Journal of Korean
Data & Information Science Society
2005, Vol. 16, No. 2, pp. 419~428

An Improved Estimator of PPV from the Screening Test[1)]

Sang-Gue Park[2)] · Jiyun Choi[3)]

Abstract

The screening test is increasingly being used for predicting future disease in the person screened and has raised concerns about reliability of the result of its procedure. We propose an improved estimator of the confidence interval for the positive predictive value(PPV) in screening test by simply taking inverse $\sinh$ transformation comparing to Gastwirth(1987) estimator and show its efficiency through the simulation study.

Keywords : positive predictive value, screening test, sensitivity, specificity

1. Introduction

Screening test refers to the application of a medical procedure or test to people who as yet have no symptoms of a particular disease, for the purpose of determining their likelihood of having the disease. The screening test procedure itself does not diagnose the illness. Those who have a positive result from the screening test will need further evaluation with subsequent diagnostic tests or procedures. The goal of screening is to reduce morbidity or mortality from the disease by detecting diseases in their earliest stages, when treatment is usually more successful.

The accuracy of a diagnostic test is quantified by two different conditional probabilities, namely the sensitivity and the specificity of the test. The sensitivity

---

1) This research was supported by the Chung-Ang Research Grants in 2004.

2) Professor, Department of Statistics, Chung-Ang University, Seoul, 156-756, Korea
E-mail : spark@cau.ac.kr

3) Ph.D Candidate, Department of Statistics, Chung-Ang University, Seoul, 156-756, Korea
E-mail : jiyun1023@hotmail.com

(η) of the test is defined as the probability that a person who truly has the disease(D) correctly receives a positive test result(+). The specificity(θ) of the test is defined as the probability that a person who is truly healthy($\overline{D}$) correctly receives a negative test result(−). Since sensitivity and specificity are probabilities, their values will always fall between 0 and 1. The closer the sensitivity is to 1, the more accurate the test is in identifying diseased individuals, and the closer the specificity is to 1, the more accurate the test is in identifying healthy individuals. While perfect diagnostic tests (with a sensitivity of 1 and a specificity of 1) are the ideal in theory, they are not realized in practice. The definitions of sensitivity and specificity in conditional probability notation are :

$$\eta = P(+ \mid D), \tag{1.1}$$

$$\theta = P(- \mid \overline{D}). \tag{1.2}$$

Sensitivity and specificity pose a problem in describing the accuracy of diagnostic tests in the field. Sensitivity assumes that one has first identified a group of diseased individuals, and then seeks to see what percentage the diagnostic test can correctly identify as diseased. Likewise, specificity assumes that one has first identified a group of healthy individuals, and then seeks to see what percentage the diagnostic test can correctly identify as healthy. But now in practice, the "opposite" happens. A person with symptoms presents himself to a physician, the physician administers a diagnostic test, the test result comes back from the lab, and then the physician asks: "Given the test result is positive, what is the probability that my patient really has the disease?" Or, "given the test result is negative, what is the probability that my patient really is free of the disease?" The answers to these two questions are, respectively, the positive predictive value(PPV : Ψ_P) and the negative predictive value(NPV : Ψ_N). The positive predictive value of the test is defined as the probability that a person who receives a positive test result truly has the disease. The negative predictive value of the test is defined as the probability that a person who receives a negative test result is truly healthy. Again, since positive predictive value and negative predictive value are probabilities, their values will always fall between 0 and 1. The closer the positive predictive value is to 1, the more accurate the test is in identifying diseased individuals, and the closer the negative predictive value is to 1, the more accurate the test is in identifying healthy individuals. The definitions of the positive predictive value and the negative predictive value in conditional probability notation are :

$$\Psi_P = P(D \mid +) = \frac{\eta\pi}{p_+}, \tag{1.3}$$

$$\Psi_N = P(\overline{D} \mid -) = \frac{\theta(1-\pi)}{1-p_+}, \tag{1.4}$$

where the prevalence $\pi = P(D)$ is defined in the population of interest and $p_+ = P(+)$ is the probability of positive response in the screening test.

Remark 1. Equations (1.3) and (1.4) can be written as other form when prevalence is not available from other source.

$$\Psi_P = P(D \mid +) = \frac{\eta(p_+ + \theta - 1)}{p_+(\eta + \theta - 1)}, \tag{1.5}$$

$$\Psi_N = P(\overline{D} \mid -) = \frac{\theta(\eta - p_+)}{(1-p_+)(\eta + \theta - 1)}, \tag{1.6}$$

since the prevalence can be rewritten as

$$\pi = \frac{p_+ + \theta - 1}{\eta + \theta - 1}. \tag{1.7}$$

Schoonmaker, Bagley and Scanlan(2002) recently discussed FDA's current regulation of assessing the screening medical device invented for some disease. They emphasized the importance of PPV calculation for its safety and needs to develop better and robust estimator.

