IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 13, Issue 1 Ver. II (Jan. - Feb. 2017), PP 43-55
www.iosrjournals.org

Multi-Stage Adaptive Pool Testing Model with Test Errors; Improved Efficiency

Okoth Annette W.[1], Nyongesa kennedy L. [2] , Kwach Bonface O.[3]
[1;2] *Department of Mathematics, Masinde Muliro University of Science and Technology*
[3] *Department of Mathematics, Kibabii University*

Abstract: *Testing units one by one for presence or absence of a trait is expensive and time consuming. This study presents a multi-stage adaptive pool testing estimator ^pn of prevalence of a trait in the presence of test errors, since errors in experiments are inevitable. Pool testing is more efficient, less expensive and less time consuming. An increase in the number of stages improves the efficiency of the estimator, hence construction of a multi-stage model. The study made use of the Maximum Likelihood Estimate (MLE) method and Martingale method to obtain the adaptive estimator and Cramer-Rao lower bound method to determine the variance of the constructed estimator. Mat lab and R, statistical softwares were used for Monte-carlo simulation and verification of the model, then analysis and discussion of properties of the constructed estimator in comparison with the non-adaptive estimator in the literature of pool testing done alongside provision of the con dence interval of the estimator. Results have shown that as the number of stages increases, the efficiency of the multi-stage adaptive estimator in the presence of test errors also increases in comparison with the non-adaptive estimator in the presence of test errors. This makes the multi-stage adaptive estimator better than the corresponding non-adaptive estimator in the literature of pool testing.*
Keywords: *Pool testing, Adaptive estimator, Confidence interval*

I. Introduction

Prevalence of defective units in a large population from accurate diagnostic tests is a fundamental risk assessment and management factor. Estimation of defective units one-by-one is inefficient and uneconomical, considering that in a given population only a few individuals may be defective. It is against this background that pool testing comes in handy because it is more effective, less time consuming and less expensive [4]. Pool testing occurs when units from a population are pooled and tested as a group for the presence or absence of a particular trait. It also reduces the Mean Squared Error (MSE) of the estimates, hence it is more efficient, as was established by Sobel and Ellashoff, [11]. There are two forms of pool testing namely

(i) Non-adaptive pool testing scheme

(ii) Adaptive pool testing scheme

1.1 Non-adaptive testing scheme

In this testing scheme, a large population is divided in to n groups which are then subjected to testing [4]. When tested, a group can either test positive or negative and the outcome of the test aids in constructing the non-adaptive model.

1.2 Adaptive testing scheme

In this scheme a population is divided in to n groups, which are partitioned depending on the number of stages to be considered. Predetermined parameters are used to partition the groups and the number of partitioning parameters depends on the number of stages [8]. Partitioned groups are then tested at various stages for the presence or absence of a trait and the results used to construct the adaptive model.

1.3 Introduction of the model

In this study we obtain a multi-stage adaptive estimator $\hat{p}_n$ of prevalence of a trait in the presence of test errors,using the maximum likelihood estimate (MLE) method and investigate its properties. The adaptive testing scheme involves testing groups in stages and updating group sizes from one stage to the next, with the group size at a stage depending on the outcome of the test(s) at the preceding stage(s). That is testing n_1 groups each of size k_1 at stage one; n_2 groups each of size k_2 at stage two; n_3 groups each of size k_3 at stage three and so on; where k_3 depends on both k_1 and k_2 while k_2 depends on k_1. For a general adaptive scheme, at stage i n_i groups each of size k_i, where k_i depends on $k_{i-1}, k_{i-2}, k_{i-3}, ...k_1$ are constructed. The constructed groups are then subjected to testing, where a group yields either a positive or a negative result. The number of groups, n_i is determined before the experiment is carried out while $k_i's$ are sequentially determined as the experiment progresses.

II. Literature Review

Pool testing has been recognized as a sampling scheme that can provide substantial benefits [9]. Early application of pool testing include tests for prevalence of plant virus transmissions by insects [12] and [11] and this was one of the pioneering applications of this concept. In [4] statistical and mathematical concepts of pool testing are introduced and used to estimate the proportion of individuals infected with some disease among the US conscripts. He also derived optimum group sizes assuming that the population was large enough for the application of the binomial model and consequently realized significant savings by reducing the number of tests required. In [11] estimation in the pool testing procedure is discussed. In the subsequent years this concept has had relevant applications in various clinical studies including psychopathology, public health and plant quarantine [1] and [3]. Alternatively, positively pooled samples can be partitioned into relatively smaller subsets there by reducing on cost and effort, which provides obvious motive for pooling samples [8]. In [7] an estimation model based on pool testing with retesting pools that test negative is developed. Pool testing need not only be applied to population where retesting is needed [10], like in identification of disease infected individuals in a human population, but also on other populations with no intentions of retesting the individuals contributing to positive pooled samples. For instance if a bunch of food items is being tested for contamination, there may be no interest in identifying the particular items which are affected. The aim may instead be on estimating the proportion of defective items in a population or deciding that the number of positive pooled samples justifies removing a food product from the market. In another related study, bacteriological testing of egg laying hens of salmonella in Great Britain was carried out using organ cultures pooled five at a time. Individual samples contributing to positive pooled samples are not tested again . A population comprised of birds in a hen house. If the infection was confirmed they were destroyed and compensation paid for the number of birds estimated to be uninfected [10]. In this procedure maximum likelihood estimation is applied to estimate the proportion and Cramer-Rao lower bound method is used to determine the variance of the estimator. In this paper, we present a multi-stage adaptive pool testing model with imperfect tests, that is applicable to real life situations where there is a possibility of errors due to test kits.

