Computation of Moments in Group Testing with Re-testing and with Errors in Inspection

Cox Lwaka Tamba *clwaka@yahoo.com*
Faculty of Science/Department of Mathematics/Division of Statistics
Egerton University
P.O Box 536, Egerton, Kenya

Martin Wafula Nandelenga *nmartin200884@gmail.com*
Faculty of Arts and Social Sciences/Department of Economics
Egerton University
P.O Box 536, Egerton, Kenya

Abstract

Screening of grouped urine sample was suggested during the Second World War as a method for reducing the cost of detecting syphilis in U.S. soldiers. Grouping has been used in epidemiological studies for screening of human immunodeficiency virus HIV/AIDS antibody to help curb the spread of the virus in recent studies. It reduces the cost of testing and more importantly it offers a feasible way to lower the misclassifications associated with labeling samples when imperfect tests are used. Furthermore, misclassifications can be reduced by employing a re-testing design in a group testing procedure. This study has developed a computational statistical model for classifying a large sample of interest based on a proposed design of group testing with re-testing. This model permits computation of moments on the number of tests and misclassification arising in this design. Simulated data from a multinomial distribution (specifically a trinomial distribution) has been used to illustrate these computations. From our study, it has been established that re-testing reduces misclassifications significantly and more so, it is stable at high rates of probability of incidences as compared to Dorfman procedure although re-testing comes with a cost i.e. increase in the number of tests. Re-testing considered reduces the sensitivity of the testing scheme but at the same time it improves the specificity.

Keywords: Group, Re-test, Specificity, Sensitivity, Multinomial, Misclassifications.

1. INTRODUCTION

Pooling refers to the process of putting together individuals to form a group and then testing the group rather than testing each individual for evidence of the characteristic of interest. Pool testing began during World War II as an economical method of testing blood samples of army inductees in order to detect the presence of infection (Dorfman, 1943). The basic idea in pooling testing is that a test is done on a pool and a good reading indicates that the group contains no defective items and a defective reading indicates the presence of at least one defective. There are two objectives of pool testing: classification of the units of a population as either defective or non-defective (Dorfman, 1943) and estimation of the prevalence of a disease in a population (Sobel and Elashoff, 1975). Pool testing reduces the cost of testing when the prevalence rate is low. This is because if a pool tests negative, it implies all its constituent members are non-defective and hence it is not necessary to test each member of the pool. An algorithm of classifying a population of interest into defective and non-defective when each unit *i* of the population has a different probability p_i of being defective (which is called a generalized binomial group test, GBGT) problem has been studied (Hwang, 1975).

In situation where all the units have the same probability *p* of being defective, the generalized binomial group test problem reduces to a binary pool testing problem which is the Dorfman, (1943) procedure. Hwang (1976) has considered pool testing model in the presence of dilution effect i.e. a pool containing a few defective items may be misidentified as a pool containing no such items, especially when the size of the pool is large.

Pool testing has been used in testing the population for the presence of HIV/AIDs antibody (Kline *et al.*, 1989 and Monzon et al., 1992). The cost effectiveness of pooling algorithm for the objective of identifying individuals with the trait has also been studied using hierarchical procedures (Johnson *et al.,* 1992). In this procedure, each pool that test positive is divided into two equal groups, which are tested, groups that tested positive are further subdivided and tested and so on. This work has been extended by considering pooling algorithms when there are errors and showed that some of these algorithms can reduce the error rates of the screening procedures (the false positives and false negatives) compared to individual testing (Litvak *et al.,* 1994). Computational statistics has been used in pool testing to compute the statistical measures when perfect and imperfect tests are used has been considered (Nyongesa and Syaywa, 2011; Nyongesa and Syaywa, 2010; Tamba et al., 2012).

The applications of pool testing are vast (Sobel and Groll, 1966). Pooling has been applied industries (Mundel, 1984), and recently it has been applied in screening the population for the presence of HIV antibody (Kline *et al.*, 1989 and Manzon *et al.*, 1992). Pool testing has been used in screening HIV antibody to help curb the further spread of the virus (Litvak *et al.*, 1994). It has been established that pooling offers a feasible way to lower the error rates associated with labeling samples when screening low risk HIV population. For instance, given the limited precision of the available test kits, it has been shown that screening pooled sera can be used to reduce the probability that a sample labeled negative in fact has antibodies since each test has a certain sensitivity and specificity.

