

Interval Estimates for Predictive Values in Diagnostic Testing with Three Outcomes

Scott Clark, Lauren Mondin, Courtney Weber and Jessica Winborn
Advisor: Melinda Miller Holt

Abstract: In disease testing, patients and doctors are interested in estimates for positive predictive value (PPV) and negative predictive value (NPV). The PPV of a test is the probability that a patient actually has the disease, given a positive test result. The NPV is the probability that a patient actually does not have the disease, given a negative test result. Here we consider diagnostic tests in which the disease state remains uncertain, so the uncertain predictive value (UPV) is also of interest. UPV is the probability that, given an uncertain test result, follow-up testing will remain inconclusive. We derive classical Wald-type and Bayesian interval estimates of PPV, NPV, and UPV. Performance of these intervals is compared through simulation studies of interval coverage and width.

1. Introduction

The positive predictive value (PPV) and negative predictive value (NPV) of a diagnostic test are important to both patients and doctors. The PPV of a test is the probability that, given a positive test result, the patient actually has the disease. Similarly, the NPV of a test is the probability that, given a negative test result, the patient is actually disease-free. Mercaldo et al. (2007) derived Wald-type interval estimates of both PPV and NPV for binary diagnostic tests in case-control studies where the true disease prevalence is assumed known. Stamey and Holt (2010) derived analogous Bayesian interval estimates that allow for uncertainty relative to disease prevalence. Here we consider studies in which both the diagnostic test under consideration and the best test available may both provide inconclusive results, such as those

reported in Khatami et al. (2007). We extend the results of Mercaldo et al. (2007) and Stamey and Holt (2010) to diagnostic tests with three possible outcomes: positive, negative, and uncertain. To address the possibility of inconclusive results, we derive an estimator for what we call the uncertain predictive value (UPV). This is the probability that, given an uncertain initial diagnostic result, the truth would remain unknown under the best test available, which we call the “gold standard” for clarity. Because such follow-up procedures can be invasive and costly, the UPV may be helpful in determining whether the patient should undergo more testing.

Other authors have considered similar situations, such as Feinstein (1990), which considers a similar design with three results of the surrogate, or diagnostic, test but only two for the gold standard test. As far as we know, there is no other work associated with three possible outcomes for both the diagnostic test and the gold standard, as experienced by Khatami et al. (2007).

To begin, Section 2 describes the experimental design and associated estimators developed by Mercaldo et al. (2007) and Stamey and Holt (2010). Wald-type interval estimators for PPV, NPV, and UPV are derived under an expanded design in Section 3.1. Section 3.2 presents analogous Bayesian interval estimators. In Section 4, we provide a performance analysis comparing coverage and width of these intervals. Conclusions are offered in Section 5.

2. Binary Response Studies

2.1 Design

Consider first the study design described in Table 1. Let $i = 1$ represent the existence of disease and $i = 2$ represent absence of disease. Similarly, let $j = 1$ represent a positive test result and $j = 2$ represent a negative result. Thus x_{ij} represents the number of patients in disease state i

and test result j . Likewise n_i represents the total number of people in disease state i . Mercaldo et al. (2007) and Stamey and Holt (2010) described studies of this type that were not designed to reflect the true disease prevalence in the target population. The same situation is considered here. Thus, we assume that these values are available from sources external to the current study and that p_1 represents the known population probability of disease and $p_2 = 1 - p_1$ reflects the known probability that a person is disease-free.

	Test +	Test -	Total
Disease +	x_{11}	x_{12}	n_1
Disease -	x_{21}	x_{22}	n_2

Table 1: Binary Response Study Design

For such studies, researchers and clinicians are interested in estimating the sensitivity (Se) of the test,

$$\text{Se} = P(\text{Test+} | \text{Disease+})$$

and the specificity (Sp),

$$\text{Sp} = P(\text{Test-} | \text{Disease-}).$$

Through Bayes Theorem, the sensitivity and specificity then lead to evaluation of the positive predictive value

$$\text{PPV} = P(\text{Disease+} | \text{Test+}) = \frac{p_1 \cdot \text{Se}}{p_1 \cdot \text{Se} + p_2(1 - \text{Sp})}$$

and the negative predictive value

$$\text{NPV} = P(\text{Disease-} | \text{Test-}) = \frac{p_2 \cdot \text{Sp}}{p_2 \cdot \text{Sp} + p_1(1 - \text{Se})}.$$

The data in Table 1 yield the following maximum likelihood estimators (MLE):

$$\widehat{\text{Se}} = \frac{x_{11}}{n_1} \quad \text{and} \quad \widehat{\text{Sp}} = \frac{x_{22}}{n_2}.$$

Similarly, the false positive (Fp) and false negative (Fn) rates are estimated by

$$\widehat{Fp} = \frac{x_{21}}{n_2} \quad \text{and} \quad \widehat{Fn} = \frac{x_{12}}{n_1}.$$

Based on the known prevalence values and the estimators above, estimators for the PPV and NPV become

$$\widehat{PPV} = \frac{p_1 \cdot \widehat{Se}}{p_1 \cdot \widehat{Se} + p_2 \cdot \widehat{Fp}} \quad (1)$$

$$\widehat{NPV} = \frac{p_2 \cdot \widehat{Sp}}{p_1 \cdot \widehat{Fn} + p_2 \cdot \widehat{Sp}}. \quad (2)$$

Mercaldo et al. (2007) derived Wald-type interval estimators for both PPV and NPV, based on equations (1) and (2). Likewise, Stamey and Holt (2010) derived analogous Bayesian estimators. These will be discussed in Sections 2.2 and 2.3, respectively.

