Vol.6.Issue.3.2018 (July-Sept.)

http://www.bomsr.com
Email:editorbomsr@gmail.com

RESEARCH ARTICLE

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

INTERNATIONAL STANDARD SERIAL NUMBER 2348-0580

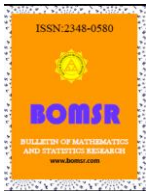
BAYESIAN ESTIMATION OF PROBABILITY OF CONTAMINATION UNDER GROUP SCREENING DESIGN

ANDREA OTWANDE[1*], ERICK OKUTO[1], JOSEPH OTTIENO[2]

[1]School of Mathematics and Actuarial Science, JaramogiOgingaOdinga University of Science and Technology, P.O Box 210-40601,Bondo Kenya.

[2]School of Mathematics and Actuarial Science, University of Nairobi, P.O Box 30197-0100, Nairobi, Kenya.

*Email of the corresponding author: otwandeandrea@gmail.com



ABSTRACT

In health area the prevalence rate p is considered to be a fraction that show positive results out of the entire sampled population when tests are done to establish infection of a disease. Tests can be done on individual basis or in a group with individual testing being more costly for large samples of individuals. Using pooled sampling such challenges can be addressed. Group screening was pioneered by Dorfman in 1943 who found it to be more economical in testing blood samples of army inductees to detect syphilis infection. Effective group screening however require choice of optimum values of parameters to guard against obtaining inflated bias in their estimation. Beta distribution was considered as prior distribution in Bayesian estimation of p. When analyzed for group testing the results indicated that Bayes estimates perform much better than maximum likelihood estimates (MLE) especially when prevalence rates are low, while MLE performed better with large values of p. It was also noted that combinations of parameters that lower than the optimum values still resulted in more cost effective performance than individual testing.

Key words: Group screening, prevalence rate, optimum values, MLE and Bayes estimate.

1.0 Introduction

In experimental designs to compare treatment of diseases the analysis can be done through comparison of the estimates of the disease transmission probabilities. These prevalence estimates are important in planning health services and policies. Many diseases can be detected in blood

samples. The most straight forward method of testing blood could involve one at a time testing for N individuals, thus forming N individual tests. The prevalence rate p therefore can be estimated as a fraction of the individuals showing positive results out of the entire population sample N. In terms of the cost it follows that the cost of testing is directly proportional to the number of tests done, thus making this design to be more costly especially for large samples of individuals under screening. The other challenge associated with individual testing is when the diagnostic tests are imperfect which can be caused by differences in laboratory techniques, individual characteristics and other aspects of study design and implementation. Estimation of p using pooled sampling is an area of study of great interest as it tends to address the challenges encountered by individual testing. Group screening design was pioneered by Dorfman[5] in 1943 as an economical method of testing blood samples of army inductees to detect those with venereal disease, syphilis. He proposed that a sample of blood from each member could be taken, pooled together and tested. If the test is negative then the pooled sample tested would be considered to be free from infection and if pooled sample tested positive then the individual samples would be re-tested to determine which inductees are infected. The group testing design applied by Dorfman has been applied to many other epidemiological research areas including; HIV[6],hepatitis[12] and many others. This paper compares group testing design to estimate parameter p using Bayesian technique with Maximum likelihood estimate method.Group testing designs can be based on classification and/or on parameter estimation. In both cases the main advantage is that there is a chance of making savings in efficiency and number of trials. This is possible especially if the prevalence rate p is small enough so that many groups would test negative leading to substantial savings in the total number of tests done. A good group testing design will have appropriate group size k, that minimizes mean square error (MSE) of estimate $\hat{p}$ and provides estimate p with MSE smaller than would be obtained in case of individual testing. Having too large bias can make estimator to have largely inflated MSE. Bias largely depends on the group size k. Therefore k must carefully be chosen to avoid large bias. Choosing the value of k that provides small bias can pose a great challenge especially if the value of p is not specific. It is therefore advisable to select a group testing design that though not optimal can have reasonable bias and prove to be better than individual test design. Swallow[18]came up with a design that could effectively be applied to choose the value of k that can minimize MSE $\hat{P}$ especially when the value of N is already determined. Swallow came up with a table indicating various optimal sizes of k used to estimate prevalence rate of disease infection based on both group testing and one to one test. The result indicated better performance with group testing than with individual test even with varying values of the sample sizes. Therefore group testing design can be more efficient and cost effective than one to one testing technique.

