Turk J Med Sci
35 (2005) 21-24

EXPERIMENTAL / LABORATORY STUDIES

Estimation of Sensitivity and Specificity for Clustered Data

Yasemin GENÇ[1], Derya GÖKMEN[1], Ersöz TÜCCAR[1], Banu YAĞMURLU[2]

[1]Department of Biostatistics, Faculty of Medicine, Ankara University, 06100 Ankara - Turkey

[2]Department of Radiodiagnostics, Faculty of Medicine, Ankara University, 06590 Ankara - Turkey

Received: July 12, 2004

Abstract: The most commonly used measurements for evaluating the accuracy of diagnostic tests in binary data are sensitivity and specificity. Sensitivity refers to the ability of a test to detect patients with some specific disease. Specificity describes how well test abnormality is restricted to those persons who have the disease in question. The variance estimators of these measurements are usually obtained using a binomial estimator, although there are several other methods. However, under certain conditions more than one observation can be taken from a subject and analyses are performed on these observations. This type of data structure is called "clustered data". For calculating sensitivity, specificity and related variances in clustered data, conventional statistical methods with the assumption that all observations are independent will not be valid. The most commonly used methods for clustered data are "ratio estimator", "within-cluster correlation estimator" and "weighted estimator". In this paper, for calculating diagnostic measurements and their variances, the above-mentioned methods are described. In the example of magnetic resonance angiography, which is used for the diagnosis of renal artery stenosis, the results of a binomial estimator are compared with those of the other 3 methods. Although there were no significant differences between sensitivity estimates, variance estimates obtained from the binomial estimator were higher than the other estimates.

Key Words: Sensitivity, Specificity, Clustered data

Introduction

Diagnostic tests are used for revealing the true conditions of the subjects in a heterogeneous population consisting of diseased and disease-free individuals. The accuracy of a diagnostic test can be measured by comparing the test results to the true condition of the patient. The true condition of the patient (diseased/disease-free) can be determined by means of a "gold standard" test whose accuracy has been confirmed.

The most commonly used measurements in evaluating the accuracies of diagnostic tests in binary data are sensitivity and specificity. Sensitivity refers to the ability of the test to detect patients with some specific disease. Specificity describes how well test abnormality is restricted to those persons who have the disease in question. The variance estimators of these measurements are usually obtained by using binomial estimator, although there are several other methods.

However, in some studies more than one measurement can be obtained from a subject and analyses are carried out on these measurements. This type of data structure is called "clustered data" and in this type of data, subjects form the clusters. Examples of clustered data can be given from studies in the fields of radiology, periodontology and ophthalmology. For example, in order to assess the accuracy of positron emission tomography (PET) for diagnosing hyperparathyroidism, Neumann et al. (1) conducted measurements on 51 parathyroid glands taken from 21 patients (1 to 4 glands were taken from every patient). In this study, patients constituted the clusters and the measurements that were carried out on the glands of the same patient constituted the diagnostic units of the study (DUOS) within each cluster.

Conventional statistical methods which assume that all observations are independent will not be valid for

calculating sensitivity, specificity and their variances in clustered data. Because ignoring the existing relationship between the several measurements obtained from a subject will result in a bias in variance estimators for sensitivity and specificity. When the correlation between such measurements for a subject is positive (negative), then the variance will be underestimated (overestimated) (2). While calculating these measurements for clustered data, methods taking into account the correlation should be used in order to overcome the problem of biased estimators. The most common methods used for this purpose are "ratio estimator", "within-cluster correlation estimator" and "weighted estimator". There are several methods for estimating sensitivity, specificity and their variances (e.g., correlated binomial model, generalized estimated equations and beta-binomial estimator) in addition to the estimators mentioned previously. However, these methods obtain the sensitivity, specificity and their variances by iterative ways. Furthermore, they require a model assumption and a cluster size of more than 30, and, therefore, cannot be used very often.

In this paper, "ratio estimator", "within-cluster correlation estimator" and "weighted estimator" are introduced for calculating sensitivity, specificity and their variances for clustered data. With the example of magnetic resonance angiography (MRA) used for the diagnosis of renal artery stenosis (RAS), the results of a binomial estimator are compared with those of the 3 other methods.

Materials and Methods

Method for Independent Data

Binomial Estimator: For independent measurements, the most commonly used method for estimating sensitivity and specificity is the binomial probability model. Assuming that the number of true positive test results is represented by y, the number of diseased sites by n, the number of true negative test results by z and the number of disease-free sites by k, the sensitivity, specificity and their variance estimators can be given as follows:

$$\text{Sensitivity (Sens)} = \frac{y}{n}, \quad \text{Var}_{BK}\,(\text{Sens}) = \frac{\text{Sens}\,(1\text{-Sens})}{n} \qquad (1)$$

$$\text{Specificity (Spe)} = \frac{z}{k}, \quad \text{Var}_{BK}\,(\text{Spe}) = \frac{\text{Spe}\,(1\text{-Spe})}{k} \qquad (2)$$

Methods for Clustered Data

The main notations for the methods discussed below are given in Table 1.

