IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 13, Issue 4 Ver. I (Jul. – Aug. 2017), PP 79-85
www.iosrjournals.org

Estimation of Prevalence Using Ordered Pool-Testing Scheme In The Presence Of Testing Errors

*Collins Musavi Amuhaya[1]
[1](MasindeMuliro University of Science and Technology)
Corresponding Author: Collins Musavi Amuhaya

Abstract: *Pool-testing prevalence estimators are poor estimators of the actual prevalence of a traitthat is they have large bias and Mean Square Error (MSE) and this is a drawback tostatistical inference. This paper proposes a method of minimizing bias and MSE when weallow errors in inspection. The probability of declaring a group as defective is derived byassumption of law of total probability. Using Maximum Likelihood Estimators (MLE)method we constructed a prevalence estimator of the ordered probabilities. Theseprobabilities are used to order the groups via the method of Pool Adjacent ViolatorsAlgorithm (PAVA) in increasing order of prevalence. The weights used in PAVA areobtained by Appropriate Lagrange Multiplier (ALM). The prevalence of the ordered groupsis constructed. The combined unbiased estimator based on ordered groups is obtained usingBest Linear Unbiased Estimator (BLUE) method. The properties of the prevalenceestimator such as Bias and MSE are studied via Monte Carlo simulation. Simulations ofMSE and bias for un-ordered are carried out for comparison. It is established that theprevalence estimator based on ordered scheme has small bias and MSE compared to estimator based on un-ordered in pool-testing schemes.*

Keywords: *Best Linear Unbiased Estimator (BLUE),Lagrange Multiplier,Maximum Likelihood Estimators(MLE), Pool AdjacentViolators Algorithm (PAVA), Weights*

Date of Submission: 24-07-2017 Date of acceptance: 05-08-2017

I. Introduction

Consider a sample of a population of size N with a purpose of estimating a prevalence p ofa trait. When N is large, it is easier to pool the items into batches (pool) of say, equal sizesand subject the items to a single test. This procedure is called pool-testing strategy. The strategy involves pooling units into pools, testing them and classifying each pool as defectiveor non-defective. A tested unit will be declared defective if the test results indicate thepresence of a specific characteristic otherwise declared non-defective. This concept ofpool-testing originated from Dorfman, (1943) who during the Second World War proposedan economical method for detecting Syphilis in US soldiers. Since then, there is abundantliterature on the subject for instance (see Sobel and Elashoff, 1975; Nyongesa, 2011 andBrookmeyer, 1999).

When testing for a trait for example Syphilis, a mixture of urine of different persons iscreated and a test on the sample may result into an error due to the dilution of the sampleas seen in Hwang, (1976) who made discussions on types of errors such as dilution andconcentration. In this study, we discuss pool-testing procedure with errors in inspection.Pool-testing is two-fold, the first being classification of samples in a population while thesecond is estimating the prevalence.

In estimating prevalence using pool-testing procedure, restricted maximum likelihood is applied and this is the problem we are concerned with. This is achieved by ordering the pools in an increasing rate of prevalence as discussed inSection 2. We derive probabilities to be used in the paper in Section 3, while in Section 4construction the estimator based on un-ordered groups is done. Ordering the probability ofclassifying a stratum as defective is carried out in Section 5 leading to derivation of therestricted maximum likelihood estimate of the prevalence. In Section 5 investigation of theproperties of the constructed estimator is shown. Construction of the unbiased combinedestimator is derived in Section 6 followed by investigation of its properties in Section 7. In section 8 wegenerate the asymptotic variance. Simulations and findings are discussed in Section 9 thenconcluded in Section 10

II. Ordered scheme

Suppose we have population divided into groups, which are further then divided intosubgroups that share a specific characteristic. This subgroups we refer to them as strata. Letthere be a total of k strata and individuals are then pooled into groups within each stratum. We construct the i^{th} stratum containing $n_i$pools/groups of different sizes s_i; $i = 1, 2 \ldots k$.In pool-testing strategy each unit is assumed to represent an independent Bernoulli randomvariable where the probability that a selected subject possesses the characteristic of interestis p_i.