2.2 Wald-Type Intervals

Mercaldo et al. (2007) developed Wald-type interval estimators for PPV and NPV by employing the delta method to determine the variances of PPV and NPV. Assuming that the responses follow a binomial distribution and based on the asymptotic normality of the MLEs, the authors proposed an approximate interval for PPV of the form

$$\widehat{PPV} \pm z_{(1-\alpha/2)} \sqrt{\text{var}(\widehat{PPV})},$$

where $z_{(1-\alpha/2)}$ is the $(1-\alpha/2)^{\text{th}}$ quantile of the standard normal distribution and

$$\text{var}(\widehat{PPV}) = \frac{\frac{(p_1 \cdot p_2 \cdot \widehat{Fp})^2 \cdot \widehat{Se} \cdot \widehat{Fn}}{n_1 + 4} + \frac{(p_1 \cdot p_2 \cdot \widehat{Se})^2 \cdot \widehat{Sp} \cdot \widehat{Fp}}{n_2 + 4}}{(p_1 \cdot p_2 \cdot \widehat{Fp} \cdot \widehat{Se})^4}.$$

Similarly, the estimator for NPV was

$$\widehat{\text{NPV}} \pm z_{(1-\alpha/2)} \sqrt{\text{var}(\widehat{\text{NPV}})},$$

where $z_{(1-\alpha/2)}$ is the $(1-\alpha/2)^{\text{th}}$ quantile of the standard normal distribution and

$$\text{var}(\widehat{\text{NPV}}) = \frac{\frac{(p_1 \cdot p_2 \cdot \widehat{\text{Sp}})^2 \cdot \widehat{\text{Se}} \cdot \widehat{\text{Fn}}}{n_1 + 4} + \frac{(p_1 \cdot p_2 \cdot \widehat{\text{Fn}})^2 \cdot \widehat{\text{Sp}} \cdot \widehat{\text{Fp}}}{n_2 + 4}}{(p_1 \cdot p_2 \cdot \widehat{\text{Fn}} \cdot \widehat{\text{Sp}})^4}.$$

Mercaldo et al. (2007) also proposed alternate intervals based on an Agresti-type adjustment (Agresti and Caffo, 2000). Here the adjusted estimates of Se and Sp became

$$\widetilde{\text{Se}} = \frac{n_1 \cdot \widehat{\text{Se}} + \frac{k^2}{2}}{n_1}$$

and

$$\widetilde{\text{Sp}} = \frac{n_2 \cdot \widehat{\text{Sp}} + \frac{k^2}{2}}{n_2},$$

so that the adjusted PPV and adjusted NPV calculations become

$$\widetilde{\text{PPV}} = \frac{p_1 \cdot \widetilde{\text{Se}}}{p_1 \cdot \widetilde{\text{Se}} + p_2 \cdot \widetilde{\text{Fp}}} \quad (3)$$

and

$$\widetilde{\text{NPV}} = \frac{p_2 \cdot \widetilde{\text{Sp}}}{p_1 \cdot \widetilde{\text{Fn}} + p_2 \cdot \widetilde{\text{Sp}}}, \quad (4)$$

respectively.

Mercaldo et al. (2007) compared the performance of these proposed interval methods, as well as intervals based on their logit transformations, using simulations of confidence interval width and coverage probability. They concluded that, when it provided a solution, the unadjusted logit transformation performed best. Otherwise, they recommended the adjusted standard interval. Stamey and Holt (2010) found, however, that the logit transformed solution often does not exist. Thus, the research herein focuses on adjusted standard Wald-type intervals.

2.3 Bayesian Intervals

Stamey and Holt (2010) derived Bayesian interval estimates of PPV and NPV for case-control studies described in Table 1. Because such designs require *a priori* knowledge of disease prevalence, the authors considered two Bayesian models: one that took disease prevalence to be known and one that placed a prior distribution on prevalence.

Like Mercaldo et al. (2007), Stamey and Holt (2010) assumed binomial data so that

$$x_{11} \sim \text{binomial}(n_1, \text{Se}) \quad \text{and} \quad x_{22} \sim \text{binomial}(n_2, \text{Sp}).$$

They assumed conjugate beta priors so that

$$\text{Se} \sim \text{beta}(\alpha_{\text{Se}}, \beta_{\text{Se}}) \quad \text{and} \quad \text{Sp} \sim \text{beta}(\alpha_{\text{Sp}}, \beta_{\text{Sp}}),$$

which yielded posterior distributions for Se and Sp of

$$\text{Se} | \mathbf{d} \sim \text{beta}(x_{11} + \alpha_{\text{Se}}, n_1 - x_{11} + \beta_{\text{Se}}) \quad \text{and} \quad \text{Sp} | \mathbf{d} \sim \text{beta}(x_{22} + \alpha_{\text{Sp}}, n_2 - x_{22} + \beta_{\text{Sp}}), \quad (5)$$

where $\mathbf{d} = (x_{11}, x_{22}, n_1, n_2)$.

Taking p_1 and p_2 to be known, Monte Carlo sampling from the posteriors in (5) provided estimates of Se and Sp. Plugging these values into equations (1) and (2) yielded estimates of the posterior distribution for the unadjusted estimates of PPV and NPV. Adjusted estimates of PPV and NPV were similarly obtained, using equations (3) and (4).

The second model assumed that p_1 is in fact estimated from previous survey data or expert opinion. In this case, the value of p_1 in (1) and (2) follows the posterior distribution

$$p_1 | \mathbf{d} \sim \text{beta}(y + \alpha_{p_1}, n - y + \beta_{p_1})$$

where the data vector is now $\mathbf{d} = (x_{11}, x_{22}, y, n_1, n_2, n)$ and y represents the number of patients with the disease in a previous survey of n patients. Here Se , Sp , and p_1 (and thus p_2) are each obtained through Monte Carlo sampling in order to estimate PPV and NPV.