III. Model Description Formulation and Analysis

We describe a multi-stage adaptive scheme with imperfect tests as it is the backbone of this study and thereafter perform comparison analysis with other existing estimators, both in the presence and absence of test errors. For a multi-stage adaptive scheme, we set $n_1 = \lambda_1 n$, $n_2 = \lambda_2 n$, $n_3 = \lambda_3 n$, ..., $n_n = (1 - \lambda_1 n - \lambda_2 n - .... - \lambda_{n-1} n)$; where λ_1, λ_2,..... , λ_{n-1} are parameters used to partition the pools; k_2 depends on the outcome at stage 1, k_3 depends on the outcomes at stages 1 and 2 and k_n depends on the outcomes at stages $1, 2, 3, ..., n-1$. Each constructed group at each stage is then subjected to testing, yielding either a positive or negative result. This is shown in Figure 3.2 below:

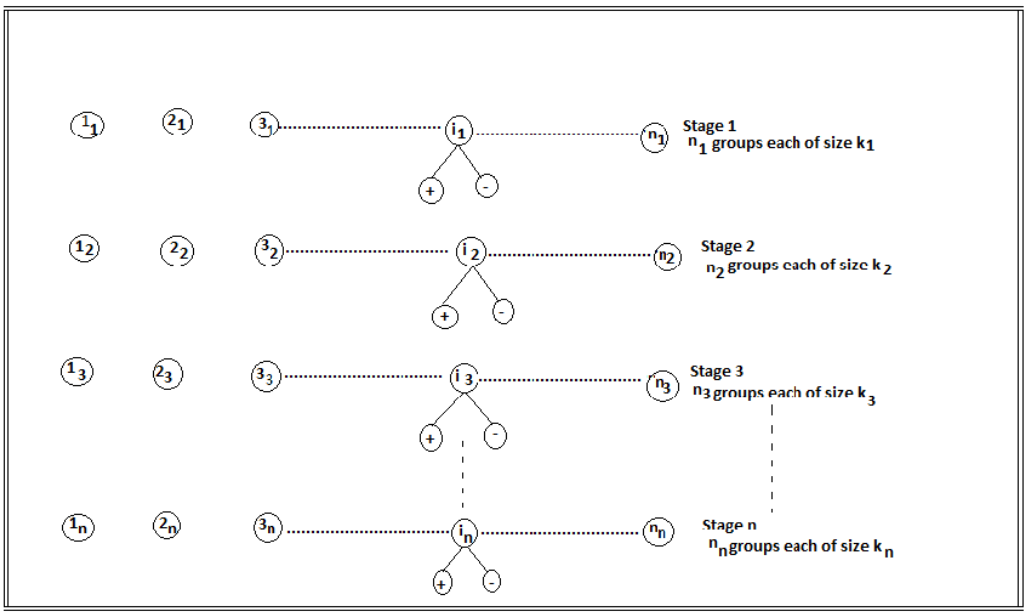

Figure 3.2: Multi-stage adaptive pool testing.

To achieve the construction of the multi-stage adaptive model in the presence of test errors, we consider two stage, three stage and four stage adaptive models in the presence of test errors and there after generalize to obtain the multi-stage model.

3.1 Two stage adaptive model

In this scheme, the population is divided into two sets of groups n_1 and n_2 which are tested in two stages, with n_1 groups tested at stage one and n_2 groups tested at stage two. We set $n_1 = \lambda n$ and $n_2 = (1 - \lambda n)$, where n is the number of groups constructed initially. k_1 which is the group size at stage one is determined by

$$k_1 = argmin_l [Var(\hat{p})]|_{p=p_0}, \tag{1}$$

Suppose X_1 groups test positive on the test at stage-one, then

$$X_1 \sim Binomial(\lambda n, \pi(p)|_{k=k_1}). \tag{2}$$

where λ is the parameter used to partition the pools while $\pi(p) is the probability that a group is defective a$ $\eta[1-(1-p)^k] + (1-\phi)(1-p)^k (3)$ Using this model we obtain the prevalence estimator at stage one as

$$\hat{p}_1 = 1 - \left[\frac{\eta - \frac{X_1}{\lambda n}}{\eta + \phi - 1}\right]^{\frac{1}{k_1}}, \tag{4}$$

The variance of Equation (4) is similar to the variance of the non-adaptive estimator in the absence of test errors, $\hat{p}$ except for K_1 in place of p. This variance is given by

$$\begin{aligned} Var(\hat{p}) &= \frac{\pi(p)(1-\pi(p))}{nk^2(1-p)^{2k-2}(\eta+\phi-1)^2} \\ &= \frac{(1-p)^{2-2k}\pi(p)(1-\pi(p))}{nk^2(\eta+\phi-1)^2} \\ &= \frac{(1-p)^2\pi(p)(1-\pi(p))(1-p)^{-2k}}{nk^2(\eta+\phi-1)^2} \end{aligned} \tag{5}$$

For the estimator at stage two, $\hat{p}_2$, we have λn groups each of size k_1 tested at stage one and $1-\lambda n$ groups each of size k_2 tested at stage two. k_2 is determined by

$$k_2 = argmin_l[Var(\hat{p}_1)]|_{p_1=p}, \tag{6}$$

Suppose that out of the $(1-\lambda)n$ groups each of size k_2 tested at stage two, X_2 groups test positive on the test, then for fixed X_1 we have

$$X_2|X_1 \sim Binomial((1-\lambda)n, \pi_{2|1}(p)) \tag{7}$$

Using this model, the estimator at stage two can be obtained as the solution to

$$\begin{aligned} &\frac{k_1X_1q^{k_1}[(1-\phi)-\eta]}{\eta-(\eta+(1-\phi))q^{k_1}} + \frac{k_2(X_1)X_2q^{k_2(X_1)}[(1-\phi)-\eta]}{\eta-(\eta+(1-\phi))q^{k_2(X_1)}} \\ = &\frac{k_1q^{k_1}(\lambda n-X_1)(\eta+(1-\phi))}{1-[\eta-(\eta+(1-\phi))q^{k_1}]} + \frac{k_2(X_1)q^{k_2(X_1)}[(1-\lambda)n-X_2][\eta+(1-\phi)]}{1-[\eta-(\eta+(1-\phi))q^{k_2(X_1)}]}. \end{aligned} \tag{8}$$

and using cramer-Rao lower bound, its variance is obtained as

$$Var(\hat{p}_2) = \frac{\pi_1(p)\pi_2(p)(1-\pi_1(p))(1-\pi_2(p))}{A}, \tag{9}$$

where A is defined in the appendices.

