GLOBAL HYPOTHESIS TEST TO COMPARE THE LIKELIHOOD RATIOS OF MULTIPLE BINARY DIAGNOSTIC TESTS WITH IGNORABLE MISSING DATA

Ana E. Marín Jiménez[1], José A. Roldán Nofuentes[2],

[1] Department of Statistic and O. R., Campus of Ceuta, University of Granada, Spain . anamarin@ugr.es

[2] Biostatistics, School of Medicine, University of Granada, Spain. jaroldan@ugr.es

Universidad de Granada

There are many parameters to assess the accuracy of a binary diagnostic test, the fundamental are sensitivity and specificity. In the presence of partial disease verification, the comparison of two or more diagnostic tests cannot be carried out ignoring those individuals who are not verified with the gold standard, since the results obtained are affected by verification bias. Other parameters to assess the accuracy of a binary diagnostic test are the likelihood ratios.

Positive and negative likelihood ratios are parameters to assess the accuracy of a binary diagnostic test, and only depend on the sensitivity and the specificity of diagnostic test. The likelihood ratios quantify the increase in terms of knowledge of the disease presence through the application of the diagnostic test. A global hypothesis test is studied to simultaneously compare the positive and negative likelihood ratios of more than two binary diagnostic tests in the presence of partial verification of the disease when the mechanism of missing data is ignorable.

INTRODUCTION

- ***LR+*** is the ratio between the probability of a positive test result in individuals with the disease (*Se*) and the probability of a positive result in individuals without the disease (*1-Sp*).
- ***LR-*** is the ratio between the probability of a negative test result for individuals with the disease (1-Se) and the probability of a negative test result for individuals without the disease (Sp) .

$$LR^{+}=\frac{Se}{(1-Sp)} \text{ and } LR^{-}=\frac{(1-Se)}{Sp}$$

The test through which we can determine the **true disease status** is called the ***gold standard.***

VERIFICATION PARTIAL

The **gold standard** is **not** usually **applied** to **all** of the **patients** in the sample

⟶ which leads to the problem of **partial disease verification**

The evaluation of a binary diagnostic test cannot be carried out neglecting patients whose disease status is unknown

⟶ since the estimators obtained are affected by **verification bias**

We study a global hypothesis test when the **missing data mechanism is ignorable ≈ MAR (Rubin, 1972)**

BINARY DIAGNOSTIC TESTS

•D random variable that models the result of the gold standard

- D = 1 the patient has the disease
- D = 0 the patient does not have the disease

• T random variable that models the result of the diagnostic test

- T = 1 the result of the test is positive
- T = 0 the result of the test is negative

• V random variable that models the verification process

- V = 1 the patient is verified
- V = 0 the patient is not verified

GLOBAL HYPOTHESIS TEST

Let the probabilities

$$\phi_{i_1,\ldots,i_J}=P(V=1,D=1,T_1=i_1,T_2=i_2,\ldots,T_J=i_J)$$
$$\varphi_{i_1,\ldots,i_J}=P(V=1,D=0,T_1=i_1,T_2=i_2,\ldots,T_J=i_J)$$
$$\gamma_{i_1,\ldots,i_J}=P(V=0,T_1=i_1,T_2=i_2,\ldots,T_J=i_J)\ i_j=0,1$$

Verifying that

$$\sum_{i_1,\ldots,i_J=0}^{1}\phi_{i_1,\ldots,i_J}+\sum_{i_1,\ldots,i_J=0}^{1}\varphi_{i_1,\ldots,i_J}+\sum_{i_1,\ldots,i_J=0}^{1}\gamma_{i_1,\ldots,i_J}=1$$

$$\boldsymbol{\omega}=\left(\phi_{1,\ldots,1},\ldots,\phi_{0,\ldots,0},\varphi_{1,\ldots,1},\ldots,\varphi_{0,\ldots,0},\gamma_{1,\ldots,1},\ldots,\gamma_{0,\ldots,0}\right)^{T}$$

