The Korean Communications
in Statistics Vol. 10 No. 2, 2003
pp. 567-574

Estimation of Predictive Value of a Positive Test from a Screening Test[1]

Hyun Chul Shin[2], Sang Gue Park[3], and Yong Hee Kim[4]

Abstract

The estimation problem of predictive value of a positive test(PVP), which is assessing the accuracy of a screening test is considered. Score methods discussed by Gart and Nam(1988) are proposed for constructing confidence interval for PVP. The simulation studies are conducted in evaluating the proposed methods and existing approximate ones.

Keywords : Screening tests, Specificity, Sensitivity, Predictive Value of a Positive test, Score method, Skewness-Corrected Score method.

1. Introduction

The increased use of screening tests for drug use of antibodies to the AIDS virus has raised questions about the reliability of the results of these procedures. These concerns arise because the prevalence of persons with the disease in the general population is far less than that in a pre-screened group. For instance, if there are persons in a high risk category for the disease in question, there may be high fraction of false positive classifications among the test results.

The sensitivity and specificity of the screening test has been widely used to evaluate the accuracy of the test. However, test results could be highly wrong even though these rates are quite high(Fleiss(1973)). The predictive value of a positive test result(PVP), $\Psi = \Pr$(disease condition positive | positive test), is currently regarded as the measure of accuracy of the

1) This research is supported in parts by Seoul National University BK21 Aid Program.
2) School of Pharmacy, Rutgers University, Piscataway, New Jersey, 08854-8020, USA
E-mail : hcbig@email.rci.rutgers.edu
3) Professor, Department of Statistics, Chung-Ang University, Seoul, 156-756, Korea.
E-mail : spark@cau.ac.kr
4) Department of Statistics, Chung-Ang University, Seoul, 156-756, Korea.
E-mail : smilehee@unitel.co.kr

screening test. Gastwirth(1987) discussed the issues of this problem and provided a large sample approximation to its standard error.

We propose a score method for constructing the confidence interval of PVP, which is motivated by Gart and Nam(1988). We derive the score statistic and examine the empirical coverage probabilities of the proposed method under various situations through simulation studies.

2. Notation and Model

Screening tests are aimed to determine whether a person belongs to the class(D) of people who have specific disease. The test result indicates that a person who is a member of this class will be denoted by S, and $\overline{S}$ for those who are non-members. The sensitivity(specificity) of the test is the probability that a person having(not having) the disease is correctly diagnosed. We note that the sensitivity η of the test is $P(S \mid D)$ and the specificity θ of the test is $P(\overline{S} \mid \overline{D})$. Also let π be $P(D)$, the prevalence of disease in the population tested and p be $P(S)$, the proportion of positive test. Since p could be written as

$$p = P(S) = \pi P(S \mid D) + (1-\pi) P(S \mid \overline{D}),$$

one could estimate the prevalence π by using p such as

$$\pi = \frac{p+\theta-1}{\eta+\theta-1}.$$

In this paper, our interest is to know the probability, $P(D \mid S)$ that a person whom the test indicates as having the disease actually has it. It could be written as

$$P(D \mid S) = \frac{P(D \wedge S)}{P(S)} = \frac{\pi\eta}{p}. \tag{2.1}$$

This probability $P(D \mid S)$, say Ψ, is called the predictive value of a positive test. Suppose that a sample of n persons are tested, and x individuals of n are found to be diseased, one can estimate $\hat{p}$ by x/n. One can estimate the sensitivity and the specificity of the screening test by testing individuals from two independent classes. Let x_1 and x_2 denote the numbers of positive responses and negative responses from tested persons n_1 known to have the disease and n_2 known to be disease-free, respectively. The natural estimates of η and θ are $\hat{\eta} = x_1/n_1$ and $\hat{\theta} = x_2/n_2$. Substituting $\hat{\pi}$ for π in (2.1) yields the estimator of $P(D \mid S)$; namely,

$$\Psi = \frac{\pi\eta}{p} = \frac{\eta(p+\theta-1)}{(\eta+\theta-1)p}. \tag{2.2}$$