3.2 Three stage adaptive model

Next we consider the estimator at stage three, $\hat{p}_3$, where we have $\lambda_1 n$ groups each of size k_1 tested at stage one, $\lambda_2 n$ groups each of size k_2 tested at stage two and $1-\lambda_1 n-\lambda_2 n$ groups each of size k_3 tested at stage three. k_3 is determined by

$$k_3 = argmin_l[Var(\hat{p}_2)]|_{p_2=p_1}, \tag{10}$$

If out of the $(1-\lambda_1-\lambda_2)n$ groups each of size k_3 tested at stage three, X_3 groups test positive on the test, then for fixed X_1 *and* X_2 we have

$$X_3|X_1, X_2 \sim Binomial((1-\lambda_1-\lambda_2)n, \pi_{3|1,2}(p)) \tag{11}$$

We use this model to obtain the estimator at stage three as the solution to

$$\begin{aligned} &\frac{k_1X_1q^{k_1}[(1-\phi)-\eta]}{\eta-(\eta+(1-\phi))q^{k_1}} + \frac{k_2(X_1)X_2q^{k_2(X_1)}[(1-\phi)-\eta]}{\eta-(\eta+(1-\phi))q^{k_2(X_1)}} + \frac{k_3(X_1,X_2)X_3q^{k_3(X_1,X_2)}[(1-\phi)-\eta]}{\eta-(\eta+(1-\phi))q^{k_3(X_1,X_2)}} \\ = &\frac{k_1q^{k_1}(\lambda_1 n-X_1)(\eta+(1-\phi))}{1-[\eta-(\eta+(1-\phi))q^{k_1}]} + \frac{k_2(X_1)q^{k_2(X_1)}[\lambda_2 n-X_2][\eta+(1-\phi)]}{1-[\eta-(\eta+(1-\phi))q^{k_2(X_1)}]} \\ &+\frac{k_3(X_1,X_2)q^{k_3(X_1,X_2)}[(1-\lambda_1-\lambda_2)n-X_3][\eta+(1-\phi)]}{1-[\eta-(\eta+(1-\phi))q^{k_3(X_1,X_2)}]} = 0, \end{aligned} \tag{12}$$

and its variance as

$$Var(\hat{p}_3) = \frac{\pi_1(p)\pi_2(p)\pi_3(p)(1-\pi_1(p))(p)(1-\pi_2(p))(p)(1-\pi_3(p))}{nB} \tag{13}$$

where B is defined in appendices.

3.3 Four stage and multi-stage adaptive models

Extending the notion in the above sub-sections further we have estimators at stages four and n given by solutions to

$$\begin{aligned}
&\frac{k_1X_1q^{k_1}[(1-\phi)-\eta]}{\eta-(\eta+(1-\phi))q^{k_1}} + \frac{k_2(X_1)X_2q^{k_2(X_1)}[(1-\phi)-\eta]}{\eta-(\eta+(1-\phi))q^{k_2(X_1)}} \\
&+\frac{k_3(X_1,X_2)X_3q^{k_3(X_1,X_2)}[(1-\phi)-\eta]}{\eta-(\eta+(1-\phi))q^{k_3(X_1,X_2)}} + \frac{k_4(X_1,X_2,X_3)X_4q^{k_4(X_1,X_2,X_3)}[(1-\phi)-\eta]}{\eta-(\eta+(1-\phi))q^{k_4(X_1,X_2,X_3)}} \\
= \;& \frac{k_1q^{k_1}(\lambda_1 n - X_1)(\eta+(1-\phi))}{1-[\eta-(\eta+(1-\phi))q^{k_1}]} + \frac{k_2(X_1)q^{k_2(X_1)}[\lambda_2 n - X_2][\eta+(1-\phi)]}{1-[\eta-(\eta+(1-\phi))q^{k_2(X_1)}]} \\
&+\frac{k_3(X_1,X_2)q^{k_3(X_1,X_2)}[\lambda_3 n - X_3][\eta+(1-\phi)]}{1-[\eta-(\eta+(1-\phi))q^{k_3(X_1,X_2)}]} \\
&+\frac{k_4(X_1,X_2,X_3)q^{k_4(X_1,X_2,X_3)}[(1-\lambda_1-\lambda_2-\lambda_3)n - X_4][\eta+(1-\phi)]}{1-[\eta-(\eta+(1-\phi))q^{k_4(X_1,X_2,X_3)}]} = 0
\end{aligned} \tag{14}$$

and

$$\begin{aligned}
&\frac{k_1X_1q^{k_1}[(1-\phi)-\eta]}{\eta-(\eta+(1-\phi))q^{k_1}} + \frac{k_2(X_1)X_2q^{k_2(X_1)}[(1-\phi)-\eta]}{\eta-(\eta+(1-\phi))q^{k_2(X_1)}}\cdots \\
&+\frac{k_n(X_1,...,X_(n-1))X_nq^{k_n(X_1,...,X_(n-1))}[(1-\phi)-\eta]}{\eta-(\eta+(1-\phi))q^{k_n(X_1,...,X_(n-1))}} \\
= \;& \frac{k_1q^{k_1}(\lambda_1 n - X_1)(\eta+(1-\phi))}{1-[\eta-(\eta+(1-\phi))q^{k_1}]} + \frac{k_2(X_1)q^{k_2(X_1)}[\lambda_2 n - X_2][\eta+(1-\phi)]}{1-[\eta-(\eta+(1-\phi))q^{k_2(X_1)}]}\cdots \\
&+\frac{k_n(X_1,...,X_(n-1))q^{k_n(X_1,...,X_(n-1))}[(1-\lambda_1-\lambda_2)n - X_4][\eta+(1-\phi)]}{1-[\eta-(\eta+(1-\phi))q^{k_n(X_1,...,X_(n-1))}]} = 0,
\end{aligned} \tag{15}$$

respectively. Using Cramer-Rao lower bound method their variances are obtained as

$$Var(\hat{p}_4) = \frac{\pi_1(p)\pi_2(p)\pi_3(p)\pi_4(p)(1-\pi_1(p))(p)(1-\pi_2(p))(p)(1-\pi_3(p))(1-\pi_4(p))}{C} \tag{16}$$

and

$$Var(\hat{p}_n) = \frac{\pi_1(p)\pi_2(p)...\pi_n(p)(1-\pi_1(p))(p)(1-\pi_2(p))(p)...(1-\pi_n(p))}{D} \tag{17}$$

where C and D are given in the appendices respectively.

