frontiers in
PUBLIC HEALTH

METHODS ARTICLE
published: 07 October 2013
doi: 10.3389/fpubh.2013.00039

Improved confidence intervals of a small probability from pooled testing with misclassification

Chunling Liu[1], Aiyi Liu[2], Bo Zhang[3] and Zhiwei Zhang[4]*

[1] *Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, PR China*
[2] *Biostatistics and Bioinformatics Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Rockville, MD, USA*
[3] *Biostatistics Core, School of Biological and Population Health Sciences, Oregon State University, Corvallis, OR, USA*
[4] *Division of Biostatistics, Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, MD, USA*

Edited by:
Dihua Xu, National Institute of Health, USA

Reviewed by:
Yifan Wang, NIH/NICHD, USA
Aijun Ye, Eunice Kennedy Shriver National Institute of Child Health and Human Development, USA
Yunlong Xie, National Institutes of Health, USA

Correspondence:
Zhiwei Zhang, Division of Biostatistics, Center for Devices and Radiological Health, Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993, USA
e-mail: zhiwei.zhang@fda.hhs.gov

This article concerns construction of confidence intervals for the prevalence of a rare disease using Dorfman's pooled testing procedure when the disease status is classified with an imperfect biomarker. Such an interval can be derived by converting a confidence interval for the probability that a group is tested positive. Wald confidence intervals based on a normal approximation are shown to be inefficient in terms of coverage probability, even for relatively large number of pools. A few alternatives are proposed and their performance is investigated in terms of coverage probability and length of intervals.

Keywords: confidence intervals, coverage probability, exact inference, pooling, prevalence, rare event, sensitivity, specificity

1. INTRODUCTION

Screening for subjects infected with a disease can be costly and time-consuming, especially when the disease prevalence is low. In an effort to overcome these barriers, Dorfman (1) proposed the pooling of blood samples to test for syphilis antigen. According to his procedure, blood samples from subjects under screening are pooled prior to testing. If a pool of blood samples is tested negative, then all subjects in the pool are declared free of infection. Otherwise, a positive test result on a pool indicates that at least one subject is infected and retesting of all individuals in that pool is then conducted to find the infected subjects.

Since its appearance, Dorfman's (1) pooled testing (also known as group testing) approach has drawn considerable attention. The approach has been applied to other areas of screening (than syphilis), such as human immunodeficiency virus (HIV) testing [e.g., Westreich et al. (2)]. A number of variations have been developed, and the scope has been expanded to include estimation of the prevalence of a disease (without necessarily identifying the diseased individuals) (3–24). However, there is relatively little discussion on the possibility of misclassification (i.e., that the disease status of an individual or a pool of individuals can be assessed incorrectly because the biomarker may be imperfect).

The existing literature on estimating the prevalence of a disease using the pooled testing approach is focused on point estimation (3, 5–7, 13–15, 17, 20, 22, 23). Construction of confidence intervals for the prevalence of a disease has been discussed by Hepworth (10–12) and Tebbs and Bilder (21). These authors assumed that the disease status of a subject can be accurately determined, which may be unrealistic in practice. For example, Weiss et al. (25) reported 97.7% sensitivity and 92.6% specificity for detecting HIV infection with an enzyme-linked immunosorbent assay, and Deitz et al. (26) reported 94% sensitivity for determining the status of *N*-acetyltransferase 2 with a commonly used 3-single nucleotide polymorphism genotyping assay.

This article focuses on construction of efficient confidence intervals for the prevalence of a rare disease using Dorfman's pooled testing procedure when the disease status is determined by an imperfect biomarker subject to misclassification. We investigate the unified approach of Tu et al. (27), which produces a confidence interval for the disease prevalence by converting a confidence interval for the probability of a pool being tested positive. We then demonstrate that Wald confidence intervals based on a normal approximation are inefficient in that they have a repetitive up-and-down behavior in the coverage probability, similar to that of the classical normal approximation binomial confidence interval discovered by Brown et al. (28). In the present context, this up-and-down behavior persists even when the number of pools is relatively large. We derive alternative confidence intervals by extending the methods of Wilson (29), Clopper and Pearson (30), Agresti and Coull (31), and Blaker (32). Simulation studies are conducted to compare the performance of the proposed methods in terms of coverage probability and mean length. The methods are applied to a real example concerning the seroprevalence of HIV among newborns.

2. INTERVAL ESTIMATION UNDER POOLED TESTING

Suppose one wants to estimate the prevalence of a disease in a population, $p = P(D = 1)$, where D denotes the disease status of a

subject in the population, with $D = 1$ if the subject is infected with the disease. We assume that the disease status is determined using an imperfect biomarker M, taking values 0 and 1, and a subject is classified as infected if $M = 1$. The accuracy of the biomarker is measured by its specificity $\pi_0 = P(M = 0|D = 0)$ and sensitivity $\pi_1 = P(M = 1|D = 1)$. For the biomarker to be of practical use we assume that $1/2 < \pi_0, \pi_1 \leq 1$; otherwise a random assignment of the disease status would perform better than the biomarker.