The authors compared the interval width and true coverage probability of both Bayesian intervals to that of the adjusted standard interval and the unadjusted logit interval of Mercaldo et al. (2007) through simulation studies. They found that the Bayesian intervals maintained coverage close to the nominal level, while the adjusted Wald-type interval dipped far below nominal for large values of Sp . Widths were generally comparable for the methods that assumed known prevalence, with the exception of large values of Sp . In that case, the Bayesian interval was wider because it maintained coverage closer to nominal. The Bayesian procedure that assumed unknown prevalence produced intervals that were generally wider than the other methods, but maintained coverage quite close to nominal for all parameter values.

3. Multinomial Response Studies

3.1 Design

Now consider studies with the design described below in Table 2. Patients either have the disease, do not have the disease, or their true state remains inconclusive, as determined by a gold standard test. The diagnostic test under consideration may return results that are positive, negative, or uncertain. Let $i = 1$ represent the existence of disease, $i = 2$ represent absence of

disease, and $i = 3$ represent true uncertainty. Similarly, let $j = 1$ represent a positive test result, $j = 2$ represent a negative test result, and $j = 3$ represent an inconclusive test result. Thus x_{ij} represents the number of patients in disease state i and test result j . Likewise n_i represents the total number of people in disease state i .

	<i>Test +</i>	<i>Test –</i>	<i>Test 0</i>	<i>Total</i>
<i>Disease +</i>	x_{11}	x_{12}	x_{13}	n_1
<i>Disease –</i>	x_{21}	x_{22}	x_{23}	n_2
<i>Disease 0</i>	x_{31}	x_{32}	x_{33}	n_3

Table 2. Multinomial Response Study Design

The data in Table 2 yield the following maximum likelihood estimators (MLE) of sensitivity (Se), specificity (Sp), and uncertainty (U):

$$\widehat{\text{Se}} = \frac{x_{11}}{n_1}, \quad \widehat{\text{Sp}} = \frac{x_{22}}{n_2}, \quad \text{and} \quad \widehat{\text{U}} = \frac{x_{33}}{n_3}.$$

Similarly, the false positive (Fp) and false negative (Fn) rates are

$$\widehat{\text{Fp}} = \frac{x_{21}}{n_2} \quad \text{and} \quad \widehat{\text{Fn}} = \frac{x_{12}}{n_1}.$$

Using this design, we must also consider the probability of an uncertain test result among patients with the disease (Yu), the probability of an uncertain test result among patients without the disease (Nu), the probability of a positive test that cannot be confirmed or refuted by the gold standard (U_1), and the probability of a negative test that cannot be confirmed or refuted by the gold standard (U_2). These values are estimated by the following:

$$\widehat{\text{Yu}} = \frac{x_{13}}{n_1}, \quad \widehat{\text{Nu}} = \frac{x_{23}}{n_2}, \quad \widehat{U}_1 = \frac{x_{31}}{n_3}, \quad \text{and} \quad \widehat{U}_2 = \frac{x_{32}}{n_3}$$

Again we assume that p_1 represents the known population probability of disease; p_2 reflects the known probability that a person is disease-free; and p_3 is equal to the known probability of undeterminable results. Just as Mercaldo et al. (2007) and Stamey and Holt (2010) assumed a binomial distribution, we assume that $x_1 = (x_{11}, x_{12}, x_{13})$, $x_2 = (x_{21}, x_{22}, x_{23})$, and $x_3 = (x_{31}, x_{32}, x_{33})$ are independent and follow a multinomial distribution. Thus, the probability mass function for x_1 is

$$f(x_{11}, x_{12}, x_{13}) = \frac{n_1!}{x_{11}!x_{12}!x_{13}!} \text{Se}^{x_{11}} \text{Fn}^{x_{12}} \text{Yu}^{x_{13}}.$$

We let $t_1 = (\text{Se}, \text{Fn}, \text{Yu})$, and write $x_1 \sim \mathcal{M}(n_1, t_1)$. Similarly, $t_2 = (\text{Fp}, \text{Sp}, \text{Nu})$ and $t_3 = (\text{U}_1, \text{U}_2, \text{U})$ so that $x_2 \sim \mathcal{M}(n_2, t_2)$, and $x_3 \sim \mathcal{M}(n_3, t_3)$.

Applying Bayes Theorem produces the following estimators for PPV, NPV, and UPV:

$$\widehat{\text{PPV}} = \frac{p_1 \cdot \widehat{\text{Se}}}{p_1 \cdot \widehat{\text{Se}} + p_2 \cdot \widehat{\text{Fp}} + p_3 \cdot \widehat{\text{U}}_1} \quad (6)$$

$$\widehat{\text{NPV}} = \frac{p_2 \cdot \widehat{\text{Sp}}}{p_1 \cdot \widehat{\text{Fn}} + p_2 \cdot \widehat{\text{Sp}} + p_3 \cdot \widehat{\text{U}}_2} \quad (7)$$

$$\widehat{\text{UPV}} = \frac{p_3 \cdot \widehat{\text{U}}}{p_1 \cdot \widehat{\text{Yu}} + p_2 \cdot \widehat{\text{Nu}} + p_3 \cdot \widehat{\text{U}}} \quad (8)$$

Based on the equations (6) through (8), we extend the results of Mercaldo et al. (2007) and Stamey and Holt (2010) to situations in which both the diagnostic test being evaluated and the gold standard produce inconclusive results.