These probabilities of the subgroups p_i are supposed to be ordered in an increasing manner.If they are not ordered we require some prior knowledge to order them. A fundamentalmethod that will be useful in ordering probabilities in ordered scheme known as Pooladjacent violators algorithm (PAVA). This is the most commonly used algorithm forcomputing the isotonic regression for a simple order, as studied by Robertson et al. (1988).

The PAVA method will be applied in this study to ensure that the probabilities are in anincreasing order. The probability of randomly selecting a subject not possessing the characteristic of interestis$(1 - p_i)$. Thus the probability of randomly selecting a group of $s_i$subjects all of whom donot possess the characteristic of interest is$(1 - p_i)s_i$. Let the probability that an i^{th} stratumpossess the characteristic of interest be denoted by $\theta_i$and it will be given by

$$\theta_i = 1 - (1 - p_i)^{s_i} \quad (1)$$

Notice that Equation (1) does not involve errors and our major contribution is to introducean error component in the model. This is realistic since in the field experimental errorssuch as human and manufacturers errors are unavoidable. To introduce errors into themodel we shall require the theory of indicator functions as provided in the next Section.

III. Derivation of Probabilities

In this section we derive probabilities that will be useful in the subsequent development. Forthis purpose, the theory of indicator functions that we now define will be useful.

$$T_i = \begin{cases} 1, \text{if the test on the ith group is positive} \\ 0, \quad \text{otherwise} \end{cases}$$

$$D_i = \begin{cases} 1, \text{if the test on the ith group is positive} \\ 0, \quad \text{otherwise} \end{cases}$$

These indicator functions will help us introduce the error element in our model. In terms ofindicator function Equation (1) becomes,

$$\Pr[D_i = 1] = \theta_i = 1 - (1 - p_i)^{s_i}$$

The errors in our model are based on the manufacturers testing kit specifications that issensitivity and specificity. By sensitivity we mean the probability of classifying a defectivesample correctly, herein denoted byψ . While specificity, is the probability of classifying anon-defective sample correctly, herein denoted byϕ. These two parameters in terms of theindicator functions are

$$\psi = \Pr[T_i = 1 | D_i = 1] \quad (2)$$

and

$$\phi = \Pr[T_i = 0 | D_i = 0] \quad (3)$$

respectively. In this development, the error under consideration is due to manufacturer'sspecifications while other errors such as human errors will be assumed held constant. Discussions on other types of errors such as dilution and concentration can be found forinstance in Hwang, (1976). With this in mind, if we introduce an error component inEquation (1) we have

$$\theta_i = \Pr[T_i = 1]$$

and using the law of total probability, we have

$$\theta_i = \Pr[T_i = 1, D_i = 1] + Pr[T_i = 1, D_i = 0],$$

and upon simplifying gives

$$\theta_i = Pr[D_i = 1]\Pr[T_i = 1 | D_i = 1] + Pr[D_i = 0]Pr[T_i = 1 | D_i = 0], \quad (4)$$

With the definition of sensitivity and specificity defined in terms of our indicator functions, using Equations (1), (2), (3) and (4), we have

$$\theta_i = (1 - (1 - p_i)^{s_i})\psi + ((1 - \phi)(1 - p_i)^{s_i}) \quad (5)$$

IV. Estimation of prevalence in unordered scheme

Let $Y_{ij} = 1$ if the j^{th}group in the i^{th} stratum possesses the characteristic of interest, and$Y_i = 0$ otherwise, $i = 1, 2, ..., j, j = 1, 2, ..., n_i$. If $X_i = \sum_{j=1}^{n_i} Y_{ij}$ groups test positive onthe test then X_i~Binomial(n_i, θ_i) the joint distribution of the$n_i$groups can be given as

$$Pr[X_i = x_i | n_i, \theta_i, \psi, \phi] \propto \theta_i^{x_i}(1 - \theta_i)^{n_i - x_i} \quad (6)$$

assuming Equation (6) is continuous with respect to p_i while other parameters are heldconstant the likelihood function of (6) can simply be written as

$$L[p_i | n_i, \theta_i, \psi, \phi] \propto \theta_i^{x_i}(1 - \theta_i)^{n_i - x_i} \quad (7)$$

Maximum Likelihood Estimator of (7) has been derived see Brookmeyer, (1999) andNyongesa, (2011) as

$$\widehat{p_i} = 1 - \left[\frac{\psi - \frac{x_i}{n_i}}{\psi + \phi - 1}\right]^{\frac{1}{s_i}} \qquad (8)$$

