

# Likelihood Based Confidence Intervals for the Difference of Proportions in Two Doubly Sampled Data with a Common False-Positive Error Rate

Seung-Chun Lee<sup>1,a</sup>

<sup>a</sup>Department of Statistics, Hanshin University

---

## Abstract

Lee (2010) developed a confidence interval for the difference of binomial proportions in two doubly sampled data subject to false-positive errors. The confidence interval seems to be adequate for a general double sampling model subject to false-positive misclassification. However, in many applications, the false-positive error rates could be the same. On this note, the construction of asymptotic confidence interval is considered when the false-positive error rates are common. The coverage behaviors of nine likelihood based confidence intervals are examined. It is shown that the confidence interval based Rao score with the expected information has good performance in terms of coverage probability and expected width.

Keywords: Profile likelihood, Rao score, information, double sampling.

---

## 1. Introduction

Suppose that binary observations are obtained by classifying experimental or sampling units into two mutually exclusive categories. Usually a researcher uses inerrant device for the classification. However, when the cost of precise classification is expensive, a researcher often uses an inexpensive but fallible classifier with a supplementary inerrant classifier. For instance, the case-control study of Hildesheim *et al.* (1991) aimed to examine that invasive cervical cancer can affect exposure to Herpes Simplex Virus(HSV). For the study, western blot procedure, which was known to be relatively inaccurate in detecting the infection of HSV, was applied to about two thousand women in case and control groups. Since the western blot procedure is fallible, it may classify an infected woman as normal (false-negative) and vice versa (false-positive). That is, the observations were exposed to measurement error, and the error rates as well as the true proportion of infection in each group are unestimable. To make the problem estimable, an additional data was necessary. A small subsample from each group was further investigated by refined western procedure which is the inerrant device.

The sampling scheme employed in the case-control study is so-called double sampling. There are numerous examples using the advantages of the double sampling scheme; see Geng and Asano (1989), York *et al.* (1995), Moors *et al.* (2000), Barnett *et al.* (2001), Raats and Moor (2003) and Boese *et al.* (2006). For instance, York *et al.* (1995) illustrated the advantage of the double sampling scheme in estimating the proportion of infants born with Down's syndrome nationwide.

Some fallible devices may have only one type of misclassification. For example, Lie *et al.* (1994) considered the case that the false-negative counts were corrected using multiple fallible classifiers

---

This work was supported by a Hanshin University research grant.

<sup>1</sup> Professor, Department of Statistics, Hanshin University, 411 Yangsan-Dong, Osan, Kyunggi-Do 449-791, Korea.  
E-mail: seung@hs.ac.kr

and gave the ML estimators. The same model was considered by York *et al.* (1995). They estimated the proportion of Down's syndrome in Norway from the Bayesian perspective. Moors *et al.* (2000) analyzed an auditing data with no observed false-negative count. They put the corresponding error rate equal to zero a priori, and gave one-sided confidence intervals for the population proportion. Boese *et al.* (2006) gave five likelihood-based confidence intervals in the false-positive misclassification model. Among them, they recommended the interval based on the Rao score function. The recommended interval seems to have a good frequentist property in that it has the coverage probability close to the nominal level. Under the same model, Lee and Byun (2008) gave a Bayesian confidence interval which slightly outperforms the confidence interval recommended by Boese *et al.* (2006).

Recently a two-sample problem was considered by Lee (2010). He investigated the interval estimation for the difference of population proportions using two doubly sampled data. In this paper, we consider the same problem, but assume the false-positive error rates are common. It is shown that the score statistic scaled by the expected information provide an efficient interval estimate. If the false-positive error rate is common in two data sets, then the confidence interval is more efficient than the corresponding confidence interval considered in Lee (2010).

## 2. Doubly Sampling Model with a Common False-Positive Error Rate

In what follows, we will use the same notation of Lee (2010). That is, for each unit tested by the inerrant device, let  $T_i = 1$ , if  $i^{th}$  unit is recorded positive (or a success), and  $T_i = 0$ , otherwise. Likewise, for each unit tested by the fallible device, define  $F_i = 1$ , if  $i^{th}$  unit is classified as positive, and  $F_i = 0$ , otherwise. Then, the proportion of positive can be written as:

$$p = \Pr[T_i = 1],$$

while the false-positive error rate incurred by the fallible device are defined to be

$$\phi = \Pr[F_i = 1|T_i = 0].$$

The false-negative error is assumed to be zero in this paper. Also we assume that the misclassification errors are independent from sampling unit to sampling unit.