In this study, we discuss the computation of moments on number of tests and misclassifications based on a proposed group testing with re-testing strategy. To the authors knowledge no article has appeared in the literature of group-testing based on Monzon et al. (1992) design that has discussed the procedure in computational aspect. The rest of the paper is arranged as follows: Section 2 discusses the re-testing scheme whereas the model of this study is discussed in Section 3. The central moments and the number of tests are discussed in Section 4. Misclassifications in the group testing with re-testing scheme are discussed in Section 5. Section 6 provides the discussion and conclusion of the study.

2. THE RE-TESTING SCHEME

Suppose we have a large population; say of size $N \to \infty$ with the purpose of testing the constituent members to detect the defective ones. To achieve this, Dorfman (1943) group testing procedure is employed as follows: subdivide *N* into *n* portions herein referred to as groups each of equal size say *k*. each of the n constructed groups is subjected to testing. Since the test kits employed in the study are not perfect, we employ repeated testing to recover some lost sensitivity (c.f Nyongesa 2011). In this testing strategy, if a group tests negative it is dropped from further investigation while if tests positive, it is re-tested and if it tests positive on the duplicate test, its constituent members are tested to identify the defective members. The testing procedure is represented in Figure 1.

Groups

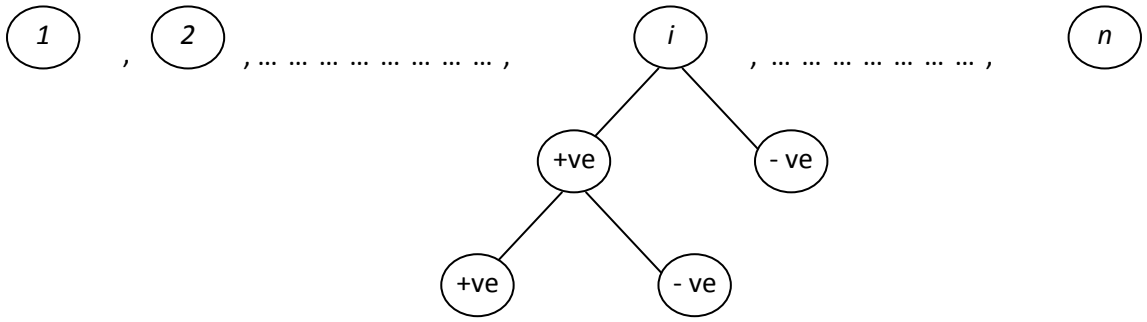

FIGURE 1: Group Testing with Re-testing Strategy.

The figure shows the *n* constructed groups and the test result on the i^{th} group, for i=1, 2, …, *n*. The analysis in this study will require the following indicator functions:

Let

$$T_i = \begin{cases} 1; & if\ the\ i^{th}\ group\ tests\ positive\ on\ the\ test\ kit \\ 0; & otherwise \end{cases}$$

$$T'_i = \begin{cases} 1; & if\ the\ i^{th}\ group\ test\ is\ positive\ on\ the\ re-test\ on\ the\ test-kit \\ 0; & otherwise \end{cases}$$

$$D_i = \begin{cases} 1; & if\ the\ i^{th}\ group\ is\ positive \\ 0; & otherwise \end{cases}$$

$$T_{ij} = \begin{cases} 1; & if\ the\ j^{th}\ individual\ in\ an\ i^{th}\ group\ tests\ positive\ on\ the\ test\ kit \\ 0; & otherwise \end{cases}$$

and

$$\delta_{ij} = \begin{cases} 1; & if\ the\ j^{th}\ individual\ in\ the\ i^{th}\ group\ is\ positive\ with\ probability\ p \\ 0; & otherwise \end{cases}$$

The indicator functions provided above are essential in the subsequent developments. The observations of the constituent members of the i^{th} group will be represented by $(\delta_{i1}, \delta_{i2}, ..., \delta_{ij}, ..., \delta_{ik})$ or simply $\{\delta_{ij}\}_{j=1}^{k}$. Clearly,

$$\Pr(D_i = 0) = \Pr(\delta_{i1} = 0, \delta_{i2} = 0, ..., \delta_{ij} = 0, ..., \delta_{ik} = 0)$$

by definition. For analysis purposes, we shall assume that the constituent member of a group act independently of each other, hence

$$\Pr(D_i = 0) = (1-p)^k, \text{ where, } p \text{ is the prevalence rate.}$$

3. THE MODEL

From Figure 1, let X_1 be the number of groups that test positive on the initial test, X_2 test negative on the initial test. Let X_{11} and X_{12} be the number of groups that test positive and