From the equation (2.2), a natural estimator of PVP is

$$\hat{\Psi} = \frac{\hat{\pi}\hat{\eta}}{\hat{p}} = \frac{\hat{\eta}(\hat{p}+\hat{\theta}-1)}{(\hat{\eta}+\hat{\theta}-1)\hat{p}} \tag{2.3}$$

and its variance is

$$Var(\hat{\Psi}) = [\frac{\eta(1-\theta)}{p(\eta+\theta-1)}]^2 \frac{p(1-p)}{p^2 n} + [\frac{\pi(1-\theta)}{p(\eta+\theta-1)}]^2 \frac{\eta(1-\eta)}{n_1} + [\frac{\eta(1-\pi)}{p(\eta+\theta-1)}]^2 \frac{\theta(1-\theta)}{n_2}.$$

Gastwirth showed that $\hat{\Psi}$ has asymptotic normal distribution. Here we provide an approximate confidence limits for Ψ which is based on a normal approximation using (2.3). We could adopt well-known Fieller(1954) type confidence limits for Ψ. The quantity $\{\eta(p+\theta-1)\} - \Psi\{(\eta+\theta-1)p\}$ is asymptotically normally distributed with mean 0 and variance

$$(1+\Psi^2)[\eta^2\frac{\theta(1-\theta)}{n} + p^2\frac{\eta(1-\eta)}{n_1}] + \Psi^2[\theta^2\frac{\theta(1-\theta)}{n_2} + \theta^2\frac{p(1-p)}{n} + \frac{p(1-p)}{n}] + \eta^2\frac{\theta(1-\theta)}{n_2} + \theta^2\frac{\eta(1-\eta)}{n_1} + \frac{\eta(1-\eta)}{n_1}. \tag{2.4}$$

One could obtain the quadratic equation of Ψ, which is based on (2.4), and construct the confidence limits of Ψ.

3. Approximate Confidence Limits Based on Score Statistic

In this section, we suggest an interval estimator of predictive value of a positive test based on score statistic. For simplicity, let $\theta_c = 1-\theta$ and $\hat{\theta_c} = x_{2c}/n_2$, where $x_{2c} = n_2 - x_2$.

Consider the log-likelihood function L, which is the sum of the logarithms of three independent binomials with parameters p, η, and θ_c. If we let $p = \eta\theta_c/(\Psi\theta_c + \eta - \Psi\eta)$, the likelihood may be written as function of Ψ, η and θ_c where Ψ is the parameter of interest :

$$L = x\ln p + (n-x)\ln(1-p) + x_1\ln\eta + (n_1-x_1)\ln(1-\eta) + x_{2c}\ln\theta_c + (n_2-x_{2c})\ln(1-\theta_c). \tag{3.1}$$

The score for Ψ is then

$$S_\Psi(\Psi,\eta,\theta_c) = \frac{\partial L(\Psi,\eta,\theta_c)}{\partial\Psi} = \frac{(x-np)}{p(1-p)}\frac{\partial p}{\partial\Psi}. \tag{3.2}$$

Also we have that

$$\frac{\partial L(\Psi,\eta,\theta_c)}{\partial\eta} = \frac{x_1-n_1\eta}{\eta(1-\eta)} + \frac{(x-np)}{p(1-p)}\frac{\partial p}{\partial\eta}, \tag{3.3}$$

and

$$\frac{\partial L(\Psi,\eta,\theta_c)}{\partial\theta_c}=\frac{x_{2c}-n_2\theta_c}{\theta_c(1-\theta_c)}+\frac{(x-np)}{p(1-p)}\frac{\partial p}{\partial\theta_c}. \tag{3.4}$$

We note that some partial derivatives are

$$P=\frac{\partial p}{\partial\Psi}=\frac{\eta^2\theta_c-\theta_c^2\eta}{(\Psi\theta_c+\eta-\eta\Psi)^2},$$

$$Q=\frac{\partial p}{\partial\eta}=\frac{\theta_c^2\Psi}{(\Psi\theta_c+\eta-\eta\Psi)^2},$$

$$R=\frac{\partial p}{\partial\theta_c}=\frac{\eta^2-\eta^2\Psi}{(\Psi\theta_c+\eta-\eta\Psi)^2}.$$