3.2 Wald-type Intervals

Consider first the estimation of the PPV. We derive a $100(1-\alpha)\%$ Wald-type interval by applying the delta method. The interval takes the form:

$$\widehat{\text{PPV}} \pm z_{(1-\alpha/2)} \sqrt{\text{var}(\widehat{\text{PPV}})}, \quad (9)$$

where $z_{(1-\alpha/2)}$ is the $(1-\alpha/2)^{\text{th}}$ quantile of the standard normal distribution. Because $\widehat{\text{Se}}$, $\widehat{\text{Fp}}$, and $\widehat{\text{U}}_1$ are asymptotically normal with means Se , Fp and U_1 , respectively, we take $\widehat{\text{PPV}}$ to be asymptotically normally distributed with mean PPV . To derive the variance of $\widehat{\text{PPV}}$, we must first consider the variances of $\widehat{\text{Se}}$, $\widehat{\text{Fp}}$, and $\widehat{\text{U}}_1$. By the properties of the multinomial distribution, namely that the marginal distribution of each individual cell count is binomial, we obtain

$$\text{var}(\widehat{\text{Se}}) = \frac{\widehat{\text{Se}}(1-\widehat{\text{Se}})}{n_1}, \quad \text{var}(\widehat{\text{Fp}}) = \frac{\widehat{\text{Fp}}(1-\widehat{\text{Fp}})}{n_2}, \quad \text{and} \quad \text{var}(\widehat{\text{U}}_1) = \frac{\widehat{\text{U}}_1(1-\widehat{\text{U}}_1)}{n_3}.$$

Because x_1 , x_2 , and x_3 are independent, the covariances of $\widehat{\text{Se}}$, $\widehat{\text{Fp}}$, and $\widehat{\text{U}}_1$ are equal to zero. In order to obtain the variance estimator for $\widehat{\text{PPV}}$, we apply the delta method and use the partial derivative of PPV with respect to Se , Fp , and U_1 . For the full derivation, see Appendix A. This provides a variance estimator for $\widehat{\text{PPV}}$ in (9):

$$\text{var}(\widehat{\text{PPV}}) = \frac{\left[(p_1 \cdot p_3 \cdot \widehat{\text{U}}_1) + (p_1 \cdot p_2 \cdot \widehat{\text{Fp}}) \right]^2 \text{var}(\widehat{\text{Se}}) + (p_1 \cdot p_3 \cdot \widehat{\text{Se}})^2 \text{var}(\widehat{\text{U}}_1) + (p_1 \cdot p_2 \cdot \widehat{\text{Se}}) \text{var}(\widehat{\text{Fp}})}{(p_1 \cdot \widehat{\text{Se}} + p_2 \cdot \widehat{\text{Fp}} + p_3 \cdot \widehat{\text{U}}_1)^4}.$$

Similarly,

$$\text{var}(\widehat{Sp}) = \frac{\widehat{Sp}(1-\widehat{Sp})}{n_1}, \quad \text{var}(\widehat{Fn}) = \frac{\widehat{Fn}(1-\widehat{Fn})}{n_2}, \quad \text{and} \quad \text{var}(\widehat{U}_2) = \frac{\widehat{U}_2(1-\widehat{U}_2)}{n_3},$$

which lead to

$$\text{var}(\widehat{NPV}) = \frac{\left[(p_2 \cdot p_3 \cdot \widehat{U}_2) + (p_1 \cdot p_2 \cdot \widehat{Fn}) \right]^2 \text{var}(\widehat{Sp}) + (p_2 \cdot p_3 \cdot \widehat{Sp})^2 \text{var}(\widehat{U}_2) + (p_1 \cdot p_2 \cdot \widehat{Sp}) \text{var}(\widehat{Fn})}{(p_2 \cdot \widehat{Fn} + p_2 \cdot \widehat{Sp} + p_3 \cdot \widehat{U}_2)^4}$$

and the interval estimator

$$\widehat{NPV} \pm z_{(1-\alpha/2)} \sqrt{\text{var}(\widehat{NPV})}. \quad (10)$$

Lastly,

$$\text{var}(\widehat{Yu}) = \frac{\widehat{Yu}(1-\widehat{Yu})}{n_1}, \quad \text{var}(\widehat{Nu}) = \frac{\widehat{Nu}(1-\widehat{Nu})}{n_2}, \quad \text{and} \quad \text{var}(\widehat{U}) = \frac{\widehat{U}(1-\widehat{U})}{n_3},$$

which lead to

$$\text{var}(\widehat{UPV}) = \frac{\left[(p_1 \cdot p_3 \cdot \widehat{Yu}) + (p_2 \cdot p_3 \cdot \widehat{Nu}) \right]^2 \text{var}(\widehat{U}) + (p_2 \cdot p_3 \cdot \widehat{U})^2 \text{var}(\widehat{Nu}) + (p_1 \cdot p_3 \cdot \widehat{U})^2 \text{var}(\widehat{Yu})}{(p_1 \cdot \widehat{Yu} + p_2 \cdot \widehat{Nu} + p_3 \cdot \widehat{U})^4}$$

and the interval estimator

$$\widehat{UPV} \pm z_{(1-\alpha/2)} \sqrt{\text{var}(\widehat{UPV})}. \quad (11)$$

As do Mercaldo et al. (2007), we consider an Agresti type adjustment (2000) to improve performance of the estimators for proportions near 0 or 1. To do so, the data in Table 2 is adapted to become those presented below in Table 3.

	<i>Test +</i>	<i>Test -</i>	<i>Test 0</i>	<i>Total</i>
<i>Disease +</i>	$x_{11}+k$	$x_{12}+k$	$x_{13}+k$	n_1+3k
<i>Disease -</i>	$x_{21}+k$	$x_{22}+k$	$x_{23}+k$	n_2+3k
<i>Disease 0</i>	$x_{31}+k$	$x_{32}+k$	$x_{33}+k$	n_3+3k

Table 3. Study Design with Adjustments

We represent estimators based on this adjustment with ‘ \sim ’. For example, $\widetilde{Se} = \frac{x_{11} + k}{n_1 + 3k}$. For

binomial proportions, Agresti and Caffo (2000) recommend using $k = \frac{(z_{(1-\alpha/2)})^2}{2}$. To our

knowledge, however, no such recommendations exist for estimation of multinomial proportions.