negative on the re-test respectively. Then X_{11} and X_{12} are random variables. Utilizing these random variables, we derive the probability of declaring a group as negative on the initial tests; $\pi_1 = \Pr(T_i = 0)$ as,

$$\pi_1 = (1-p)^k \phi + [1-(1-p)^k](1-\eta). \quad (1)$$

where, η is the sensitivity and ϕ is the specificity of the test kits. By sensitivity, we mean the probability of correctly classifying a positive group and individual while specificity means the probability of correctly classifying a negative group and individual. The probability, $\pi_2 = \Pr(T_i = 1, T'_i = 0)$ of declaring a group as negative on re-testing a group initially classified as positive is

$$\pi_2 = \phi(1-\phi)(1-p)^k + \eta(1-\eta)[1-(1-p)^k]. \quad (2)$$

With (1) and (2) at hand one can easily obtain the probability of classifying a group as positive on re-testing a group classified as positive on the initial test i.e. $\pi_3 = \Pr(T_i = 1, T'_i = 1)$ as,

$$\pi_3 = 1 - \pi_2 - \pi_1 \quad (3)$$

Equation (3) can be deduced as,

$$\pi_3 = (1-\phi)^2(1-p)^k + \eta^2[1-(1-p)^k].$$

The probabilities $\pi_1, \pi_{2,} and\ \pi_3$ can be used to define the model of group testing with re-testing. The joint probability distribution of X_2, X_{12} and X_{11} is a multinomial model given by

$$f_{X_2 X_{11} X_{12}}(x_2, x_{11}, x_{12}) = \binom{n}{x_2 x_{11} x_{12}} \pi_1^{x_2} \pi_2^{x_{11}} (1-\pi_1-\pi_2)^{n-x_2-x_{11}}. \quad (4)$$

In this retesting strategy, π_2 is regarded as a measure that filters out negative groups from the groups that were initially classified as positive. The covariance matrix of the random variables X_2, X_{11} and X_{12} is

$$Cov(X_2, X_{12}, X_{11}) = \begin{pmatrix} n\pi_1(1-\pi_1) & -n\pi_1\pi_2 & -n\pi_1\pi_3 \\ -n\pi_1\pi_2 & n\pi_2(1-\pi_2) & -n\pi_2\pi_3 \\ -n\pi_1\pi_3 & -n\pi_2\pi_3 & n\pi_3(1-\pi_3) \end{pmatrix}.$$

(c.f Nyongesa, 2011).

4. CENTRAL MOMENTS AND THE NUMBER OF TESTS

In this section, we provide the number of tests based on the proposed re-testing design as presented in Figure 1 described by Model (4) and the central moments of the number of tests. Let *Z* be the number of tests in this proposed group testing scheme. Therefore,

$$Z = n + X_1 + kX_{11}. \quad (5)$$

where, X_1 is the number of groups that test positive on the initial test. To obtain the expected number of tests and the variance of the number of tests, we employ the martingale theory. The expected number of tests is

$$E\left[Z\right]=1+n+n\pi+kn\pi_3. \tag{6}$$

where, π is $\Pr(T_i=1)$, the probability of classifying a group as positive by the initial test and is given by

$$\pi=\eta\left[1-(1-p)^k\right]+(1-\phi)(1-p)^k \tag{7}$$

The variance of the number of test is

$$Var(Z)=n\pi\left(1-\pi\right)+2kn\pi_3\left(1-\pi\right)+k^2n\pi_3(1-\pi_3) \tag{8}$$

from which, the standard deviation is $\sqrt{n\pi\left(1-\pi\right)+2kn\pi_3\left(1-\pi\right)+k^2n\pi_3(1-\pi_3)}$. Next, we consider the skewness and kurtosis of the number of tests. In general, using the theory of moment generating function of a multinomial distribution the central moments of X_1 and X_{11} can be obtained as follows:

$$\begin{aligned}
E\left(X_1-n\pi\right)^2&=n\pi\left(1-n\pi\right)\\
E\left(X_1-n\pi\right)^3&=n\pi\left[1-3\pi+3n\pi+2\pi^2-3n\pi^2+n^2\pi^2\right]\\
E\left(X_1-n\pi\right)^4&=n\pi\left\{1-7\pi+7n\pi+12\pi^2-18n\pi^2+6\left(n\pi\right)^2\right.\\
&\qquad\left.-6\pi^3+11n\pi^3-6n^2\pi^3+\left(n\pi\right)^3\right\}
\end{aligned} \tag{9}$$

