A Multi-Stage Adaptive Pool Testing Model with Test Errors Vis-a-Vis the Non-Adaptive Model without Test Errors

Okoth Annette W.
Department of Mathematics
Masinde Muliro University of Science and Technology
P.O Box 190-50100, Kakamega, Kenya

Abstract

Pool testing for presence or absence of a trait is less expensive, less time consuming and therefore more cost effective. This study presents a multi-stage adaptive pool testing estimator $\hat{p}_n$ of prevalence of a trait in the presence of test errors. 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. Matlab 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, notably efficiency in comparison with the non-adaptive estimator in the absence of test errors in the literature of pool testing done alongside provision of the confidence interval of the estimator. Results have shown that the efficiency of the multi-stage adaptive estimator in the presence of test errors is higher than that of the non-adaptive estimator in the absence of test errors. This efficiency also increases with increase in sensitivity and specificity of the test kits. This makes the multi-stage adaptive estimator in the presence of test errors better than the non-adaptive estimator in the absence of test errors, especially so that errors in experiments in our day to day encounters are inevitable.

Keywords: Pool testing, Adaptive estimator, Test errors, Confidence interval

1 Introduction

Prevalence of defective units in a large population from accurate diagnostic tests is a fundamental risk assessment and management factor [6]. 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 [3]. 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, [9]. 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 [3]. 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 [6]. 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 obtained 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 investigated its effeciency in comparison with the non-adaptive estimator in the absence of test errors. 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 [7].

2 Literature Review

Pool testing has been recognized as a sampling scheme that can provide substantial benefits [7]. Early application of pool testing include tests for prevalence of plant virus transmissions by insects [10] and [9] and this was one of the pioneering applications of this concept. In [3] 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 [9] 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 [2]. 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 [6]. In [5] 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 [8], 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 [8].

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 and compare its efficiency with that of the non-adaptive model with perfect tests

3 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,

in the 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 3.1 below:

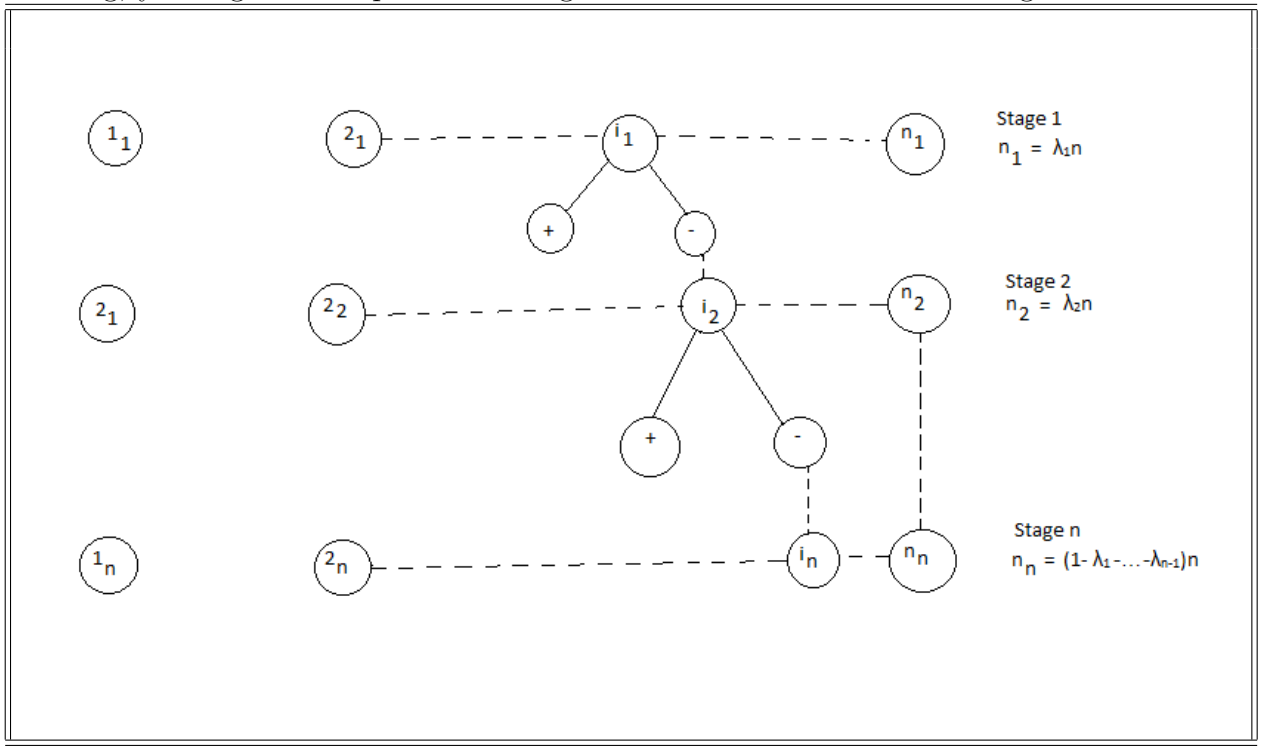

Figure 3.1: Multi-stage adaptive pool testing.