Supposed a random sample of size nk, where n and k are positive integers, is available from the target population. The conventional approach to estimating p is based on individually observed values of M, say $M_1, \ldots, M_{nk}$. Dorfman's procedure for estimating p is carried out by randomly assigning the nk individuals into n pools with k individuals in each pool and testing for positivity of the biomarker for each pool. Inference on p is then based on the number of pools that are tested positive. For this purpose, further testing for biomarker positivity for each individual in the pool is not necessary. Thus, instead of observing $M_1, \ldots, M_{nk}$, the biomarker values of the k individuals in a pool, we observe $\widetilde{M} = \max\{M_1, \cdots, M_k\}$. If $\widetilde{M} = 0$, then $M_i = 0$ for each $i = 1, \ldots, k$. If $\widetilde{M} = 1$, then $M_i = 1$ for at least one i in the pool. Throughout we assume that pooling will not affect the misclassification of the disease status by the biomarker. Let $\delta = P(\widetilde{M} = 1)$ be the probability that a pool is tested positive. Then it follows from Tu et al. (23) that

$$\delta = \pi_1 - r(1-p)^k, \quad (r = \pi_0 + \pi_1 - 1). \tag{1}$$

Consequently,

$$p = 1 - \left(\frac{\pi_1 - \delta}{r}\right)^{1/k}.$$

For fixed k, π_0, and π_1, the value of δ as a function of p increases as p increases. Because $0 \leq p \leq 1$, we have

$$1 - \pi_0 \leq \delta \leq \pi_1. \tag{2}$$

Using the relationship given by equation (1) along with the constraint equation (2), a unified (and straightforward) approach (27) to constructing a confidence interval for p is as follows. Suppose $[\delta_L, \delta_U]$ is a confidence interval for δ with level $1 - \alpha$. Define

$$p_L = 1 - \left(\frac{\pi_1 - \delta_L}{r}\right)^{1/k}, \quad p_U = 1 - \left(\frac{\pi_1 - \delta_U}{r}\right)^{1/k}. \tag{3}$$

Then $[p_L, p_U]$ is a confidence interval for p with level $1 - \alpha$.

3. CONSTRUCTING CONFIDENCE INTERVALS FOR δ

In this section we propose a few methods to construct a confidence interval for δ. Once derived, the interval can then be converted into a confidence interval for p, as indicated in the previous section. Let $\widetilde{M}_i$, $i = 1, \ldots, n$, be the test result for the ith pool. The $\widetilde{M}_i$ are independent and identically distributed Bernoulli variables with $P(\widetilde{M}_i = 1) = \delta \in [1 - \pi_0, \pi_1]$. Thus a confidence interval for δ can be constructed using methods developed for a binomial probability. However the constraint equation (2) must be taken into account. In what follows we extend several popular methods for binomial confidence intervals to construct confidence intervals for δ, taking the constraint equation (2) into consideration.

3.1. THE WALD CONFIDENCE INTERVAL

The Wald confidence interval is based on the fact that the estimator of δ, $\hat{\delta} = S/n$, is asymptotically normally distributed with mean δ and variance $\delta(1-\delta)/n$, where $S = \sum_{i=1}^{n} \widetilde{M}_i$ is the number of pools that are tested positive. Without the constraint equation (2), the Wald confidence interval is given by

$$\widetilde{\mathcal{I}} = \hat{\delta} \pm Z_{1-\alpha/2} \left\{\hat{\delta}(1-\hat{\delta})\right\}^{1/2} / n^{1/2},$$

where $Z_{1-\alpha/2}$ is the $100(1-\alpha/2)$th percentile of the standard normal distribution. With the constraint, we define the Wald confidence interval for δ as

$$\mathcal{I} = \widetilde{\mathcal{I}} \bigcap [1 - \pi_0, \pi_1], \tag{4}$$

where

$$[a_1, b_1] \bigcap [a_2, b_2] \equiv [\max(a_1, a_2), \min(b_1, b_2)].$$

3.2. THE WILSON CONFIDENCE INTERVAL

Without any constraints on the binomial probability, the Wilson confidence interval (29) is

$$\widetilde{\mathcal{I}}_W = \frac{S + Z^2_{1-\alpha/2}/2}{n + Z^2_{1-\alpha/2}} \pm \frac{Z_{1-\alpha/2} n^{1/2}}{n + Z^2_{1-\alpha/2}} \left\{\hat{\delta}(1-\hat{\delta}) + Z^2_{1-\alpha/2}/(4n)\right\}^{1/2}.$$

Accounting for the constraint equation (2), the modified Wilson confidence interval for δ is given by

$$\mathcal{I}_W = \widetilde{\mathcal{I}}_W \bigcap [1 - \pi_0, \pi_1]. \tag{5}$$