Each unit in the subsample belongs to one of three mutually disjoint categories  $\{(t, f)|(0, 0), (0, 1), (1, 1)\}$  with probabilities  $(1 - p)(1 - \phi)$ ,  $(1 - p)\phi$  and  $p$ , respectively. Let  $n_{tf}$  denote the number of units in  $(t, f)$ . Among  $N - n$  units tested only by a fallible device, let  $x$  be the number of units tested positively and  $y$  be the number of units tested negatively. Then, the joint likelihood of  $p$  and  $\phi$  is

$$L(p, \phi; \mathcal{Y}) = C(\mathcal{Y})(1 - p)^{n_0 + y} p^{n_{11}} (1 - \phi)^{n_{00} + y} \phi^{n_{01}} \pi^x,$$

where  $C(\mathcal{Y}) = n!/(n_{00}!n_{01}!n_{11}!)\binom{N-n}{x}$ ,  $n_{t\cdot} = n_{t0} + n_{t1}$ ,  $\pi = \Pr[F_i = 1] = p + (1 - p)\phi$  and  $\mathcal{Y}$  represents  $(n_{00}, n_{01}, n_{11}, x, y)$ .

A two-sample double sampling model consists of two data sets  $\mathcal{Y}_1 = (n_{100}, n_{101}, n_{111}, x_1, y_1)$  and  $\mathcal{Y}_2 = (n_{200}, n_{201}, n_{211}, x_2, y_2)$ , where each  $\mathcal{Y}_i$  is sampled from  $L(p_i, \phi_i; \mathcal{Y}_i)$  independently. Let  $\lambda = p_1 - p_2$ , then the joint likelihood of  $\lambda$  and  $\Theta^* = (p_2, \phi_1, \phi_2)$  can be written as:

$$L(\lambda, \Theta^*; \mathcal{Y}_1, \mathcal{Y}_2) = L(\lambda + p_2, \phi_1; \mathcal{Y}_1)L(p_2, \phi_2; \mathcal{Y}_2). \quad (2.1)$$

Model (2.1) was considered by Lee (2010). He derived six asymptotic confidence intervals for  $\lambda$ , and compared the performance of confidence intervals in terms of coverage probability and expected

width. It was shown that Bayesian confidence intervals outperformed the other five likelihood-based confidence intervals.

(2.1) is adequate for the inference of  $\lambda$  in general. However, if we note that the false-positive error rate is a characteristic of a fallible device, not a characteristic of population, there are cases in which it may be logical to assume  $\phi_1 = \phi_2$ . For instance, if the same fallible device is applied to obtain  $\mathcal{Y}_1$  and  $\mathcal{Y}_2$  as the case-control study of Hildesheim *et al.* (1991), then the data sets probably have a common error rate and the joint likelihood of  $\lambda$  and  $\Theta = (p_2, \phi)$  is

$$L(\lambda, \Theta; \mathcal{Y}_1, \mathcal{Y}_2) = L(\lambda + p_2, \phi; \mathcal{Y}_1) L(p_2, \phi; \mathcal{Y}_2). \quad (2.2)$$

The assumption of a common false-positive error rate can reduce the dimension of the parameter space. However, the reduction of dimension does not mean that the model becomes more tractable. On the contrary, the reduction requires more computational expense. For instance, (2.2) does not admit the closed form maximum likelihood estimators. Nonetheless, we will see that the computational expense could be compensated by the efficiency of the inferential method.

### 3. Likelihood and Profile Likelihood

#### 3.1. Maximum likelihood estimates

Taking logarithm on (2.2), we have

$$\begin{aligned} \ell(\lambda, \Theta) = & (n_{10\cdot} + y_1) \log(1 - \lambda - p_2) + n_{111} \log(\lambda + p_2) + (n_{20\cdot} + y_2) \log(1 - p_2) + n_{211} \log p_2 + \\ & (n_{100} + n_{200} + y_1 + y_2) \log(1 - \phi) + (n_{101} + n_{201}) \log \phi + x_1 \log \pi_1 + x_2 \log \pi_2, \end{aligned}$$

where  $\pi_1 = (1 - \lambda - p_2)\phi + (\lambda + p_2)$  and  $\pi_2 = (1 - p_2)\phi + p_2$ . The maximum likelihood estimates are the solutions of the following likelihood equations:

$$0 = -\frac{n_{10\cdot} + y_1}{1 - \lambda - p_2} + \frac{n_{111}}{\lambda + p_2} + \frac{(1 - \phi)x_1}{\pi_1} \quad (3.1)$$

$$0 = -\frac{n_{10\cdot} + y_1}{1 - \lambda - p_2} + \frac{n_{111}}{\lambda + p_2} - \frac{n_{20\cdot} + y_2}{1 - p_2} + \frac{n_{211}}{p_2} + (1 - \phi) \left( \frac{x_1}{\pi_1} + \frac{x_2}{\pi_2} \right) \quad (3.2)$$

$$0 = -\frac{n_{100} + n_{200} + y_1 + y_2}{1 - \phi} + \frac{n_{101} + n_{201}}{\phi} + \frac{(1 - \lambda - p_2)x_1}{\pi_1} + \frac{(1 - p_2)x_2}{\pi_2}. \quad (3.3)$$

Given the value of  $\lambda$ , the last two equations, (3.2) and (3.3) form the profile likelihood equations.