3.4 Confidence Interval(CI) of $\hat{p}_n$

Next we provide the confidence interval for our multi-stage estimator, $\hat{p}_n$. This confidence interval is given by

$$\hat{p}_n \pm Z_{\frac{\alpha}{2}}\sqrt{var(\hat{p}_n)}, \tag{18}$$

where $Z_{\frac{\alpha}{2}} \sim Normal(0,1)$. and $\hat{p}_n$ and $var(\hat{p}_n)$ are provided by the solution to 15 and Equation 17 respectively. It follows from Equation 18 that

$$p \in [\hat{p}_n - Z_{\frac{\alpha}{2}}\sqrt{var(\hat{p}_n)}, \hat{p}_n + Z_{\frac{\alpha}{2}}\sqrt{var(\hat{p}_n)}]$$

and by the law of Central Limit Theorem (CLT) we have

$$\sqrt{n}(\hat{p}_n - p) \xrightarrow{l} Normal(0, \sqrt{var(\hat{p}_n)})$$

or

$$\sqrt{n}\frac{(\hat{p}_n - p)}{\sqrt{var(\hat{p}_n)}} \xrightarrow{l} Normal(0, 1).$$

IV. Discussion of Results, Conclusion and Recommendations

In this section we discuss the results as provided by Tables 4.1, 4.2 and 4.3 and Figures 4.1, 4.2 and 4.3. The highlights of the results will enable us make a detailed conclusion to this study.

4.1 Discussion

Here we highlight our findings in this study. We estimated prevalence, p of a trait using the Multi-stage adaptive pool testing scheme. We accomplished this by employing the Maximum Likelihood Estimate (MLE) procedure. For us to recommend the suitability of the Multi-stage adaptive estimator, it would be in order to first compare with the non-adaptive estimator in the presence of test errors, as advanced by [7]; and then do an inter-stage comparison. Our measure of comparison herein is the computation of Asymptotic Relative Efficiency (ARE) values for different values of η and ϕ at various stages. For simplicity of comparison and understanding, ARE values were computed for stages two to four. Upon careful analysis of the estimators at these stages, we had a good basis to make a generalization about the multi-stage estimator in the presence of test errors.

4.1.1 Comparing the adaptive estimators with the non-adaptive estimator in the presence of test errors

The ARE values for stages two, three and four were obtained by dividing Equations (5) by (9), (13) and (16) respectively. Upon simplification we obtained

$$ARE_{\hat{p}_2} = \frac{H}{\pi_1(p)\pi_2(p)(1-\pi_1(p))(1-\pi_2(p))(1-p)^{2k}}, \tag{19}$$

where H is defined in the appendices. Similarly, the Asymptotic Relative efficiencies of $\hat{p}_3$ and $\hat{p}_4$ are given as

$$ARE_{\hat{p}_3} = \frac{J}{\pi_1(p)\pi_2(p)\pi_3(1-\pi_1(p))(1-\pi_2(p))(1-\pi_3(p))(1-p)^{2k}}, \tag{20}$$

and

$$ARE_{\hat{p}_4} = \frac{L}{\pi_1(p)\pi_2(p)(1-\pi_1(p))(1-\pi_2(p))(1-p)^{2k}}, \tag{21}$$

respectively, where J and L are defined in the appendices. respectively. Using these Equations and R-Gui software Tables 4.1, 4.2 and 4.3 were generated.

p	$\eta = \phi = 0.99$	$\eta = \phi = 0.98$	$\eta = \phi = 0.97$	$\eta = \phi = 0.96$	$\eta = \phi = 0.90$
0.1	11.7739	11.8024	11.8304	11.8577	12.0095
0.2	6.1184	6.4223	6.7152	6.9974	8.4904
0.3	3.6109	4.2835	4.8918	5.4448	7.9386
0.4	3.0823	4.3658	5.3725	6.1880	9.0662
0.5	4.4239	6.3134	7.4865	8.3030	10.5715
0.6	7.3875	9.0611	9.9042	10.4289	11.6667
0.7	10.2335	11.1440	11.5420	11.7672	12.2284
0.8	11.9115	12.1858	12.2856	12.3372	12.4329
0.9	12.4422	12.4616	12.4681	12.4713	12.4772

Table 4.1: ARE values of $\hat{p}_2$ relative to $\hat{p}$ for specified p, η and ϕ

p	$\eta = \phi = 0.99$	$\eta = \phi = 0.98$	$\eta = \phi = 0.97$	$\eta = \phi = 0.96$	$\eta = \phi = 0.90$
0.1	22.7824	22.8485	22.9133	22.9766	23.3267
0.2	15.3291	15.8547	16.3494	16.8157	19.1305
0.3	11.6186	13.0352	14.2083	15.2006	19.0300
0.4	11.6141	14.2989	16.0410	17.2898	20.9478
0.5	15.4718	18.2260	19.6402	20.5355	22.7513
0.6	200354	21.6224	22.3539	22.7912	23.7781
0.7	22.7669	23.4427	23.7314	23.8933	24.2221
0.8	24.0129	24.0000	24.2679	24.3030	24.3680
0.9	24.3747	24.3878	24.3920	24.3942	24.3981

Table 4.2: ARE values of $\hat{p}_3$ relative to $\hat{p}$ for specified p, η and ϕ

p	$\eta = \phi = 0.99$	$\eta = \phi = 0.98$	$\eta = \phi = 0.97$	$\eta = \phi = 0.96$	$\eta = \phi = 0.90$
0.1	43.4486	43.5040	43.5580	43.6109	43.9034
0.2	37.5497	37.9749	38.3745	38.7509	40.6123
0.3	34.5777	35.7352	36.6908	37.4967	40.5809
0.4	34.6078	36.8073	38.2250	39.2350	42.1479
0.5	37.8059	40.0398	41.1693	41.8752	43.5804
0.6	41.5386	42.7721	43.3239	43.6477	44.3618
0.7	43.6724	44.1484	44.3474	44.4580	44.6805
0.8	44.5490	44.6706	44.7145	44.7373	44.7793
0.9	44.7841	44.7922	44.7950	44.7967	44.7988

Table 4.3: ARE values of $\hat{p}_4$ relative to $\hat{p}$ for specified p, η and ϕ

