

Article

Bayesian Inference for the Difference of Two Proportion Parameters in Over-Reported Two-Sample Binomial Data Using the Doubly Sample

Dewi Rahardja [†]

U.S. Department of Defense, Fort Meade, MD 20755, USA; rahardja@gmail.com

[†] Disclaimer Statement: This research represents the author's own work and opinion. It does not reflect any policy nor represent the official position of the U.S. Department of Defense nor any other U.S. Federal Agency.

Received: 9 December 2018; Accepted: 3 February 2019; Published: 11 February 2019

Abstract: We construct a point and interval estimation using a Bayesian approach for the difference of two population proportion parameters based on two independent samples of binomial data subject to one type of misclassification. Specifically, we derive an easy-to-implement closed-form algorithm for drawing from the posterior distributions. For illustration, we applied our algorithm to a real data example. Finally, we conduct simulation studies to demonstrate the efficiency of our algorithm for Bayesian inference.

Keywords: Bayesian inference; binary data; double sampling; misclassification; two-sample; proportions difference

1. Introduction

Misclassifications can occur in binomial data due to human errors or imprecise diagnostic procedures. Typically, misclassifications of both types (false-positive or false-negative) are possible. In a one-sample case, for instance, with an imperfect blood test, a healthy patient can be mistakenly classified as sick and vice versa. However, in some cases, only one type of misclassification exists. For example, Moors *et al.* [1] presented auditing data where only false-negative (under-reported) errors occurred. Perry *et al.* [2] showed blood testing data which had only false-positive (over-reported) errors. In a two-sample case, for example, we may wish to obtain the odds ratio for lung cancer (or traffic accidents) between males and females. The grouping variables (males/females) can typically be inerrant while the response variable (cancer/not) can sometime be misclassified (errant). In this lung cancer (two-sample) case, under-reporting (false-negative) occurs when a person (male/female) dies due to lung cancer but this fact is not reflected in the death certificate. As another example, in the traffic accident case, over-reporting (false-positive) occurs when a driver has lied to a police officer about not wearing a seatbelt, while the hospital record examination showed otherwise, for an example see the U.S. National Highway Traffic Safety Administration, NHTSA [3].

Among many researchers, Bross [4] reported that classical estimators ignoring misclassification can be extremely biased for certain applications.

Due to economic viability reasons, in some cases, the binary outcome variable can be obtained using an errant classifier/device. In such cases, classical inferential methods are testing and estimating proportion parameters corresponding to the errant classifier/device, but not to an inerrant device (gold standard). As a consequence, these classical methods are biased and invalid. For example, Tenenbein [5] derived a formula quantifying the bias of the classical estimator of a single proportion in misclassified binomial data. Therefore, such misclassified binomial data need to

be somehow augmented to make valid statistical inferences on the proportion parameters corresponding to the inerrant device.

The most frequently used data-augmenting method for misclassified binomial data is the double-sampling scheme, pioneered by Tenenbein [5] for the one-sample problem. Such a scheme assumes an inerrant device is available only to a small number of experimental units due to economic viability reasons. The double-sampling scheme works in the following way. First, either a random subsample from the original misclassified data or another new sample is chosen. In the first case, where a random subsample from the original data is selected, we use the inerrant device to obtain the inerrant measurement of the binary outcome variable for the subsample. In the second case, where another new sample is selected, we use both the errant and the inerrant devices to obtain the binary outcome variable for this new sample. In both cases, we name the sample classified using only the errant device for the binary outcome variable as the original study and the sample classified using both the errant and inerrant devices for the binary outcome variable as the validation sub-study. Statistical inference is made based on the augmented overall data, combining the original study and the validation sub-study.

The following is the literature regarding several statistical inference methods for one-sample and two-sample misclassified binomial data. For one-sample data with only one type of misclassification error (false-positive or false-negative error), Moors *et al.* [1] derived a one-sided exact confidence interval (CI) and Boese *et al.* [6] derived several likelihood-based confidence intervals for the proportion parameters corresponding to the inerrant device (the true proportion parameter). Raats and Moors [7] and Lee and Byun [8] reported Bayesian credible intervals for the true proportion parameter with one type of misclassification only. For one-sample data with both types of misclassification errors, Raats and Moors [7] derived both an exact confidence interval and a Bayesian credible interval for the true proportion parameter, and Lee [9] reported a Bayesian interval estimation for the true binomial proportion parameter. For two-sample data with two types of misclassification errors, Prescott and Garthwaite [10] proposed a Bayesian credible interval, and Morrissey and Spiegelman [11] derived likelihood-based confidence intervals for the odds ratio. However, their methods are computationally burdensome, not guaranteed to converge to the true parameter value, and hard to reproduce by other practitioners.