Note when  $n_{itf} = 0$  for some  $(i, t, f)$ , the maximum likelihood estimates cannot be defined; see Tenenbein (1970) for further detail. Similarly, the profile log-likelihood does not admit unique maximum. A customary remedy to prevent the undefined problem is to add a small number, say 0.005 to null observed counts. Thus we will add a small number when necessary for the calculation of likelihood or profile likelihood equations.

Suppose  $\hat{\phi}$  is the unique solution of a nonlinear equation on  $\phi \in (0, 1)$

$$\frac{n_{101} + n_{201}}{\phi} - n_{111} \frac{1 - \tilde{p}(\phi; n_{111}, x_1, N_1)}{\tilde{p}(\phi; n_{111}, x_1, N_1)} - n_{211} \frac{1 - \tilde{p}(\phi; n_{211}, x_2, N_2)}{\tilde{p}(\phi; n_{211}, x_2, N_2)} = 0,$$

where

$$\tilde{p}(\phi; n_{11}, x, N) = \frac{(x + n_{11} - (N + n_{11})\phi) + \sqrt{(x + n_{11} - (N + n_{11})\phi)^2 + 4n_{11}N\phi(1 - \phi)}}{2N(1 - \phi)}.$$

Then, it can be shown that  $\hat{\phi}, \hat{p}_2 = \tilde{p}(\hat{\phi}; n_{211}, x_2, N_2)$  and  $\hat{\lambda} = \tilde{p}(\hat{\phi}; n_{111}, x_1, N_1) - \hat{p}_2$  satisfy the likelihood equations. Thus, a simple numerical algorithm such as the bisection method can be applicable to obtain the maximum likelihood estimates of  $\lambda, p_2$  and  $\phi$ .

### 3.2. Profile likelihood and information

Next, we consider the solutions of profile likelihood equations which are essential in what follows. Since  $\pi_1 = \pi_2 + \lambda(1 - \phi)$ , (3.3) is equivalent to

$$A(\phi)\pi_2^2 - (x_1 + x_2 - \lambda(1 - \phi)A(\phi))\pi_2 - \lambda(1 - \phi)x_2 = 0,$$

where  $A(\phi) = (N_1 - n_{111} + N_2 - n_{211}) - (n_{101} + n_{201})/\phi$ . Note that due to (3.2),  $A(\phi) > 0$ . Then, given  $\lambda$  and  $\phi$ ,

$$\tilde{\pi}_2^\lambda(\phi) = \frac{x_1 + x_2 - \lambda(1 - \phi)A(\phi) + \sqrt{(x_1 + x_2 - \lambda(1 - \phi)A(\phi))^2 + 4\lambda(1 - \phi)A(\phi)x_2}}{2A(\phi)}$$

satisfies (3.3), and the profile likelihood equations are equivalent to

$$(1 - \tilde{p}_2^\lambda(\phi))\left(N_1 - \frac{n_{111}}{\tilde{p}_1^\lambda(\phi)} - \frac{x_1}{\tilde{\pi}_1^\lambda(\phi)}\right) + (1 - \tilde{p}_1^\lambda(\phi))\left(N_2 - \frac{n_{211}}{\tilde{p}_2^\lambda(\phi)} - \frac{x_2}{\tilde{\pi}_2^\lambda(\phi)}\right) = 0, \quad (3.4)$$

where  $\tilde{p}_1^\lambda(\phi) = (\tilde{\pi}_2^\lambda(\phi) + \lambda(1 - \phi) - \phi)/(1 - \phi)$ ,  $\tilde{p}_2^\lambda(\phi) = (\tilde{\pi}_1^\lambda(\phi) - \phi)/(1 - \phi)$  and  $\tilde{\pi}_1^\lambda(\phi) = \tilde{\pi}_2^\lambda(\phi) + \lambda(1 - \phi)$ . Thus, we can get a nontrivial nonlinear equation of  $\phi$  for a given  $\lambda \in (-1, 1)$ , which can solve the profile likelihood equations.