Table 1. Display of clustered binary data.

Cluster	Number of True Positives (y_i)	Number of True Negatives (z_i)	Number of Diseased Sites (n_i)	Number of Disease-free Sites (k_i)
1	y_1	z_1	n_1	k_1
2	y_2	z_2	n_2	k_2
…	…	…	…	…
N	y_N	z_N	n_N	k_N

Ratio Estimator: In this method, which was proposed by Rao and Scott (3), the estimators of sensitivity and specificity are the same as those of the binomial estimator. Under the assumption that N is large and the average cluster size is constant, the variance estimators of sensitivity and specificity are as follows (4):

$$\text{Var}_{RE}\,(\text{Sens}) = \frac{N\sum_{i=1}^{N}(y_i - n_i\,\text{Sens})^2}{(N-1)\left(\sum_{i=1}^{N} n_i\right)^2}, \qquad \text{Var}_{RE}\,(\text{Spe}) = \frac{N\sum_{i=1}^{N}(z_i - k_i\,\text{Spe})^2}{(N-1)\left(\sum_{i=1}^{N} k_i\right)^2} \qquad (3)$$

Within-Cluster Correlation Estimator: In this method, proposed by Donner and Klar (5), the estimators of sensitivity and specificity are the same as those obtained from the binomial and ratio estimators. However, a different approach is proposed for calculating the variance estimator. According to this approach, variance is estimated by taking into account the correlation between observations within each cluster.

Let us assume that $\hat{\rho}$ is the within-cluster correlation coefficient. Donner and Klar (5) suggested the use of an analysis of variance estimator for the estimation of the within-cluster correlation coefficient. This estimate can be obtained by treating the outcome of each patient, coded as 0 or 1, as a continuous variable. Then let L_n be the number of patients with m sites. Thus, the variance estimators for sensitivity and specificity are

$$Var_{IC}(Sens) = \frac{B_{Sens}\, Sens\,(1\text{-}Sens)}{\sum_{i=1}^{N} n_i},$$

$$B_{Sens} = \frac{\sum nL_n \left[1 + (n-1)\ \hat{p}_{Sens}\right]}{\sum nL_n} \qquad (4)$$

$$\hat{p}_{Sens} = 1 - \frac{\sum_{i=1}^{N} \frac{y_i\,(n_i - y_i)}{n_i}}{N\,(\bar{n} - 1)\,Sens\,(1\text{-}Sens)}$$

$$Var_{IC}(Spe) = \frac{B_{Spe}\, Spe\,(1\text{-}Spe)}{\sum_{i=1}^{N} k_i},$$

$$B_{Spe} = \frac{\sum nL_n \left[1 + (n-1)\ \hat{p}_{Spe}\right]}{\sum nL_n} \qquad (5)$$

$$\hat{p}_{Spe} = 1 - \frac{\sum_{i=1}^{N} \frac{z_i\,(k_i - z_i)}{k_i}}{N\,(\bar{k} - 1)\,Spe\,(1\text{-}Spe)}$$

A positive correlation coefficient indicates that if one observation tests as a true positive (negative) test result in that cluster, the probability of testing other observations for the same cluster as true positives (negatives) will increase. A zero within-cluster correlation coefficient indicates that one observation testing positive (negative) does not affect the probabilities of other observations testing true positive (negative).

Weighted Estimator: Although the within-cluster correlation estimator demonstrates that a patient with k sites reveals less information than k patients each with one site, the ratio estimator assigns the same weight to both of these cases regardless of the number of patients they include. To overcome the problem of loss of efficiency resulting from similar weighting, Lee and Dubin (2) suggested the use of a weighted estimator. Letting w_i be the weight assigned to cluster i, the estimators of sensitivity and specificity are obtained as follows:

$$Sens_{WE} = \frac{1}{N}\sum_{i=1}^{N} w_i n_i\, Sens_i \ , \ Spe_{WE} = \frac{1}{N}\sum_{i=1}^{N} w_i k_i\, Spe_i \qquad (6)$$