2.0 Methods of Estimation

2.1 Maximum Likelihood Estimation (MLE)

For group screening experiment the observed outcomes are considered to be independently and identically distributed Bernoulli (p) random variables from a sample of N individuals to be tested. If N is divided into g groups each of which is size k (where $k > 1$) then $N = kg$.

A group is defective if it has at least one defective (positive) individual. The probability of an individual being defective is p varying according to a certain distribution.

Let $s =$ the number of defective items in a group of size k. Therefore

$$P(S = s) = \binom{k}{s} p^s (1-p)^{k-s} \qquad s = 0,1,2,\ldots k \qquad \text{.....[1]}$$

$$P(S=0)=\binom{k}{0}p^0(1-p)^{k-0}$$

$=(1-p)^k$ which is the expected fraction of the uninfected group of k individuals. Thus the probability of a group being defective is

$$P(S\geq s)=\sum_{s=1}^{k}\binom{k}{s}p^s(1-p)^{k-s}$$

$$=1-p(S=0)$$

$$=1-(1-p)^k$$

Let

$$p^*=1-(1-p)^k \qquad \text{.....[2]}$$

If $R=r=$ the number defective groups out of g groups then $\frac{r}{g}$ is the observed proportion of defective groups. Therefore $\frac{r}{g}$ is an estimator of $1-(1-p)^k$ which can be expressed as $1-(\widehat{1-p})^k=\frac{r}{g}$.

To estimate p we have

$$1-(1-p)^k=\frac{r}{g}$$

$$-(1-p)^k=-1+\frac{r}{g}$$

$$(1-p)^k=1-\frac{r}{g}$$

$$(1-p)=\left(1-\frac{r}{g}\right)^{\frac{1}{k}}$$

$$p=1-\left(1-\frac{r}{g}\right)^{\frac{1}{k}} \qquad \text{.....[3]}$$

Therefore

$$\hat{P}_k=1-\left(1-\frac{r}{g}\right)^{\frac{1}{k}} \quad or \quad 1-(1-p^*)^{\frac{1}{k}}$$

which is the maximum likelihood estimate of p .

2.2 Bayes Estimate

Here the Bayes inference is about the parameter p which is a random variable in the binomial mixture. This parameter p is considered to have a beta prior distribution since it is a conjugate to the mixed distribution and it has a domain that is fairly small (0,1) and is given as

$$g(p)=\frac{1}{B(a,b)}p^{a-1}(1-p)^{b-1}\ 0<p<1. \qquad \text{.....[4]}$$

Let $r=$ the number of defective groups out of g groups. Then the distribution of r takes binomial distribution. Thus

$$P(R=r)=\binom{g}{r}p^{*r}\left(1-p^{*}\right)^{g-r} \quad \text{.....[5]}$$

The joint distribution of (r,p) becomes

$$f(r,p)=\frac{1}{B(a,b)}p^{a-1}(1-p)^{b-1}\binom{g}{r}p^{*r}\left(1-p^{*}\right)^{g-r}$$

$$=\frac{1}{B(a,b)}p^{a-1}(1-p)^{b-1}\binom{g}{r}\left[1-(1-p)^{k}\right]^{r}\left[1-\left(1-(1-p)^{k}\right)\right]^{g-r}$$

$$=\frac{1}{B(a,b)}p^{a-1}(1-p)^{b-1}\binom{g}{r}\left[1-(1-p)^{k}\right]^{r}\left[(1-p)^{k}\right]^{g-r}$$

$$=\frac{\binom{g}{r}}{B(a,b)}p^{a-1}(1-p)^{b-1}(1-p)^{gr-kr}\left[1-(1-p)^{k}\right]^{r}$$

$$f(r,p)=\frac{\binom{g}{r}}{B(a,b)}p^{a-1}(1-p)^{gk--kr+b-1}\left[1-(1-p)^{k}\right]^{r} \quad \text{.....[6]}$$