Tables 4.1, 4.2 and 4.3 provide generated ARE values for given p, η and ϕ. It is evident from the tables that ARE values are high across all stages. That is all the ARE values are greater than one, meaning all the adaptive estimators analysed in this study, $\hat{p}_2$, $\hat{p}_3$ and $\hat{p}_4$ are more efficient than the non-adaptive estimator in the presence of test errors. A closer look at ARE values reveals that ARE is high when p is small and decreases as p increases, attaining the minimum at $p = 0.3$ across the board, except for $\eta = \phi = 0.99$ where the minimum is attained at $p = 0.4$. The ARE again improves as p moves away from 0.3 for $\eta, \phi < 0.99$. A similar scenario is observed for the case of $\eta = \phi = 0.99$ where ARE improves as p moves away from 0.4. It is also clear from the tables that ARE values increase with increase in the number of stages; the adaptive estimator at stage two having the lowest ARE values while the estimator at stage four has the highest ARE values. This is an important pointer to the fact that the adaptive testing scheme gets better as the number of stages increases. To depict these observations graphically, see Figures 4.1, 4.2 and 4.3

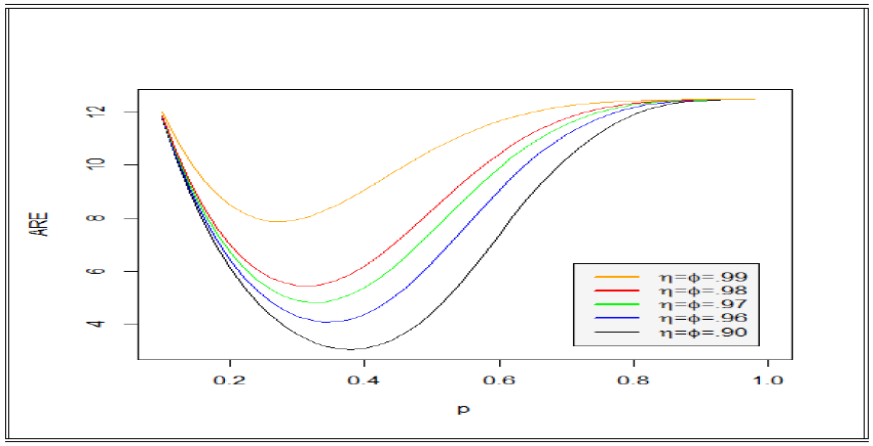

Figure 4.1: ARE of $\hat{p}_2$ vs probability, p

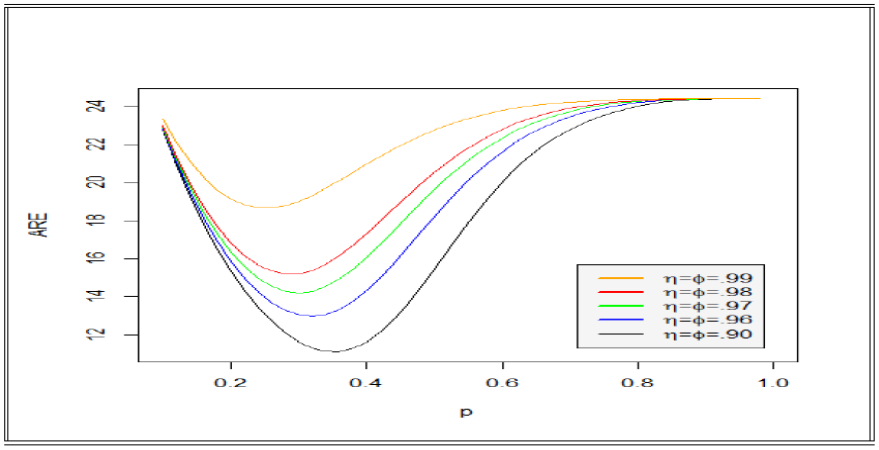

Figure 4.2: ARE of $\hat{p}_3$ vs probability, p

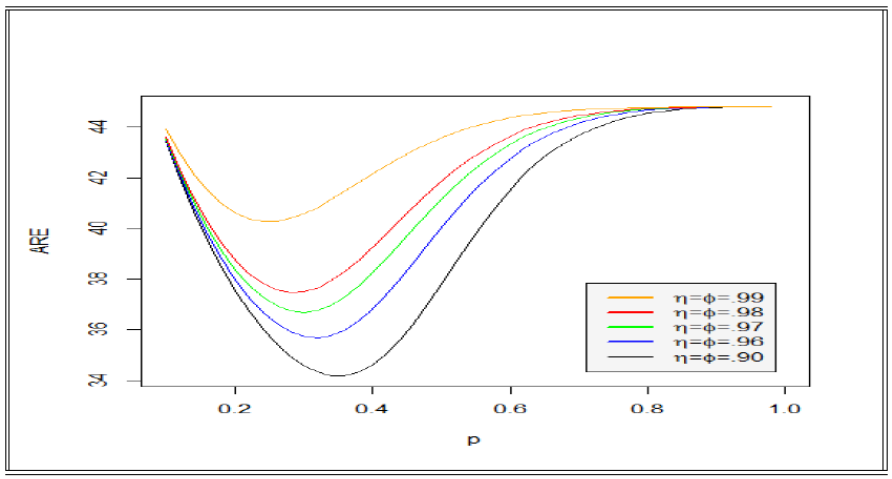

Figure 4.3: ARE of $\hat{p}_4$ vs probability, p

Figures 4.1, 4.2 and 4.3 represent ARE values plotted against prevalence p values at stages two, three and four respectively. Clearly, as noted in Tables 4.1, 4.2 and 4.3, the ARE drops as p increases up to the value $p = 0.3$, then it improves as p moves away from 0.3. From Figures 4.1, 4.2 and 4.3, it is evident that the adaptive estimators outperform the non-adaptive estimator in the presence of test errors as the sensitivity and specificity of the test kit decreases. Hence in cases where the test kits have low sensitivity and specificity, the multi-stage adaptive testing scheme is preferred for more efficient results.

