-)$.

Generally, the measure of correlation is determined by dimension of the matrix M used to employ the chi-square statistic. When M = [2 × 2], then MCC statistic is used. When M = [n × n], n > 2, then Pearson Contingency Coefficient $C = \sqrt{\dfrac{\chi^2}{n + \chi^2}} = \sqrt{\dfrac{\gamma*\psi}{\gamma*\psi+1}}$ is used, and when M = [m × n], m ≠ n, then Cramer's V is used.

Corollary 3: When the determinant is not positive, then a test is worthless or even invalid, and the test can be possibly valuable only when the determinant is positive. More precisely, there are three detection-prediction cases, depending on the value of the confusion matrix determinant:

(i) detM < 0 determines the sufficient condition for a test to be invalid; formally:

detM < 0 < = > (γ, ψ < 0, DOR < 1, PPV < NPV)

(ii) detM = 0 indicates that the study test is uninformative, i.e. it is like tossing a coin to decide; formally:

detM = 0 < = > (γ = ψ = 0, DOR = 1, PPV = NPV)

(iii) detM > 0 is the necessary condition for the test to be possibly useful for diagnostic purposes; formally:

detM > 0 < = > (γ, ψ > 0, DOR > 1, PPV > NPV).

Corollary 4: Youden index plays a specific role in the concept of AUC (area under the receiver operating characteristic curve). Plotting the values of Se, Sp on the graph with the 1 - Sp on the x-axis and Se on the y-axis, it is obtained a trapezoid with vertices O(0, 0), A(1, 0), B(1, 1), C(1 - Sp, Se). The area of the trapezoid OBCD is equal to the sum of areas of two triangles with the base 1 both, and Se, Sp as their altitudes. Therefore the AUC for a single test [5,6] is equal to $\dfrac{\gamma + 1}{2}$.

The distance between the vertex C (1 - Sp, Se) of the trapezoid OBCD and the point (Se, Se) of the straight line y = x is equal to γ and it achieves its maximum. The point C lies on the parallel straight line y = x + γ. It can be seen "anywhere you look, there γ is a cook" (it sounds like a poetry, and it is reality!).

Discussion

The sufficient codition for a diagnostic test to be useless was established as detM ≤ 0. Unfortunately, for the test to be possibly useful only the necessary condition detM > 0 can be determined. Sufficiency depends upon the aim of applying the diagnostic test results. As it is known, if it is ruling out a target condition, then high sensitivity is required; if it is ruling in the condition, then high specificity is needed.

Note that PPV, NPV are not intrinsic to the test and they depend also on Pr. Applying Bayes' Theorem to the formula PPV = P(D+/T+), we obtainwell-known adjusted formula for positive predictive value:

$$PPV = \frac{Se * Pr}{Se * Pr \cdot + (1 - Sp)(1 - Pr)}.$$

If the sample sizes do not reflect the real prevalence of the disease, then PPV should be calculated using the adjusted formula. As it can be checked, we also have

(I) if detM > 0, then PPV > Pr;

(II) the parttial derivative $\frac{\partial PPV}{\partial Pr} > 0$ for all fixed Se, Sp, and PPV is increasing function of Pr. Roughly, the lower Pr, the smaller PPV; the higher Pr, the greater PPV. When Pr is low, then a greater Sp is needed to achieve a higher PPV.

The adjusted formula for negative predictive value is known as

$$NPV = \frac{Sp * (1 - Pr)}{(1 - Se) * Pr \cdot + Sp(1 - Pr)}.$$

Similarly as above, if the sample sizes do not reflect the real prevalence of the disease, then NPV should be calculated using the adjusted formula. Furthermore,

(I) if detM > 0, then NPV > 1 - Pr;

(II) the parttial derivative $\frac{\partial NPV}{\partial Pr} < 0$ for all fixed Se, Sp, and NPV is monotonically decreasing function of Pr. Roughly saying, the higher Pr, the smaller NPV, and the lower Pr, the greater PPV. When a disease is common (Pr is high), then a greater Se is needed to achieve a higher NPV. The illustration of the effect of disease prevalence on PPV and NPV can be found in [7].

In the paper are shown connections between linear algebra on one side and test statistics on the other side. Especially, if a contingency matrix is singular, i.e. it has a determinant of 0, then the test is uninformative; it happens when rows (columns) are proportional. Matrix singularity is the critical level of any diagnostic test. If the test to be possibly effective, its matrix determinant must be positive. This criterion may be useful for decision to do or not to do the test evaluation.

As far as we know, the criterion concluded in Corollary 3 and relationships (3) (4), (5) and (6) are novel in medical literature. Thankfully, a lot of statistical tools are nowadays available and clinicians are not forced to do the math. But thanks to math three of the Five Ws are known: what, when and why.

Future Research

The concepts of the odds and matrix determinant can be also applied to other terms-to likelihood ratios for positive and for negative test, to relative risks, i.e. positive predictive rate and negative predictive rate. By this attitude the criterion for a test to be invalid, uninformative or possibly useful can be extended to those terms.

Disclosure Statement

The authors declare no conflicts of interest.

Funding Statement

Author Contributions

Both authors had equal access to the final study report, made contributions to the development of the manuscript, had final responsibility for the decision to submit, and approved the submitted version.

References

1. Linn S, Grunau PD (2006) New patient-oriented summary measure of net total gain in certainty for dichotomous diagnostic tests. Epidemiol Perspect Innov 3: 11.

2. Mayer D (2010) Essential evidence-based medicine. Cambridge University Press.

3. Norton EC, Dowd BE, Maciejewski ML (2018) Odds ratios-current best practice and use. JAMA 320: 84-85.

4. Chicco D, Jurman G (2020) The advantages of the Matthews Correlation Coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 21: 6.

5. Sokolova M, Japkowicz N, Szpakowicz S (2006) Beyond Accuracy, F-Score and ROC: A family of discriminant measures for performance evaluation. In Proceedings of the AAI'06, 1015-1021.

6. García V, Mollineda RA, Sánchez JS (2009) Index of Balanced Accuracy: A Performance Measure for Skewed Class Distributions. In: Pattern Recognition and Image Analysis. IbPRIA, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 5524: 441-448.

7. Wong HB, Lim GH (2011) Measures of diagnostic accuracy: Sensitivity, Specificity, PPV and NPV. Proceedings of Singapore Healthcare 20: 316-318.