Marginal distribution of r becomes

$$f(r)=\int_{0}^{1}f(r,p)dp$$

$$=\frac{\binom{g}{r}}{B(a,b)}\int_{0}^{1}p^{a-1}(1-p)^{gk-kr+b-1}\left[1-(1-p)^{k}\right]^{r}dp$$

$$=\frac{\binom{g}{r}}{B(a,b)}\int_{0}^{1}p^{a-1}(1-p)^{k(g-r)+b-1}\sum_{l=0}^{r}(-1)^{l}\binom{r}{l}(1-p)^{kl}\,dp$$

$$=\frac{\binom{g}{r}}{B(a,b)}\sum_{l=0}^{r}(-1)^{l}\binom{r}{l}\int_{0}^{1}p^{a-1}(1-p)^{gk+kl-kr+b-1}dp$$

$$\therefore f(r)=\frac{\binom{g}{r}}{B(a,b)}\sum_{l=0}^{r}(-1)^{l}\binom{r}{l}B\left(a,k(g+l-r)+b\right) \quad \text{.....[7]}$$

Posterior distribution is the probability of varying parameter p given the random variable r. It tells much about parameter p given that the sample data has occurred.

The posterior distribution becomes

$$f(p\mid r)=\frac{f(r,p)}{f(r)}$$

$$=\frac{\binom{g}{r}B(a,b)^{-1}p^{a-1}(1-p)^{k(g-r)+b-1}\left[1-(1-p)^{k}\right]^{r}}{\binom{g}{r}B(a,b)^{-1}\sum_{l=0}^{r}(-1)^{l}\binom{r}{l}B\left(a,k(g+l-r)+b\right)}$$

$$= \frac{p^{a-1}(1-p)^{k(g-r)+b-1}\left[1-(1-p)^k\right]^r}{\sum_{l=0}^{r}(-1)^l\binom{r}{l}B(a,k(g+l-r)+b)} \quad \text{.....[8]}$$

And the j[th] moment of the posterior distribution becomes

$$E(P^j) = \frac{\int_0^1 p^{j+a-1}(1-p)^{k(g-r)+b-1}\left[1-(1-p)^k\right]^r dp}{\sum_{l=0}^{r}(-1)^l\binom{r}{l}B(a,k(g+l-r)+b)}$$

$$= \frac{\sum_{l=0}^{r}(-1)^l\binom{r}{l}\int_0^1 p^{j+a-1}(1-p)^{gk+kl-gr+b-1}dp}{\sum_{l=0}^{r}(-1)^l\binom{r}{l}B(a,kg+kl-kr+b)}$$

$$E(P^j) = \frac{\sum_{l=0}^{r}(-1)^l\binom{r}{l}B(j+a,gk+kl-gr+b)}{\sum_{l=0}^{r}(-1)^l\binom{r}{l}B(a,gk+kl-gr+b)}. \quad \text{.....[9]}$$

Then the expected value of parameter p is

$$E(P \mid X) = \hat{p}_k = \frac{\sum_{l=0}^{r}(-1)^l\binom{r}{l}B(1+a,gk+kl-gr+b)}{\sum_{l=0}^{r}(-1)^l\binom{r}{l}B(a,gk+kl-gr+b)} \quad \text{.....[10]}$$

2.3 Estimates of $p,\ a\ \ and\ \ b$

Estimation of the parameters $p,\ a\ \ and\ \ b$ can be achieved based on the theorem below.