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 and is given by

$$\pi(p) := \eta[1-(1-p)^k] + (1-\phi)(1-p)^k \tag{3}$$

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 presence 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

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

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

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

and

$$\begin{aligned}
&\frac{k_1 X_1 q^{k_1}[(1-\phi)-\eta]}{\eta-(\eta+(1-\phi))q^{k_1}}+\frac{k_2(X_1)X_2 q^{k_2(X_1)}[(1-\phi)-\eta]}{\eta-(\eta+(1-\phi))q^{k_2(X_1)}}....\\
&+\frac{k_n(X_1,...,X_(n-1))X_n q^{k_n(X_1,...,X_(n-1))}[(1-\phi)-\eta]}{\eta-(\eta+(1-\phi))q^{k_n(X_1,...,X_(n-1))}} \qquad (15)\\
=\ &\frac{k_1 q^{k_1}(\lambda_1 n-X_1)(\eta+(1-\phi))}{1-[\eta-(\eta+(1-\phi))q^{k_1}]}+\frac{k_2(X_1)q^{k_2(X_1)}[\lambda_2 n-X_2][\eta+(1-\phi)]}{1-[\eta-(\eta+(1-\phi))q^{k_2(X_1)}]}...\\
&+\frac{k_n(X_1,...,X_(n-1))q^{k_n(X_1,...,X_(n-1))}[(1-\lambda_1-\lambda_2)n-X_4][\eta+(1-\phi)]}{1-[\eta-(\eta+(1-\phi))q^{k_n(X_1,...,X_(n-1))}]}=0,
\end{aligned}$$

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

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

and

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

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

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

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

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

4 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 absence of test errors. 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.

4.2 Comparing the Adaptive estimator in the presence of test errors with the non-adaptive estimator in the absence of test errors

We now compare our constructed estimators at stages two, three and four, that is $\hat{p}_2$, $\hat{p}_3$ and $\hat{p}_4$ with the non-adaptive estimator in the absence of test errors. According to [**?**], for the non-adaptive estimator in the absence of test errors, say, $\hat{p}_e$ its variance is defined as

$$var(\hat{p}_e) = \frac{1-(1-p)^k}{nk^2(1-p)^{k-2}} \tag{19}$$

ARE values for stages two, three and four are obtained by dividing Equation (19) by Equations (9), (13) and (16) respectively. That is,

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

where $Var(\hat{p}_e)$, $Var(\hat{p}_2)$, $Var(\hat{p}_3)$ and $Var(\hat{p}_4)$ are given by Equations (19), (9), (13) and (16) respectively. On simplifying we obtain

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

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

$$ARE_{\hat{p}_4} = \frac{U}{k^2(1-p)^{k-2}\pi_1(p)\pi_2(p)\pi_3(p)(1-\pi_1(p))(1-\pi_2(p))(1-\pi_3(p))},$$

where S, T and U are described in the appendices. Using these Equations and R-Gui software Tables 4.1, 4.2 and 4.3 was generated.

p	$\eta = \phi = 0.99$	$\eta = \phi = 0.98$	$\eta = \phi = 0.97$	$\eta = \phi = 0.96$	$\eta = \phi = 0.90$
0.1	0.6201	0.5937	0.5679	0.5429	0.9733
0.2	0.9348	0.8920	0.8507	0.8107	1.0439
0.3	1.5885	1.4827	1.3853	1.2953	1.3985
0.4	3.0934	2.7607	2.4843	2.2492	2.0817
0.5	6.9677	5.7850	4.9357	4.2828	3.3723
0.6	18.6193	13.8911	10.9879	8.9982	5.9110
0.7	59.3263	36.9918	26.4215	20.2549	11.3545
0.8	202.1403	105.234	69.6775	51.2275	26.4611
0.9	922.4898	449.351	290.6582	211.1664	106.6068

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	0.6979	0.6686	0.6400	0.6121	0.4590
0.2	0.9565	0.9058	0.8578	0.8120	0,5784
0.3	1.4487	1.3245	1.2169	1.1220	0.7158
0.4	2.4726	2.1414	1.8917	1.6911	0.9626
0.5	4.4726	4.0575	3.4339	2.9648	1.4774
0.6	12.6180	9.3723	7.4005	6.0547	2.5406
0.7	39.6622	24.7168	17.6506	13.5297	4.8532
0.8	134.8077	70.1796	46.466	34.1625	11.2807
0.9	615.0049	299.5731	193.7758	140.7802	45.4864

