Listy Biometryczne – Biometrical Letters
Vol. 41 (2004), No. 1, 15-24

Sequential selection for the optimal set of multiple diagnostic tests

Dariusz Radomski[1], Andrzej Jakubiak[2], Piotr Brzeski

[1]Institute of Radioelectronics, [2]Institute of Telecommunications,
Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warszawa

SUMMARY

This paper presents the sequential statistical test used for selection of the most accurate subset of diagnostic tests. The proposed method is based on iterative application of Z-test which compares areas under ROC (AUC) assessed for multiple diagnostic tests. Usage of the nonparametric method for estimation of AUC allows to apply this method for small numbers of patients.

KEY WORDS: diagnostic test, receiver operating characteristic, area under curve, sequential statistical test.

1. Introduction

In the eighties of last century Mac Master introduced the principles of medicine based on evidences. The basic assumptions of this trend say that all clinical decisions are made on the basis of scientific circumstances. It develops application of statistical methods to clinical practices and gives a background for a new medical branch which is called *clinical epidemiology*. One of the main tasks for clinical epidemiologists is to search most effective diagnostic methods for a given disease in respect of health and economic costs. Solving these problems requires the objective assessment which could evaluate accuracy of diagnostic tests. The most popular method for this purpose is utilization of probabilistic parameters which express diagnostic precision such as sensitivity and specificity of a test.

The development of the diagnostic equipment and laboratory tests makes a new problem in diagnostic medicine – finding an optimal set of diagnostic examinations

This work was financed by KBN grant no. 4T11E 007 25

which ensure the best accuracy. The goal of our article is to propose the sequential methods for selection of such a set. In Section 2 we describe the basic model for diagnostic accuracy estimation, in Section 3 we expand this model for multiple tests and in Section 4 we present our strategy for selection of the optimal set ensuring the possible best diagnostic accuracy.

2. Basic probabilistic model for estimation of diagnostic accuracy

2.1. The parameters for diagnostic accuracy measurement

The basic model is used for estimation of accuracy of a single diagnostic test. It also allows comparing diagnostic accuracy between two tests.

Let's assume that D is a dichotomous random variable which represents a disease status, i.e. $D = 0$ expresses the absence of a given disease and $D = 1$ describes the existence of the disease. Similarly, let T is a dichotomous random variable which represents results of a given diagnostic test, with $T = 0$ for negative results and $T = 1$ for positive results. The values of the T variable are denominated based on measurements of new test activity. The "true" status of a disease is determined by the other, reference diagnostic test with possible highest diagnostic accuracy, called the "gold standard test". However, the clinical application of the reference diagnostic test is usually impossible because of its high health or economic costs.

The empirical joint distribution of (D,T) could be presented in the decision table. It is easy to see that the table content could be expressed by the four conditional probabilities: $P(T = 1|D = 1), P(T = 0|D = 0), P(T = 1|D = 0), P(T = 0|D = 1)$. The first one is known as a diagnostic test sensitivity (denoted by S_e) and the second one is its specificity (denoted by S_p).

Table 1. The decision table for a single diagnostic test; (n_{kl}) denotes the number of patients in given subgroups

Disease	Test	
	$T = 0$	$T = 1$
$D = 0$	n_{11}	n_{12}
$D = 1$	n_{21}	n_{22}

These are the most important parameters describing accuracy of a diagnostic test. The remaining parameters are known as false negative and false positive probabilities. Sometimes it is more comfortable to use one parameter which is a combination of

sensitivity and specificity. Thus, we can apply *odds ratio* defined as

$$OR = \frac{S_e}{1 - S_e} : \frac{S_p}{1 - S_p}. \tag{1}$$

The empirical marginal distributions of D and T let us find the estimators of sensitivity and specificity. They have the following form:

$$\hat{S}_e = \frac{n_{22}}{n_{21} + n_{22}}, \qquad \hat{S}_p = \frac{n_{11}}{n_{11} + n_{12}}. \tag{2}$$

The variance of specificity and sensitivity can be expressed in the following manner (Zhou *et al.* 2001)

$$Var(\hat{S}_e) = \frac{S_e\,(1 - S_e)}{n_{21} + n_{22}}, Var(\hat{S}_p) = \frac{S_p\,(1 - S_p)}{n_{11} + n_{12}}. \tag{3}$$

The estimators for variances are obtained by inserting (2) into (3).

