B04901098 電機三 陳立富

統計學報告

Measure of Diagnostic Accuracy

INTRODUCTION

Diagnostic tests are performed to determine the presence or absence of diseases. By means of a gold standard that tells us the true condition status of the patient, we can evaluate the diagnostic accuracy of a test that tells us how good the test is at identifying subjects with and without a disease or condition. This paper describes a variety of diagnostic accuracy measures regarding the intrinsic accuracy, the prevalence of the disease, the consequences of the test's misleading, and the impact of the cognitive and perceptual abilities of the interpreting clinician on the diagnosis.

SUMMARY

	Test Result		
True Condition Status	**Positive(T=1)**	**Negative(T=0)**	**Total**
Present(D=1)	s_1	s_0	n_1
Absent(D=0)	r_1	r_0	n_0
Total	m_1	m_0	N

Term	Formula	Definition	Property	Advantage	Disadvantage
Sensitivity (Se)	$\frac{s_e}{n_1}$	the ability of a test to detect the condition when it is present	a.intrinsic b.affected by the spectrum of disease c.connected to Sp	*	*
Specificity (Sp)	$\frac{r_0}{n_0}$	the ability of a test to exclude the condition in patients without the condition	a.intrinsic b.affected by the spectrum of disease c.connected to Se	*	*
False negative rate (FNR)	$\frac{s_0}{n_1}$	the probability of a test to indicate the absence of the condition in a patient who has it	a.intrinsic b.Se+FNR=1	*	a.delay treatment b.provide false reassurance

Term	Formula	Definition	Property	Advantage	Disadvantage
False positive rate (FPR)	$\frac{r_1}{n_0}$	the probability of a test to indicate the presence of the condition in a patient who doesn't have it	a.intrinsic b.Sp+FPR=1	*	a.lead to unnecessary, perhaps risky, confirmatory test b.incorrect treatment c.false labeling of patients
Probability of a correct test result	$\frac{s_1 + r_0}{N}$	the probability of a test to make a correct diagnosis	a.not intrinsic b.incorporate Sp and Se c.weighted average of Se and Sp with weights equaling to the sample prevalence	a.simple computation	a.affected by the sample prevalence(Gilbert's example) b.report only single pairs of Sp and Se c.FP and FN are considered equally undesirable
Odds ratio	$\frac{S_e/(1-S_e)}{(1-S_p)/S_p}$	the amount of certainty gained after a positive test result divided by the amount of certainty gained after a negative test result	a.intrinsic b.if odds ratio is greater/less than 1, then the odds of a positive test result is greater /less for patients with the condition	*	a.depend on only one decision threshold b.FP and FN are considered equally undesirable
Youden's index	$S_e + S_p - 1$	the difference between the sensitivity and the false positive rate	a.intrinsic b.reflect the likelihood of a positive result among patients with versus without the condition	*	a.depend on only one decision threshold b.FP and FN are considered equally undesirable
ROC curve	*	a plot of a test's sensitivity versus its FPR with each point generated by different decision thresholds	a.intrinsic b.as the Se increases, the Sp decreases c.affected by the spectrum of disease	a.visual representation of accuracy data b.invariant to monotonic transformation c.provide different pairs of Se and Sp with respect to different decision threshold	*

Term	Formula	Definition	Property	Advantage	Disadvantage
ROC area	*	the area bounded by the ROC curve and the coordinate axes	a.intrinsic b.a test's inherent ability to discriminate between patients with versus without the condition 3.global measure of intrinsic accuracy	*	a.may not reflect the relevantly local information
Sensitivity at a fixed FPR	*	the sensitivity at a fixed FPR, written as $Se_{(FPR=e)}$	a.intrinsic	a.good for evaluating a test for a particular application	a.reported Se are often at different FPR b.if FPR is selected after data are examined, some bias can be introduced
Partial ROC area	*	the area under a portion of the ROC curve between two FPRs, written as $A_{(e_1 \leq FPR \leq e_2)}$	a.intrinsic	a.compromise between the ROC curve area and the Se at a fixed FPR b.the variance for its distribution is less than that of $Se_{(FPR=e)}$	a.reported Se are often at different FPR b.if FPR is selected after data are examined, some bias can be introduced c.the variance for its distribution is larger than that of the ROC area
Likelihood ratio	$\frac{P(T=t \mid D=1)}{P(T=t \mid D=0)}$	the ratio of the probability of a particular test result among patients with the condition to the probability of that test result among patients without the condition	a.intrinsic b.reflect the magnitude of evidence that a particular test result provides in favor of the presence of the condition relative to the absence of the condition	*	a.hard to find its standard error and distribution
PPV	$P(D=1 \mid T=1)$	the probability of the disease given a positive test result	a.not intrinsic b.post-test probability	*	a.require the knowledge about the prior probability
NPV	$P(D=0 \mid T=0)$	the probability of the health given a negative test result	a.not intrinsic b.post-test probability	*	b.post-test probability