If we take our sensitivity and specificity as 100% i.e.$\psi = \phi$=100% as assumed by Thompson, (1962); that is, the test is free of errors Equation (8) reduces to

$$\widehat{p_i} = 1 - \left[1 - \frac{x_i}{n_i}\right]^{\frac{1}{s_i}} (9)$$

hence a result similar to Thompson, (1962). Also the asymptotic variance of (8) is shownas

$$var(\widehat{p_i}) = s_i^{-2}(1 - p_i)^{2-2s_i}\ \theta_i(1 - \theta_i)(\psi + \phi - 1)^{-2}\ (10)$$

see Nyongesa, (2011). For further discussion of the properties of the estimator (8) see forexample Brookmeyer, (1999) and Hepworth, (2009) and Nyongesa, (2011). Equations (9)and (10) will be useful in our subsequent development of ordered testing scheme andconstruction of confidence intervals. The next section discusses the primary objective ofthis study; that is, ordered pool-testing scheme.

V. Ordering ofθ

Consider ani^{th} stratum whose sample prevalence's are isotonic with$p_1 \le p_2 \ldots\ldots \le p_i$.Hence, θ are $\theta_1 \le \theta_2 \ldots \le \theta_i$and are increasing functions of p. The estimators of p thatis$\widehat{p_1}$,$\widehat{p_2}$, , . . . ,$\widehat{p_i}$, might not be isotonic in practice, Tebbs,(2003) i.e., $\widehat{p_1}$,$\widehat{p_2}$, , . . . , $\widehat{p_i}$, might notbe ordered and in such circumstances $\widehat{\theta_1} \le \widehat{\theta_2} \ldots \le \widehat{\theta_i}$might not also be ordered. To estimate$\theta_i's$ the method of MLE is applied by maximizing

$$L[p_i\,|n_i\,, \theta_i, \psi, \phi] \propto \prod_{j=1}^{n_i} \theta_i^{x_i}(1 - \theta_i)^{n_i - x_i}$$

$\theta = (\theta_1, \theta_2 \ldots, \theta_i)$ subject to $\theta_1 \le \theta_2 \ldots \le \theta_i$. For $\hat{\theta}$ to be isotonic we need to put weightson the $\widehat{p_i}$ and this leads to estimating θ by MLE under the condition

$$\{\, \theta :\ 0 \le \theta_i \le 1, \theta_1 \le \theta_2 \ldots \le \theta_i \}\,,$$

and this is accomplished by computing

$$\widehat{\theta^*} = (\widehat{\theta_1^*}, \widehat{\theta_2^*}\ \ \ldots \widehat{\theta_i^*})\,,$$

the isotonic regression of $\theta = (\theta_1, \theta_2 \ldots, \theta_i)$ with weights $w_1\,, w_2\ \ldots\ldots w_i$. The method ofPAVA is used in ordering $\widehat{\theta^*}$. For example, if $\widehat{\theta^*} = (\widehat{\theta_1^*} \le \widehat{\theta_2^*} \ldots\ldots \le \widehat{\theta_i^*})$ is the isotonic regression of $\hat{\theta} = (\widehat{\theta_1} < \widehat{\theta_2} \ldots < \widehat{\theta_i}.)$ with weights $n_1\,, n_2\ \ldots\ldots n_i$ where n_i is the number ofgroups/pools in ani^{th} stratum.

VI. Restricted maximum likelihood estimate of prevalence

In the preceding discussion, we have discussed how $\theta_i's$can be ordered and even how theycan be estimated. We are now in a position to estimate the prevalence of the trait. Utilizing(5) and replacing $\theta_i's$with $\theta_i^{*}s$after ordering, we have

$$\theta_i^* = (1 - (1 - p_i)^{s_i})\psi + ((1 - \phi)(1 - p_i)^{s_i})$$
$$\psi - \theta_i^* = (1 - p_i)^{s_i}(\psi + \phi - 1)$$
$$(1 - p_i)^{s_i} = \frac{\psi - \theta_i^*}{(\psi + \phi - 1)}$$
$$(1 - p_i) = \left[\frac{\psi - \theta_i^*}{\psi + \phi - 1}\right]^{\frac{1}{s_i}}$$

hencep_i^* becomes

$$p_i^* = 1 - \left[\frac{\psi - \theta_i^*}{\psi + \phi - 1}\right]^{\frac{1}{s_i}} \qquad (11)$$