The sensitivity and specificity are used for characterization of a new diagnostic test. From clinical point of view, it is very important to assess the predictable values of diagnostic tests which describe probability of a disease based on the results of a diagnostic test. Such probabilities depend on disease prevalence in a given population, denoted by$P\,(D)$. Applying the Bayes theorem, we obtain:

$$\begin{aligned} P\,(D = 1|T = 1) &= \frac{S_e P\,(D)}{S_e P\,(D) + (1 - S_p)(1 - P\,(D))}, \\ P\,(D = 0|T = 0) &= \frac{Sp P\,(D)}{(1 - S_p)\,P\,(D) + S_e(1 - P\,(D))}. \end{aligned} \tag{4}$$

The first expression is known as a positive predictive value (PPV) and the second one is a negative predictive value (NPV). The estimators of the predictive values have the forms:

$$\widehat{PPV} = \frac{n_{22}}{n_{12} + n_{22}}, \widehat{NPV} = \frac{n_{11}}{n_{11} + n_{21}}. \tag{5}$$

They express the so called "post test" parameters which strongly depend on a study population.

2.2. Estimation and comparison of ROC curves

The expressions (2) and (4) were presented under 'the assumption that a diagnostic test gives two values, for example existence or lack of a given clinical symptom. In many applications the values are on the ordinal scale, i.e. we estimate sensitivity and specificity of a diagnostic test for some of its values, e.g. $t_1, t_2, ..., t_k$.

The relation between parameters values and test values is described by the set of the pairs $[S_e(t_i), 1 - S_p(t_i)]$ for $i = 1, ..., k$. This set is called a *receiver operating cha-*

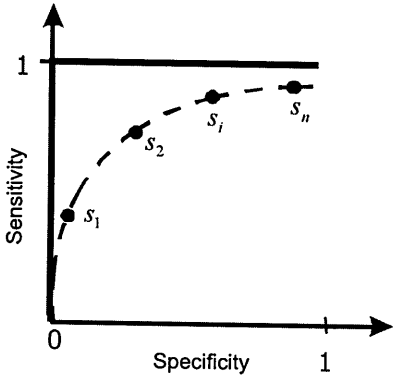

Fig. 1. The example of a ROC curve. The thick line denotes an ideal curve.

racteristic (ROC). When we approximate it by a given curve, we will obtain the ROC curve. This approximation may be performed on the base of traditional methods (e.g. LMS) (Grey *et al.*,1972). However, such approximation is not an efficient estimator of the ROC. Therefore, we recommend the methods introduced by Dorfman and Alf (1968). They assume two latent random variables $\mathcal{T}_0, \mathcal{T}_1$ which represent the values of a diagnostic test obtained for healthy and ill patients, respectively. For a large sample of patients the distributions of $\mathcal{T}_0, \mathcal{T}_1$ are normal, i.e. $\mathcal{T}_i \sim \mathcal{N}(\mu_i, \sigma_i), i = 0.1$. Thus, the estimator of ROC has the following form:

$$[1 - \Phi(t), 1 - \Phi(\hat{b}t - \hat{a})] - \infty < t < \infty, \tag{6}$$

where Φ is the cumulative normal distribution and $\hat{b} = \hat{\sigma}_0/\hat{\sigma}_1$, $\hat{a} = (\hat{\mu}_1 - \hat{\mu}_0)/\hat{\sigma}_1$.These parameters standardize the distribution of the latent variables. The example of ROC curve is shown in Fig.1.

The parameters of the ROC curve and their variances are obtained by ML methods. The algorithm for estimation of their variance (as inverse of Fisher matrix) is presented by Collett (2003).