Several values of k were considered and tested. We recommend $k = 0.5$ because it produced the best interval coverage and width in simulation-based performance analysis.

3.3 Bayesian Intervals

Next, we derive Bayesian estimators for PPV, NPV, and UPV. In the Bayesian model, we consider the Dirichlet distribution because it is the conjugate prior for multinomial data. Thus, the joint prior distribution for $t_1 = (Se, Fn, Yu)$, the parameters in the distribution of x_1 , is

$$g(t_1) = \frac{\prod_{i=1}^3 \Gamma(\alpha_{1i})}{\Gamma(\sum_{i=1}^3 \alpha_{1i})} Se^{\alpha_{11}-1} Yu^{\alpha_{12}-1} Fn^{\alpha_{13}-1}.$$

The joint posterior for Se, Fn, and Yu then becomes

$$g(t_1 | x_1) = \frac{\prod_{i=1}^3 \Gamma(\alpha_{1i} + x_{1i})}{\Gamma(\sum_{i=1}^3 \alpha_{1i} + x_{1i})} Se^{\alpha_{11} + x_{11} - 1} Fn^{\alpha_{12} + x_{12} - 1} Yu^{\alpha_{13} + x_{13} - 1},$$

written $t_1 | x_1 \sim \mathcal{D}(\alpha_{11} + x_{11}, \alpha_{12} + x_{12}, \alpha_{13} + x_{13})$. Similarly, the posterior distributions for $t_2 =$

(Fp, Sp, Nu) and $t_3 = (U_1, U_2, U)$ can be written $t_2 | x_2 \sim \mathcal{D}(\alpha_{21} + x_{21}, \alpha_{22} + x_{22}, \alpha_{23} + x_{23})$ and $t_3 |$

$x_3 \sim \mathcal{D}(\alpha_{31} + x_{31}, \alpha_{32} + x_{32}, \alpha_{33} + x_{33})$, respectively. For the purposes of the analysis herein, we

consider noninformative Dirichlet priors with parameter values equal to 0.5. This makes the Bayesian analysis analogous to the Wald-type analysis in that no information is available relative to t_1 , t_2 , and t_3 beyond that in Table 3.

Here again, the prevalence vector, $p = (p_1, p_2, p_3)$, is assumed known, so the posterior distributions for PPV, NPV, and UPV can be approximated via Monte Carlo sampling by B values from the posteriors of t_1 , t_2 , and t_3 . At each iteration, the generated values of $\text{Se}^{(j)}$, $\text{Fn}^{(j)}$, $\text{Yu}^{(j)}$, $\text{Fp}^{(j)}$, $\text{Sp}^{(j)}$, $\text{Nu}^{(j)}$, $\text{U}_1^{(j)}$, $\text{U}_2^{(j)}$, and $\text{U}^{(j)}$ are plugged into

$$\text{PPV}^{(j)} = \frac{p_1 \cdot \text{Se}^{(j)}}{p_1 \cdot \text{Se}^{(j)} + p_2 \cdot \text{Fp}^{(j)} + p_3 \cdot \text{U}_1^{(j)}} , \quad (12)$$

$$\text{NPV}^{(j)} = \frac{p_2 \cdot \text{Sp}^{(j)}}{p_1 \cdot \text{Fn}^{(j)} + p_2 \cdot \text{Sp}^{(j)} + p_3 \cdot \text{U}_2^{(j)}} , \quad (13)$$

and

$$\text{UPV}^{(j)} = \frac{p_3 \cdot \text{U}^{(j)}}{p_1 \cdot \text{Yu}^{(j)} + p_2 \cdot \text{Nu}^{(j)} + p_3 \cdot \text{U}^{(j)}} , \quad (14)$$

where the superscript (j) denotes that this is the j^{th} iteration in the Monte Carlo approximation. The posteriors for PPV, NPV, and UPV are then approximated using these B values. The 2.5th and 95.5th percentiles of the distribution then provide the 95% Bayesian credible sets.

4. Performance Analysis

This section provides a simulation-based comparison of the 95% adjusted Wald-type and Bayesian interval estimates of PPV through both coverage probability and interval width. Consider $n_1 = n_2 = n_3 = 25$, $\text{Sp} = 0.5$, and let Se vary from 0.5 to 0.80 by 0.10. For this analysis, $\text{Yu} = \text{Nu} = 0.1$, $\text{U}_1 = 0.8$, $\text{U} = \text{U}_2 = 0.1$, $\text{Fp} = 0.4$, and Fn varies with Se . In each boxplot, values

of p_1 vary from 0.50 to 0.85 by 0.05, while fixing $p_3 = 0.10$ so that p_2 varies from 0.05 to 0.40 by 0.05. For each set of parameter values, 10,000 samples were simulated leading to 10,000 Wald (W) and Bayesian (B) intervals. Coverage probability and width are estimated by averaging over the 10,000 samples. Figure 1 compares resulting estimates of coverage probability, while Figure 2 compares interval width.