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	0.7753	0.7424	0.7102	0.6789	0.5076
0.2	0.9490	0.8888	0.8332	0.7816	0.5347
0.3	1.2518	1.11482	1.0049	0.9131	0.5555
0.4	1.8535	1.5650	1.3624	1.2051	0.6672
0.5	3.3923	2.7251	2.2866	1.9634	0.9654
0.6	8.1539	6.0084	4.7282	3.8611	1.6141
0.7	24.9753	15.5415	11.0927	8.5007	3.0478
0.8	84.3418	43.9045	29.0689	21.3714	7.0569
0.9	384.399	187.2438	121.116	87.9926	28.4306

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 specified values of p, η and ϕ at stages two, three and four. It can be noticed from the tables that the adaptive estimators at stages two, three and four, that is $\hat{p}_2$, $\hat{p}_3$ and $\hat{p}_4$ register ARE values slightly less than 1 at $p < 0.3$, except for $\eta = \phi = .90$ where the efficiency is slightly less than 1 for $p = 0.1$ at stage two, for $p < 0.5$ at stage three and for $p < 0.6$ at stage four. This means that when prevalence is low the adaptive estimators in the presence of test errors are slightly less efficient than the non-adaptive estimator in the absence of test errors. However, the scenario changes for $p > 0.2$ across all stages for $\eta = \phi > .90$ and for $p > 0.4$ for $\eta = \phi = .90$ at stages three and four, as efficiency of the adaptive estimators significantly improves. Therefore, as prevalence increases, the adaptive estimators in the presence of test errors outperform the non-adaptive estimator in the absence of test errors especially when the sensitivity and specificity of test kits are high. 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