Be a vector sized $3\cdot 2^J$ whose components aret the probabilities

Tabla 1. Frecuencies observed to compare three *BDTs* in the presence of partial disease verification

	$T_1=1$				$T_1=0$			
	$T_2=1$		$T_2=0$		$T_2=1$		$T_2=0$	
	$T_3=1$	$T_3=0$	$T_3=1$	$T_3=0$	$T_3=1$	$T_3=0$	$T_3=1$	$T_3=0$
$V=1$								
$D=1$	s_{111}	s_{110}	s_{101}	s_{100}	s_{011}	s_{010}	s_{001}	s_{000}
$D=0$	r_{111}	r_{110}	r_{101}	r_{100}	r_{011}	r_{010}	r_{001}	r_{000}
$V=0$	u_{111}	u_{110}	u_{101}	u_{100}	u_{011}	u_{010}	u_{001}	u_{000}
Total	n_{111}	n_{110}	n_{101}	n_{100}	n_{011}	n_{010}	n_{001}	n_{000}

Assuming that the mechanism of missing data is ignorable, the *LR+* and the *LR-* of the *j*-th diagnostic test are written in terms of the the previous probabilities, as

$$LR_j^{+}=\frac{(1-p)\left(\sum_{\substack{i_1,\ldots,i_J=0\\ i_j=1}}^{1}\frac{\phi_{i_1,\ldots,i_J}\eta_{i_1,\ldots,i_J}}{\phi_{i_1,\ldots,i_J}+\varphi_{i_1,\ldots,i_J}}\right)}{p\left((1-p)-\sum_{\substack{i_1,\ldots,i_J=0\\ i_j=0}}^{1}\frac{\varphi_{i_1,\ldots,i_J}\eta_{i_1,\ldots,i_J}}{\phi_{i_1,\ldots,i_J}+\varphi_{i_1,\ldots,i_J}}\right)} \text{ and } LR_j^{-}=\frac{(1-p)\left(p-\sum_{\substack{i_1,\ldots,i_J=0\\ i_j=1}}^{1}\frac{\phi_{i_1,\ldots,i_J}\eta_{i_1,\ldots,i_J}}{\phi_{i_1,\ldots,i_J}+\varphi_{i_1,\ldots,i_J}}\right)}{p\left(\sum_{\substack{i_1,\ldots,i_J=0\\ i_j=0}}^{1}\frac{\varphi_{i_1,\ldots,i_J}\eta_{i_1,\ldots,i_J}}{\phi_{i_1,\ldots,i_J}+\varphi_{i_1,\ldots,i_J}}\right)}$$

The logarithm of the likelihood function is $l=\sum_{i_1,\ldots,i_J=0}^{1}s_{i_1,\ldots,i_J}\log\left(\phi_{i_1,\ldots,i_J}\right)+\sum_{i_1,\ldots,i_J=0}^{1}r_{i_1,\ldots,i_J}\log\left(\varphi_{i_1,\ldots,i_J}\right)+\sum_{i_1,\ldots,i_J=0}^{1}u_{i_1,\ldots,i_J}\log\left(\gamma_{i_1,\ldots,i_J}\right)$

The maximum likelihood estimators of the probabilities are

$$\hat{\phi}_{i_1,\ldots,i_J}=\frac{s_{i_1,\ldots,i_J}}{n},\quad \hat{\varphi}_{i_1,\ldots,i_J}=\frac{r_{i_1,\ldots,i_J}}{n} \text{ and } \hat{\gamma}_{i_1,\ldots,i_J}=\frac{u_{i_1,\ldots,i_J}}{n}$$

Where $p=\sum_{i_1,\ldots,i_J=0}^{1}\frac{\phi_{i_1,\ldots,i_J}\eta_{i_1,\ldots,i_J}}{\phi_{i_1,\ldots,i_J}+\varphi_{i_1,\ldots,i_J}}$ is the prevalence of the disease and $\eta_{i_1,\ldots,i_J}=\phi_{i_1,\ldots,i_J}+\varphi_{i_1,\ldots,i_J}+\gamma_{i_1,\ldots,i_J}$