In this paper we dealt with two-sample binomial data with one type of misclassification error obtained using double sampling, via a Bayesian approach. Our research objectives are to conduct Bayesian inference for the difference of two proportion parameters in over-reported two-sample binomial data using the doubly sample. The remainder of the paper is organized as follows. In Section 2 we describe the data. In Section 3 we develop a Bayesian point and interval estimation. Next, we illustrate our Bayesian method using real data, in Section 4. A discussion follows in Section 5.

2. Data

In this section, we consider doubly-sampled data containing two samples in both the original study and the validation sub-study subject to one type of misclassification error. Without loss of generality (WLOG), we assume that the data were obtained using an errant classification procedure that yields false-positive but not false-negative counts. For sample i ($i = 1, 2$), we let m_i and n_i be the number of items in the original and the validation study, respectively. We then defined $N_i = m_i + n_i$ as the total sample size for Sample i . For the j th item in the i th sample, where $i = 1, 2$, and $j = 1, \dots, N_i$, we let F_{ij} and T_{ij} be the binary response variables obtained by the errant and inerrant devices, respectively. Next, we denote $F_{ij} = 1$ if the result is positive by the errant device and $F_{ij} = 0$ otherwise. The binary variable T_{ij} is similarly defined for the inerrant device. Note that F_{ij} is observed for all items in both the original study and the validation sub-study, while T_{ij} is observed for items in the validation sub-study but not for items in the original study. Clearly, misclassification error occurs when F_{ij} differs from T_{ij} .

In the validation sub-study, for $i = 1, 2$, $j = 0, 1$, and $k = 0, 1$, we use n_{ijk} to denote the number of items in Sample i classified as j and k by the inerrant device and the errant device, respectively. In the

main study, we let x_i and y_i be the number of positive and negative classifications in Sample i obtained by the errant device. The summary frequencies in both the main study and the validation sub-study for sample i are displayed in Table 1.

Table 1. Sample i ($i = 1, 2$) binary data subject to false-positive misclassification and obtained using double sampling.

Study	Inerrant device	Errant device		
		0	1	Total
Validation	0	n_{i00}	n_{i01}	$n_{i0\cdot}$
	1	NA	n_{i11}	n_{i11}
	Total	n_{i00}	$n_{i\cdot 1}$	n_i
Original	NA	Y_i	X_i	M_i

NA: not available.

Next, we define the parameters for sample i as follows. We let the true proportion parameter of interest be $p_i = \Pr(T_{ij} = 1)$, the proportion parameter of the errant device be $\pi_i = \Pr(F_{ij} = 1)$, and the false-positive rate of the errant device be $\varphi_i = \Pr(F_{ij} = 1 | T_{ij} = 0)$. Note that π_i is not an additional unique parameter because it is obtainable by using all other parameters. In particular, by the law of total probability, we have

$$\pi_i = \Pr(T_{ij} = 1) \Pr(F_{ij} = 1 | T_{ij} = 1) + \Pr(T_{ij} = 0) \Pr(F_{ij} = 1 | T_{ij} = 0) = p_i + q_i \varphi_i, \quad (1)$$

where $q_i = 1 - p_i$. For the summary frequencies displayed in Table 1, the corresponding cell probabilities are shown in Table 2.

Table 2. Cell probabilities for sample i ($i = 1, 2$).

Study	Inerrant device	Errant device		
		0	1	Total
Validation	0	$q_i (1 - \varphi_i)$	$q_i \varphi_i$	q_i
	1	NA	p_i	p_i
Original	NA	$1 - \pi_i$	π_i	1

NA: not available.