Similarly the central moments for X_{11} are given by

$$\begin{aligned}
E\left(X_{11}-n\pi_3\right)^2&=n\pi_3\left(1-n\pi_3\right)\\
E\left(X_{11}-n\pi_3\right)^3&=n\pi_3\left[1-3\pi_3+3n\pi_3+2{\pi_3}^2-3n{\pi_3}^2+\left(n\pi_3\right)^2\right]\\
E\left(X_{11}-n\pi_3\right)^4&=n\pi_3\left\{1-7\pi_3+7n\pi_3+12{\pi_3}^2-18n{\pi_3}^2+6\left(n\pi_3\right)^2\right.\\
&\qquad\left.-6{\pi_3}^3+11n{\pi_3}^3-6n^2{\pi_3}^3+\left(n\pi_3\right)^3\right\}
\end{aligned} \tag{10}$$

With the aid of Equations (9) and (10), we derive skewness and kurtosis of the random variable Z. First, by definition, the skewness, γ_1 of Z is

$$\gamma_1=\frac{a}{b} \tag{11}$$

where,

$$a = E\left\{X_1 - n\pi\right\}^3 \left\{1 + 3k\frac{\pi_3}{\pi} + 3k^2\left(\frac{\pi_3}{\pi}\right)^2\right\} + E\left\{X_1 - n\pi\right\}^2 \left\{3k^2\frac{\pi_3}{\pi}\left(1 - \frac{\pi_3}{\pi}\right)\right\} +$$

$$k^3 E\left\{X_{11} - n\pi_3\right\}^3$$

$$b = \left\{n\pi\left(1-\pi\right) + 2kn\pi_3\left(1-\pi\right) + k^2 n\pi_3(1-\pi_3)\right\}^{3/2}$$

Next is the computation of kurtosis, γ_2 and is given by,

$$\gamma_2 = \frac{c}{d} \tag{12}$$

where,

$$c = E\left\{X_1 - n\pi\right\}^4 \left\{1 + 4k\frac{\pi_3}{\pi} + 6\left(k\frac{\pi_3}{\pi}\right)^2 + 4\left(k\frac{\pi_3}{\pi}\right)^3\right\} +$$

$$E\left\{X_1 - n\pi\right\}^3 \left\{6k^2\frac{\pi_3}{\pi}\left(1 - \frac{\pi_3}{\pi}\right) + 12k^3\frac{\pi_3^2}{\pi^2}\left(1 - \frac{\pi_3}{\pi}\right) + 4\left(k\frac{\pi_3}{\pi}\right)^3\right\} +$$

$$E\left\{X_1 - n\pi\right\}^2 \left\{6k^2 n\pi_3\left(1 - \frac{\pi_3}{\pi}\right) + 4k^3\frac{\pi_3}{\pi}\left(1 - \frac{\pi_3^2}{\pi^2}\right)\right\} + k^4 E\left\{X_{11} - n\pi_3\right\}^4$$

$$d = \left\{n\pi\left(1-\pi\right) + 2kn\pi_3\left(1-\pi\right) + k^2 n\pi_3(1-\pi_3)\right\}^2.$$

5. MISCLASSIFICATIONS IN THE GROUP TESTING STRATEGY WITH RE-TESTING

In this study, we modeled the model of interest with errors of inspection through sensitivity and specificity of the test kits. Thus allowing errors in inspection, misclassifications are bound to arise and this is the subject of this section. There are two possible misclassifications namely: false negative and false positive. A false- positive refers to a non- defective item being classified as defective whereas a false- negative means that a defective item is classified as non-defective. First, we derive sensitivity of the re-testing scheme, Sensitivity= $\Pr(T_i = 1, T'_i = 1, T_{ij} = 1 \mid \delta_{ij} = 1)$ and by the assumption of independence in the tests used, we have the sensitivity as

$$Sensitivity = \eta^3. \tag{13}$$

Thus the false positive probability of the scheme is

$$f_p = 1 - \eta^3. \tag{14}$$

Note that $\eta^3 < \eta$ since $0 \le \eta \le 1$ therefore the re-testing procedure lowers the sensitivity, thus this calls for re-testing of groups that were classified as negative in order to recover some lost sensitivity. Similarly $\eta^3 \le \eta^2$, and hence the sensitivity of this re-testing procedure is less than that of pool testing strategy without re-testing (cf Tamba et al., 2012). Now the specificity of this testing procedure is given by

$$\begin{aligned} Specificity &= \Pr(T_i = 0 \mid D_{ij} = 0) + \Pr(T_i = 1, T_i^{'} = 0 \mid D_{ij} = 0) + \\ &\quad \Pr(T_i = 1, T_i^{'} = 1, T_{ij} = 0 \mid D_{ij} = 0). \\ &= 1 - (1-\phi)\left\{(1-\phi)^2(1-p)^{k-1} + \eta^2(1-(1-p)^{k-1})\right\}. \end{aligned} \quad (15)$$