Since the sensitivity and specificity of the screening test usually are not perfect and estimated from some finite samples, the calculation of the PPV has to depend on the sample variations and the prediction based on the sample estimator could mislead the result. Thus the estimation of Ψ_P and Ψ_N on the basis of the screening test has long been of interest to many scientists since Gastwirth(1987) suggested that the standard error of the PPV estimator should be considered to interpret the efficacy and reliability of the screening test.

Recently, Brown, Cai and DasGupta(2001) discussed the general problem of constructing confidence interval for a binomial proportion and suggested the commonly used methods, which are normally given in standard statistical textbooks, are very dangerous and other forms of approximated methods should be provided. Since Gastwirth(1987) estimator is based on several independent binomial proportions, his estimator is also at risk and could be implemented in various ways. This is the motivation of the present study. We would discuss the current standard estimation methods of Ψ_P and propose a simple, but more efficient estimator of Ψ_P in section 2. In conclusion, we would present a comparative simulation study and concluding remark for the newly proposed estimator in section 3.

2. Interval Estimation of Ψ_P

2.1 The case that the prevalence is known

Natural estimators of Ψ_P in (1.3) with the known prevalence from other source is

$$\hat{\Psi}_P = \frac{\hat{\eta}\pi}{\hat{p_+}}, \tag{2.1}$$

where $\hat{p_+}$ is obtained from the patients with positive response, among whom n_1 patients actually receive the screening test, $\hat{\eta}$ is estimators of the sensitivity of screening test based on the sample size n_2. Further assume that all these two samples are drawn independently.

This is the same problem of constructing of so-called relative risk $\psi_{\eta,p_+} = \eta/p_+$ from two independent samples. Computing confidence intervals for the relative risk has been important inferential procedures in statistical analysis of categorical data and so many methods have been proposed in the literature. However, if the proportions are close to 0 or 1, then all the methods have their own problems even in large samples. Gart and Nam(1988) gave an extensive discussion about this problem.

The simplest $100(1-\alpha)\%$ confidence interval for ψ_{η,p_+} is

$$\widehat{\psi_{\eta,p_+}} \pm z_{\alpha/2}\widehat{\psi_{\eta,p_+}}\sqrt{\frac{1}{n_1\hat{p_+}} + \frac{1}{n_2\hat{\eta}} - \frac{1}{n_1} - \frac{1}{n_2}}, \tag{2.2}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ percentile point of the standard normal distribution. This is called the Wald type confidence interval for ψ_{η,p_+}, since it results from inverting the Wald test for ψ_{η,p_+}.

One might get a similar approximate confidence interval for ψ_{η,p_+} by hiring a log-transformation such as

$$\exp\left\{ln\ \widehat{\psi_{\eta,p_+}} \pm z_{\alpha/2}\sqrt{\frac{1}{n_1\hat{p_+}} + \frac{1}{n_2\hat{\eta}} - \frac{1}{n_1} - \frac{1}{n_2}}\right\}. \tag{2.3}$$

This is called logit Wald type confidence interval for ψ_{η,p_+}. Gart and Nam(1988) proposed several approximate confidence intervals based on the likelihood score method better than (2.2) and (2.3), but they required an iterative calculation that

could be computationally laborious, and could be difficult with certain data sets. Thus these would not be discussed further in this paper.

Recently Newcombe(2001) proposed a method to construct an approximate confidence interval for single proportion by the inverse $\sinh$ transformation. He derived a $100(1-\alpha)\%$ confidence interval of a logit of two proportions like

$$logit\ p^* \pm 2\sinh^{-1}\left(\frac{z_{\alpha/2}}{2}\, s.e.(logit\ p^*)\right),$$

where $s.e.$ means a standard error of estimator $logit\ p^*$ and $z_{\alpha/2}$ is the upper $\alpha/2$ percentile point of the standard normal distribution. From this one can get an approximate confidence interval for ψ_{η,p_+} is

$$\exp\left\{\ln\ \widehat{\psi_{\eta,p_+}} \pm 2\ \sinh^{-1}\left(\frac{z_{\alpha/2}}{2}\sqrt{\frac{1}{n_1\widehat{p_+}}+\frac{1}{n_2\hat{\eta}}-\frac{1}{n_1}-\frac{1}{n_2}}\right)\right\}. \tag{2.4}$$