Our goal is to conduct Bayesian inference and we are interested in estimating and testing if a difference exists among the two true proportion parameters p_1, p_2 . In particular, the statistical hypotheses are

$$H_0 : p_1 = p_2 \text{ versus } H_1 : p_1 \neq p_2. \quad (2)$$

We develop Bayesian algorithms to estimate all the parameters in Expression (1) and constructed a Bayesian credible interval (CI) to see whether the two true proportion parameters p_1, p_2 are equal (i.e., if zero is in the CI). We note that the proportion difference, δ can be expressed as follows:

$$\delta = p_1 - p_2. \quad (3)$$

Therefore, our objective is ultimately to construct a $100(1-\alpha)\%$ Bayesian CI for δ and hence, simultaneously test (2) by checking if zero is in the $100(1-\alpha)\%$ Bayesian CI for δ , where α is the type-I error rate.

3. Model

We derive the Bayesian inference for the data described in the previous section. Specifically, we develop a closed-form algorithm for sampling from the exact joint posterior distribution of all the parameters given the data. Once a posterior sample is drawn for p_1 and p_2 then a posterior sample for the difference, δ , in Expression (3) can be obtained. Finally, we construct a $100(1-\alpha)\%$ CI for δ .

WLOG, for Sample i in Table 1, the observed counts $(n_{i00}, n_{i01}, n_{i11})$ of the validation sub-study have a trinomial distribution with total size n_i and probabilities displayed in an upper right 2×2 submatrix in Table 2, i.e.,

$$(n_{i00}, n_{i01}, n_{i11}) | p_i, \varphi_i \sim \text{Trinomial}[n_i, (q_i(1-\varphi_i), q_i \varphi_i, p_i)]. \quad (4)$$

In addition, the observed counts (X_i, Y_i) in the main study have the following binomial distribution:

$$(X_i, Y_i) | p_i, \varphi_i, \theta_i \sim \text{Bin}(M, \pi_i). \quad (5)$$

Since $(n_{i00}, n_{i01}, n_{i11})$ and (X_i, Y_i) are independent for group i and these cell counts are independent across groups, the sampling distribution of the vector of all data

$$\mathbf{d} = (n_{100}, n_{101}, n_{111}, X_1, Y_1, n_{200}, n_{201}, n_{211}, X_2, Y_2) \quad (6)$$

given the vector of all parameters

$$\boldsymbol{\eta} = (p_1, \varphi_1, p_2, \varphi_2)$$

is

$$f(\mathbf{d} | \boldsymbol{\eta}) \propto \prod_{i=1}^2 \{ [q_i(1-\varphi_i)]^{n_{i00}} (q_i \varphi_i)^{n_{i01}} p_i^{n_{i11}} \pi_i^{X_i} (1-\pi_i)^{Y_i} \}. \quad (7)$$

To put it into a Bayesian framework, a non-informative proper prior for $\boldsymbol{\eta}$ is chosen. In particular, a uniform prior for each component of $\boldsymbol{\eta}$ is chosen, where each parameter (i.e., component of $\boldsymbol{\eta}$) is constrained to lay in an interval of $(0, 1)$, and we assume that these priors are independent, i.e., the joint prior distribution is

$$p(\boldsymbol{\eta}) = 1. \quad (8)$$

The following reasons explain why the uniform priors are chosen. First, typically in most applications, the prior information is unknown (non-informative); therefore, a uniform prior which is a special case of Beta distribution, i.e., Beta $(1, 1)$ equivalence, would be appropriate since the distribution is flat or constant (non-informative) for all possible values within $(0, 1)$ interval constraints. Second, the availability of the validation sub-study provides information about all model parameters and therefore an informative prior is not needed any more. Third, as Yang and Berger [12] mentioned, Bayesian analysis with non-informative priors is increasingly recognized as a method for classical statisticians to obtain good classical procedures. In this paper, it is also desirable to aim for such a frequentist matching approach to develop non-informative (flat or constant) priors (i.e., most suitably, uniform priors) to ensure that we have Bayesian credible sets with good/matching frequentist properties for our forthcoming future research so as to find good/matching frequentist confidence intervals.

Combining Equations (7) and (8), we obtain the following joint posterior distribution:

$$f(\boldsymbol{\eta} | \mathbf{d}) \propto \prod_{i=1}^2 \{ [q_i(1-\varphi_i)^{n_{i00}} (q_i\varphi_i)^{n_{i01}} p_i^{n_{i11}} \pi_i^{X_i} (1-\pi_i)^{Y_i} \}, \quad (9)$$

which has the same functional form as the sampling distribution in Equation (7).