3.3. THE CLOPPER–PEARSON CONFIDENCE INTERVAL

The Clopper–Pearson confidence interval is often referred to as the exact confidence interval due to its derivation based on the binomial distribution rather than the normal approximation. Note that S follows a binomial distribution with size n and probability δ. Let s be the observed value of S. If there are no constraints on δ, then the lower bound δ_L and the upper bound δ_U of the Clopper–Pearson interval can be derived by solving the equations:

$$P_{\delta_L}(S \geq s) = 1 - B(s-1; n, \delta_L) = \alpha/2, \text{ and}$$
$$P_{\delta_U}(S \leq s) = B(s; n, \delta_U) = \alpha/2,$$

respectively, where $b(s;\ n,\ \delta) = P_\delta(S = s)$ is the binomial density function with size n and probability δ, and $B(s; n, \delta) = \sum_{i=1}^{s} b(s; n, \delta) = P_\delta(S \leq s)$ is the corresponding binomial distribution function. Tu et al. (27) suggested using this interval without any modification for δ. The modified Clopper–Pearson confidence interval that accounts for the constraint equation (2) is given by

$$\mathcal{I}_{CP} = [\delta_L, \delta_U] \bigcap [1 - \pi_0, \pi_1]. \tag{6}$$

3.4. THE AGRESTI–COULL CONFIDENCE INTERVAL

The Agresti–Coull confidence interval is a modification of the Wald confidence interval with $\hat{\delta}$ replaced by

$$\tilde{\delta} = \frac{S + Z_{1-\alpha/2}^2/2}{n + Z_{1-\alpha/2}^2}.$$

Thus, when δ is not constrained, the Agresti–Coull confidence interval is given by

$$\widetilde{\mathcal{I}}_{AC} = \tilde{\delta} \pm Z_{1-\alpha/2} \left\{\tilde{\delta}(1-\tilde{\delta})\right\}^{1/2} / n^{1/2}.$$

With the constraint equation (2), the Agresti–Coull confidence interval becomes

$$\mathcal{I}_{AC} = \widetilde{\mathcal{I}}_{AC} \bigcap [1-\pi_0, \pi_1]. \quad (7)$$

3.5. THE BLAKER CONFIDENCE INTERVAL

The confidence intervals of Wilson, Clopper–Pearson, and Agresti–Coull are highly recommended by Brown et al. (28). Blaker (32) proposed a method to improve the standard "exact" confidence intervals for discrete distributions, and called the resulting confidence intervals *acceptability intervals*. For the binomial case, the author showed that the acceptability interval is shorter than the Wald, Wilson, and Agresti–Coull intervals. Define

$$\gamma(\delta, s) = \min\{1 - B(s-1; n, \delta), B(s; n, \delta)\}$$

and

$$\alpha(\delta; s) = \sum_{\{i:\, \gamma(\delta,i) \le \gamma(\delta,s)\}} b(s; n, \delta)\,.$$

Then for the binomial probability δ with no constraints, by reformulating the notation of Blaker (32), the Blaker interval is given by $\widetilde{\mathcal{I}}_B = \{\delta : \alpha(\delta; s) > \alpha\}$. Blaker (32) showed that $\widetilde{\mathcal{I}}_B$ is indeed an interval and has coverage probability $1-\alpha$. When δ is constrained by equation (2), the Blaker confidence interval can be defined as

$$\mathcal{I}_B = \widetilde{\mathcal{I}}_B \bigcap [1-\pi_0, \pi_1]. \quad (8)$$

4. SIMULATIONS

There has been a large amount of research on the performance of various binomial confidence intervals for the disease prevalence p under the usual setting where independent and identically distributed Bernoulli observations of the disease status are available. However, not much research has been conducted under Dorfman's pooled testing setting, especially in the presence of misclassification. In this section we conduct simulations to compare the coverage probability and mean length of the confidence intervals for p, and to investigate the effect of the pool size k and the misclassification rates (i.e., $1-\pi_0$ and $1-\pi_1$) on the precision (coverage and length) of the intervals. It is worth noting that when a confidence interval $[\delta_L, \delta_U]$ for δ is converted into a confidence interval $[p_L, p_U,]$ for p via equation (3), the coverage probability remains unchanged because of the monotonicity of p as a function of δ.

4.1. THE OSCILLATION BEHAVIOR OF WALD CONFIDENCE INTERVALS

Brown et al. (28) investigated the performance of a number of confidence intervals for a binomial probability in the usual setting, where the individual disease status is observed without error, corresponding to $k=1$ and $\pi_0 = \pi_1 = 1$ in our setting. The authors showed a remarkable oscillation up-and-down behavior of the widely used Wald confidence intervals based on a normal approximation; the coverage probability of the interval increases from far below the nominal level of $1-\alpha$ to the nominal level and then decreases, and the pattern repeats until the sample size becomes rather large. We demonstrate here that Wald confidence intervals have the same oscillation up-and-down phenomenon under pooled testing with misclassification.