Usually continuity correction is hired when the confidence interval for relative risk is calculated. This would help the performance of extreme cases as well as approximation itself. Usual treatment is adding of 1/2 of a success and 1/2 of a failure to the sample implying the estimator

$$\widehat{\Psi_{\eta,p_+}} = \frac{\hat{\eta}^*}{\widehat{p_+}^*} = \frac{(x_2+1/2)/(n_2+1)}{(x_1+1/2)/(n_1+1)} \text{ and}$$

$$\sqrt{\frac{1}{x_1+1/2}+\frac{1}{x_2+1/2}-\frac{1}{n_1+1}-\frac{1}{n_2+1}}.$$

After constructing the confidence intervals such as (2.2), (2.3) and (2.4) multiplying π to the above resulting intervals leads to the final confidence interval for Ψ_P.

2.2 The case that the estimator of prevalence is available

Natural estimators of Ψ_P in (1.3) with the estimator of the prevalence is

$$\widehat{\Psi}_P = \frac{\hat{\eta}\hat{\pi}}{\widehat{p_+}}, \tag{2.5}$$

where $\hat{\pi}$ is an estimator of prevalence from the sample size n_3. We also assume that all these three samples are drawn independently.

The asymptotic mean and variance of $\widehat{\Psi}_P$ can be obtained by Taylor expansion

$$AE(\hat{\Psi}_P) = \Psi_P,$$
$$AVAR(\hat{\Psi}_P) = \Psi_P^2\left(\frac{1-p_+}{p_+ n_1} + \frac{1-\eta}{\eta n_2} + \frac{1-\pi}{\pi n_3}\right).$$

The asymptotic distribution of $\hat{\Psi}_P$ is well discussed by Gastwirth(1987).

The approximate $100(1-\alpha)\%$ Wald-type confidence interval for Ψ_P is

$$\hat{\Psi}_P \pm z_{\alpha/2}\hat{\Psi}_P\sqrt{\frac{1}{n_1\widehat{p_+}} + \frac{1}{n_2\hat{\eta}} + \frac{1}{n_3\hat{\pi}} - \frac{1}{n_1} - \frac{1}{n_2} - \frac{1}{n_3}}. \tag{2.6}$$

The logit Wald-type confidence interval for Ψ_P is

$$\exp\left\{\ln\hat{\Psi}_P \pm z_{\alpha/2}\sqrt{\frac{1}{n_1\widehat{p_+}} + \frac{1}{n_2\hat{\eta}} + \frac{1}{n_3\hat{\pi}} - \frac{1}{n_1} - \frac{1}{n_2} - \frac{1}{n_3}}\right\}. \tag{2.7}$$

The approximate confidence interval for Ψ_P by inverse $\sinh$ transformation is

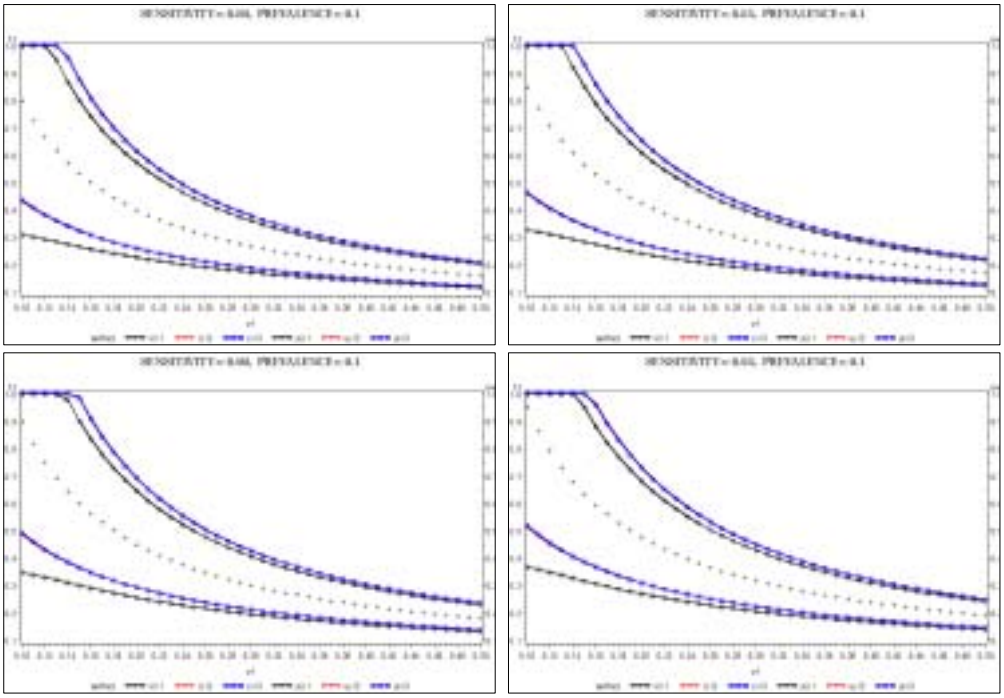
$$\exp\left\{\ln\hat{\Psi}_P \pm 2\sinh^{-1}\left(\frac{z_{\alpha/2}}{2}\sqrt{\frac{1}{n_1\widehat{p_+}} + \frac{1}{n_2\hat{\eta}} + \frac{1}{n_3\hat{\pi}} - \frac{1}{n_1} - \frac{1}{n_2} - \frac{1}{n_3}}\right)\right\}. \tag{2.8}$$