For estimation of the ROC it is necessary to introduce the next parameter which describes "global" properties of a given statistical test. It is the area under ROC curve, defined as

$$A = \int_{-\infty}^{\infty} S_e(t) d(1 - S_p(t)). \tag{7}$$

When we apply the estimator (6), the area under curve can be estimated as (McClish, 1989):

$$\widehat{A} = \Phi\left(\frac{\hat{a}}{\sqrt{1 + \hat{b}^2}}\right). \tag{8}$$

The variance of (8) is given by the formula;

$$Var(\widehat{A}) = \hat{\beta}^2 Var(\hat{a}) + \hat{\gamma}^2 Var(\hat{b}) + 2\hat{\beta}\hat{\gamma} Cov(\hat{a}, \hat{b}), \tag{9}$$

where

$$\hat{\beta} = \frac{\exp[-\hat{a}^2/2(1+\hat{b}^2)]}{\sqrt{2\pi(1+\hat{b}^2)}}, \hat{\gamma} = \frac{\hat{a}\hat{b}\exp[-\hat{a}^2/2(1+\hat{b}^2)]}{\sqrt{2\pi(1+\hat{b}^2)^3}}.$$

It is very important to note that the above equations present the parametric form of the estimator for the area under ROC. It is an efficient estimator under assumption of the large sample size. When the number of patients is small, the latent variables do not have the normal distributions. Therefore, we propose to use a nonparametric estimator obtained by Hanley and McNeil (1982):

$$\widehat{A} = \frac{1}{n_0 n_1} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \Psi(t_{1i}, t_{0j}), \tag{10}$$

where n_1, n_0 denote, respectively, the number of patients with a disease and without it and t_1, t_0 are the values of a test measured in these groups. Ψ is so called a ranking function defined as:

$$\Psi(x,y) = \begin{cases} 0 & \text{for } y > x, \\ \frac{1}{2} & \text{for } y = x, \\ 1 & \text{for } y < x. \end{cases} \tag{11}$$

Now, we concentrate on the estimation of the variance of the area under ROC curve. To simplify the equation form, let's introduce the two vectors being the components of the variance: $\mathbf{V}^o = [V^o(1) \cdots V^o(n_o)]^T$ and $\mathbf{V}^1 = [V^1(1) \cdots V^1(n_1)]^T$. Theirs coordinates are defined as (Zhou *et al.*, 2003):

$$\begin{aligned} V^1(i) &= \tfrac{1}{n_o} \textstyle\sum_{j=1}^{n_o} \Psi(t_{1i}, t_{oj}), \\ V^o(j) &= \tfrac{1}{n_1} \textstyle\sum_{i=1}^{n_1} \Psi(t_{1i}, t_{oj}). \end{aligned} \tag{12}$$

Thus, the estimator for the variance of (10) has the following form:

$$Var(\widehat{A}) = \frac{1}{n_1} S_1 + \frac{1}{n_o} S_0, \tag{13}$$

where $S_i = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} [V^i(k) - \widehat{A}]^2, i = 0.1.$

Direct comparison of two curves is often too conservative. The value of the area under ROC curve is frequently used to compare diagnostic accuracy of a new test with the reference one. Therefore, it's sufficient to find the difference between the areas under the testing and "gold standard" curve.

Let's assume that A_T denotes the area under ROC for a new test and A_G for a gold standard which were estimated for n_T and n_G patients.

The classical method used for this purpose is to test the following statistical hypothesis:

$$H_0 : A_T = A_G, \tag{14}$$

against the alternative hypothesis:

$$H_1 : A_T \neq A_G.$$

The verification of these hypothesizes is performed by the classical Z-statistic:

$$Z = \frac{A_T - A_G}{S_E[A_T - A_G]}, \tag{15}$$

which has a standard normal distribution under H_0. $S_E[A_T - A_G]$ is a standard error of a difference between two areas. For independent tests, we have $S_E[A_T - A_G] = \sqrt{\frac{1}{n_T}Var(A_T) + \frac{1}{n_G}Var(A_G)}$.

However, the modern biostatistical inference sometimes assumes that equality of two areas under the curve may be too conservative condition. Therefore, Obuchowski (1997) introduced the term: *equivalent areas*. Two areas are equivalent when the different between them belongs to the assumed interval, i.e. $\Delta A \in [\Delta_U; \Delta_L]$. This condition can be transformed to the following hypothesis:

$$H_0 : \Delta A \leq \Delta_L \quad \text{or } \Delta A \geq \Delta_U \tag{16}$$

against alternative hypothesis

$$H_1 : \Delta_U \leq \Delta A \leq \Delta_L.$$

The hypothesis (16) requires usage of two test statistics:

$$Z_1 = \frac{\Delta A - \Delta_L}{S_E[\Delta A]}, \quad Z_2 = \frac{\Delta A - \Delta_U}{S_E[\Delta A]}. \tag{17}$$

Both these statistics have standard normal distribution under H_0.