In general, it is nontrivial to sample from the posterior distribution (9). Therefore, we derive a closed-form algorithm for sampling from Equation (9) via the reparameterization of $\boldsymbol{\eta}$. Note that the term reparameterization is different than transformation. For transformation a Jacobian is needed, while in reparameterization we simply reparametrize the original parameters for algebraic convenience and hence, no Jacobian is needed. Note also that other non-informative priors than the uniform prior, such as Beta ($\frac{1}{2}$, $\frac{1}{2}$) which is the Jeffreys prior, could also be considered plausible or sensible (depending on the specific aim of the problem statement), as long as they produce proper posteriors and as long as they are in alignment with desirable good/matching frequentist properties, resulting in a good/matching frequentist likelihood function for its corresponding/counterpart frequentist confidence intervals. Also note that proper means it is a probability density and hence does integrate to 1, while improper means it is not a probability density and hence does not integrate to 1. In particular, we define

$$\lambda_i = \frac{p_i}{\pi_i}. \quad (10)$$

Using Equation (10), the posterior density in Equation (9) become

$$f(\boldsymbol{\eta}^* | \mathbf{d}) \propto \prod_{i=1}^2 \lambda_i^{n_{i11}} (1-\lambda_i)^{n_{i01}} \pi_i^{X_i+n_{i,1}} (1-\pi_i)^{Y_i+n_{i00}}, \quad (11)$$

where \mathbf{d} is defined in Equation (6) and

$$\boldsymbol{\eta}^* = (\lambda_1, \pi_1, \lambda_2, \pi_2) \quad (12)$$

is the re-parameterized parameter vector. Because the re-parameterized parameter vector of Equation (12) is now separable, we can straightforwardly draw λ_i and π_i from the posterior Equation (11) by using the following closed-form algorithm:

$$\lambda_i \sim \text{Beta}(n_{i11} + 1, n_{i01} + 1), \quad (13)$$

$$\pi_i \sim \text{Beta}(X_i + n_{i,1} + 1, Y_i + n_{i00} + 1), \quad (14)$$

with $i = 1, 2$. Next, when λ_i and π_i are available, we can obtain p_i and φ_i by solving Equations (1) and (10) so that

$$p_i = \pi_i \lambda_i, \quad (15)$$

$$\varphi_i = \frac{(1-\lambda_i)\pi_i}{q_i}. \quad (16)$$

In summary, the following is the closed-form algorithm to sample from the posterior density of Equation (11).

1. Choose a large number J , say, 10,000, as a large sample size for the posterior draw.
2. Obtain J samples of λ_i and π_i , each for $i = 1, 2$, using (13) and (14), respectively.

3. Obtain J samples of p_i and φ_i , each for $i = 1, 2$, using (15) and (16), respectively.
4. Compute J samples of each difference, δ , from Equation (3).
5. Use the median of the posterior sample of δ as a point estimator of δ . The median was chosen because the distribution of the posterior sample of δ is skewed.
6. Obtain a $100(1-\alpha)\%$ CI for δ by using the lower and upper $(\alpha/2)$ th percentile of the sample of δ , where α is the type I error rate.
7. Finally, the statistical hypothesis testing in (2) can be rejected if this CI does not contain the number zero.

4. Example

As an illustration of our proposed procedure, we utilize over-a-decade old retrospective data collected from a large introductory statistics class (1148 students) at a university in the United States of America. This is not a prospective study with randomized trials. Several teaching assistants (TA) were hired to grade the homework, quizzes, extra credits, etc. from the laboratory (lab) component of that course. Note that to alleviate extremely high volumes for homework grading, only two selected problems were chosen to be graded, but these were not pre-announced to the students. Hence, the students were still required to do all assigned homework problems. Grades in this lab portion of the course constituted only 20% of the course grade. Due to the large class size, weekly time constraints, and the need for speed grading so that assignments could be returned in a timely manner, the TA (errant device) tended to grade assignments on a pass/fail scale very leniently to avoid students coming back to bargain for better scores. This resulted in an overly high pass rate, i.e., a false-positive rate for the lab portion of the course grade. In that lab portion of the course grade, “pass” was defined as 60% or above, and was coded as 1 (Tables 1–3). Likewise, “fail” was defined as below 60% and was coded as 0. The professor (inerrant device) was interested in auditing those TAs’ grading for the quality of graded (pass/fail) lab scores. He randomly selected a smaller sample of size 434 ($144 + 23 + 47 + 169 + 13 + 38$) as a validation sub-study. This total validation sub-study size, 434 out of 1148 (37.8%), is not small. As the professor’s original intention was to draw validation samples until a sufficient number of false-positive errors were caught, he convinced himself that this kind of grading system is sustainable in the long run for many large classes. It turns out that he caught 13 out of 582 (about 2%) of false-positive errors in the female group and 23 out of 656 (about 4%) in the male groups. Both group’s error rates were satisfactory for the professor’s original intention.