The continuity correction by adding 1/2 to both of a success and a failure would be considered as well.

Figure 1 shows the 95% confidence intervals of PPV, (2.6), (2.7) and (2.8), as $\widehat{p_+}$ based on 100 sample observations increases. The prevalence is estimated 0.1 based on 1,000 sample observations under several estimated sensitivities such as 0.80, 0.85, 0.90 and 0.95, which are based on 1,000 sample observations, respectively.

All four graphs show consistent result, which is confidence interval based on (2.6) is skewed to the left comparing to confidence intervals (2.7) and (2.8) when $\widehat{p_+}$ is low. This phenomenon never changes even in high $\widehat{p_+}$ even if the interval (2.6) is close to the intervals (2.7) and (2.8) as $\widehat{p_+}$ increases. This suggests the further simulation study to examine the coverage probabilities and their mean squared errors of three discussed confidence intervals.

Schoonmaker, Bagley and Scanlan(2002) also mentioned the importance of the negative predictive value Ψ_N as assessing the screening test. The statistical inference of NPV can be done along with the above discussion.



[Figure 1] 95% Confidence intervals of PPV with $n_1 = 100, n_2 = 1000,\ n_3 = 1000$.

3. Simulation Results and Concluding Remarks

The discussed confidence intervals need to examine the coverage probabilities and their mean squared errors under the given nominal level to recommend the better method.

We assume that the sensitivities are from 0.8 to 0.9 based on the 500 observations and the prevalence are from 0.01 to 0.2 based on 500 observations. There is no particular reason that we set 500 observations in calculating both estimators, but they are usually estimated from 500 to 1,000 observations in the literature.

We set the true positive response rate from 0.1 to 0.5 and estimate the positive response rate from 500 sample observations of screening test.

We construct 95% confidence intervals based on (2.6), (2.7) and (.2.8) and check if the intervals contain the true PPV. We repeat 10,000 times and calculate the empirical coverage probabilities in each assumed parameters and their mean squared errors. We conduct the simulation study by SAS IML program.

Table 3.1 shows that usual estimator of PPV based on (2.6) does not maintain the given nominal level when the positive response rates are less than 20%

regardless of the value of prevalence. We can easily see that the usual PPV estimator starts maintaining the given nominal level when the positive responses are greater than .30. However, logit Wald type confidence interval based on (2.7) and inverse $\sinh$ transformation confidence interval based on (2.8) maintain the given nominal level regardless of the values of prevalence and positive response rate. In fact they perform very similar, but inverse $\sinh$ transformed intervals are a bit closer to the given nominal level than logit Wald type confidence intervals, especially when the prevalence and the positive response rate are low. Since prevalence of disease and positive response rate from the screening test are usually low, the proposed method should be used. We also remark that all these observations from the simulation study confirms the graphs in Figure 1.

Increasingly, medical screening test is now a standard procedure of day to day medical practice. However, an adopted screening test is not usually the most valid diagnostic method for a disease because it must be inexpensive and easy to perform. It often produces, therefore, incorrect results due to imperfect sensitivity and specificity of the test. In addition to its sensitivity and specificity, the performance of a test should be measured by its PPV or NPV. The present research proposes a simple and efficient statistical method of constructing the confidence interval for PPV of a screening test by considering inverse $\sinh$ transformation and the simulation study confirms the validity of the proposed method. However, the statistical properties of the proposed method should be further researched under smaller sample cases and more extreme cases like low prevalence and positive responses.