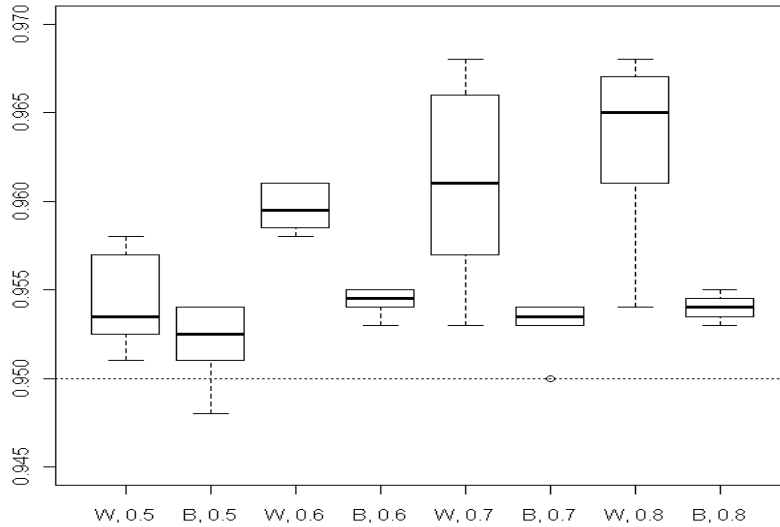


Figure 1. PPV Coverage Probability Comparison for $Sp = 0.5, Se = 0.5, 0.6, 0.7, 0.8$

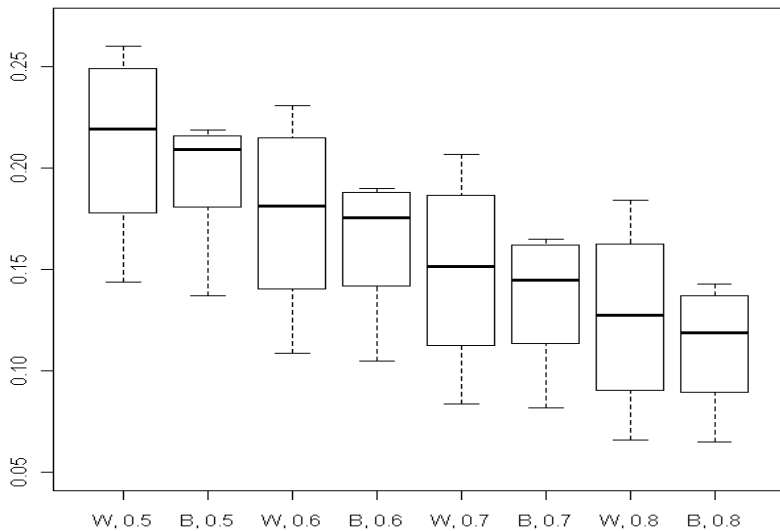


Figure 2. PPV Width Comparison for $Sp = 0.5, Se = 0.5, 0.6, 0.7, 0.8$

Figure 1 indicates that the Bayesian coverage probability remains closer to the nominal 0.95 when estimating PPV than that of the Wald-type interval for the parameters considered here, while Figure 2 suggests that the Bayesian interval is slightly narrower. Figures 3 and 4 provide the coverage and width, respectively, for estimating NPV at the same parameter values. These plots indicate generally the same performance pattern, with the Bayesian interval coverage closer to nominal and its width narrower. The behavior exhibited here by the Wald-type intervals is similar to that observed in Stamey and Holt (2010), because coverage deviates further from nominal as Se increases. Likewise, the width of the Bayesian interval increases, while coverage remains close to nominal.

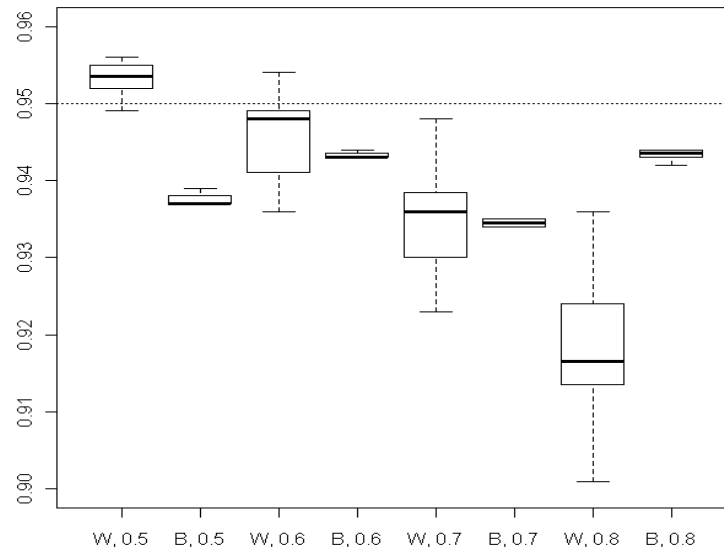


Figure 3. NPV Coverage Probability Comparison for $Sp = 0.5$, $Se = 0.5, 0.6, 0.7, 0.8$

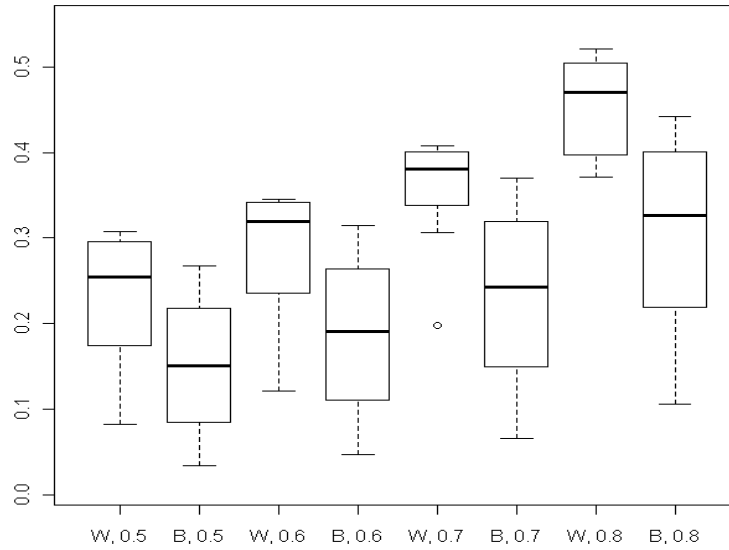


Figure 4. NPV Width Comparison for $Sp = 0.5, Se = 0.5, 0.6, 0.7, 0.8$

Figures 5 and 6 below indicate that the Bayesian coverage probabilities are much closer to nominal than those of the Wald-type interval when estimating UPV, although generally the Bayesian interval is slightly wider.