As a function of  $\phi$ , (3.4) does not admit unique solution on  $\phi \in (0, 1)$ . However, given  $\lambda \in (-1, 1)$ ,  $p_2 = (\pi_2 - \phi)/(1 - \phi)$  ranges from  $\max(0, -\lambda)$  to  $\min(1, 1 - \lambda)$ . Thus, we should find the solution on

$$\Phi_\lambda = \left\{ \phi \in (0, 1) \mid \max(0, -\lambda) < \frac{\tilde{\pi}_2^\lambda(\phi) - \phi}{1 - \phi} < \min(1, 1 - \lambda) \right\}.$$

Note that  $\Phi_\lambda = (\phi_0^\lambda, \phi_1^\lambda)$  where

$$\phi_0^\lambda = \begin{cases} \frac{\lambda(A + B - x_1) - A + x_1 + x_2 + \sqrt{(\lambda(A + B - x_1) - A + x_1 + x_2)^2 + 4\lambda(1 - \lambda)(A - x_1)B}}{2\lambda(A - x_1)}, & \lambda > 0, \\ \frac{\lambda(x_2 - A - B) - A + x_1 + x_2 + \sqrt{(\lambda(x_2 - A - B) - A + x_1 + x_2)^2 + 4\lambda(1 + \lambda)B(x_2 - A)}}{2\lambda(x_2 - A)}, & \lambda < 0, \\ \frac{B}{A - x_1 - x_2}, & \lambda = 0 \end{cases}$$

and

$$\phi_1^\lambda = \begin{cases} \frac{B - \lambda(A + B + x_2) + x_1 + x_2 + \sqrt{(B - \lambda(A + B + x_2) + x_1 + x_2)^2 + 4\lambda(1 - \lambda)A(B + x_2)}}{2(1 - \lambda)A}, & \lambda > 0, \\ \frac{(1 + \lambda)(B + x_1 + x_2) + \lambda(A - x_2) + \sqrt{((1 + \lambda)(B + x_1 + x_2) + \lambda(A - x_2))^2 - 4\lambda(1 + \lambda)A(B + x_1)}}{2(1 + \lambda)A}, & \lambda < 0 \end{cases}$$

with  $A = N_1 - n_{111} + N_2 - n_{211}$  and  $B = n_{101} + n_{201}$ . Again, a simple numerical algorithm can be applicable to obtain the unique solution of (3.4), say  $\hat{\phi}^\lambda$ . Once we have  $\hat{\phi}^\lambda$ , the profile likelihood is maximized at  $(\hat{p}_2^\lambda, \hat{\phi}^\lambda)$  where  $\hat{p}_2^\lambda = \tilde{p}_2^\lambda(\hat{\phi}^\lambda)$ .

The observed information matrix consists of minus the second-order derivatives of  $\ell(\lambda, \Theta)$ :

$$\begin{aligned} J_{\lambda\lambda} = J_{\lambda p_2} &= \frac{n_{10\cdot} + y_1}{(1-p_1)^2} + \frac{n_{111}}{p_1^2} + \frac{(1-\phi)^2 x_1}{\pi_1^2}, & J_{\lambda\phi} &= \frac{x_1}{\pi_1^2}, \\ J_{p_2 p_2} = J_{\lambda\lambda} &+ \frac{n_{20\cdot} + y_2}{(1-p_2)^2} + \frac{n_{211}}{p_2^2} + \frac{(1-\phi)^2 x_2}{\pi_2^2}, & J_{p_2\phi} &= \frac{x_1}{\pi_1^2} + \frac{x_2}{\pi_2^2}, \\ J_{\phi\phi} &= \frac{n_{100} + n_{200} + y_1 + y_2}{(1-\phi)^2} + \frac{n_{101} + n_{201}}{\phi^2} + \frac{(1-p_1)^2 x_1}{\pi_1^2} + \frac{(1-p_2)^2 x_2}{\pi_2^2}, \end{aligned}$$

where  $p_1 = \lambda + p_2$ . Replacing observed counts by their expectation, we have

$$\begin{aligned} I_{\lambda\lambda} = I_{\lambda p_2} &= \frac{n_1 + (N_1 - n_1)(1-\phi)}{(1-p_1)} + \frac{n_1}{p_1} + \frac{(1-\phi)^2(N_1 - n_1)}{\pi_1}, & I_{\lambda\phi} &= \frac{N_1 - n_1}{\pi_1}, \\ I_{p_2 p_2} = I_{\lambda\lambda} &+ \frac{n_2 + (N_2 - n_2)(1-\phi)}{(1-p_2)} + \frac{n_2}{p_2} + \frac{(1-\phi)^2(N_2 - n_2)}{\pi_2}, & I_{p_2\phi} &= \frac{N_1 - n_1}{\pi_1} + \frac{N_2 - n_2}{\pi_2}, \\ I_{\phi\phi} &= \frac{N_1(1-p_1) + N_2(1-p_2)}{1-\phi} + \frac{n_1(1-p_1) + n_2(1-p_2)}{\phi} + \frac{(1-p_1)^2(N_1 - n_1)}{\pi_1} + \frac{(1-p_2)^2(N_2 - n_2)}{\pi_2}. \end{aligned}$$

Then the observed information for  $\lambda$  is obtained from

$$J^{\lambda\lambda}(\lambda, \Theta) = J_{\lambda\lambda} - (J_{\lambda p_2}, J_{\lambda\phi}) \begin{pmatrix} J_{p_2 p_2} & J_{p_2\phi} \\ I_{p_2\phi} & J_{\phi\phi} \end{pmatrix}^{-1} \begin{pmatrix} J_{\lambda p_2} \\ J_{\lambda\phi} \end{pmatrix}. \quad (3.5)$$

Similarly, the expected information for  $\lambda$ ,  $I^{\lambda\lambda}(\lambda, \Theta)$  can be obtained by replacing  $J$  notation in (3.5) by  $I$  notation.