This design improves the specificity as compared to the Dorfman (1943) Model. One minus the specificity of the testing scheme yields the probability of false negative as

$$f_n = (1-\phi)\left\{(1-\phi)^2(1-p)^{k-1} + \eta^2(1-(1-p)^{k-1})\right\}. \quad (16)$$

To investigate the performance of this design we shall utilize Equation (14) and (16), in our computations.

6. RESULTS

To this end, we have presented formulas that can be used to compute the central moments of the number of tests in group testing with re-testing scheme. We illustrate the procedure by computing the central moment measures for various sensitivity and specificity.

Characteristics	P=0.01				P=0.05				P=0.1			
	μ	σ	γ_1	γ_2	μ	σ	γ_1	γ_2	μ	σ	γ_1	γ_2
Number of non- defective groups on the 1st test	8.9470	1.0005	-0.888	3.3359	5.9710	1.4798	-0.205	2.8798	3.5710	1.5288	0.1718	2.7379
Number of non-defective groups on re-test	0.0980	0.3250	0.9265	12.259	0.0900	0.3286	3.0334	12.051	0.1050	0.2986	3.7687	11.722
Number of defective groups on the re- test	0.9550	0.9428	0.9265	3.6347	3.9390	1.4696	0.2439	2.9038	6.3240	1.5490	-0.163	2.6714
Number of group tests	11.053	-	-	-	14.029	-	-	-	16.429	-	-	-
Total number of individual tests	9.550	9.428	0.9265	3.6347	39.390	14.696	0.2439	2.9038	63.240	15.490	-0.163	2.6714
Total number of tests	21.576	9.428	0.9265	3.6347	54.419	14.696	0.2439	2.9038	80.669	15.490	-0.163	2.6714
Total testing cost	21.576	9.428	0.9265	3.6347	54.419	14.696	0.2439	2.9038	80.669	15.490	-0.163	2.6714
Percentage savings	78.424	9.428	0.9265	3.6347	45.581	14.696	0.2439	2.9038	19.331	15.490	-0.163	2.6714

TABLE1: Various characteristics along with relative savings for group testing with retesting strategy with 1000 runs for N =100, k=10, $\eta = \phi = 99\%$.

Characteristics	P=0.01				P=0.05				P=0.1			
	μ	σ	γ_1	γ_2	μ	σ	γ_1	γ_2	μ	σ	γ_1	γ_2
Number of non- defective groups on the 1st test	20.275	1.9973	-0.514	3.2247	9.0480	2.2746	0.1715	3.0086	3.1950	1.6644	0.4933	3.0536
Number of non-defective groups on re-test	0.241	0.5119	1.7600	7.3919	0.2340	0.4607	2.0780	7.7788	0.2460	0.4865	1.9591	5.5551
Number of defective groups on the re- test	4.4840	1.9387	0.5252	3.2351	15.718	2.2820	-.1354	2.8377	21.559	1.7219	-.4776	3.0860
Number of group tests	29.725	-	-	-	40.952	-	-	-	46.805	-	-	-
Total number of individual tests	89.680	39.740	0.5252	3.2351	314.36	45.640	-.1354	2.8377	431.18	34.438	-.4776	3.0860
Total number of tests	120.41	39.740	0.5252	3.2351	356.31	45.640	-.1354	2.8377	478.99	34.438	-.4776	3.0860
Total testing cost	24.082	7.9480	0.5252	3.2351	71.262	9.1280	-.1354	2.8377	95.798	6.8876	-.4776	3.0860
Percentage savings	75.918	7.9480	0.5252	3.2351	28.738	9.1280	-.1354	2.8377	4.202	6.8876	-.4776	3.0860

TABLE 2: Various characteristics along with relative savings for group testing with retesting strategy with 1000 runs for *N* =500, k=20, $\eta = \phi = 99\%$.