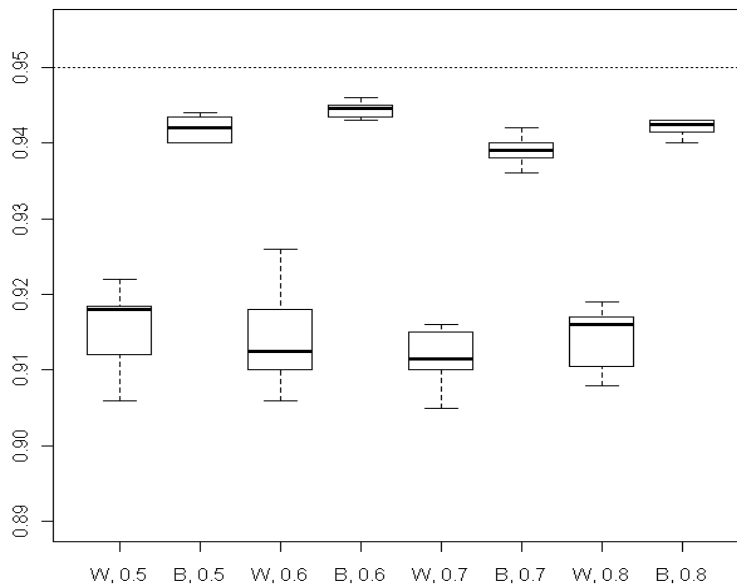


Figure 5. UPV Coverage Probability Comparison for $Sp = 0.5, Se = 0.5, 0.6, 0.7, 0.8$

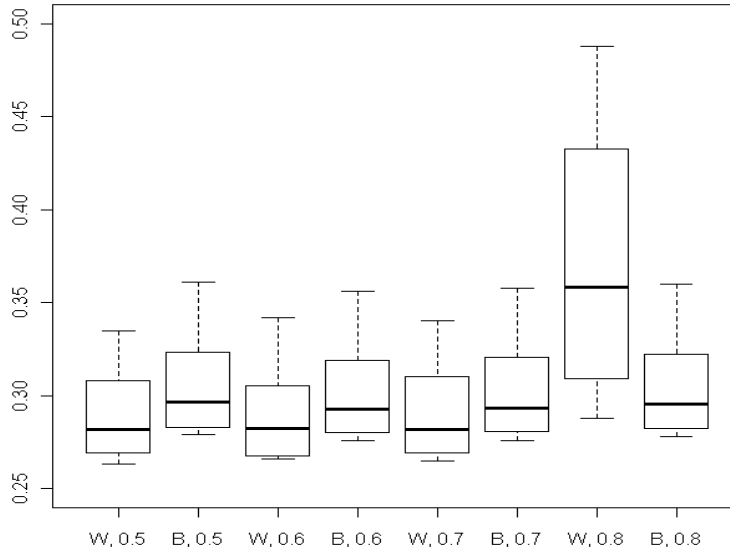


Figure 6. UPV Width Comparison for $Sp = 0.5, Se = 0.5, 0.6, 0.7, 0.8$

Several other sets of parameter values were considered, varying Sp and sample size. Those simulation results are not included for the sake of space. For varying values of Sp the behavior demonstrated above remains consistent. At sample sizes of 50 and above, the two procedures maintained very similar coverage and width. Further details about simulation results are available from the authors.

5. Example

The research herein was motivated by the work of Khatami et al. (2007). Khatami and co-authors developed a prediction model, combining the results of various blood and tissue markers, to determine whether or not a cancer originated in the ovaries. The model designated the cancer origin as ovaries, not ovaries, or uncertain. To determine its effectiveness, they compared its diagnoses against the standard histologic assessment. Table 4 describes the data obtained in their experiment:

	Test +	Test -	Test 0	Total
Disease +	118	6	6	130
Disease -	6	15	11	32
Disease 0	23	4	9	36

Table 4. Ovarian Cancer Data (Khatami et al., 2007)

In order to calculate the interval estimates of PPV, NPV, and UPV, the values of p_1 , p_2 , and p_3 must be determined. We take these to be known but base them on calculations in Khatami et al. (2007), we take p_1 to be 0.66. We take p_2 and p_3 to be 0.16 and 0.18, respectively. For the adjusted Wald-type interval, we take $a = 0.5$. For the Bayesian credible sets, t_1 , t_2 , and t_3 are each given noninformative Dirichlet (0.5, 0.5, 0.5) distributions. The resulting interval estimates are presented in Table 5.

Method	PPV Estimate	95% Interval	NPV Estimate	95% Interval	UPV Estimate	95% Interval
Wald	0.805	0.767-0.844	0.599	0.435-0.770	0.345	0.187-0.505
Adjusted Wald	0.805	0.768-0.847	0.599	0.415-0.748	0.345	0.186-0.499
Bayes	0.801	0.779-0.823	0.595	0.473-0.717	0.302	0.216-0.398

Table 5. Interval Estimates of PPV, NPV and UPV from Khatami et al. (2007)

Evaluating Table 5, we see that the interval estimates are very similar. The main difference is in the different interpretations of a confidence interval and a credible set. A confidence interval states that there is a 95% chance that the true proportion is in the interval, while the credible set states that if we run the test an infinite number of times, 95% of the time the true proportion will be in the interval.