Theorem1; The marginal density of r is

$$f(r) = k^*\binom{g}{r}\frac{a^r b^{k(g-r)}}{(a+b)^{r+k(g-r)}} \quad \text{.....[11]}$$

Proof: From Muhua.....[13]

$$f(r) = \int_0^1 \binom{g}{r}\left(B(a,b)\right)^{-1} p^{a-1}(1-p)^{kg-kr+b-1}\left[1-(1-p)^k\right]^r dp$$

For small p, $(1-p)^k \approx 1-kp$

Therefore

$$f(r) \approx \int_0^1 \binom{g}{r}\left(B(a,b)\right)^{-1} p^{a-1}(1-p)^{k(g-r)+b-1}(kp)^r dp$$

$$= k^r\binom{g}{r}\left(B(a,b)\right)^{-1}\int_0^1 p^{a+r-1}(1-p)^{k(g-r)+b-1}dp$$

$$= k^r\binom{g}{r}\left(B(a,b)\right)^{-1}B(a+r,k(g-r)+b)$$

$$= k^r \binom{g}{r} \frac{\Gamma(a+b)\Gamma(a+r)\Gamma(k(g-r)+b)}{\Gamma a \Gamma b \Gamma(a+r+k(g-r)+b)}$$

Using Abramowitz and Stegun[1] technique in which for large N $\frac{\Gamma(N+a)}{\Gamma(N+b)} \approx N^{a-b}$

therefore

$$f(r) = k^* \binom{g}{r} \frac{a^r b^{k(g-r)}}{(a+b)^{r+k(g-r)}}$$

To obtain the maximum likelihood estimate of the parameter b equation [11] is differentiated partially with respect to b and equated to zero

$$\frac{df(r)}{db} = \frac{k^r \binom{g}{r} \left[(a+b)^{r+k(g-r)} r a^r k(g-r) b^{k(g-r)-1} - a^r b^{k(g-r)} \left(r + k(g-r)\right)(a+b)^{r+k(g-r)-1} \right]}{(a+b)^{2(r+k(g-r))}} = 0$$

On further simplification

$$\hat{b} = \frac{ak(g-r)}{r} \quad \text{.....[12]}$$

which is easier to apply.

When $a = 1$ then the joint equation[6] becomes

$$f(r,p) = \binom{g}{r} \left[1-(1-p)^k\right]^r (1-p)^{k(g-r)} b(1-p)^{b-1} \quad \text{.....[13]}$$

And the marginal density is

$$f(r) = b\binom{g}{r} \int_0^1 \left[1-(1-p)^k\right]^r [1-p]^{k(g-r)+b-1} dp \quad \text{.....[14]}$$

Using transformation technique, let

$$u = (1-p)^k, \qquad p = 1-u^{\frac{1}{k}} \text{ and } dp = -(1-p)^{k-1} du$$

Therefore equation [14] becomes

$$f(r) = bk^{-1} \binom{g}{r} \int_0^1 u^{g-r+\frac{b}{k}-1} (1-u)^r du$$

$$= \frac{b\Gamma(g+1)\Gamma\left(a-r+\frac{b}{k}\right)}{k\Gamma(g-r+1)\Gamma\left(g+\frac{b}{k}\right)}$$

$$\approx \frac{b(g+1)^r}{k\left(g+\frac{b}{k}\right)^{r+1}} \; r = 0,1,2,\ldots g \, . \quad \text{.....[15]}$$

To obtain estimate of b using MLE equation [15] is differentiated with respect to b and equated to zero.

$$\frac{\frac{1}{k}\left[\left(g+\frac{b}{k}\right)^{r+1}(g+1)^r - b(g+1)^{\frac{r(r+1)}{k}}\left(g+\frac{b}{k}\right)^r\right]}{\left[\left(g+\frac{b}{k}\right)^{r+1}\right]^2} = 0$$

$$\hat{b} = \frac{gk}{r}$$

.....[16]

Posterior distribution based on special case when $a = 1$ becomes

$$f(p \mid r) = \frac{k\Gamma\left(g + \frac{\hat{b}}{k} + 1\right)}{\Gamma\left(g - r + \frac{\hat{b}}{k}\right)\Gamma(r+1)}(1-p)^{k(g-r)+\hat{b}-1}\left[1-(1-p)^k\right]^r$$

$$= \frac{k\Gamma\left(g + \frac{g}{r} + 1\right)}{\Gamma\left(g - r + \frac{g}{r}\right)\Gamma(r+1)}(1-p)^{k(g-r)+\frac{gk}{r}-1}\left[1-(1-p)^k\right]^r \quad 0 < p < 1 \qquad \text{.....[17]}$$