### 3.3. Likelihood-based confidence intervals

A large sample theory indicates that  $\hat{\lambda}$  is asymptotically normally distributed with mean  $\lambda$  and inverse variance  $I^{\lambda\lambda}(\lambda, \Theta)$ . Thus, the asymptotic distribution of  $(\hat{\lambda} - \lambda)^2 I^{\lambda\lambda}(\lambda, \Theta)$  is a  $\chi^2$ -distribution with 1 degree of freedom. However, the existence of nuisance parameters,  $\Theta$  prevent us from using this result directly. Barndorff-Nielsen and Cox (1994) suggested that  $I^{\lambda\lambda}(\lambda, \Theta)$  can be replaced by  $I^{\lambda\lambda}(\lambda, \hat{\Theta}^\lambda)$ ,  $I^{\lambda\lambda}(\hat{\lambda}, \hat{\Theta})$ ,  $J^{\lambda\lambda}(\lambda, \hat{\Theta}^\lambda)$  and  $J^{\lambda\lambda}(\hat{\lambda}, \hat{\Theta})$ . Hence four Wald-like confidence intervals can be setup as follow:

$$W_{EP} = \left\{ \lambda : (\hat{\lambda} - \lambda)^2 I^{\lambda\lambda}(\lambda, \hat{\Theta}^\lambda) \leq z_\alpha^2 \right\},$$

$$W_{EM} = \left\{ \lambda : (\hat{\lambda} - \lambda)^2 I^{\lambda\lambda}(\hat{\lambda}, \hat{\Theta}) \leq z_\alpha^2 \right\},$$

$$W_{OP} = \left\{ \lambda : (\hat{\lambda} - \lambda)^2 J^{\lambda\lambda}(\lambda, \hat{\Theta}^\lambda) \leq z_\alpha^2 \right\}$$

and

$$W_{OM} = \left\{ \lambda : (\hat{\lambda} - \lambda)^2 J^{\lambda\lambda}(\hat{\lambda}, \hat{\Theta}) \leq z_\alpha^2 \right\}.$$

Next four asymptotic confidence intervals are based on the score statistic obtained from (3.1)

$$U_\lambda(\hat{\Theta}^\lambda) = -\frac{n_{10\cdot} + y_1}{1 - \hat{p}_1^\lambda} + \frac{n_{111}}{\hat{p}_1^\lambda} + \frac{(1 - \hat{\phi}^\lambda)x_1}{\hat{\pi}_1^\lambda},$$

Table 1: Case-control data of Hildesheim *et al.* (1991) (absorbing false-negatives into true-positives)

	Inerrant device	Fallible device			
		Control group		Case group	
		0	1	0	1
Subsample	0	33	11	13	3
	1	na	32	na	23
		701	535	318	375

Table 2: 95% Confidence limits of  $\lambda$  of case-control data of Hildesheim *et al.* (1991) (absorbing false-negatives into true-positives)

Confidence interval	Common false-positive error	No restriction
$W_{EP}$	(-.0936, 0.0119)	(-.254, 0.034)
$W_{EM}$	(-.0920, 0.0120)	not available
$W_{OP}$	(-.0928, 0.0115)	not available
$W_{OM}$	(-.0918, 0.0118)	not available
$S_{EP}$	(-.0917, 0.0113)	(-.238, -0.058)
$S_{EM}$	(-.0924, 0.0112)	not available
$S_{OP}$	(-.0916, 0.0117)	not available
$S_{OM}$	(-.0926, 0.0114)	not available
$L_R$	(-.1361, -.0068)	(-.247, -0.052)

which is also known to asymptotically normally distribute with mean 0 and variance  $I^{\lambda\lambda}(\lambda, \Theta)$ . As before, four asymptotic confidence intervals can be setup as:

$$S_{EP} = \left\{ \lambda : [U_{\lambda}(\hat{\Theta}^{\lambda})]^2 [I^{\lambda\lambda}(\lambda, \hat{\Theta}^{\lambda})]^{-1} \leq z_{\alpha}^2 \right\},$$

$$S_{EM} = \left\{ \lambda : [U_{\lambda}(\hat{\Theta}^{\lambda})]^2 [I^{\lambda\lambda}(\hat{\lambda}, \hat{\Theta})]^{-1} \leq z_{\alpha}^2 \right\},$$

$$S_{OP} = \left\{ \lambda : [U_{\lambda}(\hat{\Theta}^{\lambda})]^2 [J^{\lambda\lambda}(\lambda, \hat{\Theta}^{\lambda})]^{-1} \leq z_{\alpha}^2 \right\}$$

and

$$S_{OM} = \left\{ \lambda : [U_{\lambda}(\hat{\Theta}^{\lambda})]^2 [J^{\lambda\lambda}(\hat{\lambda}, \hat{\Theta})]^{-1} \leq z_{\alpha}^2 \right\}.$$