Characteristics	P=0.01				P=0.05				P=0.1			
	μ	σ	γ_1	γ_2	μ	σ	γ_1	γ_2	μ	σ	γ_1	γ_2
Number of non- defective groups on the 1st test	8.6440	1.1700	-.5423	2.9606	5.8830	1.5252	0.0222	2.8162	3.6320	1.5477	0.1486	2.8169
Number of non-defective groups on re-test	0.4970	0.6982	1.2849	4.5230	0.4490	0.6842	1.1915	4.1657	0.4260	0.6766	1.2065	4.8986
Number of defective groups on the re- test	0.8590	0.9300	0.8361	3.3976	3.6680	1.5228	0.0719	3.0232	5.9420	1.5957	-.1297	2.8060
Number of group tests	11.356	-	-	-	14.117	-	-	-	16.368	-	-	-
Total number of individual tests	8.590	9.300	0.8361	3.3976	36.680	15.228	0.0719	3.0232	59.420	15.957	-.1297	2.8060
Total number of tests	20.946	9.300	0.8361	3.3976	51.797	15.228	0.0719	3.0232	76.788	15.957	-.1297	2.8060
Total testing cost	20.946	9.300	0.8361	3.3976	51.797	15.228	0.0719	3.0232	76.788	15.957	-.1297	2.8060
Percentage savings	79.054	9.300	0.8361	3.3976	48.203	15.228	0.0719	3.0232	23.212	15.957	-.1297	2.8060

TABLE 3: Various characteristics along with relative savings for group testing with retesting strategy with 1000 runs for *N* =100, k=10, $\eta = \phi = 95\%$.

Characteristics	P=0.01				P=0.05				P=0.1			
	μ	σ	γ_1	γ_2	μ	σ	γ_1	γ_2	μ	σ	γ_1	γ_2
Number of non- defective groups on the 1st test	19.777	2.0304	-.1293	2.8915	9.3350	2.4371	0.0460	2.8004	4.0630	1.8311	0.3887	3.2041
Number of non-defective groups on re-test	1.1230	1.0704	0.9512	2.8250	1.1620	1.0599	0.8855	3.0411	1.2420	1.0706	0.8703	3.3969
Number of defective groups on the re- test	4.1000	1.8146	0.3253	2.9687	14.503	2.5441	0.0251	2.8278	19.695	1.9869	-.3472	3.1641
Number of group tests	30.223	-	-	-	40.665	-	-	-	45.937	-	-	-
Total number of individual tests	82.000	36.292	0.3253	2.9687	290.06	50.882	0.0251	2.8278	393.90	39.738	-.3472	3.1641
Total number of tests	113.22	36.292	0.3253	2.9687	331.73	50.882	0.0251	2.8278	440.84	39.738	-.3472	3.1641
Total testing cost	22.645	7.2584	0.3253	2.9687	66.345	10.176	0.0251	2.8278	88.167	7.9476	-.3472	3.1641
Percentage savings	77.355	7.2584	0.3253	2.9687	33.655	10.176	0.0251	2.8278	11.833	7.9476	-.3472	3.1641

TABLE 4: Various characteristics along with relative savings for group testing with retesting strategy with 1000 runs for *N* =500, k=20, $\eta = \phi = 95\%$.

Probability, p	*N* =100, k=10				*N* =500,k=20				*N* =1000,k=20			
	μ	σ	γ_1	γ_2	μ	σ	γ_1	γ_2	μ	σ	γ_1	γ_2
0.01	0.0298	0.1699	5.5352	28.6422	0.1427	0.3721	2.5277	5.9728	0.2913	0.5317	1.7691	2.9258
0.02	0.0602	0.2416	3.8927	14.1656	0.2926	0.5328	1.7652	2.9131	0.5893	0.7562	1.2439	1.4465
0.03	0.0870	0.2905	3.2375	9.7984	0.4428	0.6555	1.4350	1.9250	0.8841	0.9262	1.0156	0.9642
0.04	0.1188	0.3396	2.7700	7.1731	0.5861	0.7541	1.2472	1.4542	1.1762	1.0683	0.8804	0.7247
0.05	0.1435	0.3732	2.5206	5.9395	0.7344	0.8442	1.1142	1.1606	1.4654	1.1924	0.7888	0.5817
0.1	0.2930	0.5332	1.7642	2.9095	1.4624	1.1912	0.7896	0.5829	2.9246	1.6846	0.5584	0.2915
0.15	0.4361	0.6505	1.4460	1.9546	2.1784	1.4539	0.6470	0.3913	4.3606	2.0570	0.4573	0.1955

TABLE 5: Number of false positives in the group testing with retesting strategy for different group sizes for $\eta = \phi = 99\%$.