The expected value of p becomes

$$E\left(\hat{P} \mid X\right) = \int_0^1 p f(p \mid r, b)\, dp$$

$$= \int_0^1 \frac{k\Gamma\left(g + \frac{\hat{b}}{k} + 1\right)}{\Gamma\left(g - r + \frac{\hat{b}}{k}\right)\Gamma(r+1)} p(1-p)^{k(g-r)+\hat{b}-1}\left[1-(1-p)^k\right]^r$$

Using change of variable technique, let

$$u = (1-p)^k, \qquad p = 1 - u^{\frac{1}{k}} \text{ and } dp = -(1-p)^{k-1}\, du$$

$$= \frac{k\Gamma\left(g + \frac{\hat{b}}{k} + 1\right)}{\Gamma\left(g - r + \frac{\hat{b}}{k}\right)\Gamma(r+1)} \int_0^1 \frac{1}{k}\left(1 - u^{\frac{1}{k}}\right) u^{g-r+\frac{\hat{b}}{k}-1}(1-u)^r\, du$$

$$= \frac{\Gamma\left(g + \frac{\hat{b}}{k} + 1\right)}{\Gamma\left(g - r + \frac{\hat{b}}{k}\right)\Gamma(r+1)} \left\{ \int_0^1 u^{g-r+\frac{\hat{b}}{k}-1}(1-u)^r\, du - \int_0^1 u^{g-r+\frac{\hat{b}}{k}+\frac{1}{k}-1}(1-u)^r\, du \right\}$$

$$= \frac{\Gamma\left(g + \frac{\hat{b}}{k} + 1\right)}{\Gamma\left(g - r + \frac{\hat{b}}{k}\right)\Gamma(r+1)} \left\{ \frac{\Gamma\left(g + \frac{\hat{b}}{k} - r\right)\Gamma(r+1)}{\Gamma\left(g + 1 + \frac{\hat{b}}{k}\right)} - \frac{\Gamma\left(g + \frac{\hat{b}}{k} + \frac{1}{k} - r\right)\Gamma(r+1)}{\Gamma\left(g + \frac{\hat{b}}{k} + 1\right)} \right\}$$

$$= 1 - \frac{\Gamma\left(g + \frac{\hat{b}}{k} + 1\right)\Gamma\left(g - r + \frac{\hat{b}}{k} + \frac{1}{k}\right)}{\Gamma\left(g - r + \frac{\hat{b}}{k}\right)\Gamma\left(g + \frac{\hat{b}}{k} + \frac{1}{k} + 1\right)}$$

And using Abramowitz and Stegun technique

$$\hat{p} = 1 - \left(\frac{g-r}{g+1}\right)^{\frac{1}{k}} \qquad \text{.....[18]}$$

as obtained by Muhua[13].

For example, In a study conducted by Liu et al. in 1997 the results of 1875 blood donors screened for anti HCV at the blood transfusion in China were tested individually (k=1) to examine the effectiveness of pooling and compared with a group of size k=5 and g=375, they got r=37.

Using Bayes estimate technique, from equation [17] we have

$$\hat{b} = \frac{gk}{r}.$$

$$\frac{1875}{37} = 50.6757$$

And the expected value of p using equation [18] becomes,

$$\hat{p} = 1 - \left[1 - \frac{375-37}{376}\right]^{\frac{1}{5}}$$
$$= 1 - 0.97891$$
$$= 0.021083$$

2.4 Optimal choice of Parameter b and the Mean Squared Error

Since the choice of the parameter b has strong influence over the bias and MSE in estimating the value p, an appropriate value of b must be considered in order to obtain bias that is as small as possible. From Xing[22] it is noted that the Bayes estimate p decrease as parameter b increases and vise versa. Thus with a low parameter b the posterior mean overestimates the parameter p and with large b it underestimates the parameter p.