Fixing specificity $\pi_0 = 0.85$ and sensitivity $\pi_1 = 0.90$, we computed via simulation the coverage probability of the Wald confidence interval for p, in a variety of scenarios with $k=2$, 5, $10 \le n \le 150$, and $p = 0.01$, 0.10. For each configuration of (k, n, p), 10,000 simulations were conducted. For each simulation, we generated a random observation from the binomial distribution with probability $\delta = \pi_1 - r(1-p)^k$ and size n, and constructed the 95% Wald confidence interval for δ according to equation (4). This confidence interval for δ was then converted into a confidence interval for p using equation (3). The average coverage probability of the confidence interval is the proportion of the 10,000 intervals that contain the true value of p.

Figure 1 presents the simulation results. It is clear that, for each configuration, the coverage probability as a function of n starts with very low coverage, usually below 85%, and then gradually increases as n gets larger to a value near the nominal level of 95%. Then it quickly decreases to a low coverage probability. The trend then repeats until n is large enough to stabilize the coverage probability. Therefore, unless n is sufficiently large, the Wald confidence interval does not provide the desired coverage and should not be recommended. This unfortunate observation is consistent with that of Brown et al. (28) for the classical binomial confidence intervals.

4.2. COMPARISON OF CONFIDENCE INTERVALS

Using simulations again we compared the precision of the four alternative confidence intervals, the Wilson, Clopper–Pearson, Agresti–Coull, and Blaker intervals, along with the Wald interval, in terms of mean length and coverage probability. To set up the simulation we considered various representative configurations of (p, n, k, π_0, π_1) with $p = 0.001$, 0.1, 0.3, $n = 10$, 20, 50, $k = 2$, 5, 10, and $(\pi_0, \pi_1) = (0.85, 0.95)$. A total of 10,000 simulations were conducted, and the coverage probability of each interval was estimated the same way as for the Wald interval. The mean length of each interval was estimated by averaging over the 10,000 simulated intervals.

Table 1 shows the estimated overage probability and average length of each confidence interval in various scenarios with $p = 0.001$ (the results for $p = 0.1$ and $p = 0.3$ are similar and therefore not shown). In almost all cases considered, the four alternative confidence intervals (i.e., the Wilson, Clopper–Pearson, Agresti–Coull, and Blaker intervals) provide satisfactory coverage probability around the 95% nominal level of confidence. The Wald intervals are quite unstable, with poor precision when n is small.

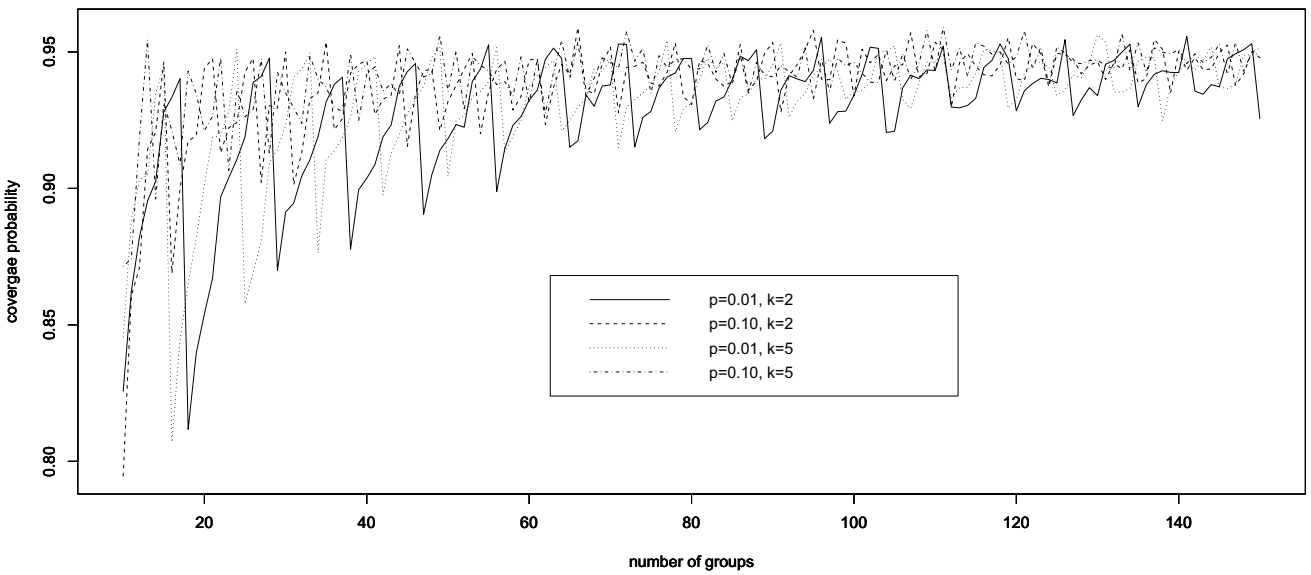

FIGURE 1 | The oscillation up-and-down behavior of the Wald confidence interval under pooled testing with misclassification.

The Clopper–Pearson and Agresti–Coull intervals tend to be more conservative, producing higher coverage probability, followed by the Blaker interval and then the Wilson interval. However, conservatism usually comes with the price of longer intervals, as shown in **Table 1**.