Probability, p	*N* =100,k=10				*N* =500,k=20				*N* =1000,k=20			
	μ	σ	γ_1	γ_2	μ	σ	γ_1	γ_2	μ	σ	γ_1	γ_2
0.01	0.1623	0.3730	1.9160	1.9137	0.8085	0.8326	0.8585	0.3841	1.6098	1.1748	0.6084	0.1929
0.02	0.2988	0.5061	1.4121	1.0395	1.4606	1.1191	0.6387	0.2126	2.9355	1.5865	0.4505	0.1058
0.03	0.4168	0.5978	1.1957	0.7453	2.1019	1.3424	0.5324	0.1478	4.2150	1.9010	0.3760	0.0737
0.04	0.5555	0.6901	1.0357	0.5591	2.7227	1.5279	0.4678	0.1141	5.4826	2.1681	0.3297	0.0567
0.05	0.6822	0.7648	0.9346	0.4553	3.3942	1.7059	0.4190	0.0915	6.7955	2.4138	0.2961	0.0457
0.1	1.3217	1.0645	0.6714	0.2350	6.6171	2.3819	0.3001	0.0469	13.2175	3.3664	0.2123	0.0235
0.15	1.9276	1.2856	0.5560	0.1611	9.8323	2.9034	0.2462	0.0316	19.5610	4.0953	0.1745	0.0159

TABLE 6: Number of false positives in the group testing with retesting strategy for different group sizes for $\eta = \phi = 95\%$.

Probability, p	*N* =100, k=10				*N* =500,k=20				*N* =1000,k=20			
	μ	σ	γ_1	γ_2	μ	σ	γ_1	γ_2	μ	σ	γ_1	γ_2
0.01	0.0840	0.2897	3.4459	11.8539	0.8439	0.9179	1.0858	1.1749	1.6878	1.2981	0.7678	0.5874
0.02	0.1599	0.3995	2.4949	6.2043	1.5314	1.2356	0.8043	0.6428	3.0626	1.7473	0.5687	0.3214
0.03	0.2283	0.4772	2.0856	4.3293	2.0905	1.4428	0.6871	0.4681	4.1797	2.0400	0.4860	0.2341
0.04	0.2897	0.5375	1.8494	3.3994	2.5414	1.5900	0.6223	0.3831	5.0812	2.2482	0.4401	0.1916
0.05	0.3444	0.5858	1.6947	2.8510	2.9018	1.6983	0.5816	0.3341	5.8008	2.4011	0.4114	0.1671
0.1	0.5419	0.7339	1.3461	1.7900	3.8251	1.9475	0.5048	0.2504	7.6480	2.7538	0.3570	0.1252
0.15	0.6433	0.7990	1.2327	1.4961	3.9910	1.9884	0.4935	0.2389	7.9783	2.8113	0.3490	0.1195

TABLE 7: Number of false negatives in the group testing with retesting strategy for different group sizes for $\eta = \phi = 99\%$.

Probability, p	*N* =100, k=10				*N* =500,k=20				*N* =1000,k-20			
	μ	σ	γ_1	γ_2	μ	σ	γ_1	γ_2	μ	σ	γ_1	γ_2
0.01	0.3971	0.6289	1.5773	2.4675	3.9282	1.9741	0.4985	0.2445	7.8566	2.7918	0.3525	0.1222
0.02	0.7447	0.8597	1.1455	1.2918	7.0875	2.6429	0.3674	0.1309	14.1722	3.7373	0.2598	0.0655
0.03	1.0594	1.0236	0.9556	0.8925	9.6540	3.0760	0.3122	0.0933	19.3125	4.3507	0.2207	0.0466
0.04	1.3424	1.1505	0.8449	0.6931	11.7331	3.3833	0.2811	0.0749	23.4679	4.7849	0.1988	0.0374
0.05	1.5948	1.2522	0.7718	0.5747	13.4060	3.6095	0.2615	0.0642	26.8125	5.1047	0.1849	0.0321
0.1	2.5157	1.5640	0.6040	0.3428	17.7311	4.1278	0.2233	0.0455	35.4220	5.8343	0.1580	0.0228
0.15	2.9931	1.6998	0.5475	0.2766	18.5761	4.2162	0.2168	0.0423	37.1598	5.9632	0.1533	0.0212

TABLE 8: Number of false negatives in the group testing with retesting strategy for different group sizes for $\eta = \phi = 95\%$.