References

1. Brown, L.D., Cai, T.T. and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, 16, 101-133.
2. 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.
3. Gastwirth, J.L. (1987). The statistical precision of medical screening procedures : Application to polygraph and AIDS antibodies test data. *Statistical Science*, 2, 213-238.
4. Newcombe, R.G. (2001). Logit confidence intervals and the inverse $\sinh$ transformation. *The American Statistician*, 55, 200-202.
5. Schoonmaker, M., Bagley, G.P., Scanlan, M.K.(2002). Coordination of federal regulation and payment for new diagnostic tests; A proposed new approach. *Food and Drug Law Journal*, 57, 2, 195-204.

[Table 3.1] Empirical Coverage probabilities and their MSEs in case sensitivity 0.9

Prevalence	Positive	PPV	Wald type CI		Logit Wald type CI		Inverse sinh CI	
			Coverage probabilities	MSE	Coverage probabilities	MSE	Coverage probabilities	MSE
0.01	0.100	0.0950	0.9293	0.061	0.9699	0.078	0.9580	0.072
	0.200	0.0475	0.9366	0.153	0.9614	0.145	0.9518	0.144
	0.300	0.0317	0.9400	0.270	0.9571	0.265	0.9482	0.265
	0.400	0.0238	0.9479	0.378	0.9577	0.374	0.9504	0.374
	0.500	0.0190	0.9450	0.483	0.9485	0.480	0.9496	0.480
0.05	0.100	0.4750	0.9297	0.241	0.9604	0.214	0.9565	0.210
	0.200	0.2375	0.9399	0.078	0.9601	0.074	0.9545	0.074
	0.300	0.1583	0.9443	0.222	0.9544	0.218	0.9519	0.218
	0.400	0.1188	0.9519	0.345	0.9559	0.343	0.9540	0.343
	0.500	0.0950	0.9521	0.458	0.9523	0.457	0.9505	0.457
0.1	0.100	0.9500	0.9189	0.235	0.9530	0.154	0.9481	0.152
	0.200	0.4750	0.9382	0.067	0.9526	0.061	0.9503	0.060
	0.300	0.3167	0.9487	0.173	0.9554	0.169	0.9544	0.169
	0.400	0.2375	0.9507	0.314	0.9551	0.312	0.9551	0.312
	0.500	0.1900	0.9497	0.437	0.9548	0.435	0.9540	0.435
0.2	0.200	0.9500	0.9369	0.044	0.9515	0.060	0.9499	0.060
	0.300	0.6333	0.9437	0.094	0.9488	0.096	0.9497	0.096
	0.400	0.4750	0.9486	0.257	0.9440	0.255	0.9490	0.255
	0.500	0.3800	0.9501	0.398	0.9522	0.397	0.9509	0.397

[Table 3.2] Empirical Coverage probabilities and their MSEs in case sensitivity 0.85

Prevalence	Positive	PPV	Wald type CI		Logit Wald type CI		Inverse sinh CI	
			Coverage probabilities	MSE	Coverage probabilities	MSE	Coverage probabilities	MSE
0.01	0.100	0.0950	0.9293	0.061	0.9699	0.078	0.9580	0.072
	0.200	0.0475	0.9366	0.153	0.9614	0.145	0.9518	0.144
	0.300	0.0283	0.9434	0.273	0.9567	0.268	0.9488	0.260
	0.400	0.0213	0.9423	0.380	0.9555	0.377	0.9479	0.370
	0.500	0.0170	0.9430	0.484	0.9559	0.482	0.9474	0.480
0.05	0.100	0.4250	0.9252	0.221	0.9595	0.203	0.9540	0.199
	0.200	0.2125	0.9423	0.087	0.9534	0.082	0.9510	0.080
	0.300	0.1417	0.9492	0.229	0.9578	0.227	0.9556	0.227
	0.400	0.1063	0.9487	0.351	0.9554	0.349	0.9532	0.345
	0.500	0.0850	0.9456	0.463	0.9477	0.462	0.9480	0.462
0.1	0.100	0.8500	0.9214	0.246	0.9542	0.173	0.9503	0.170
	0.200	0.4250	0.9441	0.062	0.9520	0.058	0.9500	0.057
	0.300	0.2833	0.9434	0.185	0.9527	0.182	0.9503	0.182
	0.400	0.2125	0.9499	0.323	0.9531	0.321	0.9519	0.321
	0.500	0.1700	0.9476	0.443	0.9503	0.442	0.9500	0.442
0.2	0.200	0.8500	0.9391	0.054	0.9485	0.049	0.9493	0.049
	0.300	0.5667	0.9412	0.110	0.9486	0.109	0.9499	0.109
	0.400	0.4250	0.9509	0.271	0.9532	0.269	0.9525	0.269
	0.500	0.3400	0.9485	0.408	0.9544	0.407	0.9543	0.407

[received date : Feb. 2005, accepted date Apr. 2005]