The global hypothesis test for the simultaneous comparison of the predictive values of more than two binary diagnostic tests is

$$H_0: LR_1^{+}=LR_2^{+}=\ldots=LR_J^{+} \text{ and } LR_1^{-}=LR_2^{-}=\ldots=LR_J^{-}$$
$$H_1: \text{at least one equality is not.}$$

or $H_0:\boldsymbol{\psi\eta}=0$ vs $H_1:\boldsymbol{\psi\eta}\neq 0$

$\boldsymbol{\psi}$ → Matriz sized (J-1) x J

For J=2 $\boldsymbol{\psi}=\begin{pmatrix}1&-1&0&0\\0&0&1&-1\end{pmatrix}$

For J=3 $\boldsymbol{\psi}=\begin{pmatrix}1&-1&0&0&0&0\\0&1&-1&0&0&0\\0&0&0&1&-1&0\\0&0&0&0&1&-1\end{pmatrix}$

The **contrast statistic** for the **global hypothesis test** is

$$\hat{\boldsymbol{\eta}}^{T}\boldsymbol{\psi}^{T}\left(\boldsymbol{\psi}\hat{\Sigma}_{\hat{\boldsymbol{\eta}}}\boldsymbol{\psi}^{T}\right)^{-1}\boldsymbol{\psi}\hat{\boldsymbol{\eta}}\xrightarrow[n\to\infty]{}\chi^2_{2(J-1)}$$

In the case of the overall test being significative to the error α , the investigation of the causes of this significance is carried out by comparing the positive (negative) likelihood ratio for each pair of diagnostic tests and applying any method for multiple comparison (for example, the Holm method or Hochberg's method).

SIMULATION EXPERIMENTS

Monte Carlo simulation experiments were carried out to study the type I error and the power of the global hypothesis test and a comparison was other method of multiple comparison, when simultaneously comparing the LRs of two and of three binary diagnosctic tests respectively. These experiments consisted in the generation of 5000 random samples with multinomial distributions of different sizes.

$Se \text{ and } Sp=\{0.70,0.75,\ldots,0.95\}$ and $p=\{0.10,0.20,0.30,0.40,0.50\}$

If J=2 the verification probabilities are

$(\lambda_{11}=0.70,\lambda_{10}=\lambda_{01}=0.40,\lambda_{00}=0.10)$ $(\lambda_{11}=0.95,\lambda_{10}=\lambda_{01}=0.60,\lambda_{00}=0.30)$

If J=3 the verification probabilities are

$(\lambda_{111}=0.70,\lambda_{110}=0.40,\lambda_{101}=0.40,\lambda_{100}=0.25,\lambda_{011}=0.40,\lambda_{010}=0.25,\lambda_{001}=0.25,\lambda_{000}=0.05)$

$(\lambda_{111}=1,\lambda_{110}=0.80,\lambda_{101}=0.80,\lambda_{100}=0.40,\lambda_{011}=0.80,\lambda_{010}=0.40,\lambda_{001}=0.40,\lambda_{000}=0.20)$

From the results obtained in the simulation experiments the following conclusions are reached:
The global hypothesis test based on the chi-square distribution has the behavior of an asymptotic hypothesis test (from a sample size, the type I error fluctuates around the nominal error). Generally, the type I error fluctuates around the nominal error (especially for $n\geq 1000$) and the type I error is lower than the nominal error for samples of a smaller size.
In general terms, the power of the global test is very high (higher than 80%-90%), depending on the prevalence and the verification probabilities.