Remark 1: In all the above tables we have; $\mu = mean, \sigma = s\tan dard\ deviation, \gamma_1 = skewness, \gamma_2 = kurtosis$

7. DISCUSSIONS AND CONCLUSION

This study has presented a computational group testing strategy with re-testing. It has been shown from the results; Tables 1, 2, 3 and 4 that when the group size and prevalence rate are small, significant savings are realized. This is an empirical result since group testing is only feasible when the prevalence rate is small otherwise individual testing is preferred. Similarly large groups are prone to increase the dilution effect and hence increase the misclassifications. It has been established that re-testing groups that were initially classified as positive increases the cost of testing however, the false negatives significantly reduces as compared to the Dorfman procedure when imperfect tests are used (Tamba *et al*., 2011). The results in Tables 5, 6, 7 and 8 show that the higher the efficiency of the tests, the lower the misclassifications. This implies that group testing should be carried out when specificity and sensitivity of the testing procedure are high. It has also been noted that this re-testing strategy improves the specificity of the testing procedure making it viable in screening the population for presence of HIV/AIDS. Misclassifications are high when the prevalence rate is high and the efficiency of the test kits is low.

8. REFERENCES

[1] C. L. Tamba, K. L. Nyongesa and J. W. Mwangi. “Computational Pool-Testing Strategy” Egerton University Journal, 11: pp 51-56, 2012.

[2] E. Litvak, , X. M. Tu, and M. Pagano. “Screening for the presence of a disease by grouping sera samples”. Journal of the America statistical Association, 89, pp 424-434, 1994.

[3] F. K. Hwang. “Group testing with a dilution effect”. Biometrika 63, pp 611-613, 1976.

[4] F.K. Hwang. “A Generalized Binomial Group Testing Problem”. Journal of the American Statistical Association, 70, pp 923- 926, 1975.

[5] G. Hepworth and R. Watson. “Debiased estimation of proportions in group testing”. Royal Statistical Society, 58, pp 105–121, 2008.

[6] L. K. Nyongesa. “Dual Estimation of Prevalence and Disease Incidence in Group-Testing Strategy”. Communication in Statistics Theory and Method. Vol. (1): Issue (1), 2010.

[7] L. K. Nyongesa. “Hierarchical Screening with Retesting in a low Prevalence Population”. The Indian Journal of Statistics.66, pp 779-790, 2005.

[8] L. K. Nyongesa. “Multistage group testing procedure (Group screening)”. Communication in Statistics-Simulation and computation, 33, pp 621-637, 2004.

[9] L.K. Nyongesa and J. P.Syaywa. “Block Testing Strategy with Imperfect Tests and its Improved Efficient Testing Model for Donor Blood”. Communication in Statistics-Computational Statistics, 2011.

[10] L.K. Nyongesa, and J. P. Syaywa. “Group Testing with Test Errors Made Easier”. International Journal of Computational Statistics. Volume (1): Issue (1), 2010.

[11] M. Sobel and R. M. Elashoff. “Group-testing with a new goal, Estimation”. Biometrika, 62, 181-193,1975.

[12] M. Sobel, and P. A. Groll. “Binomial Group-Testing with an Unknown Proportion of Defectives”. American Statistical Association and American Society for Quality, 8, pp 631-656, 1966.

[13] N. L. Johnson, S. Kotz and X. Wu. “Inspection errors for attributes in quality control”. London; Chapman and Hall, 1992

[14] O.T. Monzon, F. J. E. Palalin, E. Dimaal, A.M. Balis, C. Samson and S. Mitchel. “Relevance of antibody content and test format in HIV testing of grouped sera”. AIDS, 6, pp 43-48, 1992.

[15] R. B. Hunt, L.L. Ronald and M. R. Jonathan. “A Guide to MATLAB for Beginners and Experienced Users”. Cambridge University Press, pp.101-119, 2004.

[16] R. Dorfman. “The detection of defective members of large population”. Annals of Mathematical Statistics 14, pp 436-440, 1943.

[17] R.L. Kline, T. Bothus, R. Brookmeyer, S. Zeyer, and T. Quinn. “Evaluation of Human Immunodeficiency Virus seroprevalence in population surveys using grouped sera”. Journal of clinical microbiology, 27, pp 1449-1452,1989.

[18] S. Maheswaran, V. Haragopal, and S.N. Pandit. Group-testing using block testing strategy. Journal of statistical planning and inference, 2008.