If we consider the log-likelihood, the information matrix has elements as follows :

$$I_{\Psi\Psi}=E(-\frac{\partial^2L}{\partial\Psi^2})=\frac{n}{p(1-p)}P^2,$$

$$I_{\Psi\eta}=E(-\frac{\partial^2L}{\partial\Psi\partial\eta})=\frac{n}{p(1-p)}PQ,$$

$$I_{\eta\eta}=E(-\frac{\partial^2L}{\partial\eta^2})=\frac{n_1}{\eta(1-\eta)}+\frac{n}{p(1-p)}Q^2,$$

$$I_{\eta\theta_c}=E(-\frac{\partial^2L}{\partial\eta\partial\theta_c})=\frac{n}{p(1-p)}QR,$$

$$I_{\theta_c\theta_c}=E(-\frac{\partial^2L}{\partial\theta^2})=\frac{n_2}{\theta_c(1-\theta_c)}+\frac{n}{p(1-p)}R^2,$$

$$I_{\Psi\theta_c}=E(-\frac{\partial^2L}{\partial\Psi\partial\theta_c})=\frac{n}{p(1-p)}PR.$$

It follows from above results that its variance is given by

$$Var\{S_\Psi(\Psi,\eta,\theta_c)\}=I_{\Psi\Psi}-\frac{1}{I_{\eta\eta}I_{\theta_c\theta_c}-I^2_{\eta\theta_c}}\times(I^2_{\eta\Psi}I_{\theta_c\theta_c}+I^2_{\theta_c\Psi}I_{\eta\eta}-2I_{\Psi\eta}I_{\eta\theta_c}I_{\theta_c\Psi})$$

$$=nn_1n_2P^2\times\{n_1n_2(1-p)+nn_2Q^2\eta(1-\eta)+nn_1R^2\theta_c(1-\theta_c)\}^{-1}.$$

Thus the approximate $100(1-\alpha)\%$ confidence limits of the predictive value of positive test Ψ are the two solutions to the equation

$$\frac{S_\Psi(\Psi,\tilde{\eta},\tilde{\theta_c})^2}{Var(S_\Psi(\Psi,\tilde{\eta},\tilde{\theta_c}))}=z^2_{\alpha/2}. \tag{3.5}$$

The equation (3.5) can be solved by iteration. To eliminate the nuisance parameters η and θ_c, we use the MLEs, $\tilde{\eta}$ and $\tilde{\theta_c}$, which are obtained by NEWTON-RAPHSON method by using (3.3) and (3.4).

Although the performance of the score method is better than other methods, it might have the skewed tail probabilities as Gart and Nam(1988) suggested. So we shall consider correcting the score method for skewness. The general theory of score method permits the correcting of the resulting confidence coefficients for both bias and skewness. Consider the following normal deviate

$$Z(\Psi) = \frac{S_\Psi(\Psi, \eta, \theta_c)}{\sqrt{Var(S_\Psi(\Psi, \eta, \theta_c))}} .$$

Bartlett(1953) considered the statistic

$$T(\cdot) = S_\Psi(\cdot) - (I_{\Psi\eta} \quad I_{\Psi\theta_c}) \begin{pmatrix} I_{\eta\eta} & I_{\eta\theta_c} \\ I_{\theta_c\eta} & I_{\theta_c\theta_c} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial L}{\partial \eta} \\ \frac{\partial L}{\partial \theta_c} \end{pmatrix}.$$

The third moment of $T(\cdot)$ is

$$\mu_3[T(\cdot)] = \frac{n(1-2p)}{p^2(1-p)^2} X^3 - \frac{n_1(1-2\eta)}{\eta^2(1-\eta)^2} Y^3 - \frac{n_2(1-2\theta_c)}{\theta_c^2(1-\theta_c)^2} Z^3,$$

where

$$X = P - \frac{Q}{DET}(I_{\Psi\eta} I_{\theta_c\theta_c} - I_{\Psi\theta_c} I_{\theta_c\eta}) - \frac{R}{DET}(I_{\Psi\theta_c} T_{\eta\eta} - I_{\Psi\eta} I_{\eta\theta_c}),$$

$$Y = \frac{1}{DET}(I_{\Psi\eta} I_{\theta_c\theta_c} - I_{\Psi\theta_c} I_{\theta_c\eta}),$$

$$Z = \frac{1}{DET}(I_{\Psi\theta_c} I_{\eta\eta} - I_{\Psi\eta} I_{\eta\theta_c}),$$

$$DET = I_{\eta\eta} I_{\theta_c\theta_c} - I^2_{\eta\theta_c}.$$

Since

$$\mu_3[S_\Psi(\cdot)] = \mu_3[T(\cdot)],$$

the skewness is

$$\gamma_1[S_\Psi(\cdot)] = \frac{\mu_3[S_\Psi(\cdot)]}{Var[S_\Psi(\cdot)]^{3/2}} = \frac{\mu_3[T(\cdot)]}{Var[T(\cdot)]^{3/2}} . \qquad (3.6)$$