The effect of misclassification on the Wald interval does not seem to be clear due to its oscillation up-and-down behavior. For the other intervals, it appears that the coverage probability increases as the sensitivity π_1 increases or as the specificity decreases. For fixed misclassification rates (π_0, π_1), including more samples in a pool seems to improve the coverage probability, up to certain pool size. This latter observation seems to agree with that of Tu et al. (23) and Liu et al. (33), who found that in presence of misclassification the efficiency of estimation increases with the pool size up to a certain point.

5. EXAMPLE

We now illustrate the methods by applying them to a real example concerning the seroprevalence of HIV among newborns in the State of New York (34). The data were obtained by testing blood specimens from all infants born in this state during a 28-month period (from November 30, 1987 through March 31, 1990). The test was targeted at serum antibodies produced by the immune system in response to HIV infection. A positive test result indicates HIV infection in the mother but not necessarily in the child. To illustrate the methods, we focus on the Manhattan area, where 50,364 newborns were tested with 799 positive results.

Because the study did not involve pooled testing, we create pools in a *post hoc* manner by grouping subjects randomly into pools of a given size ($k = 5$ or 10). With $k = 10$, for instance, we obtain 5,036 pools of size 10, ignoring the four additional subjects. The test result for each pool is taken to be the maximum of all individual test results in the pool; that is, a pool is declared positive if and only if it contains one or more infants with positive test results. To account for possible misclassification, estimation of HIV seroprevalence requires knowledge of the sensitivity and specificity of the test. Because the true values of these performance measures are not known precisely, we perform a sensitivity analysis that covers a range of plausible values for the sensitivity and the specificity of the test. The reasoning of Tu et al. (27) and the numerical results in their **Table 1** suggest that the specificity of the HIV test in this study is at least 99%. Accordingly, our sensitivity analysis includes the values 99, 99.5, and 99.9% for the specificity. The appears to be less information about the sensitivity of the HIV test in this study, and we therefore consider a wider range (95, 97.5, 99, and 99.9%) for the sensitivity. For each pair of sensitivity and specificity values and each value of k, we apply the five methods described earlier to the pooled dataset to obtain five 95% confidence intervals for the individual-level HIV seroprevalence rate, in addition to a point estimate (common to all five methods).

Table 2 presents the results of our sensitivity analysis (with different combinations of sensitivity and specificity values) for each value of k (5 or 10). It appears that the results are more sensitive to the specificity of the HIV test than to the sensitivity of the test. The point estimate and the confidence limits (for all five methods) tend to decrease with the sensitivity of the test and increase with the specificity of the test, as predicted by theory. Intuitively, increased sensitivity means fewer false negatives, and increased specificity means fewer false positives, and these are reflected in the estimates in **Table 2**. Between the different pool sizes (5 and 10), which lead to different datasets, there are some numerical differences, especially at lower values of the sensitivity and the specificity. However, when the sensitivity and the specificity are both high (say, 99.9%), there is remarkable agreement between the estimates based on $k = 5$ and those based on $k = 10$. In any case, the five confidence intervals are generally similar to each other, perhaps as a result of the large sample size.

Table 1 | Empirical comparison of confidence intervals in terms of coverage probability and average length (see Section 4.2 for details).