4.1.2 Comparing the adaptive estimators at various stages

Next we compared Asymptotic Relative Efficiencies among the constructed estimators in this study. Here we compared $\hat{p}_2$ with $\hat{p}_1$, $\hat{p}_3$ with $\hat{p}_2$ and $\hat{p}_4$ with $\hat{p}_3$. This analysis would enable us to make important generalizations about the efficiency of the multi-stage adaptive estimator, $\hat{p}_n$. ARE values for adaptive estimators at stages two, three and four relative to adaptive estimators at stages one, two and three respectively were obtained by dividing Equation (5), where $\hat{p} = \hat{p}_1$ and $k = k_1$ by Equation (9), Equation (9) by Equation (13) and Equation (13) by Equation (16) respectively.

$$\frac{Var(\hat{p}_1)}{Var(\hat{p}_2)}, \frac{Var(\hat{p}_2)}{Var(\hat{p}_3)} \; and \; \frac{Var(\hat{p}_3)}{Var(\hat{p}_4)},$$

Upon simplification we obtain

$$ARE_{\hat{p}_2} = \lambda {k_1}^2 + \frac{\pi_1(1-\pi_1)(1-\lambda){k_2}^2(1-p)^{2k_2}}{(1-p)^{2k_1}\pi_2(1-\pi_2)} \tag{22}$$

$$ARE_{\hat{p}_3} = \frac{M}{\pi_2(1-\pi_2)\lambda {k_1}^2(1-p)^{2k_1-2} + \pi_1(1-\pi_1)(1-\lambda){k_2}^2(1-p)^{2k_2-2}} \tag{23}$$

and

$$ARE_{\hat{p}_4} = \frac{N}{R} \tag{24}$$

where M, N and R are defined in the appendices. Using these Equations and R-Gui software Tables 4.4, 4.5 and 4.6 were generated .

p	$\eta = \phi = 0.99$	$\eta = \phi = 0.98$	$\eta = \phi = 0.97$	$\eta = \phi = 0.96$	$\eta = \phi = 0.90$
0.1	9.7163	9.7158	9.7153	9.7149	9.7126
0.2	9.7369	9.7383	9.7396	9.7409	9.7480
0.3	9.7712	9.7775	9.7836	9.7893	9.8188
0.4	9.8384	9.8627	9.8846	9.9043	9.9913
0.5	10.0142	10.1098	10.1865	10.2492	10.4700
0.6	10.7070	11.0878	11.3347	11.5076	11.9708
0.7	14.7585	16.0357	16.6444	17.0000	17.7520
0.8	47.4125	50.0919	51.0753	51.5856	52.5348

Table 4.4: ARE values of $\hat{p}_2$ relative to $\hat{p}_1$ for specified p, η and ϕ

p	$\eta = \phi = 0.99$	$\eta = \phi = 0.98$	$\eta = \phi = 0.97$	$\eta = \phi = 0.96$	$\eta = \phi = 0.90$
0.1	3.6211	3.6160	3.6111	3.6063	3.5801
0.2	6.7871	6.6413	6.5045	6.3761	5.7484
0.3	17.6647	16.0685	14.7652	13.6813	9.7387
0.4	48.5664	37.0247	30.0569	25.3924	13.6778
0.5	102.7462	61.4910	44.1744	34.6400	15.7793
0.6	140.1473	73.3601	50.0271	38.1510	16.4663
0.7	150.1342	76.1234	51.3380	38.9296	16.6292
0.8	151.4241	76.4962	51.5258	39.0479	16.6611

Table 4.5: ARE values of $\hat{p}_3$ relative to $\hat{p}_2$ for specified p, η and ϕ

p	$\eta = \phi = 0.99$	$\eta = \phi = 0.98$	$\eta = \phi = 0.97$	$\eta = \phi = 0.96$	$\eta = \phi = 0.90$
0.1	3.9481	3.9349	3.9221	3.9096	3.8420
0.2	9.8857	9.4430	9.0475	8.6920	7.1475
0.3	31.9137	26.4622	22.6951	19.9353	11.9794
0.4	85.1181	54.4296	40.3510	32.2515	15.3742
0.5	135.2677	72.6353	50.1207	38.4952	16.8860
0.6	154.7372	79.1945	53.6152	40.7423	17.4759
0.7	160.3179	81.1190	54.6702	41.44267	17.6922
0.8	161.4478	81.5535	54.9309	41.6278	17.7616

Table 4.6: ARE values of $\hat{p}_4$ relative to $\hat{p}_3$ for specified p, η and ϕ

As clearly shown from tables 4.4, 4.5 and 4.6, ARE values are high across all stages. That is, the adaptve estimator at stage two is is more efficient than that at stage one, the adaptive estimator at stage three is more efficient than that at stage two and the estimator at stage four is more efficient than that at stage three. This means that $\hat{p}_4$ is a better estimator than $\hat{p}_3$, $\hat{p}_3$ is a better estimator than $\hat{p}_2$ and $\hat{p}_2$ is a better estimator than $\hat{p}_1$. A closer look at table 4.4 shows that ARE values for $\hat{p}_2$ relative to $\hat{p}_1$ rise by very small margins at values of $p < 0.8$ and steadily shoot when $p = 0.8$, while for higher estimators ARE values rise steadily at $p < 0.6$ and almost level off at $p = 0.6, 0.7, 0.8$. It is also evident from tables 4.5 and 4.6 that ARE values increase with increase in sensitivity and specificity of the test kits. As pointed out earlier, this analysis shows that the adaptive testing scheme improves as the number of stages increases.

These observations can be graphically depicted by Figures 4.4, 4.5 and 4.6.