In this study, the interest is to estimate the overall score/performance of male and female students, in terms of their proportion difference. Summary statistics of data are provided in Table 3.

Table 3. The audited pass/fail data using double sampling with false-positive misclassification.

Sample	Study	Inerrant Device	Errant Device	
			0	1
Male	Validation	0	144	23
		1	NA	47
	Original	NA	443	123
Female	Validation	0	169	13
		1	NA	38
	Original	NA	480	102

NA: not available; 1 = pass; 0 = fail.

Note that, since there is no personally identifiable information (PII) of any human subject (such as the person’s name, university name, date of birth or DOB, social security number or SSN, numeric grades, etc.) involved, and only summary statistics (such as counts and proportions of pass/fail lab portion, in Table 3) are displayed, an International Review Board (IRB) approval or an exemption to collect such data and publish results based on Table 3 is not required. Moreover, this was a student

class assignment which was a sole exception for classroom activities, and hence did not fall under the jurisdiction of the IRB and did not require IRB application, approval, or oversight (<http://research-compliance.umich.edu/human-subjects/human-research-protection-program-hrpp/hrpp-policies/class-assignments-irb-approval>).

From Expression (2), the null hypothesis in this context is to test if the proportions of failing the lab portion of the course between male and female students are the same, i.e., the δ in Expression (3) is equal to zero; with the alternative hypothesis that the two proportions are not the same, i.e., δ is not equal to zero. Following the algorithm from the previous section, the median of the posterior sampling distribution of the proportion difference δ is 0.024, i.e., the proportion difference of failing the course for male students is about 0.024 points more than female students, i.e., we see that the male students had about a 2.4% higher chance of failing the lab portion of the course than female students. The resulting 90% Bayesian CI for the difference δ is $(-0.014, 0.063)$. Because this CI contains zero, we conclude that the study did not provide enough statistical evidence in support of the outcome (pass/fail) difference between the two genders (male/female students). Table 4 displays the parameter estimates and 90% CI for the audited pass/fail data.

Table 4. Parameter estimates and 90% Credible Interval (CI) for the audited pass/fail data.

Parameters	Estimates	90% CI
p_1	0.165	(0.137, 0.193)
p_2	0.141	(0.116, 0.166)
δ	0.024	$(-0.014, 0.063)$

5. Simulation

We conduct simulation studies to evaluate and compare the performance of our algorithms under various scenarios, such as evaluating the impact on the CIs of δ by varying sample sizes, the ratio of validation and original study, and the false-positive rate. We considered a two-sided nominal type-I error level of $\alpha = 0.1$. Although not required by our algorithm, in order to simplify the conduct of the simulation studies and the presentation of the simulation results, we let the total sample sizes be $N_1 = N_2 = N$, sub-study sample sizes be $n_1 = n_2 = n$, and false-positive rates be $\varphi_1 = \varphi_2 = \varphi$. For each simulation scenario a total of $K = 10,000$ datasets is generated.

We consider 32 simulation scenarios resulting from combinations of the following values:

1. True proportion parameters of interest $(p_1, p_2) = (0.1, 0.2), (0.4, 0.6)$.
2. False-positive rate $\varphi = 0.1, 0.2$.
3. Ratio of sub-study sample size versus the total sample size $r = n/N = 0.2, 0.4$.
4. Total sample size $N = 100, 200, 300, 400$.