APPLICATION

The results obtained were applied to the diagnosis of coronary stenosis, a disease that consists of the obstruction of the coronary artery and its diagnosis can be made through a dobutamine echocardiography, a stress echocardiography or through a *CT* scan, and as the gold standard a coronary angiography is used. As the coronary angiography can cause different reactions in individuals (thrombosis, heart attack, infections, etc.), not all of the individuals are verified with the coronary angiography. We were applying the three diagnostic tests and the gold standard (T1 : dobutamine ecocardiography; T2: stress echocardiography; T3: *CT* scan) to a sample of 2455 males over 45 and when applying the coronary angiography (*D*) only to a subset of these individuals. This study was carried out in two phases: in the first phase, the three diagnostic tests were applied to all of the individuals; and in the second phase, the coronary angiography was applied only to a subset of these individuals depending only on the results of the three diagnostic tests. Therefore, in this example it can be assumed that the missing data mechanism is ignorable, and therefore the results previous can be applied. The values of the estimators of the *LRs* are ,

$$LR_1^{+}=5.31,\ LR_2^{+}=3.04,\ LR_3^{+}=7.61,\ LR_1^{-}=0.13,\ LR_2^{-}=0.33 \text{ and } LR_3^{-}=0.09$$

and therefore we reject the joint equality of the *LRs*. In order to investigate the causes of the significance, the marginal hypothesis tests are solved and it holds that

$H_0: LR_1^{+}=LR_2^{+},\ z=6.24,\ two\ sided\ p-value=4.47\times10^{-13}$
$H_0: LR_1^{+}=LR_3^{+},\ z=3.30,\ two\ sided\ p-value=0.001$
$H_0: LR_2^{+}=LR_3^{+},\ z=7.29,\ two\ sided\ p-value=3.06\times10^{-13}$
$H_0: LR_1^{-}=LR_2^{-},\ z=7.53,\ two\ sided\ p-value=5.15\times10^{-14}$
$H_0: LR_1^{-}=LR_3^{-},\ z=1.77,\ two\ sided\ p-value=0.077$
$H_0: LR_2^{-}=LR_3^{-},\ z=9.19,\ two\ sided\ p-value=0,$

and applying the Bonferroni, Holm or Hochberg methods, it holds that the three positive likelihood ratios are different and the largest is that of the *CT* scan, followed by that of the dobutamine echocardiography and finally that of the stress echocardiography. Regarding the negative likelihood ratios, no significant differences were found between that of the dobutamine echocardiography and that of the *CT* scan; whilst the negative likelihood ratio of the stress echocardiography is significantly larger than that of the dobutamine echocardiography and that of the *CT* scan.

CONCLUSIONS

We propose the following method for simultaneously comparing the predictive values of multiple binary diagnostic tests with ignorable missing data:

1. Solve the global hypothesis test based on the chi-square distribution to an error rate α.
2. In the case of the overall test being significative to the error α, the investigation of the causes of this significance is carried out by comparing the positive (negative) likelihood ratios for each pair of diagnostic tests solving the marginal hypothesis tests and applying any method for multiple comparison (for example, the Holm method or Hochberg's method).

REFERENCES

1. Begg, C.B., Greenes, R.A., (1983). *Assessment of diagnostic tests when disease verification is subject to selection bias*. Biometrics 39, 207-215.
2. Harel, O., Zhou, X.H., (2007). *Multiple imputation for the comparison of two screening tests in two-phase Alzheimer studies*. Statistics in Medicine 26, 2370–2388.
3. Roldán Nofuentes, J.A., Luna del Castillo, J.D., Marín Jiménez, A.E..(2010) . *Comparison of the accuracy of multiple binary tests in the presence of partial disease verification.* Journal of Statistical Planning and Inference 140, 2504-2519.
4. Torrance-Rynard, V.L., Walter, S.D., (1997). *Effects of dependent errors in the assessment of diagnostic test performance*. Statistics in Medicine 16, 2157-2175.
5. Zhou, X.H., (1993). *Maximum likelihood estimators of sensitivity and specificity corrected for verification bias*. Communication in Statistics – Theory and Methods 22, 3177-3198.
6. Zhou, X., (1994). *Effect of verification bias on positive and negative predictive values.* Statistics in Medicine 13, 1737-1745.