k	π_0	π_1	Empirical coverage					Average length				
			Wald	Wilson	C-P	A-C	Blaker	Wald	Wilson	C-P	A-C	Blaker
$n = 10$												
2	0.85	0.85	0.796	0.951	0.989	0.989	0.951	0.162	0.252	0.297	0.285	0.270
2	0.85	0.95	0.791	0.951	0.989	0.989	0.951	0.136	0.214	0.250	0.240	0.229
2	0.95	0.85	0.407	0.911	0.987	0.987	0.987	0.061	0.204	0.233	0.221	0.230
2	0.95	0.95	0.412	0.908	0.987	0.987	0.987	0.054	0.179	0.204	0.194	0.201
5	0.85	0.85	0.804	0.946	0.990	0.990	0.946	0.077	0.115	0.139	0.132	0.123
5	0.85	0.95	0.802	0.948	0.990	0.990	0.948	0.063	0.095	0.113	0.108	0.102
5	0.95	0.85	0.424	0.901	0.986	0.986	0.986	0.029	0.089	0.104	0.098	0.101
5	0.95	0.95	0.425	0.903	0.986	0.986	0.986	0.025	0.077	0.089	0.084	0.087
10	0.85	0.85	0.808	0.939	0.989	0.989	0.939	0.044	0.061	0.076	0.072	0.065
10	0.85	0.95	0.809	0.940	0.989	0.989	0.940	0.034	0.050	0.060	0.057	0.053
10	0.95	0.85	0.449	0.980	0.980	0.980	0.980	0.017	0.047	0.055	0.052	0.053
10	0.95	0.95	0.452	0.982	0.982	0.982	0.982	0.014	0.040	0.047	0.045	0.046
$n = 20$												
2	0.85	0.85	0.825	0.978	0.978	0.978	0.978	0.118	0.164	0.181	0.172	0.172
2	0.85	0.95	0.820	0.978	0.978	0.978	0.978	0.101	0.141	0.154	0.147	0.148
2	0.95	0.85	0.651	0.918	0.980	0.980	0.980	0.051	0.123	0.132	0.127	0.137
2	0.95	0.95	0.652	0.922	0.983	0.983	0.983	0.046	0.109	0.117	0.112	0.121
5	0.85	0.85	0.826	0.974	0.974	0.974	0.974	0.052	0.071	0.079	0.075	0.075
5	0.85	0.95	0.836	0.974	0.993	0.974	0.974	0.045	0.061	0.067	0.064	0.064
5	0.95	0.85	0.655	0.981	0.981	0.981	0.981	0.022	0.052	0.056	0.054	0.058
5	0.95	0.95	0.671	0.979	0.979	0.979	0.979	0.021	0.047	0.050	0.048	0.052
10	0.85	0.85	0.835	0.974	0.993	0.974	0.974	0.027	0.037	0.041	0.039	0.039
10	0.85	0.95	0.846	0.972	0.993	0.972	0.972	0.024	0.032	0.035	0.033	0.033
10	0.95	0.85	0.698	0.972	0.994	0.972	0.972	0.013	0.028	0.030	0.029	0.031
10	0.95	0.95	0.700	0.973	0.995	0.973	0.973	0.011	0.024	0.026	0.025	0.027
$n = 50$												
2	0.85	0.85	0.943	0.957	0.974	0.957	0.957	0.074	0.093	0.098	0.095	0.096
2	0.85	0.95	0.942	0.953	0.972	0.953	0.953	0.065	0.081	0.085	0.082	0.083
2	0.95	0.85	0.925	0.955	0.986	0.955	0.955	0.038	0.064	0.066	0.063	0.069
2	0.95	0.95	0.927	0.956	0.986	0.956	0.956	0.034	0.057	0.059	0.056	0.061
5	0.85	0.85	0.937	0.945	0.968	0.945	0.945	0.032	0.040	0.042	0.040	0.041
5	0.85	0.95	0.946	0.949	0.971	0.949	0.949	0.028	0.035	0.036	0.035	0.035
5	0.95	0.85	0.933	0.949	0.983	0.983	0.983	0.017	0.027	0.028	0.027	0.029
5	0.95	0.95	0.933	0.948	0.983	0.983	0.983	0.015	0.024	0.025	0.024	0.026
10	0.85	0.85	0.901	0.947	0.967	0.967	0.947	0.017	0.021	0.022	0.021	0.021
10	0.85	0.95	0.901	0.947	0.971	0.971	0.971	0.015	0.018	0.019	0.018	0.018
10	0.95	0.85	0.941	0.974	0.974	0.974	0.974	0.009	0.014	0.015	0.014	0.015
10	0.95	0.95	0.803	0.974	0.993	0.974	0.974	0.008	0.013	0.013	0.013	0.014

6. DISCUSSION

In this article we proposed a few approaches to constructing a confidence interval for the disease prevalence under pooled testing with misclassification. These approaches share a common feature in that they are all obtained by converting a valid confidence interval for the probability of a pool being tested positive. Our investigation of the coverage probability and mean length of the confidence intervals indicates that caution needs to be taken in using the Wald interval when the sample size is not large enough. From our overall evaluation it appears that the Clopper–Pearson and Agresti–Coull intervals, though somewhat conservative, tend to be more valid than the Wilson and Blaker intervals, especially when the disease probability and the sample size are relatively small.

Misclassification of the disease status clearly impacts the precision of the confidence intervals, as demonstrated by the simulation results in **Figure 1** and **Table 1**. In this article, the misclassification is assumed to be independent of the pool size, which seems to be a reasonable assumption in some situations. However, as noted by Cahoon-Young (35), this assumption may be violated when the pool size gets larger. It remains to be seen how the performance of a confidence interval might be affected by pool-size-dependent misclassification.

Table 2 | Analysis of HIV seroprevalence data (see Section 5 for details).