Therefore the skewness-corrected interval is based on the

$$Z_S(\Psi) = Z(\Psi) - \frac{\tilde{\gamma}_1(\Psi)(Z^2_{\alpha/2} - 1)}{6},$$

where $\tilde{\gamma}_1$ is γ_1 in (3.6) with $\tilde{\eta}$ and $\tilde{\theta}_c$ substituted for η and θ_c. The skewness corrected limits are the two solutions to the equation

$$Z_S(\Psi) = \pm Z_{\alpha/2},$$

where the plus and minus signs indicate the lower and upper limits.

4. Simulation Studies and Concluding Remark

We conduct a Monte Carlo simulation study to evaluate the proposed score method of constructing an approximate confidence interval of PVP. We simulate with the score method suggested by Gart and Nam (1988) as well as Gastwirth(1987)'s method and well-known Fieller (1954) type method.

We carry out the simulation study as following setting ;

(1) (n, n_1, n_2) : 100 to 1000

(2) (η, θ) : 0.7 to 0.95

(3) π : 0.01 to 0.05

We present the results of the simulation study calculating actual coverage probabilities. These results are based on 10000 repetitions by IMSL FORTRAN subroutines.

Table 1 shows the empirical coverage probabilities of PVP under the relatively high sensitivity and specificity(i.e., 0.9 to 0.95). From the table we could observe that the score method and Skewness-corrected score method maintain the nominal levels well. However, Gastwirth's method tends to underestimate the nominal levels and Fieller type method does not maintain the given nominal levels.

Table 2 shows the empirical coverage probabilities of PVP under relatively now sensitivity and specificity(i.e., 0.7 to 0.8). The proposed score methods maintain the nominal levels well, but Gastwirth's method and Fieller type method still do not maintain the nominal levels.

In this paper, we proposed the score methods for constructing and approximate confidence interval for the predictive value of the positive test result from a screening test. Since the current methods do not maintain the nominal levels under the moderate sample sizes, it could lead the wrong conclusions from their statistical inferences. However, the confidence limits based on the proposed scored methods ensure the nominal tail percentages of the confidence interval and it would lead to recommend the use of the score methods under moderate sample sizes.

It might be also useful to note that there are some Bayesian methodologies for estimation of accuracy of tests. Johnson and Gastwirth(1991) have discussed Bayesian inference for screening tests and Gastwirth et al(1991) have also discussed Bayesian methodology for analyzing of AIDS data. One might be interested in comparing the performances between the score method and Bayesian method.

Reference

[1] Bartlett, M. S. (1953). Approximate confidence intervals, II. More than one unknown parameter. *Biometrika,* 40, 306-317.

[2] Fieller, E. C. (1954). Some problem in interval estimation. *Journal of Royal Statistical Society, Series B*, 16, 175-185.

[3] Fleiss, J. L. (1973). Statistical Methods for Rates and Proportions. *John Wiley and Sons, New York.*

[4] Gart, J .J. and Buck, A. A. (1966). Comparison of a screening test and a reference test in epidemiologic studies : II. A probabilistic model for the comparison of diagnostic tests. *American Journal of epidemiology*, 83, 593-602.

[5] Gart, J. J. and Nam, J. (1988). Approximate interval estimation of the ratio of binomial parameters :A Review and Corrections for Skewness. *Biometrics,* 44, 323-338.

[6] Gastwirth, J. L. (1987). The statistical precision of medical screening process : Application to polygraph and Aids antibodies test data. *Statistical Science*, 2, 213-238.

[7] Gastwirth, J. L., Johnson, W. O. and Reneau, D. M. (1991). Bayesian Analysis of Screening Data: Application to AIDS in Blood Donors. *The Canadian Journal of Statistic*s, 19, 135-150.

[8] Johnson, W. O. and Gastwirth, J. L. (1991). Bayesian Inference for Medical Screening Tests:Approximations used for the Analysis of Acquired Immune Deficiency Syndrome. Journal of *the Royal Statistical Society. Series B*, 53, 427-439.