Notice by Invariance property in estimation theory (Lehmann and Cassella , 1994) that p_i^*in (11) is the restricted MLE for p under the group testing model subject to the constraintsthat p is isotonic, then its ultimate estimator $\hat{p}_i^*$ is given by

$$\hat{p}_i^* = (\hat{p}_1^*, \hat{p}_2^*, \ldots\ldots \hat{p}_{n_i)}^* \qquad (12)$$

where each $\hat{p}_i^*$is as given in (11). Now that we have derived the estimator, it is customaryto discuss its properties namely biasness, MSE and its variance to measure its efficiency.

VII. Bias and MSE of Restricted MLE Estimator $\widehat{P}_i^*$

After deriving the restricted maximum likelihood estimator as provided in (12) we can nowdiscuss its properties. To compute the expectation and variance of $\hat{p}_i^*$,we need to computethe expected value and variance of each $\hat{p}_i^*$, the expectation is upon simplification using (9)becomes by letting $x_i = l$

$$E[\hat{p}_i^*] = 1 - \sum_{l=1}^{n_i} \left[\frac{\psi - \frac{1}{n_i}}{\psi + \phi - 1}\right]^{\frac{1}{s_i}} \binom{n_i}{l} \theta_i^{*l}[1 - \theta_i^*]^{n_i - l} \qquad (13)$$

If we take our sensitivity and specificity as 100% i.e.ψ=ϕ = 100% as assumed by Thompson, (1962); that is, the test is free of errors Equation (13) reduces to

$$E[\hat{p}_i^*] = 1 - \sum_{l=1}^{n_i} \left[1 - \frac{l}{n_i}\right]^{\frac{1}{s_i}} \binom{n_i}{l} \theta_i^{*l}[1 - \theta_i^*]^{n_i - l}$$

as derived by Swallow, 1985, using different notation. The variance of each $\hat{p}_i^*$ is given by

$$\begin{aligned} Var(\hat{p}_i^*) &= E[(\hat{p}_i^* - E[\hat{p}_i^*])^2] \\ &= E[(\hat{p}_i^* - E[\hat{p}_i^*] - 1 + 1)^2] \\ &= E[(1 - E[\hat{p}_i^*] - (1 - \hat{p}_i^*)^2] \end{aligned}$$

which simplifies to

$$Var(\hat{p}_i^*) = E[1 - \hat{p}_i^*]^2 - 1 - \ E[\hat{p}_i^*]^2 \qquad (14)$$

Substituting Equation (13) in (14) we get

$$Var(\hat{p}_i^*) = 1 - \sum_{l=1}^{n_i} \left[\frac{1 - \frac{l}{n_i}}{\psi + \phi - 1}\right]^{\frac{2}{s_i}} \binom{n_i}{l} \theta_i^{*l}[1 - \theta_i^*]^{n_i - l} \ - [1 - \ E[\hat{p}_i^*]]^2 (15)$$

Now that we have found the expected value and variance of each $\hat{p}_i^*$ we compute varianceand expected value of $\hat{p}_i^*$as follows

$$\hat{p}_i^* = \sum_{i=1}^{n_i} w_i \hat{p}_i^*$$

where$w$ is chosen so that it gives unbiased estimator and at the same time minimizevariance on the condition that$\sum_{i=1}^{n_i} w_i = 1$. To find these weights $w_i$we construct alagrangean function

$$L(w, \lambda) = E\left[\sum_{i=1}^{n_i} w_i (p_i^* - p^*)\right]^2 + \lambda(1 - \sum_{i=1}^{n_i} w_i \,,$$

where$w = (w_i, \ldots . w_{n_i})$and find a saddle point of L(w, λ) (a relative maximum with respectto the weight w and a relative minimum with respect to λ). Since we do not have anyinequality or sign restrictions on the choice of variables we have

$$\frac{\partial L(w,\lambda)}{\partial_{w_i}} = 2 \sum_{i=1}^{n_i} w_i \, E[(p_i^* - p^*)(p_i^* - p^*)] - \lambda. \ (16)$$