Specificity	Sensitivity	Pt. est.	95% Confidence interval for p (%)									
π_0 (%)	π_1 (%)	$\hat{p}$ (%)	Wald		Wilson		Clopper–Pearson		Agresti–Coull		Blaker	
$k = 5$												
99	95	1.48	1.36	1.60	1.36	1.60	1.36	1.60	1.36	1.60	1.36	1.60
99	97.5	1.44	1.32	1.55	1.33	1.56	1.33	1.56	1.32	1.55	1.33	1.56
99	99	1.42	1.30	1.53	1.31	1.53	1.30	1.53	1.30	1.53	1.31	1.53
99	99.9	1.40	1.29	1.51	1.29	1.52	1.29	1.52	1.29	1.51	1.29	1.52
99.5	95	1.58	1.46	1.70	1.47	1.70	1.47	1.70	1.46	1.70	1.47	1.70
99.5	97.5	1.54	1.43	1.66	1.43	1.66	1.43	1.66	1.43	1.66	1.43	1.66
99.5	99	1.52	1.40	1.63	1.41	1.63	1.41	1.63	1.40	1.63	1.41	1.63
99.5	99.9	1.50	1.39	1.61	1.39	1.62	1.39	1.62	1.39	1.61	1.39	1.62
99.9	95	1.67	1.55	1.78	1.55	1.79	1.55	1.79	1.55	1.78	1.55	1.79
99.9	97.5	1.62	1.51	1.74	1.51	1.74	1.51	1.74	1.51	1.74	1.51	1.74
99.9	99	1.60	1.48	1.71	1.49	1.71	1.49	1.71	1.48	1.71	1.49	1.71
99.9	99.9	1.58	1.47	1.69	1.47	1.70	1.47	1.70	1.47	1.69	1.47	1.70
$k = 10$												
99	95	1.59	1.47	1.71	1.47	1.72	1.47	1.72	1.47	1.71	1.47	1.72
99	97.5	1.55	1.43	1.67	1.43	1.67	1.43	1.67	1.43	1.67	1.43	1.67
99	99	1.52	1.41	1.64	1.41	1.64	1.41	1.64	1.41	1.64	1.41	1.64
99	99.9	1.51	1.39	1.62	1.40	1.62	1.39	1.63	1.39	1.62	1.40	1.62
99.5	95	1.64	1.52	1.77	1.53	1.77	1.52	1.77	1.52	1.77	1.53	1.77
99.5	97.5	1.60	1.48	1.72	1.48	1.72	1.48	1.72	1.48	1.72	1.48	1.72
99.5	99	1.57	1.46	1.69	1.46	1.69	1.46	1.69	1.46	1.69	1.46	1.69
99.5	99.9	1.56	1.44	1.67	1.45	1.67	1.44	1.67	1.44	1.67	1.44	1.67
99.9	95	1.69	1.56	1.81	1.57	1.81	1.57	1.81	1.56	1.81	1.57	1.81
99.9	97.5	1.64	1.52	1.76	1.53	1.76	1.52	1.76	1.52	1.76	1.52	1.76
99.9	99	1.61	1.50	1.73	1.50	1.73	1.50	1.73	1.50	1.73	1.50	1.73
99.9	99.9	1.60	1.48	1.71	1.49	1.71	1.48	1.71	1.48	1.71	1.48	1.71

ACKNOWLEDGMENTS

The authors thank Dr. Paul Albert for helpful discussion. The views expressed in this article do not represent the official position of the U.S. Food and Drug Administration. This research was supported in part by the Intramural Research Program of the National Institutes of Health, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, and by the Long-Range Research Initiative of the American Chemistry Council. Dr. Chunling Liu's research was partially supported by the Hong Kong research grant BQ25U.

REFERENCES

1. Dorfman R. The detection of defective members of large populations. *Ann Math Stat* (1943) **14**:436–40. doi:10.1214/aoms/1177731363
2. Westreich DJ, Hudgens MG, Fiscus SA, Pilcher CD. Optimizing screening for acute human immunodeficiency virus infection with pooled nucleic acid amplification tests. *J Clin Microbiol* (2008) **46**:1785–92. doi:10.1128/JCM.00787-07
3. Brookmeyer R. Analysis of multistage pooling studies of biological specimens for estimating disease incidence and prevalence. *Biometrics* (1999) **55**:608–12. doi:10.1111/j.0006-341X.1999.00608.x
4. Burns KC, Mauro CA. Group testing with test error as a function of concentration. *Commun Stat Theory Methods* (1987) **16**:2821–37. doi:10.1080/03610928708829544
5. Chen CL, Swallow WH. Using group testing to estimate a proportion, and to test the binomial model. *Biometrics* (1990) **46**:1035–46. doi:10.2307/2532446
6. Farrington C. Estimation prevalence by group testing using generalized linear models. *Stat Med* (1992) **11**:1591–7. doi:10.1002/sim.4780111206
7. Gastwirth JL, Hammick PA. Estimation of prevalence of a rare disease, preserving anonymity of subjects by group testing: application to estimating the prevalence of AIDS antibodies in blood donors. *J Stat Plan Inference* (1989) **22**:15–27. doi:10.1016/0378-3758(89)90061-X
8. Gastwirth J, Johnson W. Screening with cost-effective quality control: potential applications to HIV and drug testing. *J Am Stat Assoc* (1994) **89**:972–81. doi:10.1080/01621459.1994.10476831
9. Graff LE, Roeloffs R. Group testing in the presence of test errors: an extension of the Dorfman procedure. *Technometrics* (1972) **14**:113–22. doi:10.1080/00401706.1972.10488888
10. Hepworth G. Exact confidence limits for proportions estimated by group testing. *Biometrics* (1996) **52**:1134–46. doi:10.2307/2533075
11. Hepworth G. Mid-p confidence intervals based on the likelihood ratio for proportions estimated by group testing. *Aust N Z J Stat* (2004) **46**:391–405. doi:10.1111/j.1467-842X.2004.00338.x
12. Hepworth G. Confidence intervals for proportions estimated by group testing with groups of unequal size. *J Agric Biol Environ Stat* (2005) **10**:478–97. doi:10.1198/108571105X81698
13. Hughes-Oliver JM, Swallow WH. A two-stage adaptive group-testing procedure for estimating small proportions. *J Am Stat Assoc* (1994) **89**:982–93. doi:10.1080/01621459.1994.10476832
14. Hughes-Oliver JM, Rosenberger WF. Efficient estimation of the prevalence of multiple rare traits. *Biometrika* (2000) **87**:315–27. doi:10.1093/biomet/87.2.315