A measure is intrinsic if it doesn't change as the the prevalence differs.
*** will be discussed in the next topic.**

DISCUSSION

The concepts of the sensitivity and specificity in diagnosis tests can be interpreted in terms of the hypothesis testing. The interesting analogies between those two enable us to apply the approaches used in the hypothesis testing to deal with the measures of diagnostic accuracy and provide more solid insight when analyzing the data. As a consequence, in the following discussion, we will focus on the similarities and connections between the diagnostic accuracy measures and those corresponding terms in the hypothesis testing.

In a binary hypothesis test, there are two hypothetical probability models, H_0 and H_1, and two possible conclusions: accept H_0 as the true model, or accept H_1 as the true model.

In the hypothesis test, two kinds of errors are possible. Statisticians refer to them as the type I errors and the type II errors with the following definitions:

Type I error (False Rejection): reject H_0 when H_0 is true.

Type II error (False Acceptance): accept H_0 when H_0 is false.

If we define the null H_0 and alternative H_1 as follows:

H_0: the condition is not present H_1: the condition is present

Then, the type I error rate is analogous to the FPR and the type II error rate is analogous to the FNR. Statistical power, 1-the type II error rate, is analogous to sensitivity.

To determine which hypothesis to be accepted, we divides S, the sample space of the experiment, into an event space consisting of an acceptance set A_0 and a rejection set $A_1 = A_0^c$. If the observation $s \in A_0$, we accept H_0 . If $s \in A_1$, we reject the hypothesis H_0 and accept H_1.

Thus, $P(A_1 | H_0)$ corresponds to the probability of a Type I error, which is the FPR of a diagnostic test. Similarly, $P(A_0 | H_1)$ corresponds to the probability of a Type II error, which is the FNR of a diagnostic test.

There is a trade-off between $P(A_0 | H_1)$ and $P(A_1 | H_0)$. To understand the trade-off, consider an extreme example in which $A_0 = S$ and $A_1 = \phi$ is the empty set. In this case, $P(A_1 | H_0) = 0$ and $P(A_0 | H_1) = 1$. Now let A_1 expand to include an increasing proportion of the outcomes in S. As A_1 expands, $P(A_1 | H_0)$ increases and $P(A_0 | H_1)$ decreases. At the other extreme, $A_0 = \phi$ and $A_1 = S$, $P(A_1 | H_0) = 1$ and $P(A_0 | H_1) = 0$. Connecting the result to the FPR and FNR, we have that as the FPR increases, the FNR will decrease, in other words, Se=1-FNR increases, which can be seen obviously in the ROC curve. Actually, the ROC curve is the plot of $P(A_1 | H_0)$ versus 1-$P(A_0 | H_1)$. To understand the concept, consider that S=the set of all possible test results(in magnitude), $A_0(t)$=the set of the test results whose magnitude is less than the decision threshold t, $A_1(t) = A_0(t)^c$. As the decision threshold t varies, $P(A_1(t) | H_0)$ and 1-$P(A_0(t) | H_1)$ will vary accordingly, that is, FPR and Se varies with respect to the decision threshold t. The graph of FPR($P(A_1(t) | H_0)$)and Se (1-$P(A_0(t) | H_1)$ is the ROC curve.

Based on the concept, we can explain the area under the ROC curve more clearly.

Suppose $f_1(t_1)=f(t_1 \mid H_1), f_0(t_0)=f(t_0 \mid H_0),\ X_1\ and\ X_0$ denote the random variables of test results among patients with and without the condition respectively.

$$ROC\ Area = \int_0^1 \mathrm{Se}(FPR)dFPR = \int_{\infty}^{-\infty} \mathrm{Se}(t_0)\mathrm{FPR}'(t_0)\,dt_0 = \int_{-\infty}^{\infty}\int_{t}^{\infty} f_1(t_1)f_0(t_0)\,dt_1\,dt_0 = P(X_1 > X_0)$$

Thus, the area of the ROC curve can be interpreted as the probability that the test result of a patient with the condition is more suspicious than the test result of a patient without the condition.