Now, using the fact that independence implies zero covariance, we obtain $\frac{\partial L}{\partial_{w_i}} = 0$ whichimplies

$$\frac{\partial L(w,\lambda)}{\partial_{w_i}} = 2var(\hat{p}^*) = \lambda \ (17)$$

Solving for$w$ by using Equation (17), we have $w_i = \frac{\lambda}{2var(\hat{p}^*)}$, then substituting this into

$$\frac{\partial L(w, \lambda)}{\partial_{w_i}} = 1 - \sum_{i=1}^{n_i} w_i$$

and solving for λ gives $1 = \sum_{i=1}^{n_i} w_i = n_i \frac{\lambda}{2var(\hat{p}^*)}$and thus $\lambda = \frac{2var(\hat{p}^*)}{n_i}$. We obtain theoptimal weights for $\hat{p}^*$as $w_i = \frac{1}{n_i}$. The expectation of $\hat{p}^*$ is thus given by

$$E(\hat{p}^*) = E\left(\sum_{i=1}^{n_i} w_i \hat{p}_i^*\right)$$

$$= \sum_{i=1}^{n_i} w_i \hat{p}^*$$

The variance of $\hat{p}^*$is

$$\begin{aligned} var(\hat{p}^*) &= E[(\hat{p}^* - E[\hat{p}^*])^2]. \ (18) \\ &= E\left[\sum_{i=1}^{n_i} w_i \, (p_i^* - p^*)\right]^2. \end{aligned}$$

The bias of the estimator $\hat{p}^*$ is

$$Bias(\hat{p}^*) = E[\hat{p}^*] - \sum_{i=1}^{n_i} \hat{p}^*. \quad (19)$$

Having found the variance and bias of $\hat{p}^*$the MSE can be stated as

$$\text{MSE}[\hat{p}^*] = Variance[\hat{p}^*] + Bias[\hat{p}^*]^2 (20).$$

VIII. Asymptotic variance of $\widehat{p}^*$

Since we have constructed an estimator it will be of interest that the confidence interval and asymptotic variance be established. The statistic $\hat{p}^*$is strongly consistent for$p$, the true prevalence. However, in general$\sqrt{n_i}(\hat{p}^* - p)$ converges to a standard normal with mean 0 and variance v_i distribution as n tends to infinity, where $i = 1, 2, \ldots, n_i$. We know that p is isotonic, a consistent estimator and v_i is given by

$v_i = s_i(1-p_i^*)^{2-2s_i}\ \theta_i^*(1-\theta_i^*)(\psi+\phi-1)^{-2}$, see Nyongesa, (2011). From Equation (16), using the fact that independence implies zero covariance, the asymptotic variance of $\hat{p}^*$ isgiven by

$$var(\hat{p}^*) = \begin{bmatrix} v_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & v_{n_i} \end{bmatrix}$$

For further discussion on the above subject see Tebbs et al.,(2003), for the case with no errors in inspection for an uncombined case. We know that $\sqrt{n_i}(\hat{p}^* - p)$converges to a $N(0, v_i)$ distribution as n_i tends to infinity. Hence the Wald Interval $=\hat{p}_i^* \pm Z_{1-\frac{\alpha}{2}}\sqrt{\left(\frac{\hat{v}_i}{s_i}\right)}$, serves as $100(1-\alpha)\%$ confidence interval forp_i^* . The value$Z_{1-\frac{\alpha}{2}}$is the $1-\frac{\alpha}{2}$ quantile of the standard normal distribution. Using the Wald Interval some computer simulations have been generated as shown in Tables 7, 8 and 9.

IX. Discussion

In this Section we provide highlights of our findings in this paper. To enrichen thediscussions herein we have Monte Carlo simulations for bias and Mean Squared Error(MSE) for various group sizes for given sensitivity and specificity. In Table 1 we haveresults of simulated bias and MSE for various group sizes with sensitivity and specificity of99%.from the simulated results, we observe that bias of the unordered estimator increasewith increase in group sizes but vice versa for MSE that is MSE reduce with increase ingroup size. We note that the group of size 5provides the minimal bias and the observationenrich with that of Swallow, (1985) who recommend that relatively small groups size to beused to obtain optimal results.