3. The estimation of diagnostic accuracy in multiple tests

Considerations presented in the second section were concerned with the estimation of diagnostic accuracy for a single medical test. However, modern diagnostic procedures usually contain several, different types of examinations. Thus, there is a need to estimate diagnostic accuracy of multiple tests performed on the same patients.

We limit our description to two multiple tests but the presented method is valid for any number of tests. Thus, we must distinguish two diagnostic rules:

1. the diagnosis is positive when both tests A **and** B are positive; in this case the sensitivity and specificity of multiple two tests will be denoted by $S_{e_{A\cap B}}, S_{p_{A\cap B}}$,
2. the diagnosis is positive when either test A **or** test B are positive; analogously, the sensitivity and specificity will be denoted by $S_{e_{A\cup B}}, S_{p_{A\cup B}}$.

Zhou *et al.* (2002) showed that these parameters could be expressed by the sensitivity and specificity of a single test in the following manner:

$$\begin{aligned} S_{e_{A\cap B}} &= S_{e_A} S_{e_B}, & S_{p_{A\cap B}} &= S_{p_A} + S_{p_B} - S_{p_A} S_{p_B}, \\ S_{e_{A\cup B}} &= S_{e_A} + S_{e_B} - S_{e_A} S_{e_B}, & S_{p_{A\cup B}} &= S_{p_A} S_{p_B} \end{aligned} \tag{18}$$

Moreover, the above formulas do not depend on a diagnostic scheme, i.e., in both the parallel and in the serial scheme the combined parameters are expressed by (18).

The estimation of the sensitivity and specificity of multiple tests allows to find a "global" ROC curve.

4. Selection of the optimal set of diagnostic tests

Let's assume that our problem relies on finding the optimal set of diagnostic tests in the hierarchical diagnostic tree (Fig. 2). This model is common for many clinical situations, for example in ovarian cancer diagnosis. In the beginning a patient has a clinical examination. If there are clinical circumstances, the USG will be performed. If the USG shows an ovarian tumour, a biochemical test will be applied. This model fulfils the rule (i).

Let's assume also that we have m tests with known specificities and sensitivities at each level of diagnosis, estimated in a group consisting of N patients. We want to find the most accurate subset of these tests. The proposed strategy uses the estimation of the multiple test parameters and possibility of hypothesis testing. The strategy goes as follows:

1. Find the best test at a clinical level in the sense of the area under ROC denoted by A_C.
2. Select one of diagnostic tests at level two. In the case of USG examination, it could be the set of picture features.
3. Estimate the "global" ROC for the selected clinical symptoms and the USG test, denoted by $A_{C\cap U}(i)$, $i = 1, ..., m$..

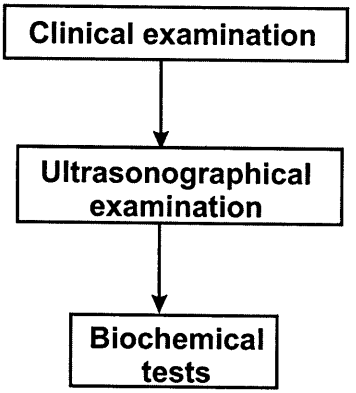

Fig. 2. The three-layer diagnostic model

4. Test the following hypothesis:

$$\begin{aligned} H_0 &: \hat{A}_{C\cap U}(i) = \hat{A}_C \\ H_1 &: \hat{A}_{C\cap U}(i) > \hat{A}_C \end{aligned} \quad ,i = 1, ..., m. \tag{19}$$

5. Use the one-side statistical test:

$$Z = \frac{\hat{A}_{C\cap U}(i) - \hat{A}_C}{S_E[A_{C\cap U}(i) - A_C]}, i = 1, ..., m. \tag{20}$$

Because the results of these diagnostic tests are mutually dependent, the denominator of (20) is described by the formula:

$$S_E[A_{C\cap U} - A_C] = \sqrt{\frac{1}{N}[Var(A_{C\cap U}) + Var(A_C)] + \frac{2}{N^2}Cov(A_{C\cap U}, A_C)}.$$