When the weight $w_i = N / \sum_{i=1}^{N} n_i$ is used for sensitivity and the weight $w_i = N / \sum_{i=1}^{N} k_i$ is used for specificity, the estimators of sensitivity and specificity are the same as those obtained from other methods. Lee and Dubin (2) suggested the use of the weight $w_i = 1/n_i$ for sensitivity and $w_i = 1/k_i$ for specificity. When these weights are used, sensitivity and specificity can be obtained as

$$Sens_{WE} = \frac{\sum_{i=1}^{N} Sens_i}{N}\ , \ Sens_{WE} = \frac{\sum_{i=1}^{N} Spe_i}{N} \qquad (7)$$

being the simple average of positive (negative) rates of each cluster. This weighting system avoids the dominance of a few large clusters in the sample. Then the variances of sensitivity and specificity can be computed as follows:

$$Var_{WE}(Sens) = \frac{\sum_{i=1}^{N} (Sens_i - Sens_{WE})^2}{N\,(N-1)}\ ,$$

$$Var_{WE}(Spe) = \frac{\sum_{i=1}^{N} (Spe_i - Spe_{WE})^2}{N\,(N-1)} \qquad (8)$$

Example

In a study conducted by the Ankara University School of Medicine Department of Radiodiagnostics, 30 renal arteries of 21 patients (each patient had 1 to 4 arteries) diagnosed with RAS were examined by 3-dimensional contrast enhanced dynamic renal MR angiography (MRA) and catheter angiography (DSA).

In this study, while DSA was taken as the gold standard test, patients formed the clusters and the renal arteries thereof formed the diagnostic units of study. The sensitivity of the MRA diagnostic test and its variance was obtained by **"ratio estimator"**, **"within-cluster correlation estimator"** and **"weighted estimator"** and these were compared to the results of a binomial

Table 2. The distribution of data.

Cluster No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
y_i	1	1	1	1	1	1	1	2	1	1	2	1	3	1	1	2	2	1	2	1	0
n_i	1	1	1	1	1	1	1	2	1	1	2	1	4	1	1	2	2	2	2	1	1

Table 3. Estimates of sensitivity for MRA diagnostic test.

Method	Sensitivity	Variance
Binomial Estimator	0.900	0.0030
Ratio Estimator	0.900	0.0025
Within-Cluster Correlation Estimator	0.900	0.0017
Weighted Estimator	0.917	0. 0028

estimator. The number of renal arteries diagnosed by DSA as stenotic (n_i) and the number of true positive test results (y_i) in a sample of 21 patients are given in the table above:

While there is no significant difference between sensitivity estimates, the variance estimate obtained from the binomial estimator is higher than the other estimates because of the negative within-cluster correlation coefficient ($\hat{p}$ = -0.543).

Discussion

In certain studies conducted in the field of medicine, it has been observed that, more than one measurement is obtained from each subject; however, analyses are conducted by using standard methods assuming that these measurements are independent of each other. Yet it has been shown that not taking into account the within-cluster correlation caused variance estimates with biases.

In the analysis of clustered data, while the ratio estimator and within-cluster correlation estimator estimate the sensitivity (specificity) like the binomial estimator, a different approach is used in the weighted estimator. In the ratio estimator, the variance estimate of sensitivity (specificity) is based on the approach used in cluster sampling by Cochran (4). While in the within-cluster correlation estimator correlation between observations within each cluster is taken into consideration, variance is estimated by weighting proportional to cluster sizes in the weighted estimator method.

As a result of a simulation study performed by Ahn (6), it was recommended to use the binomial estimator when the within-cluster correlation coefficient ($\hat{p}$) was zero, and to use the ratio estimator, within-cluster correlation estimator or weighted estimator when $\hat{p}$ = 0.2 or $\hat{p}$ = 0.4. When $\hat{p}$ = 0.6, the weighted estimator was recommended.

Although there are several methods used in clustered data to obtain measurements and their variances, the methods introduced in this paper are preferred to complex methods because of their ease of use.

Corresponding author:
Derya GÖKMEN
Department of Biostatistics,
Faculty of Medicine,
Ankara University,
06100 Ankara - Turkey
E-mail: dgokmen2001@yahoo.com

References

1. Neumann DR, Esselstyn CB, Go RT et al. Comparison of double-phase ^{99m}Tc-sestamibi with ^{123}I-^{99m}Tc-sestamibi subtraction spect in hyperparathyroidism. Am J Roentgenol 169: 1671-1674, 1997.
2. Lee E, Dubin N. Estimation and sample size considerations for clustered binary responses. Stat Med 13: 1241-1252, 1994.
3. Rao JNK, Scott AJ. A simple method for the analysis of clustered binary data. Biometrics 48: 577-585, 1992.
4. Cochran WG: Sampling Techniques, Wiley. New York 1977, pp: 247-49.
5. Donner A, Klar N. Confidence interval construction for effect measures arising from cluster randomization trials. J Clin Epidemiol 46: 123-131, 1993.
6. Ahn C. Statistical methods for the estimation of sensitivity and specificity of site-specific diagnostic tests. Journal of Periodontal Research 32: 351-354, 1997.