	Bias				MSE			
s_i	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$
1	−45.3	−84.0	−64.9	−63.6	7.47	12	17	21
5	2.37	4.74	7.33	10	1.16	2.29	3.48	4.75
10	2.69	5.85	9.76	15	0.59	1.27	2.08	3.10
15	2.91	6.8	12	20	0.41	0.95	1.72	2.85
25	3.33	9.07	20	43	0.27	0.76	1.76	3.77
40	4.05	15	44	50	0.21	0.83	1.84	3.98

Table 1: The bias,$* 10^4$ and mean squared error (MSE) $* 10^4$of$\hat{p}$ for various group sizes, $s_i$and with prevalence $p = (0.02, 0.04, 0.06, 0.08)$ and $\psi = \phi = 99\%$

	Bias				MSE			
s_i	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$
1	-79.7	−97.0	−61.6	−69.2	10	15	19	24
5	2.64	5.06	7.73	11	1.30	2.4	3.67	4.97
10	2.88	6.13	10	15	0.63	1.33	2.18	3.24
15	3.07	7.10	13	21	0.43	1.00	1.80	3.03
25	3.47	9.52	22	50	0.29	0.80	1.74	5.06
40	4.23	16	55	70	0.22	0.93	3.43	0.57

Table2: The bias,$* 10^4$ and mean squared error (MSE) $* 10^4$of$\hat{p}$ for various group sizes,s_i and with prevalence $p = (0.02, 0.04, 0.06, 0.08)$ and $\psi = \phi = 98\%$

	Bias				MSE			
s_i	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$
1	93.0	−91.8	−96.0	−100.3	40	45	49	54
5	5.58	8.63	12	16	2.76	4.19	5.76	7.50
10	4.92	9.22	15	23	1.09	2.01	3.23	4.94
15	4.8	10	20	40	0.68	1.48	2.9	10.00
25	5.14	15	88	477	0.43	1.68	54	408
40	6.30	112	926	1531	0.34	87	866	1483

Table 3: The bias,$* 10^4$ and mean squared error (MSE) $* 10^4$of$\hat{p}$ for various group sizes,s_i and with prevalence $p = (0.02, 0.04, 0.06, 0.08)$ and $\psi = \phi = 90\%$

Similar observations as discussed above are depicted when the sensitivity and specificity ofthe test are reduced from 99% to 98% and 90% respectively. Notably the bias and meansquared error increases with decrease in testing kit accuracy as it can be seen in Tables 2and 3 respectively. Computer simulation for bias and mean squared error (MSE) for ourrestricted maximum likelihood estimator for ordered testing scheme are provided in Tables$4, 5$ and 6 with sensitivity and specificity of $99\%, 98\%$ and 90% respectively. For thisestimator, similar observations are depicted as noted in unordered testing scheme and theonly benefit of using the restricted maximum likelihood estimator is that bias and meansquared errors are relatively small as compared to unordered testing scheme.

	Bias				MSE			
s_i	$\hat{p}_1^*$	$\hat{p}_2^*$	$\hat{p}_3^*$	$\hat{p}_4^*$	$\hat{p}_1^*$	$\hat{p}_2^*$	$\hat{p}_3^*$	$\hat{p}_4^*$
1	−45.3	−84.0	−64.9	−63.6	7.47	12	17	21
5	2.37	4.74	7.33	1.30	1.10	2.29	3.48	4.75
10	2.69	1.0	8.4	15	0.59	1.26	2.08	3.10
15	2.91	5.6	12	20	0.41	0.95	1.72	2.85
25	3.33	9.07	20	43	0.27	0.76	1.76	3.77
40	4.05	15	44	50	0.21	0.83	1.84	3.98

Table 4: The bias,$* 10^4$ and mean squared error (MSE) $* 10^4$of$\hat{p}^*$ for various group sizes, s_i and with prevalence $p = (0.02, 0.04, 0.06, 0.08)$ and $\psi = \phi = 99\%$