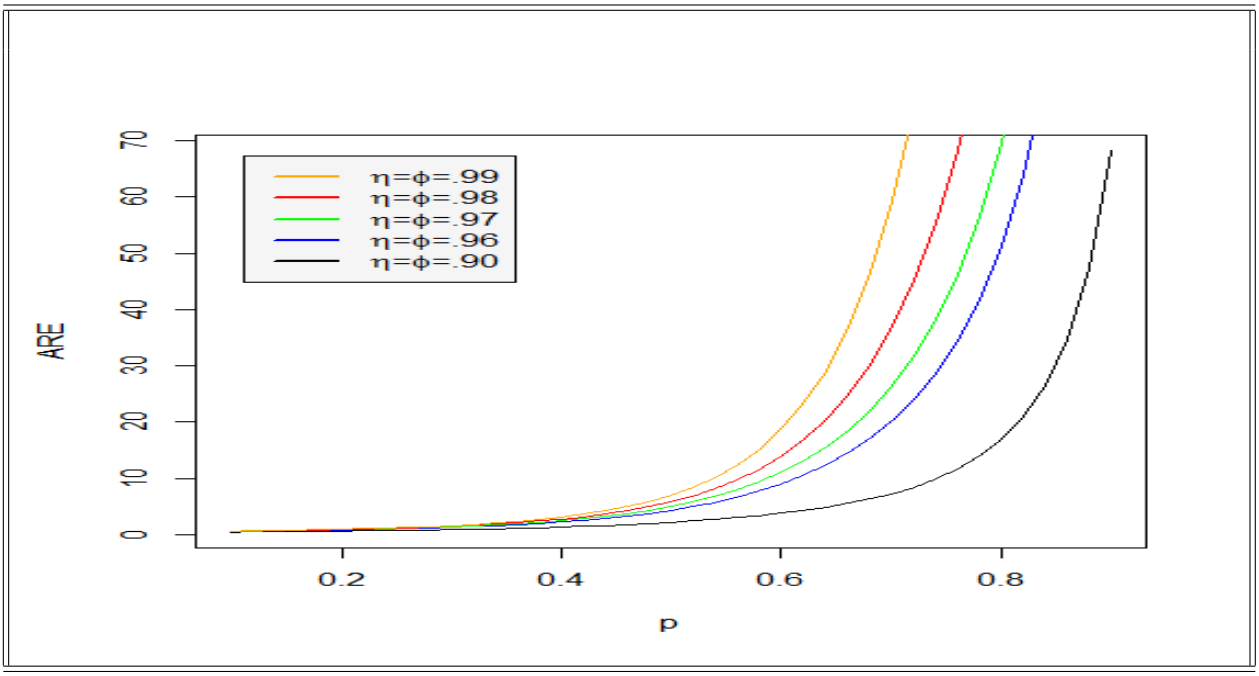

Figure 4.1: ARE of $\hat{p}_2$ vs probability, p

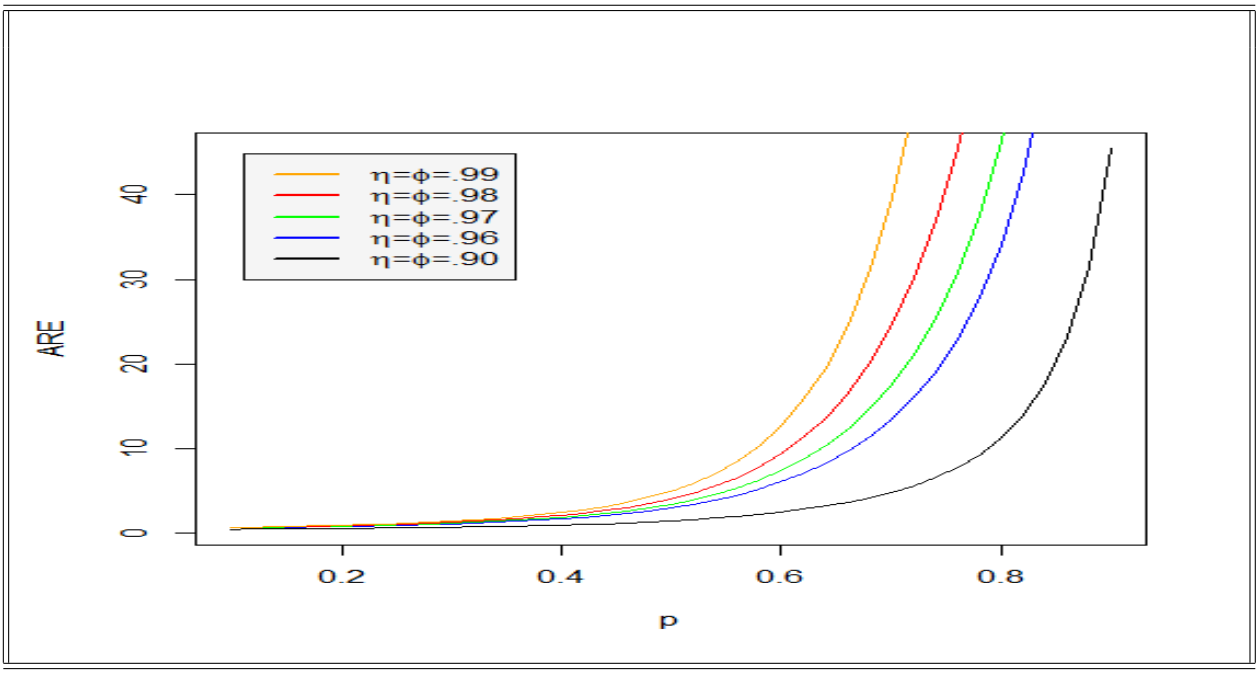

Figure 4.2: ARE of $\hat{p}_3$ vs probability, p

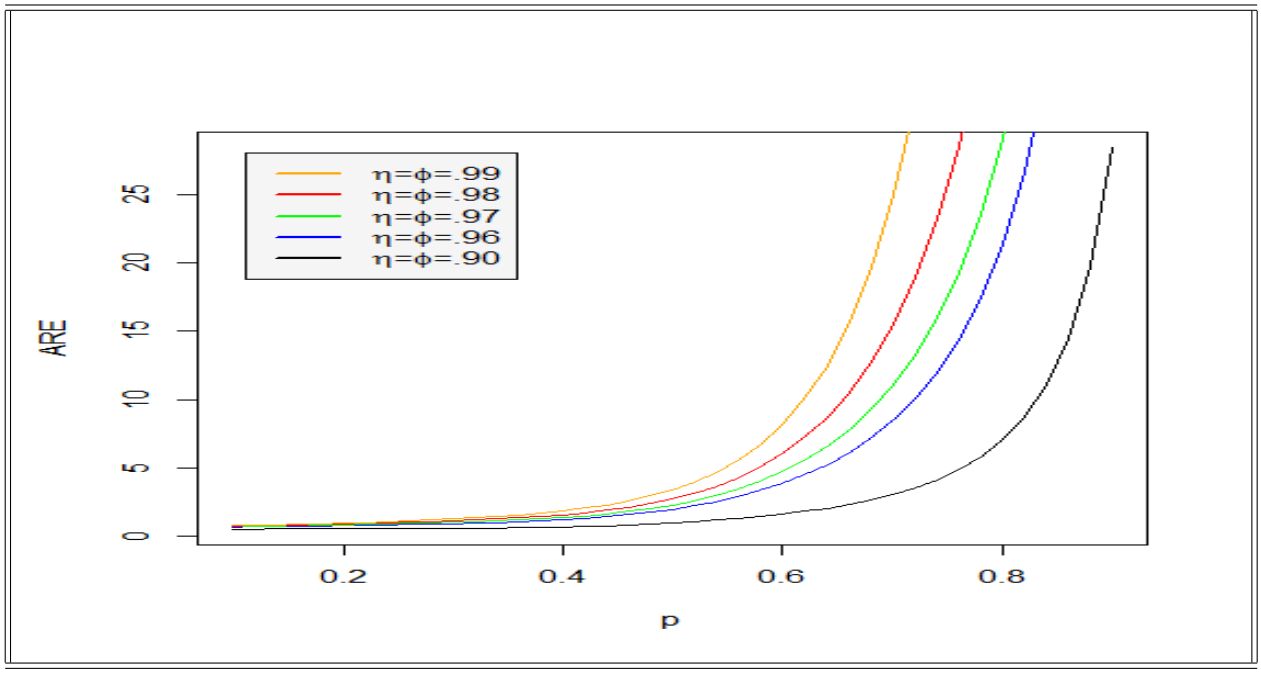

Figure 4.3: ARE of $\hat{p}_4$ vs probability, p

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 values are slightly less than 1 at low values of p but increase significantly as prevalence increases. From Figures 4.1, 4.2 and 4.3, it is clear that the adaptive estimators outperform the non-adaptive estimator in the absence of test errors as the sensitivity and specificity of the test kit increases.

4.3 Conclusion and Recommendations

From the above discussions, it is clear that the multi-stage adaptive estimator in the presence of test errors outperforms the non-adaptive estimator in the absence of test errors. A closer look at the results reveals that the multi-stage adaptive estimator is particularly better in cases where test kits have high 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] 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

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

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

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

[6] 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

[7] 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.

[8] 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.