For each simulation scenario, we simulate $K = 10,000$ datasets. Within each dataset, we draw a size of $J = 10,000$ posterior samples of δ , according to the algorithm in Section 3. We then compute the posterior median (point estimator) and a 90% CI. Finally, we generate boxplots of the K posterior medians of δ to examine their behavior around the true δ . In addition, we calculate the coverage probability (CP) and the average length (AL) of the K CIs.

In Figures 1 and 2, we present the boxplots of K posterior medians of δ against the total sample size N . The true proportion parameters of (p_1, p_2) are $(0.1, 0.2)$ and $(0.4, 0.6)$ for Figures 1 and 2, respectively. In each figure, the top two panels have $\varphi = 0.2$ and the bottom two panels have $\varphi = 0.1$. In addition, the left two panels have $n/N = 0.2$ and the right two panels have $n/N = 0.4$.

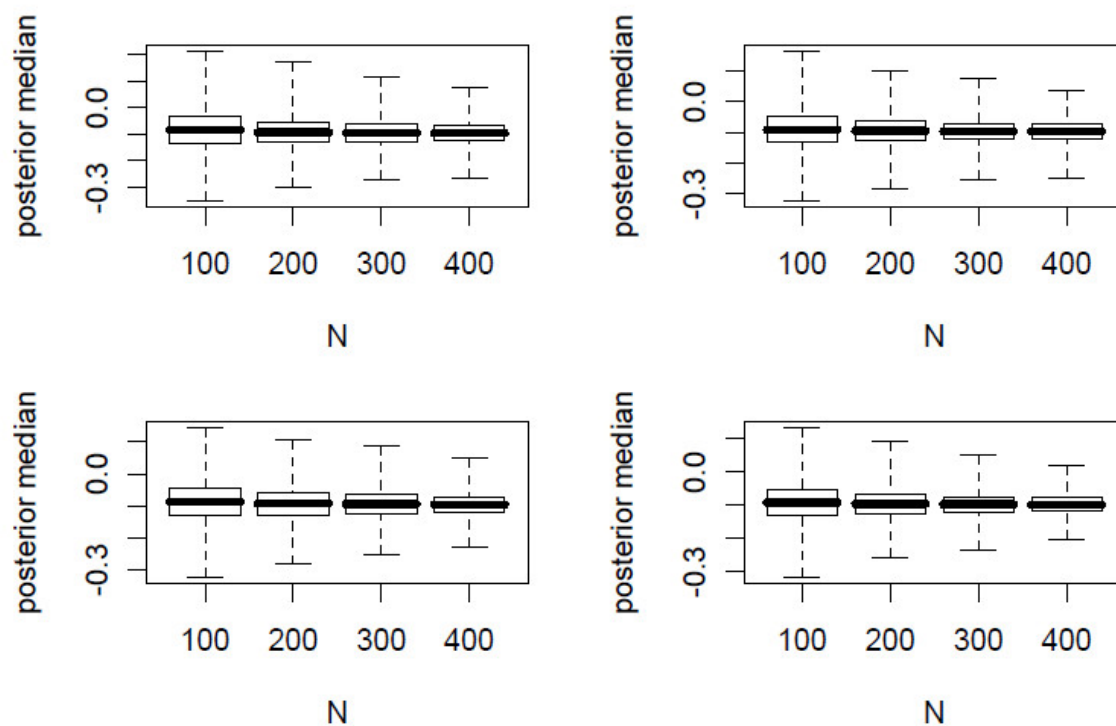


Figure 1. Boxplots of posterior medians versus total sample sizes N where $(p_1, p_2) = (0.1, 0.2)$. The top two panels have $\varphi = 0.2$ and the bottom two panels have $\varphi = 0.1$; the left two panels have $n/N = 0.2$ and the right two panels have $n/N = 0.4$.