15. Hung M, Swallow WH. Robustness of group testing in the estimation of proportions. *Biometrics* (1999) **55**:231–7. doi:10.1111/j.0006-341X.1999.00231.x
16. Hwang FK. Group testing with a dilution effect. *Biometrika* (1976) **63**:671–3. doi:10.1093/biomet/63.3.671
17. Le CT. A new estimator for infection rates using pools of variable size. *Am J Epidemiol* (1981) **14**:132–6.
18. Litvak E, Tu XM, Pagano M. Screening for the presence of a disease by pooling sera samples. *J Am Stat Assoc* (1994) **89**:424–34. doi:10.1080/01621459.1994.10476764
19. Sobel M, Groll PA. Group testing to eliminate efficiently all defectives in a binomial sample. *Bell Syst Tech J* (1959) **38**:1179–252. doi:10.1002/j.1538-7305.1959.tb03914.x
20. Sobel M, Elashoff R. Group testing with a new goal: estimation. *Biometrika* (1975) **62**:181–93. doi:10.1093/biomet/62.1.181
21. Tebbs JM, Bilder CR. Confidence limit procedures for the probability of disease transmission in multiple vector-transfer designs. *J Agric Biol Environ Stat* (2004) **9**:75–90. doi:10.1198/1085711043127
22. Tebbs JM, Swallow WH. Estimating ordered binomial proportions with the use of group testing. *Biometrika* (2003) **90**:471–7. doi:10.1093/biomet/90.2.471
23. Tu XM, Litvak E, Pagano M. On the informativeness and accuracy of pooled testing in estimating prevalence of a rare disease: application to HIV screening. *Biometrika* (1995) **82**:287–9. doi:10.1093/biomet/82.2.287
24. Xie M, Tatsuoka K, Sacks J, Young SS. Group testing with blockers and synergism. *J Am Stat Assoc* (2001) **96**:92–102. doi:10.1093/ndt/gfn454
25. Weiss SH, Goedert JJ, Sarngadharan MG, Bodner AJ. The AIDS Seroepidemiology Collaborative working Group, Gallo RC, Blattner WA. Screening tests for HTLVI-III (AIDS Agent) antibodies: specificity, sensitivity, and applications. *J Am Med Assoc* (1985) **253**:221–5. doi:10.1001/jama.253.2.221
26. Deitz AC, Rothman N, Rebbeck TR, Hayes RB, Chow WH, Zheng W, et al. Impact of misclassification in genotype-exposure interaction studies: example of N-Acetyltransferase 2 (NAT2), smoking, and bladder cancer. *Cancer Epidemiol Biomarkers Prev* (2004) **13**:1543–6.
27. Tu XM, Litvak E, Pagano M. Screening tests: can we get more by doing less? *Stat Med* (1994) **13**:1905–19. doi:10.1002/sim.4780131904
28. Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. *Stat Sci* (2001) **16**:101–28. doi:10.1214/ss/1009213286
29. Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc* (1927) **22**:209–12. doi:10.1080/01621459.1927.10502953
30. Clopper C, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* (1934) **26**:13. doi:10.1093/biomet/26.4.404
31. Agresti A, Coull BA. Approximate is better than 'exact' for interval estimation of binomial proportions. *Am Stat* (1998) **52**:26. doi:10.1080/00031305.1998.10480550
32. Blaker H. Confidence curves and improved exact confidence intervals for discrete distributions. *Can J Stat* (2000) **28**:783–98. doi:10.1093/biostatistics/kxp050
33. Liu A, Liu C, Zhang Z, Albert P. Optimality of group testing in the presence of misclassification. *Bimetrika* (2012) **99**:245–51. doi:10.1093/biomet/asr064
34. Novick LF, Glebatix DM, Stricof RL, MacCubbin PA, Lessner L, Berns DSII. Newborn seroprevalence study: methods and results. *Am J Public Health* (1991) **81**:15–21. doi:10.2105/AJPH.81.Suppl.15
35. Cahoon-Young B, Chandler A, Livermore T, Gaudino J, Benjamin R. Sensitivity and specificity of pooled versus individual sera in a human immunodeficiency virus (mY) antibody prevalence study. *J Clin Microbiol* (1989) **27**:1893–5.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 12 August 2013; paper pending published: 30 August 2013; accepted: 17 September 2013; published online: 07 October 2013.
*Citation: Liu C, Liu A, Zhang B and Zhang Z (2013) Improved confidence intervals of a small probability from pooled testing with misclassification. Front. Public Health **1**:39. doi: 10.3389/fpubh.2013.00039*
This article was submitted to Epidemiology, a section of the journal Frontiers in Public Health.