The tables 1 -2 below show optimal values of parameter b with their corresponding mean squared errors obtained using simulated data under different combinations of k, g and p when $a = 1$

Table 1: Optimal b based on the smallest MSE of Bayes estimator by using $g = 10$ withcombinations of different k, and p.

	k	Optimal Value b	MSE
P = 0.05	5	34	1.196E-03
	10	42	8.509E-04
	15	46	2.404E-03
P = 0.01	5	131	2.212E-04
	10	151	1.159E-04
	15	166	8.058E-05
P = 0.1	5	19	2.807E-03
	10	23	1.291E-02
	15	26	-
P = 0.2	5	11	1.768E-02
	10	12	0.2068155
	15	11	-

Table2: Optimal b based on the smallest MSE of Bayes estimator by using $g = 20$ withcombinations of different k, and p.

	k	Optimal Value b	MSE
P = 0.005	5	37	2.375E-03
	10	43	3.346E-04
	15	50	2.750E-04
P = 0.01	5	153	1.056E-04
	10	167	5.470E-05
	15	183	3.768E-05
P = 0.1	5	20	1.224E-03
	10	26	1.087E-03
	15	32	-

P = 0.2	5	12	3.365E-03
	10	16	6.763E-02
	15	14	-

Mean squared error (MSE) is a measure of average squared deviation between the estimated parameters and the true parameter.

$$Bias = E(\hat{P}) - P$$

$$MSE = E\left[\left(\hat{P} - P\right)^2\right]$$

Where $E(\hat{P})$ is the estimated parameter and p is the true parameter.

The tables above indicate various group sizes k showing different MSE under specific optimal values of parameter b from the prior Beta distribution which contain information about parameter p to be estimated using maximum likelihood estimate and Bayes estimate. Therefore the choice of parameter b is critical in estimating the parameter p.

The following table 3 compares the MSE of Bayes and Maximum Likelihood estimators from simulated data using statistical software R, with different groups (g), group sizes (k) and parameter (p) which are not necessarily optimal values.

Table 3

			Number of groups g			
	Group size k	MSE	5	10	15	20
$P = 0.005$	5	$(\hat{P}B)$	2.245E-06	4.036E-06	4.036E-06	4.605E-06
		$(\hat{P}mle)$	2.410E-04	6.836E-05	6.836E-05	5.400E-05
	10	$(\hat{P}B)$	3.331E-06	4.451E-06	4.990E-06	4.891E-06
		$(\hat{P}mle)$	1.184E-04	5.482E-05	3.778E-05	2.681E-05
$P = 0.05$	5	$(\hat{P}B)$	4.709E-04	4.146E-04	3.536E-04	2.989E-04
		$(\hat{P}mle)$	3.455E-03	1.172E-03	7.687E-04	5.414E-04
	10	$(\hat{P}B)$	4.210E-04	3.206E-04	2.520E-04	2.091E-04
		$(\hat{P}mle)$	1.256E-02	8.486E-04	4.516E-04	3.325E-04
$P = 0.01$	5	$(\hat{P}B)$	1.372E-05	1.775E-05	1.936E-05	2.034E-05
		$(\hat{P}mle)$	4.940E-04	2.144E-04	1.424E-04	1.099E-04
	10	$(\hat{P}B)$	1.820E-05	1.994E-05	1.910E-05	1.795E-05
		$(\hat{P}mle)$	2.677E-04	1.167E-04	7.130E-05	5.354E-05
$P = 0.1$	5	$(\hat{P}B)$	1.620E-03	1.204E-03	9.600E-04	8.091E-04
		$(\hat{P}mle)$	1.629E-02	2.990E-03	1.700E-03	1.262E-03
	10	$(\hat{P}B)$	1.225E-03	9.236E-04	6.835E-04	5.731E-04
		$(\hat{P}mle)$	9.640E-02	1.330E-02	1.618E-03	1.337E-03

The result from the table3 above indicate that MSE of Bayes estimate generally perform better than their corresponding estimates obtained using MLE when p is smaller as MLE does better for bigger values of p .It is important to note that, although the choice of optimal values of k , g and p are critical to minimize MSE,use of values smaller than the optimal still remain cost effective much more than greater values as it also minimizes MSE. However the choice of k that is larger than the optimal may produce larger bias. Swallow[18] demonstrated that most of the advantage of group testing over one to one test designs can be realized much better with group sizes smaller than the optimal size k . The paper illustrates how when the cost of increasing group size is not negligible, then using smaller size than the optimal size of k is likely to be more cost effective than using the optimal size k . This idea is more illustrated from the following tables of results of relative bias and relative efficiency for selected values of p , k and g .