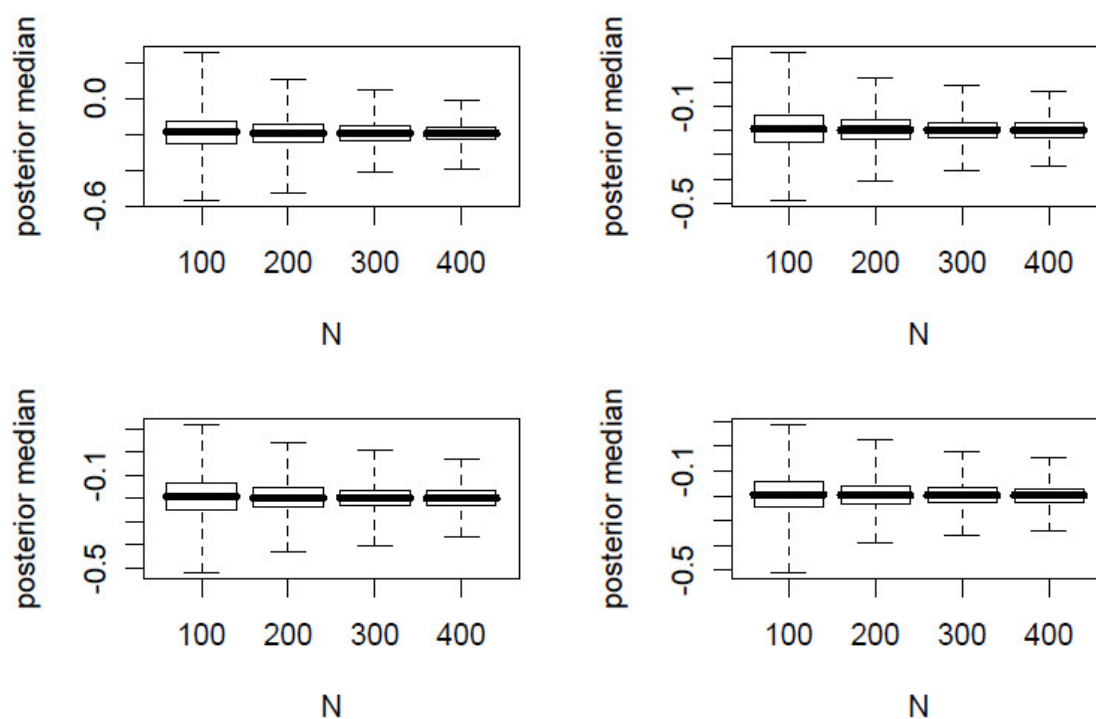


Figure 2. Boxplots of posterior medians versus total sample sizes N where $(p_1, p_2) = (0.4, 0.6)$. The top two panels have $\varphi = 0.2$ and the bottom two panels have $\varphi = 0.1$; the left two panels have $n/N = 0.2$ and the right two panels have $n/N = 0.4$.

From the 32 simulation scenarios for both figures, the posterior medians are centered around the true δ and hence is a good point estimator. Additionally we made the following observations:

1. For each panel of four boxplots, the variation of the posterior medians decreases as N increases.
2. For each figure, the variation of the posterior medians of the top two panels with a larger φ is slightly greater than that of the bottom two panels with a smaller φ .
3. For each figure, the variation of the posterior medians of the left two panels with a smaller n/N is slightly greater than that of the right two panels with a larger n/N .

Table 5. Coverage probabilities (CPs) and average lengths (ALs) of 90% CIs for risk difference δ .

Delta	(p_1, p_2)	φ	n/N	N				
					100	200	300	400
−0.1	(0.1,0.2)	0.2	0.2	CP	91.61	90.74	90.46	90.19
				AL	0.26	0.20	0.17	0.15
			0.4	CP	90.81	90.52	90.32	90.67
				AL	0.21	0.15	0.13	0.11
		0.1	0.2	CP	91.58	90.82	90.70	90.48
				AL	0.23	0.17	0.14	0.13
			0.4	CP	90.50	90.39	90.18	89.96
				AL	0.19	0.14	0.12	0.10
		0.2	0.2	CP	91.62	91.07	90.24	90.58
				AL	0.33	0.24	0.20	0.17
			0.4	CP	90.28	90.49	89.85	90.04
				AL	0.27	0.20	0.16	0.14
−0.2	(0.4,0.6)	0.2	0.2	CP	93.31	91.49	90.96	90.75
				AL	0.31	0.22	0.18	0.15
			0.4	CP	90.81	90.48	90.61	90.08
				AL	0.26	0.18	0.15	0.13

In Table 5, we show the CPs and ALs of the K CIs of δ for each simulation scenario. Since the CPs are all close to the nominal 90% level, this is an indication that our CI estimator is a good estimator. Similar to the observations made for the figures, we made the following observations in Table 5:

1. For fixed (p_1, p_2) , φ , n/N , the AL of the CIs decreases as N increases.
2. For fixed (p_1, p_2) , n/N , N , the AL of the CIs decreases as φ decreases.
3. For fixed (p_1, p_2) , φ , N , the AL of the CIs decreases as n/N increases.