	Bias				MSE			
s_i	$\hat{p}_1^*$	$\hat{p}_2^*$	$\hat{p}_3^*$	$\hat{p}_4^*$	$\hat{p}_1^*$	$\hat{p}_2^*$	$\hat{p}_3^*$	$\hat{p}_4^*$
1	-79.7	−97.0	−61.6	−69.2	10	15	19	24
5	2.64	5.06	7.73	11	1.30	2.4	3.60	4.90
10	2.88	6.13	7.4	15	0.63	1.33	2.10	3.20
15	3.07	5.1	13	21	0.43	1.00	1.80	3.03
25	0.78	9.52	22	50	0.28	0.80	1.74	5.06
40	4.23	16	55	70	0.22	0.93	3.43	0.57

Table 5: The bias,$* 10^4$ and mean squared error (MSE) $* 10^4$of$\hat{p}^*$ for various group sizes,s_i and with prevalence $p = (0.02, 0.04, 0.06, 0.08)$ and $\psi = \phi = 98\%\backslash$

	Bias				MSE			
s_i	$\hat{p}_1^*$	$\hat{p}_2^*$	$\hat{p}_3^*$	$\hat{p}_4^*$	$\hat{p}_1^*$	$\hat{p}_2^*$	$\hat{p}_3^*$	$\hat{p}_4^*$
1	-93.0	−91.8	−96.0	−100.3	40	45	49	54
5	5.58	8.63	12	16	2.76	4.19	5.76	7.50
10	4.92	9.22	1.38	28	1.09	2.01	3.21	4.94
15	4.8	0.41	20	40	0.68	1.40	2.9	10.00
25	5.14	15	88	477	0.43	1.68	54	408
40	6.30	112	926	1531	0.34	87	866	1483

Table 6: The bias,$* 10^4$ and mean squared error (MSE) $* 10^4$of$\hat{p}^*$ for various group sizes,s_i and with prevalence $p = (0.02, 0.04, 0.06, 0.08)$ and $\psi = \phi = 90\%$

To compute confidence interval (CI) for both restricted maximum likelihood estimator andthe ordinary maximum likelihood estimator, we employed Wald interval procedure withvarious confidence coefficient $(c_1, c_2, c_3, c_4) = (0.90, 0.95, 0.975, 0.99)$ which gives theprobability that the interval produced includes the true value of the parameter and resultsare present in Tables $7, 8$ and 9 for various group sizes, prevalence of the trait and forsensitivity and specificity of $99\%, 99\%$ and 90% respectively.

	Confidence coefficient				Prevalence							
s_i	c_1	c_2	c_3	c_4	$\hat{p}_1$		$\hat{p}_2$		$\hat{p}_3$		$\hat{p}_4$	
					p_l	p_u	p_l	p_u	p_l	p_u	p_l	p_u
5	0.90	0.95	0.975	0.99	0.0042	0.0393	0.0122	0.0708	0.0057	0.1167	0.0105	0.1513
10	0.90	0.95	0.975	0.99	0.0083	0.0332	0.0189	0.0621	0.0177	0.1026	0.0240	0.1355
15	0.90	0.95	0.975	0.99	0.0100	0.0308	0.0215	0.0587	0.0218	0.0975	0.0275	0.1307

Table7: Coverage probability for the Wald interval, of$\hat{p}$ for various group sizes,s_i and withprevalence rate $p = (0.02, 0.04, 0.06, 0.08)$ and $\psi = \phi = 99\%$

	Confidence coefficient				Prevalence							
s_i	c_1	c_2	c_3	c_4	$\hat{p}_1$		$\hat{p}_2$		$\hat{p}_3$		$\hat{p}_4$	
					p_l	p_u	p_l	p_u	p_l	p_u	p_l	p_u
5	0.90	0.95	0.975	0.99	0.0050	0.0421	0.0127	0.0732	0.0054	0.1194	0.0097	0.1538
10	0.90	0.95	0.975	0.99	0.0086	0.0344	0.0188	0.0631	0.0169	0.1036	0.0225	0.1364
15	0.90	0.95	0.975	0.99	0.0102	0.0315	0.0212	0.0592	0.0207	0.0980	0.0253	0.1311

Table 8: Coverage probability for the Wald interval, of$\hat{p}$ for various group sizes, s_i, and withprevalence rate $p = (0.02, 0.04, 0.06, 0.08)$ and $\psi = \phi = 98\%$