The last likelihood-based confidence interval is due to the well-known log-likelihood ratio statistic,

$$L_R = \left\{ \lambda : 2 [\ell(\hat{\lambda}, \hat{\Theta}) - \ell(\lambda, \hat{\Theta}^{\lambda})] \leq z_{\alpha/2}^2 \right\}.$$

### 3.4. An example

The case-control study of Hildesheim *et al.* (1991) examined that invasive cervical cancer can affect exposure to Herpes Simplex Virus(HSV). To explore the relationship, western blot procedure was applied to 693 women in the case group and for 1236 women in the control group to detect the infection of HSV. Since the western blot procedure is fallible, a sub-sample from each group was further investigated by refined western blot procedure, which is known to be a relatively accurate procedure. Originally the fallible procedure is exposed to the two types of error, but we assume the false-negative error rate is zero. The false-negative cases are absorbed into the true-positive. This artificial data is shown in Table 1.

The artificial data was analyzed by Lee (2010). Presumably he considered that there were no real restrictions on parameters. However, it may be logical to assume that the false-positive error rate

Table 3: Averages of 81 estimated coverage probabilities, expected widths and AMADN of MLE based confidence intervals

confidence interval																					
$\phi$	Group 1		Group 2		Average coverage probability and expected width, AMADN $\times 10,000$																
	$N_1$	$n_1$	$N_2$	$n_2$	$W_{EP}$			$W_{EM}$			$W_{OP}$			$W_{OM}$			$L_R$				
0.1	100	20	100	20	.947	.270	375	.918	.330	3169	.957	.313	721	.941	.266	949	.943	.267	757		
				30	.947	.265	332	.916	.325	3362	.955	.292	548	.942	.263	793	.944	.264	648		
			200	40	.948	.232	327	.911	.287	3872	.956	.278	653	.942	.230	823	.944	.231	673		
		60		.948	.229	232	.911	.283	3950	.954	.246	418	.943	.228	719	.944	.229	578			
		300	60	.949	.217	250	.909	.270	4108	.955	.264	611	.942	.216	814	.944	.217	655			
			90	.949	.215	244	.908	.266	4236	.954	.230	416	.943	.215	766	.944	.215	608			
	200	40	200	40	.948	.191	294	.891	.252	5886	.954	.215	512	.946	.189	496	.947	.190	420		
				60	.948	.186	251	.886	.247	6417	.952	.196	343	.947	.185	390	.947	.186	333		
			300	60	.948	.174	249	.883	.233	6660	.954	.200	445	.946	.173	434	.947	.173	369		
		90		.949	.171	262	.881	.230	6901	.952	.181	342	.947	.171	411	.947	.171	353			
		0.2	100	20	100	20	.947	.294	372	.914	.370	3558	.956	.322	686	.939	.291	1200	.940	.293	1055
						30	.947	.287	335	.911	.366	3931	.955	.306	548	.940	.285	1023	.942	.286	888
200	40				.948	.253	297	.907	.325	4277	.955	.280	566	.940	.252	1043	.941	.253	902		
	60			.948	.249	250	.904	.321	4562	.953	.260	410	.940	.248	1024	.942	.249	866			
300	60			.948	.237	267	.903	.307	4659	.954	.265	496	.939	.237	1092	.941	.238	929			
	90			.949	.233	234	.902	.303	4814	.953	.242	325	.940	.234	1056	.941	.235	887			
200	40		200	40	.948	.208	252	.884	.288	6621	.954	.223	483	.945	.208	598	.945	.208	520		
				60	.949	.201	234	.875	.285	7521	.952	.207	324	.946	.200	470	.947	.201	415		
			300	60	.949	.190	233	.875	.268	7475	.953	.203	392	.945	.189	537	.946	.190	466		
	90			.949	.186	240	.870	.266	7967	.952	.192	327	.945	.186	516	.946	.186	445			

is common, since the same fallible device was applied to both case and control groups. Under this assumption, the maximum likelihood estimate of  $\lambda$  was  $-0.040$ , which is somewhat different from  $-0.157$  of Lee (2010). Similarly the maximum likelihood estimate of  $\phi$  was  $0.1435$ , but Lee (2010) gave  $0.1677$  and  $0.1226$  for control and case groups, respectively. The limits of 95% confidence intervals for  $\lambda$  are shown in Table 2. The confidence limits calculated by Lee (2010) are also tabulated for comparison. It can be seen that the confidence intervals of a common false-positive model have significantly narrow widths than the corresponding confidence intervals of the no restriction model.