Next, we discuss the application of Bayes Theorem for clinicians in diagnosis.

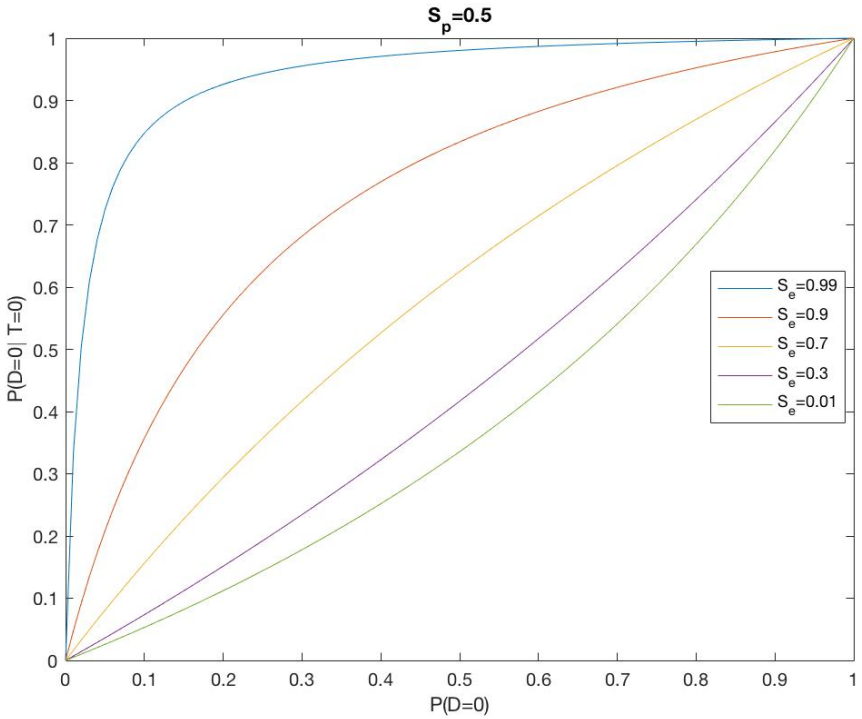
$$P(D = d \mid T = t) = \frac{P(T = t \mid D = d)P(D = d)}{P(T = t \mid D = 0)P(D = 0) + P(T = t \mid D = 1)P(D = 1)}$$

$$PPV = P(D = 1 \mid T = 1) = \frac{P(T = 1 \mid D = 1)P(D = 1)}{P(T = 1 \mid D = 0)P(D = 0) + P(T = 1 \mid D = 1)P(D = 1)} = \frac{S_e \times P(D = 1)}{(1 - S_p) \times P(D = 0) + S_e \times P(D = 1)}$$

$$NPV = P(D = 0 \mid T = 0) = \frac{P(T = 0 \mid D = 0)P(D = 0)}{P(T = 0 \mid D = 0)P(D = 0) + P(T = 0 \mid D = 1)P(D = 1)} = \frac{S_p \times P(D = 0)}{(1 - S_e) \times P(D = 1) + S_p \times P(D = 0)}$$

According to the above formula, when FPR is low, that is, 1-Sp is low, PPV will be nearly 1, which indicates that given a positive test result, the probability that the patient has the disease is much high. In contrast, when FNR is low, that is , 1-Se is low, NPV will be nearly 1, which indicates that given a negative test result, the probability that the patient doesn't have the disease is much high.

We could run the numerical stimulation to demonstrate that Se has a large impact when a test result is negative.



REFERENCES

1.Zhou, X. H., Obuchowaski, N. A., and McClish, D. K. (2002). Statistical Methods in Diagnostic Medicine. Wiley.

2.Karlijn J. van Stralen, Vianda S. Stel, Johannes B. Reitsma, Friedo W. Dekker, Carmine Zoccali, Kitty J. Jager, Diagnostic methods I: sensitivity, specificity, and other measures of accuracy,Kidney International,Volume 75, Issue 12,2009,Pages 1257-1263

3.Roy D. Yates, David J. Goodman. Probability and Stochastic Processes A Friendly Introduction for Electrical and Computer Engineers Second Edition