Tables4-5 below display the relative bias and relative efficiency respectively for selected values of p , k and g which are not necessarily optimal. Relative bias and relative efficiency can as well be used to compare performance of parameter estimators.

Relative bias of the estimate p is

$$RB\left(\hat{P}\right)=\frac{E\left(\hat{P}-P\right)}{P}$$

And the relative efficiency of p is

$$RE\left(\hat{P}\right)=\frac{MSE(\hat{P})}{MSE\left(P\right)}.$$

Table 4; Relative bias for selected values of p , k and $g=10$.

g=10 p	k	$\hat{P}mle$	(a,b)	$\hat{P}B$
0.25			(1,3)	
	1	0.00000		0.00000
	5	0.21661		0.09552
	10	1.57583		0.29291
	15	2.56691		0.37393
	20	2.88854		0.35010
0.10			(1,9)	
	1	0.00000		0.00000
	5	0.05731		0.03843
	10	0.19010		0.08031
	15	0.90474		0.14572
	20	2.39111		0.23484
0.05			(1,19)	
	1	0.00000		0.00000
	5	0.04851		0.02363
	10	0.06713		0.04351
	15	0.11444		0.06241
	20	0.29962		0.08494
0.01			(1,99)	
	1	0.00000		0.00000

	5	0.04360	0.00452
	10	0.05082	0.01180
	15	0.05451	0.01821
	20	0.05734	0.02374

Table 5; Relative Efficiency for selected values of p, k and $g=10$.

g=10 p	k	$\hat{P}mle$	(a,b)	$\hat{P}B$
0.25			$(1,3)$	
	1	1.00000		1.96000
	5	1.03260		4.69672
	10	0.99481		19.03341
	15	0.99862		35.90983
	20	0.99971		59.85294
0.10			$(1,9)$	
	1	1.00000		4.00000
	5	1.17922		1.70664
	10	1.04641		7.82342
	15	1.00283		35.66633
	20	0.99924		71.11520
0.05			$(1,19)$	
	1	1.00000		9.00000
	5	1.12490		2.19461
	10	1.19852		1.89812
	15	1.07565		5.68040
	20	1.01584		24.86763
0.01			$(1,99)$	
	1	1.00000		121.00000
	5	1.09484		9.85463
	10	1.11742		4.47010
	15	1.13311		3.14221
	20	1.14780		2.57100

From table4above the relative bias of Bayes estimators are generally lower than MLE indicating that Bayes estimators perform much better than MLE. Table 5 shows that the relative efficiency reduce with reduction in the values of p and increase in size of k.

3.0 Discussion

This article asserts the significance of group testing especially in areas such as health and other research industries. It is noted that the size of group (k) and the value of p play a critical role in determining a group screening design that is precise and consistent. Small group size with large values of p results in a relatively large bias while small values of p and k produce small bias. Mean squared error is considered as the means to determine the optimal group sizes and parameter p. It is observed that optimal group size k and p produce minimum MSE as shown in the tables 1-2.

When Maximum likelihood estimate and Bayesian approach are compared in estimating prevalence rates, Bayesian procedure seems to perform much better for low values of p for experiments assumed to have no errors. This is because estimating low prevalence using MLE require large sample to obtain estimate that is above zero, which turns out to be costly hence the need for Bayesian approach.

4.0 Conclusion

Group testing is where units are pooled together and tested as a group instead of individually. Individual testing can be too time consuming making group testing be more cost effective. Using simulated studies in the application of results from Bayesian procedure performs better when compared with MLE in estimating prevalence rates especially for cases where prevalence rate is low. This method can be relevant in health area to test for diseases that are in blood samples to reduce the cost and improve effectiveness in diagnosis and treatment of some diseases.