By analogy to (13), the covariance between two areas under ROC is computed in the following form (DeLong *et al.* 1988):

$$Cov(A_{C\cap U}, A_C) = \frac{1}{N}(C_1 + C_0), \tag{21}$$

where:

$$C_1 = \frac{1}{N-1}\sum_{l=1}^{N}[V^1_{C\cap U}(l) - \hat{A}_{C\cap U}][V^1_C(l) - \hat{A}_C],$$

$$C_0 = \frac{1}{N-1}\sum_{l=1}^{N}[V^0_{C\cap U}(l) - \hat{A}_{C\cap U}][V^0_C(l) - \hat{A}_C].$$

1. Repeat 3 and 4 m times (for all possible USG tests).
2. The above test allows to find a subset $\mathbb{T} = \{1, ..., k\}$ of diagnostic tests which significantly increase diagnostic accuracy in comparison to the clinical examinations.
3. Find the optimal USG test as:

$$i^* = \arg\max_{i\in\mathbb{T}} Z(i). \tag{22}$$

4. Repeat this procedure for biochemical test, replacing A_C by $A^*_{C,U}$ and $A_{C,U}$ by $A_{C,U,B}$ which denotes the area under ROC for the "three-layer" test, i.e. for the optimal clinical-USG test together with a new biochemical test.

The proposed criteria for the most accurate set of tests described by (22) can be expanded to other decision rules. Particularly, when we try also to minimize the economical cost of diagnosis , the "invasive cost" for a patient or the time needed to obtain the result, we can rewrite the expression (22) to the formula form:

$$i^* = \arg\max_{i\in\mathbb{T}} [c_1 Z(i) - c_2\varepsilon(i) - c_3\iota(i) - c_4\tau(i)], \tag{23}$$

where ε, ι, τ are the respective costs and $c_1; ..., c_4$ are the weights in the decision function.

5. Conclusion

The aim of the article was the presentation of the analytical methods for diagnostic accuracy analysis. We generalized the known theory of ROC curve for multiple diagnostic tests and applied it for the selection of the most accurate set of diagnostic tests. A drawback of the proposed method is its sequential character which requires multiple computation of the test statistic. However, the iterative form of our strategy allows to implement it in a SAS procedure which could use the available procedures for computation of the Z-statistic in paired version.

Moreover, the proposed expansion of the decision function allows for the multidimensional selection of optimal diagnostic tests.

References

Zhou X.H., Obuchowski N.A., McClish D.K. (2002) *Statistical Methods in Diagnostic Medicine.* Wiley.

Grey D.R, Morgan B.J.T. (1972). *Some aspects of ROC curve fitting. Normal and logistic models. J. Math. Psych.* **9**, 128-139.

Dorfman D.D, Alf E. (1968). Maximum likelihood estimation of parameters of signal detection theory – a direct solution. *Psychometrika* **33**, 117-124.

Collett D. (2003). *Modelling binary data.* CRC Press.

McClish D.K. (1989). Analyzing a portion of the ROC curve. *Med. Dec. Making* **9**, 190-195.

Hanley J.A., McNeil B.J. (1982). The meaning and use of the area under the receiver operating characteristic curve. *Radiology* **143**, 29-36.

Obuchowski N. (1997). Testing for equivalence of diagnostic tests. *Am. J. Radiol.* **168**, 13-17.

DeLong E., DeLong D., Clarke-Pearsone D. (1988). Comparing the areas under two or more correlated ROC curves. A nonparametric approach. *Biometrics* **44**, 837-845.

Received 20 May 2004; revised 15 October 2004

Sekwencyjny wybór optymalnego zbioru testów diagnostycznych

Streszczenie

W artykule zaproponowano sekwencyjny test statystyczny służący wyborowi optymalnego zestawu testów diagnostycznych, zapewniających najlepszą dokładność diagnostyczną. Metoda ta polega na iteracyjnym stosowaniu testu Z do porównania pól pod krzywymi ROC, wyznaczanymi dla grupy testów diagnostycznych. Wykorzystanie nieparametrycznego estymatora pola pod krzywą ROC umożliwia zastosowanie tej metody dla małej grupy pacjentów.

Słowa kluczowe: test diagnostyczny, charakterystyka operacyjna odbiornika, obszar pod krzywą, sekwencyjny test diagnostyczny.