6. Conclusion

In this paper, we derive both Wald-type and Bayesian interval estimates of PPV, NPV, and UPV for diagnostic procedures with three possible outcomes. We considered prevalence to be known for both the Wald-type and Bayesian analysis. Both of these methods require knowledge of the outcome of the gold standard. Simulation studies indicate that the Bayesian approach using noninformative priors typically provides better coverage and smaller width than that of the Wald-type interval with a 0.5 Agresti type adjustment. It is, therefore, our conclusion that the Bayesian method is preferred. Future research will consider Bayesian analyses that incorporate informative Dirichlet priors to reflect any available prior information on values such as the sensitivity and specificity in order to further improve performance of the Bayesian method and Bayesian methods for estimating PPV, NPV, and UPV for unknown prevalence. Additional work will also expand the second model in Stamey and Holt (2010) to account for possible uncertainty about disease prevalence values.

ACKNOWLEDGEMENTS

We wish to thank the anonymous referees for their comments that significantly improved the paper. Also, we wish to thank Dr. Ken Smith and the National Science Foundation for their support through grant #DMS-0636528.

References:

Agresti, A. and Caffo, B. (2000). "Simple and Effective Confidence Intervals for Proportions and Differences of Proportions." *The American Statistician* 54: 20-88.

Feinstein, A. (1990). "The Inadequacy of Binary Models for the Clinical Reality of Three-Zone Diagnostic Decisions." *J Clin Epidemiol* 43: 109-113.

Khatami, Z. Cross P., Stevenson M.R., Naik R. (2007). "Design of a bio-mathematical prediction model using serum tumor markers and immunohistochemistry in peritoneal carcinomatosis with ovarian involvement: a pilot study." *International Journal of Gynecological Cancer* 17:1258-1263.

Mercaldo, N. D. Lau, K. G., Zhou, X. H. (2007). "Confidence Intervals for predictive values with an emphasis to case-control studies." *Statistics in Medicine* 26:2170-2183.

Stamey, J. Holt, M. A. (2010) "Bayesian interval estimation for predictive values from case-control studies." *Communications in Statistics - Simulation and Computation* 39: 101-110.

Appendix A

To derive $var(PPV)$ we employ the delta method, so that $var(PPV) = (\nabla PPV)' \Sigma (\nabla PPV)$, where $\nabla PPV = [\partial(PPV)/\partial(Se), \partial(PPV)/\partial(U_1), \partial(PPV)/\partial(Fp)]$ and Σ is the variance/covariance matrix of $v = (Se, U_1, Fp)$. Recall that

$$PPV = \frac{p_1 \cdot Se}{(p_1 \cdot Se) + (p_2 \cdot Fp) + (p_3 \cdot U_1)}.$$

Thus,

$$var(PPV) = \begin{bmatrix} \frac{\partial(PPV)}{\partial(Se)} \\ \frac{\partial(PPV)}{\partial(Fp)} \\ \frac{\partial(PPV)}{\partial(U_1)} \end{bmatrix}^T \begin{bmatrix} var(Se) & 0 & 0 \\ 0 & var(Fp) & 0 \\ 0 & 0 & var(U_1) \end{bmatrix} \begin{bmatrix} \frac{\partial(PPV)}{\partial(Se)} \\ \frac{\partial(PPV)}{\partial(Fp)} \\ \frac{\partial(PPV)}{\partial(U_1)} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{(p_1 \cdot p_2 \cdot Fp) + (p_1 \cdot p_3 \cdot U_1)}{[(p_1 \cdot Se) + (p_2 \cdot Fp) + (p_3 \cdot U_1)]^2} & \frac{-p_1 \cdot p_2 \cdot Se}{[(p_1 \cdot Se) + (p_2 \cdot Fp) + (p_3 \cdot U_1)]^2} & \frac{-p_1 \cdot p_3 \cdot Se}{[(p_1 \cdot Se) + (p_2 \cdot Fp) + (p_3 \cdot U_1)]^2} \end{bmatrix}$$

$$\begin{bmatrix} \frac{Se(1-Se)}{n_1} & 0 & 0 \\ 0 & \frac{Fp(1-Fp)}{n_2} & 0 \\ 0 & 0 & \frac{U_1(1-U_1)}{n_3} \end{bmatrix} \begin{bmatrix} \frac{(p_1 \cdot p_2 \cdot Fp) + (p_1 \cdot p_3 \cdot U_1)}{[(p_1 \cdot Se) + (p_2 \cdot Fp) + (p_3 \cdot U_1)]^2} \\ \frac{-p_1 \cdot p_2 \cdot Se}{[(p_1 \cdot Se) + (p_2 \cdot Fp) + (p_3 \cdot U_1)]^2} \\ \frac{-p_1 \cdot p_3 \cdot Se}{[(p_1 \cdot Se) + (p_2 \cdot Fp) + (p_3 \cdot U_1)]^2} \end{bmatrix}$$

$$= \frac{[(p_1 \cdot p_2 \cdot \text{Fp}) + (p_1 \cdot p_3 \cdot \text{U}_1)]^2 \left(\frac{\text{Se}(1-\text{Se})}{n_1} \right) + (-p_1 \cdot p_2 \cdot \text{Se})^2 \left(\frac{\text{Fp}(1-\text{Fp})}{n_2} \right) + (-p_1 \cdot p_3 \cdot \text{Se})^2 \left(\frac{\text{U}_1(1-\text{U}_1)}{n_3} \right)}{[(p_1 \cdot \text{Se}) + (p_2 \cdot \text{Fp}) + (p_3 \cdot \text{U}_1)]^4}.$$

Substituting the estimators $\widehat{\text{Se}}$, $\widehat{\text{Fp}}$, and $\widehat{\text{U}}_1$, yields the estimator for $\text{var}(\widehat{\text{PPV}})$. The values of

$\text{var}(\widehat{\text{UPV}})$ and $\text{var}(\widehat{\text{NPV}})$ are derived using the same method and thus omitted.