5.0 References

[1]. Abramowitz, M. and Stegun, I.A. (1960): ''Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables. Washington, D.C.: Bureau of National Standards".

[2]. Cardoso M, Koerner K, Kubanek B.(1998) Mini-pool screening by nucleic acid testing for hepatitis B virus, hepatitis C virus, and HIV: preliminary results. *Transfusion*;38:905–907.

[3]. Chao, L.C and Swallow, W.H (1990): ''Using Group Testing to Estimate a Proportion and to Test the Binomial Model". *Biometrics*46:1035-1046.

[4]. Chaubey,Y and Li,W(1995); Comparison between MLE and Bayes Methods for estimation of binomial probability with sample compositing*. Journal of Official statistics 11*,379-390.

[5]. Dorfman, R (1943): ''The Detection of Defective Members of a large Population*". Annals of Mathematical statistics,14*,436-440.

[6]. Emmanuel JC, Bassett MT, Smith HJ, Jacobs J.A(1988). Pooling of sera for human immunodeficiency virus (HIV) testing: an economic method for use in developing countries. *Journal of Clinical Pathology*; 41:582–585.

[7]. GastwirthJ(2000). The efficiency of pooling in the detection of rare mutations. *American Journal of Human Genetics*; 67:1036–1039.

[8]. Gastwirth, J.L and Hammick, P. (1989):"Estimation of the Prevalence of a Rare Disease, Preserving the Anonymity of the Subjects by Group Testing: Application to Estimating the Prevalence of AIDS Antibodies in Blood Donors." *Journal of Statistical Planning and Inference* 22: 15-27.

[9]. Griffiths, D.A (1973):"Maximum Likelihood Estimation for the Beta-Binomial Distribution and Application to the Household Distribution of the Total Number of a Disease." *Biometrics,* 29: 637-648.

[10]. Kline, R., Brothers, T., Bookmeyer, R., Zeger, S. and Quinn, T., (1989): ''Evaluation of HIV Seroprevalence in Population Surveys Using Pooled Sera." *Journal of Clinical Microbiology* 27: 1449-1452

[11]. Lew, R.A. and Levy, P. (1989):" Estimation of Prevalence on the Basis of Screening Tests." *Statistics in Medicine* 8: 1225-1230.

[12]. Liu, P., Shi, Z.,Zhang, Y., Xu, Z., Shu, H. and Zhang, X. (1997):" A Prospective Study of a Serum-Pooling Strategy in Screening Blood Donors for Antibody to Hepatitis C Virus." *Transfusion* 37: 732-736.

[13]. MuhuaG.O(2009); Estimation problem in Group Screening designs. A PhD thesis. University of Nairobi.

[14]. Muhua,G.O and Ottieno, J.A.M(2016); On Bayesian Estimation in Group Screening Designs without errors in decision.International journal of Computational and Theoretical Statistics 3, 39-48.

[15]. Rao, C.R. (1973): ''Linear Statistical Inference and Its Applications. 2nd Edition. New York: John Wiley and Sons: p. 426"

[16]. Skellam, J.G. (1948): "A probability Distribution Derived from the Binomial Distribution by Regarding the Probability of Success as Variable Between the Sets of Trials". *Journal of Royal Statistical Society Series B* 10: 254-261

[17]. Swallow, H.W. (1985):"Group Testing for Estimating Infection Rates and Probabilities of Disease Transmission." *Am Phytopathological Society* 75: 882-889.

[18]. Swallow, H.W. (1987):"Relative Mean Squared Error and Cost Considerations in Choosing Group Size for Group Testing to Estimate Infection Rates and Probabilities of Disease Transmission" .*Phytopathology* 77: 1376-1381

[19]. Van TT, Miller J, Warshauer DM, Reisdorf E, Jernigan D, Humes R, ShultPA(2012); Pooling nasopharyngeal/throat swab specimens to increase testing capacity for influenza viruses by PCR. *Journal of Clinical Microbiology*; 50:891–896.

[20]. Xing,W(2015); Comparison of Estimators of small proportions under group testing. Master's thesis. Florida International University; Miami Florida.