6. Discussion

In this article we derived a Bayesian algorithm to conduct statistical inference on the difference of two proportion parameters for binary data subject to one type of misclassification. Our closed-form algorithm to draw from the full posterior distributions has many advantages, including the following:

- Since we draw directly from the posterior distributions, there is no need to specify initial values and there is no burn-in period or convergence problem.
- Our algorithm can handle zero counts as shown in Equations (13) and (14).
- No asymptotic (large sample) theory is involved and hence it is easy to implement, in the sense that our algorithm does not require a large sample size for the complex asymptotic (large sample) theory, such as the regularity conditions, to work out—rather, it simply draws from the joint posterior distribution.

The uniform (0, 1) prior distribution in Equation (8) is identical to a Beta (1, 1) distribution; however, to generalize this prior distribution in Expression (8) to be Beta (α , β) would complicate Equations (8) and (9). This would affect the generalized sensibility of the results in relation to the

choice of the hyper-parameter values. Still, doing so can be considered so long as the hyper-parameter choices produce sensible proper (i.e., a probability density that integrates to 1) posteriors and produce the corresponding good/matching frequentist's counterpart of confidence intervals. For example, the Uniform (0, 1) priors which are Beta (1, 1) and the Jeffreys priors which are Beta ($\frac{1}{2}$, $\frac{1}{2}$) satisfy the above two aforementioned criteria.

Funding: This research received no external funding.

Acknowledgments: The author would like to thank the two anonymous referees for their insightful comments and constructive suggestions which have improved the presentations of this manuscript.

Conflicts of Interests: The authors declare no conflict of interest.

References

1. Moors, J.J.A.; van der Genugten, B.B.; Strijbosch, L.W.G. Repeated audit controls. *Stat. Neerl.* **2000**, *54*, 3–13.
2. Perry, M.; Vakil, N.; Cutler, A. Admixture with whole blood does not explain false-negative urease tests. *J. Clin. Gastroenterol.* **2000**, *30*, 64–65.
3. United States Department of Transportation, National Highway Traffic Safety Administration, NHTSA. In *Report to Congress on the Benefits of Safety Belts and Motorcycle Helmets Based on Data from the Crash Outcome Data Evaluation System (CODES)*, NHTSA Technical Report DOT HS 808 338; NHTSA: Washington, DC, USA, 1996.
4. Bross, I. Misclassification in 2x2 tables. *Biometrics.* **1954**, *10*, 478–486.
5. Tenenbein, A. A double sampling scheme for estimating from binomial data with misclassifications. *J. Am. Statist. Assoc.* **1970**, *65*, 1350–1361.
6. Boese, D.H.; Young, D.M.; Stamey, J.D. Confidence intervals for a binomial parameter based on binary data subject to false-positive misclassification. *Comput. Stat. Data Anal.* **2006**, *50*, 3369–3385.
7. Raats, V.M.; Moors, J.J.A. Double-checking auditors: a Bayesian approach. *The Statistician* **2003**, *52*, 351–365.
8. Lee, S.C.; Byun, J.S. A Bayesian approach to obtain confidence intervals for binomial proportion in a double sampling scheme subject to false-positive misclassification. *J. Korean Stat. Soc.* **2018**, *37*, 393–403.
9. Lee, S.C. Bayesian confidence intervals of proportion with misclassified binary data. *J. Korean Stat. Soc.* **2013**, *42*, 291–299.
10. Prescott, G.J.; Garthwaite, P.H. A simple Bayesian analysis of misclassified binary data with a validation substudy. *Biometrics* **2002**, *58*, 454–458.
11. Morrissey, M.J.; Spiegelman, D. Matrix methods for estimating odds ratios with misclassified exposure data: Extensions and comparisons. *Biometrics* **1999**, *55*, 338–344.
12. Yang, R.; Berger, J.O. A Catalog of Noninformative Priors. 1998. Available Online: <http://www.stats.org.uk/priors/noninformative/YangBerger1998.pdf> (accessed on 31 January 2019).



© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).