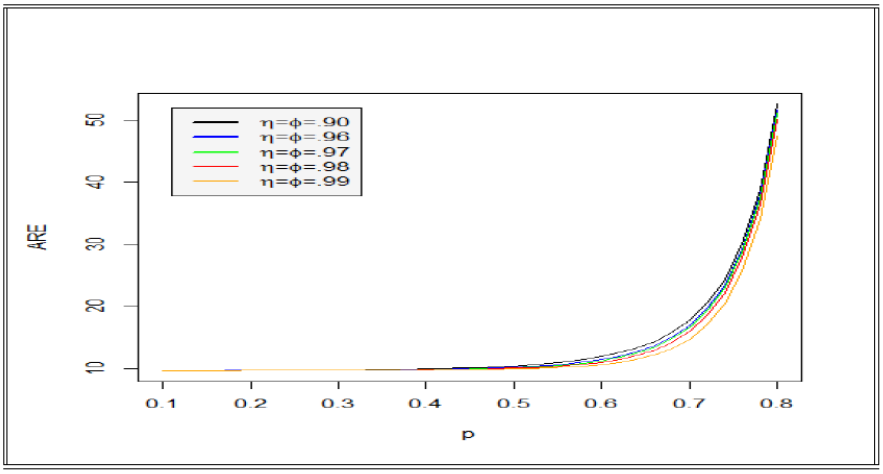

Figure 4.4: ARE of $\hat{p}_2$ relative to $\hat{p}_1$ vs probability, p

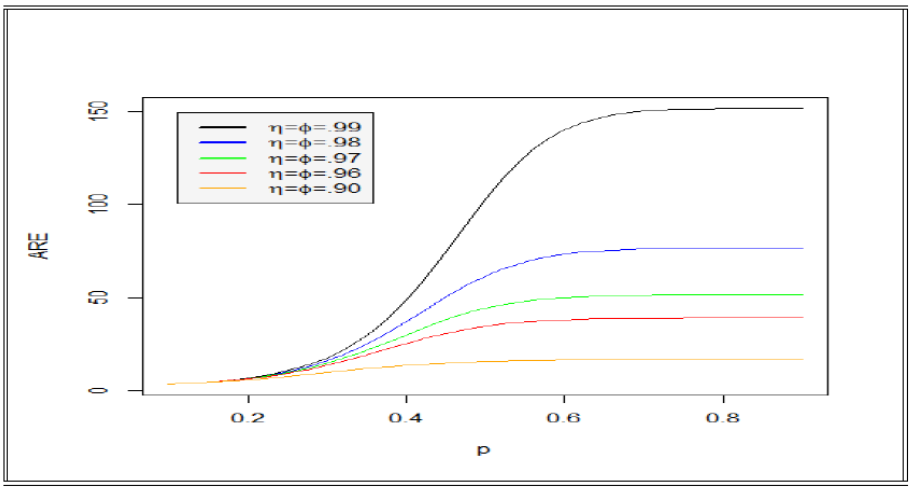

Figure 4.5: ARE of $\hat{p}_3$ relative to $\hat{p}_2$ vs probability, p

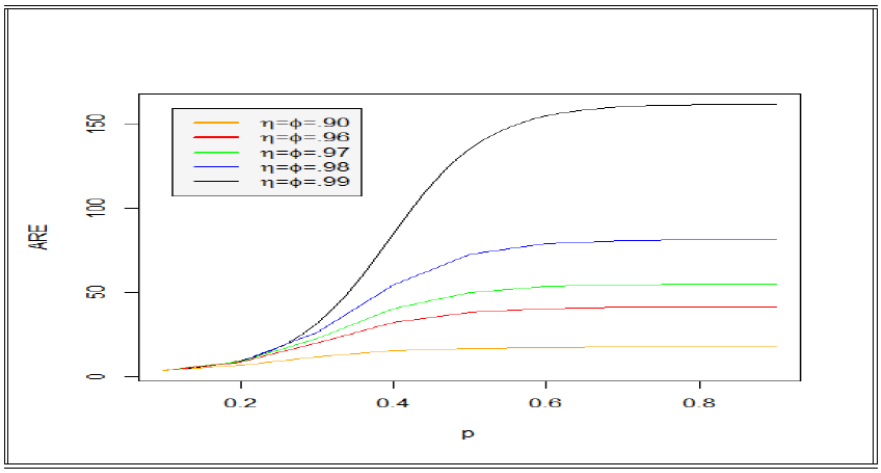

Figure 4.6: ARE of $\hat{p}_4$ relative to $\hat{p}_3$ vs probability, p

4.1.3 Conclusion and Recommendations

From the above discussions, it is clear that the multi-stage adaptive estimator outperforms the non-adaptive estimator in the presence of test errors. It is also clear that the adaptive testing scheme improves as the number of stages increases. A closer look at the results reveals that the multi-stage adaptive estimator is particularly better in cases where test kits have low sensitivity and specificity. Given that experiments are never 100% perfect, the multi-stage adaptive testing scheme is therefore more ideal in estimating prevalence of a trait.

References

[1] Bhattacharyya, G.K., Karandinos, M.G., and DeFoliart, G. R. (1979). *Point estimates and confidence intervals for infection rates using pooled organisms in epidemiologic studies.* American journal of epidemiology **109**,124-131.

[2] Brookmeyer, R. (1999). *Analysis of multistage pooling studies of biological specimens for estimating disease incidence and prevalence.* Biometrics **55**,608-612.

[3] Chiang, C. L. and Reeves, W. C. (1962). *Statistical estimation of virus infection rates in mosquito vector populations.* American journal on hygiene **75**, 377-391

[4] Dorfman R. (1943). *The detection of defective members of large populations.* Annals of mathematical statistics **14**, 436-440.

[5] Gastwirth, J.L., and Hammick,P.A. (1989). *Estimation of prevalence of a rare disease, preserving the anonymity of the subjects by group testing: Application to estimate the prevalence of AIDS antibodies in blood donors.* Journal of Statistical Planning and Inference **22**, 15-27.

[6] Mood, D. A., Graybill, G. A., Boes, D. C. (1974). Introduction to the theory of statistics. **27**, 72.

[7] Nyongesa, L. K. (2011). *Dual estimation of prevalence and disease incidence in pool testing strategy.* Communication in statistics theory and methods, **40**, 1-12

[8] Okoth, A. W. (2012). Two Stage Adaptive Pool Testing for estimating prevalence of a trait in the presence of test errors. Lambert Academic publishers, 18 and 27

[9] Oliver-Hughes J.M. and Swallow W.H. (1994), *A two-stage adaptive group testing procedure for estimating small proportions.* American statistical association **89**, 982-993.

[10] Richards, M. S. (1991). *Interpretation of the results of bacteriological testing of egg laying flocks for salmonella enteritidis.* Proceedings of the 6th international symposium on veterinary epidemiology and economics, 124-126, Ottawa, Canada.

[11] Sobel M., Ellashoff R.M. (1975).*Group testing with a new goal; estimation.* Biometrika **62**, 181-193.

[12] Swallow W.H. (1987).*Relative MSE and cost considerations in choosing group size for group testing to estimate infection rates and probability of disease infection.* Phytopathology **77**, 1376-1381.