	Confidence coefficient				Prevalence							
s_i	c_1	c_2	c_3	c_4	$\hat{p}_1$		$\hat{p}_2$		$\hat{p}_3$		$\hat{p}_4$	
					p_l	p_u	p_l	p_u	p_l	p_u	p_l	p_u
5	0.90	0.95	0.975	0.99	0.0113	0.0651	0.0160	0.0947	0.0013	0.1433	0.0203	0.1578
10	0.90	0.95	0.975	0.99	0.0111	0.0447	0.0179	0.0718	0.0093	0.1134	0.0236	0.1309
15	0.90	0.95	0.975	0.99	0.0111	0.0376	0.0182	0.0638	0.0101	0.1035	0.0202	0.1227

Table 9: Coverage probability for the Wald interval, of$\hat{p}$ for various group sizes, s_i, and withprevalence rate $p = (0.02, 0.04, 0.06, 0.08)$ and $\psi = \phi = 90\%$

From the simulated results no negative value of confidence interval was observed in thethree tables which imply that Wald interval procedure performs satisfactory. When thegroup sizes is increased at a fixed sensitivity and specificity the confidence interval sizereduces. Notably across the three tables that is Table 7, Table 8 and Table 9 is that13confidence interval increase with decrease in sensitivity and specificity.

X. Conclusion

Based on ourdiscussion we provide conclusion to our present study. In this paper we have constructed arestricted maximum likelihood estimator of prevalence in a population. The properties ofthe maximum likelihood estimator such as bias, mean squared error, asymptotic varianceand confidence interval are provided in the discussion. To justify the purpose of this study, we also discussed the estimator of un-ordered group testing. We compared the efficiency ofour developed estimator and the existing estimator by simulation of the bias and meansquared error for both procedures. When the group size $s_i > 1$ the bias in both procedures isminimal as compared to one-at-a-time testing. Similar observations are noted for meansquared error for both procedures. For fixed sensitivity and specificity the bias of restrictedmaximum likelihood estimator is less than that of un-ordered maximum likehood estimator. For example for $\psi = \phi = 99\%$and group size ($s_i = 10$) when we use restricted maximumlikelihood estimator we have bias of 1.0 and when we use ordinary maximum likelihoodestimator we have a bias of5.85. Therefore, our constructed restricted maximum likelihoodestimator yields an estimator with smaller variance as compared to ordinary maximumlikelihood estimator hence substantial improvement. A similar conclusion is arrived at inthe case of mean squared error (MSE).

References

[1] Dorfman R. (1943). The detection of defective members of large populations. Annals of mathematical statistics 14, 436-440.
[2] Sobel M., Ellashoff R.M. (1975).Group testing with a new goal, Estimation.Biometrika 62, 181-193.
[3] Nyongesa, L. K., (2011). Dual Estimation of Prevalence and Disease Incidence in Pool-Testing Strategy. Communication in Statistics Theory and Method. 40, 3218-3229.
[4] Brookmeyer, R. (1999). Analysis of Multistage Pooling Studies of Biological Specimensfor Estimating Disease Incidence and Prevalence. Biometrics 55,608-612.
[5] Hwang, F.K., (1976). Group testing with a dilution effect. Biometrika 63,611-613.
[6] T. Robertson, F. T. Wright, and R. L. Dykstra.Order Restricted Statistical Inference. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley and Sons, Chichester 1988.
[7] Thompson K.H. (1962).Estimation of the proportion of vectors in a natural populationof insects. Biometrics 18, 568-578.
[8] Hepworth, G. and Watson, R. (2009). Debiased Estimation of Proportions inGroup Testing. Appl. Stats, 58, 105-121.
[9] lehmannG.casella, Theory of Point Estimation. 2nd Edition.(Vol.3,No.2 1994) 236-239
[10] TebbsJm& Swallow Wh, (2003).Estimating ordered binomial proportions withthe use of group testing. Biometrika 90,471-477.
[11] Swallow Wh, (1985).Group testing for estimating infection rates and probabilities ofdisease transmission. Phytopathology Vol.75, N.8,568-578.

Collins Musavi Amuhaya. "Estimation of Prevalence Using Ordered Pool-Testing Scheme In The Presence Of Testing Errors." IOSR Journal of Mathematics (IOSR-JM) 13.4 (2017): 79-85.