#### 4. Comparison of Confidence Intervals

The likelihood-based confidence intervals are computationally expensive. They require several iterative numerical computations to calculate confidence limits. Practically, it is hard to compute actual coverage probability or expected width. For instance, when  $N_1 = N_2 = 100$  and  $n_1 = n_2 = 20$ , it requires  $18,711 \times 18,711 = 350,101,521$  iterations to calculate actual coverage probability for each parameter point  $(p_1, p_2)$ . Thus, we abandon comparing actual values. The comparison is done through simulations.

We estimated the coverage probability and the expected width of 95% confidence intervals at every 81 grid points of  $(p_1, p_2)$  where  $p_i = 0.1$  up to  $0.9$  with 10,000 random samples under various values of  $N_1, n_2, N_2$  and  $n_2$ , and then calculated the averages of these 81 estimated coverage probabilities and expected widths. We also computed the averages of the mean absolute deviation of coverage probabilities from nominal level (AMADN). The AMADN is multiplied by 10,000. The false-positive error rate considered in this simulation study was  $0.1$  and  $0.2$ , because it is not large in general. The results are shown in Table 3 and 4. The random samples were generated by IMSL FORTRAN Library.

Table 3 demonstrates the performance of five confidence intervals based on the maximum likelihood estimate (MLE). Among them, the coverage probability of  $W_{EP}$  is closest to the nominal level. It

Table 4: Averages of 81 estimated coverage probabilities, expected widths and AMADN of score based confidence intervals

$\phi$	Group 1		Group 2		Average coverage probability and expected width, AMADN $\times 10,000$												
	$N_1$	$n_1$	$N_2$	$n_2$	$S_{EP}$			$S_{EM}$			$S_{OP}$			$S_{OM}$			
0.1	100	20	100	20	.950	.271	200	.954	.287	451	.942	.283	892	.953	.285	410	
				30	.950	.266	166	.953	.276	347	.944	.271	695	.952	.274	309	
			200	40	.950	.233	198	.953	.241	373	.945	.257	652	.952	.240	349	
				60	.950	.230	165	.953	.236	364	.946	.235	540	.953	.235	339	
			300	60	.950	.218	167	.954	.224	370	.947	.250	510	.953	.223	349	
				90	.951	.215	177	.954	.220	447	.947	.223	442	.954	.220	418	
	200	40	200	40	.950	.191	176	.952	.194	268	.947	.199	491	.951	.193	259	
				60	.950	.186	169	.951	.188	242	.948	.189	337	.951	.187	229	
			300	60	.949	.174	170	.951	.176	234	.947	.188	416	.951	.176	234	
				90	.950	.172	195	.951	.173	249	.948	.174	360	.951	.173	249	
	0.2	100	20	100	20	.950	.295	191	.955	.311	535	.941	.301	959	.955	.310	508
					30	.949	.288	170	.954	.298	450	.942	.291	810	.953	.297	425
200				40	.950	.254	171	.954	.264	465	.943	.266	711	.953	.262	445	
				60	.950	.249	181	.954	.257	441	.944	.253	618	.954	.256	430	
300				60	.950	.238	203	.954	.246	446	.945	.257	608	.954	.245	422	
				90	.950	.234	175	.954	.241	470	.945	.238	556	.954	.240	460	
200		40	200	40	.950	.209	183	.952	.213	305	.946	.212	459	.952	.212	298	
				60	.950	.201	182	.952	.203	282	.947	.202	376	.952	.203	276	
			300	60	.950	.190	164	.952	.193	291	.946	.194	424	.952	.192	278	
				90	.950	.187	189	.952	.189	313	.947	.188	384	.952	.188	302	

seems that  $W_{EP}$  would have slightly smaller coverage probability than the nominal level, but be able to approximate the nominal level best. It also has the narrowest average expected width. Thus, we may conclude that  $W_{EP}$  is most preferable among the five MLE based confidence intervals.  $W_{OP}$  also shows a reasonable performance, little conservative with slightly wider width, but  $W_{EP}$  gives better approximation of the coverage probability. Efron and Hinkley (1978) claimed that the observed information is preferable than the expected information in general. However, their claim is not suitable for this case. In view of Boese *et al.* (2006), the preferred information seems to be the expected information for the interval estimation of a proportion. See also Lee (2010).