[9] Rogan, W. J. and Gladen, B. (1978). Estimating prevalence from the results of a screening test. *American Journal of Epidemiology*, 107, 71-76.

[Received April 2003, Accepted July 2003]

TABLE 1. Actual coverage percentage for 95% confidence interval for Ψ

		$n=1000$, $n_1=100$, $n_2=100$		$n=1000$, $n_1=250$, $n_2=250$		$n=1000$, $n_1=500$, $n_2=500$		$n=1000$, $n_1=1000$, $n_2=1000$	
	π	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05
	Score	95.2	95.5	95.3	95.7	95.1	95.3	95.2	95.1
$\eta=0.90$	Skew-adj score	95.2	95.7	95.3	95.6	95.1	95.3	95.2	95.2
$\theta=0.90$	Gastwirth	94.2	94.2	94.1	94.3	94.1	94.0	94.4	94.5
	Fieller	89.3	89.2	90.8	90.7	91.9	91.0	93.7	93.8
	Score	95.0	95.5	95.0	94.8	95.6	96.0	95.4	95.3
$\eta=0.90$	Skew-adj score	95.0	95.7	95.0	95.0	95.6	96.0	95.4	95.4
$\theta=0.95$	Gastwirth	94.0	94.2	94.1	94.2	94.2	94.3	94.4	94.4
	Fieller	90.1	90.2	91.0	91.0	92.0	91.8	93.8	93.7
	Score	95.1	95.1	95.5	95.5	94.7	94.9	95.0	95.0
$\eta=0.95$	Skew-adj score	95.1	95.1	95.5	95.5	94.7	94.9	95.0	95.1
$\theta=0.90$	Gastwirth	94.0	94.2	94.1	94.0	94.3	94.3	94.4	94.3
	Fieller	90.5	90.8	90.9	90.9	91.1	91.9	93.9	93.3
	Score	94.4	95.2	94.9	95.0	95.3	96.0	95.4	96.1
$\eta=0.95$	Skew-adj score	94.5	95.2	94.9	95.0	95.2	96.0	95.4	96.1
$\theta=0.95$	Gastwirth	93.9	93.8	93.7	93.6	94.2	94.4	94.4	94.5
	Fieller	85.4	85.6	90.7	90.9	92.1	92.6	93.9	93.7

TABLE 2. Actual coverage percentage for 95% confidence interval for Ψ

		$n=1000$, $n_1=100$, $n_2=100$		$n=1000$, $n_1=250$, $n_2=250$		$n=1000$, $n_1=500$, $n_2=500$		$n=1000$, $n_1=1000$, $n_2=1000$	
	π	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05
	Score	95.0	95.0	94.8	94.8	94.9	95.0	95.1	95.0
$\eta=0.70$	Skew-adj score	95.0	95.0	94.8	94.8	94.8	95.0	95.1	95.0
$\theta=0.70$	Gastwirth	93.7	93.9	93.9	94.0	94.2	94.3	94.3	94.3
	Fieller	89.7	89.5	90.7	90.6	91.3	91.2	93.4	93.4
	Score	95.0	95.3	95.3	95.5	95.4	95.2	94.8	95.0
$\eta=0.70$	Skew-adj score	95.0	95.3	95.2	95.5	95.4	95.2	94.8	95.0
$\theta=0.80$	Gastwirth	94.1	94.1	94.1	94.2	94..0	94.0	94.4	94.4
	Fieller	91.0	91.0	92.2	92.2	92.2	92.0	93.5	93.6
	Score	95.0	95.0	94.8	94.9	94.8	94.7	94.8	94.9
$\eta=0.80$	Skew-adj score	95.0	95.0	94.8	94.9	94.8	94.7	84.8	94.9
$\theta=0.70$	Gastwirth	93.5	93.2	94.3	94.2	94.2	94.2	94.0	94.1
	Fieller	90.5	90.4	91.4	91.4	91.1	91.0	93.3	93.2
	Score	95.2	95.3	95.3	95.1	95.1	95.4	94.9	95.1
$\eta=0.80$	Skew-adj score	95.2	95.2	95.3	95.1	95.1	95.4	94.9	95.1
$\theta=0.80$	Gastwirth	93.9	93.9	94.0	94.1	94.2	94.3	94.3	94.3
	Fieller	88.1	88.1	90.2	90.2	92.1	92.3	93.5	93.3