Appendices

$$A = (\eta+\phi-1)^2 n \left[\pi_2(p)(1-\pi_2(p))\lambda k_1^2(1-p)^{2k_1-2} + \pi_1(p)(1-\pi_1(p))(1-\lambda)k_2^2(X_1)(1-p)^{2k_2(X_1)-2}\right] \tag{26}$$

$$\begin{aligned} B &= (\eta+\phi-1)^2 \\ &* \left[\pi_2(p)(1-\pi_2(p))\pi_3(p)(1-\pi_3(p))\lambda_1 n k_1(1-p)^{k_1-2}\right. \\ &+ \pi_1(p)(1-\pi_1(p))\pi_3(p)(1-\pi_3(p))\lambda_2 n k_2(X_1)(1-p)^{k_2(X_1)-2} \\ &+ \left.\pi_1(p)(1-\pi_1(p))\pi_2(p)(1-\pi_2(p))\lambda_2 n k_3(X_2)(1-p)^{k_3(X_2)-2}\right] \end{aligned}$$

$$\begin{aligned} C &= (\eta+\phi-1)^2 n \\ &\left[\pi_2(p)(1-\pi_2(p))\pi_3(p)(1-\pi_3(p))\pi_4(p)(1-\pi_4(p))\lambda_1 n k_1(1-p)^{k_1-2}\right. \\ &+ \pi_1(p)(1-\pi_1(p))\pi_3(p)(1-\pi_3(p))\pi_4(p)(1-\pi_4(p))\lambda_2 n k_2(X_1)(1-p)^{k_2(X_1)-2} \\ &+ \pi_1(p)(1-\pi_1(p))\pi_2(p)(1-\pi_2(p))\pi_4(p)(1-\pi_4(p))\lambda_3 n k_3(X_2)(1-p)^{k_3(X_2)-2} \\ &+ \left.\pi_1(p)(1-\pi_1(p))\pi_2(p)(1-\pi_2(p))\pi_3(p)(1-\pi_3(p))(1-\lambda_1-\lambda_2-\lambda_3)n k_4(X_3)(1-p)^{k_4(X_3)-2}\right] \end{aligned} \tag{28}$$

$$\begin{aligned} D &= (\eta+\phi-1)^2 n \\ &* \left[\pi_2(p)(1-\pi_2(p))...\pi_n(p)(1-\pi_n(p))\lambda_1 n k_1(1-p)^{k_1-2}\right. \\ &+ \pi_1(p)(1-\pi_1(p))...\pi_n(p)(1-\pi_n(p))\lambda_2 n k_2(X_1)(1-p)^{k_2(X_1)-2} + ... \\ &+ \pi_1(p)(1-\pi_1(p))\pi_2(p)(1-\pi_2(p))....\pi_(n-1)(p)(1-\pi_(n-1)(p)) \\ &\left.(1-\lambda_1-\lambda_2-...-\lambda_(n-1))n k_n(X_(n-1))(1-p)^{k_n(X_3)-2}\right] \end{aligned}$$

$$H = \pi(p)(1-\pi(p)) * \left[\pi_2(p)(1-\pi_2(p))\lambda k_1^2(1-p)^{2k_1} + \pi_1(p)(1-\pi_1(p))(1-\lambda)k_2^2(X_1)(1-p)^{2k_2(X_1)}\right]$$

$$\begin{aligned} J &= \pi(p)(1-\pi(p)) \\ &* \left[\pi_2(p)\pi_3(p)(1-\pi_2(p))(1-\pi_3(p))\lambda_1 k_1^2(1-p)^{2k_1}\right. \\ &+ \pi_1(p)\pi_3(p)(1-\pi_1(p))(1-\pi_3(p))\lambda_2 k_2^2(X_1)(1-p)^{2k_2(X_1)} \\ &+ \left.\pi_1(p)\pi_2(p)(1-\pi_1(p))(1-\pi_2(p))(1-\lambda_1-\lambda_2)k_3^2(X_2)(1-p)^{2k_3(X_2)}\right] \end{aligned}$$

$$\begin{aligned} L &= \pi(p)(1-\pi(p)) \\ &* \left[\pi_2(p)\pi_3\pi_4(p)(1-\pi_2(p))(1-\pi_3(p))(1-\pi_4(p))\lambda_1 k_1^2(1-p)^{2k_1}\right. \\ &+ \pi_1(p)\pi_3(p)\pi_4(1-\pi_1(p))(1-\pi_3(p))(1-\pi_4(p))\lambda_2 k_2^2(X_1)(1-p)^{2k_2(X_1)} \\ &+ \pi_1(p)\pi_2(p)\pi_4(p)(1-\pi_1(p))(1-\pi_2(p))(1-\pi_4(p))\lambda_3 k_3^2(X_2)(1-p)^{2k_3(X_2)} \\ &+ \left.\pi_1(p)\pi_2(p)\pi_3(p)(1-\pi_1(p))(1-\pi_2(p))(1-\pi_3(p))(1-\lambda_1\lambda_2\lambda_3)k_4^2(X_3)(1-p)^{2k_4(X_3)}\right] \end{aligned}$$

$$\begin{aligned} M &= \pi_2\pi_3(1-\pi_2)(1-\pi_3)\lambda_1 {k_1}^2(1-p)^{2k_1-2} \\ &+ \pi_1\pi_3(1-\pi_1)(1-\pi_3)\lambda_2 {k_2}^2(1-p)^{2k_2-2} \\ &+ \pi_1\pi_2(1-\pi_1)(1-\pi_2)(1-\lambda_1-\lambda_2){k_3}^2(1-p)^{2k_3-2} \end{aligned}$$

$$\begin{aligned} N &= \pi_2\pi_3\pi_4(1-\pi_2)(1-\pi_3)(1-\pi_4)\lambda_1 {k_1}^2(1-p)^{2k_1-2} \\ &+ \pi_1\pi_3\pi_4(1-\pi_1)(1-\pi_3)(1-\pi_4)\lambda_2 {k_2}^2(1-p)^{2k_2-2} \\ &+ \pi_1\pi_2\pi_4(1-\pi_1)(1-\pi_2)(1-\pi_4)\lambda_3){k_3}^2(1-p)^{2k_3-2} \\ &+ \pi_1\pi_2\pi_3(1-\pi_1)(1-\pi_2)(1-\pi_3)(1-\lambda_1-\lambda_2-\lambda_3){k_4}^2(1-p)^{2k_4-2} \end{aligned}$$

$$\begin{aligned} R &= \pi_2\pi_3(1-\pi_2)(1-\pi_3)\lambda_1 {k_1}^2(1-p)^{2k_1-2} \\ &+ \pi_1\pi_3(1-\pi_1)(1-\pi_3)\lambda_2 {k_2}^2(1-p)^{2k_2-2} \\ &+ \pi_1\pi_2(1-\pi_1)(1-\pi_2)(1-\lambda_1-\lambda_2){k_3}^2(1-p)^{2k_3-2} \end{aligned}$$