$W_{EM}$  has the widest average expected width, but has the smallest average coverage probability. Note that  $W_{EM}$  is actually Wald confidence interval constructed by the maximum likelihood estimate and estimate of its asymptotic variance. It was demonstrated in many sampling designs that the performance of Wald confidence interval is poor; see for example, Agresti and Coull (1998), Agresti and Caffo (2000), Browne *et al.* (2001), Boese *et al.* (2006), and Lee (2010). The small value of coverage probability with the wide expected width suggests that the location of interval would not be proper. A shrinkage estimate such as a Bayesian estimate rather than MLE seems to be desirable for the interval estimation.  $W_{OM}$  and  $L_R$  are not interesting as well, since they are dominated by  $W_{EP}$  in the approximation and the expected width.

Similar conclusions can be reached from Table 4 which shows the performance of four score based confidence intervals. The score statistic scaled by the inverse expected information shows remarkably good performance in the approximation as well as the expected width. Even, it dominates  $W_{EP}$ . However, the observed information seems not to provide a proper scale for the score statistic, since  $S_{OP}$  has significantly smaller coverage probability than the nominal level, but has relative wider expected width than  $S_{EP}$ . Again the expected information is the preferred information.  $S_{EM}$  and  $S_{OM}$  works reasonably or even better than MLE based confidence intervals, but they are not good enough to compete with  $S_{EP}$ .



## 5. Conclusion

In the nature of double sampling design for binary data, it customarily happens that the error rates should be the same. For instance, the confidence intervals given in Lee (2010) was obtained under no restriction on parameters. However, the error rates incurred by fallible device should be the same in the case-control study of Hildesheim *et al.* (1991). Thus, the model considered in this paper is applicable correctly, if we can assume that the false-negative error rate is zero.

The assumption of common error rates can make the model simple by reducing the number of nuisance parameters, but it does not make the statistical problem more tractable. Rather, the assumption requires more computational expense. However, a methodology to handle the statistical problem is given in this paper. It was shown by the simulation study that the score statistic scaled by the inverse of the expected information has remarkable performance in terms of the approximation and the expected width. If we have assumed correctly, then  $S_{EP}$  can reduce the expected width greatly with better approximation compared with the corresponding interval under no restrictions on nuisance parameters.

## References

- Agresti, A. and Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures, *The American Statistician*, **54**, 280–288.
- Agresti, A. and Coull, B. A. (1998). Approximation is better than “exact” for interval estimation of binomial proportions, *The American Statistician*, **52**, 119–126.
- Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics*, Chapman & Hall, London.
- Barnett, V., Haworth, J. and Smith, T. M. F. (2001). A two-phase sampling scheme with applications to auditing or sed quis custodiet ipsos custodes?, *Journal of Royal Statistical Society, Serie A*, **164**, 407–422.
- Boese, D. H., Young, D. M. and Stamey, J. D. (2006). Confidence intervals for a binomial parameter based on binary data subject to false-positive misclassification, *Computational Statistics and Data Analysis*, **50**, 3369–3385.
- Brown, L. D., Cai, T. T. and DasGupta, A. (2001). Interval estimation for a binomial proportion, *Statistical Science*, **16**, 101–133.
- Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information, *Biometrika*, **65**, 457–482.
- Geng, Z. and Asano, C. (1989). Bayesian estimation methods for categorical data with misclassifications, *Communications in Statistics, Theory and Methods*, **18**, 2935–2954.
- Hildesheim, A., Mann, V., Brinton, L. A., Szklo, M., Reeves, W. C. and Rawls, W. E. (1991). Herpes simplex virus type 2: A possible interaction with human papillomavirus types 16/18 in the development of invasion cervical cancer, *International Journal of Cancer*, **49**, 335–340.
- Lee, S. C. (2010). Confidence intervals for the difference of binomial proportions in two doubly sampled data, *Communications of the Korean Statistical Association*, **3**, 301–310.
- Lee, S. C. and Byun, J. S. (2008). A Bayesian approach to obtain confidence intervals for binomial proportion in a double sampling scheme subject to false-positive misclassification, *Journal of the Korean Statistical Society*, **37**, 393–403.
- Lie, R. T., Heuch, I. and Irgens, L. M. (1994). Maximum likelihood estimation of proportion of congenital malformations using double registration systems, *Biometrics*, **50**, 433–444.

- Moors, J. J. A., van der Genugten, B. B. and Strijbosch, L. W. G. (2000). Repeated audit controls, *Statistica Neerlandica*, **54**, 3–13.
- Raats, V. M. and Moors, J. J. A. (2003). Double-checking auditors: A Bayesian approach, *The Statistician*, **52**, 351–365.
- Tenenbein, A. (1970). A double sampling scheme for estimating from binomial data with misclassifications, *Journal of the American Statistical Association*, **65**, 1350–1361.
- York, J., Madigan, D., Heuch, I. and Lie, R. T. (1995). Birth defects registered by double sampling: A Bayesian approach incorporating covariates and model uncertainty, *Applied Statistics*, **44**, 227–242.

*Received June 2010; Accepted July 2010*