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

[10] 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

$$A = (\eta+\phi-1)^2 n \\ [\pi_2(p)(1-\pi_2(p))\lambda k_1^2(1-p)^{2k_1-2} + \pi_1(p)(1-\pi_1(p))(1-\lambda)k_2^2(X_1)(1-p)^{2k_2(X_1)-2}] \tag{21}$$

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

$$C = (\eta+\phi-1)^2 n \\ \left[\pi_2(p)(1-\pi_2(p))\pi_3(p)(1-\pi_3(p))\pi_4(p)(1-\pi_4(p))\lambda_1 n k_1(1-p)^{k_1-2}\right. \\ + \pi_1(p)(1-\pi_1(p))\pi_3(p)(1-\pi_3(p))\pi_4(p)(1-\pi_4(p))\lambda_2 n k_2(X_1)(1-p)^{k_2(X_1)-2} \\ + \pi_1(p)(1-\pi_1(p))\pi_2(p)(1-\pi_2(p))\pi_4(p)(1-\pi_4(p))\lambda_3 n k_3(X_2)(1-p)^{k_3(X_2)-2} \\ + \pi_1(p)(1-\pi_1(p))\pi_2(p)(1-\pi_2(p))\pi_3(p)(1-\pi_3(p))(1-\lambda_1-\lambda_2-\lambda_3) n k_4(X_3)(1-p)^{k_4(X_3)-2} \tag{23}$$

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

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

$$
\begin{aligned}
T &= [1-(1-p)^k](\eta+\phi-1) \\
&\quad \Big[\pi_2(p)(1-\pi_2(p))\pi_3(p)(1-\pi_3(p))\lambda k_1^2(1-p)^{2k_1-2} \\
&+ \pi_1(p)(1-\pi_1(p))\pi_3(p)(1-\pi_3(p))\lambda_2 k_2^2(X_1)(1-p)^{2k_2(X_1)-2} \\
&+ \pi_1(p)\pi_2(p)(1-\pi_1(p))(1-\pi_2(p))(1-\lambda_1-\lambda_2)k_3^2(X_2)(1-p)^{2k_3(X_2)-2}\Big]
\end{aligned}
$$

$$
\begin{aligned}
U &= [1-(1-p)^k](\eta+\phi-1)\Big[\pi_2(p)\pi_3(p)\pi_4(p)(1-\pi_2(p))(1-\pi_3(p))(1-\pi_4(p))\lambda k_1^2(1-p)^{2k_1-2} \\
&+ \pi_1(p)\pi_3(p)\pi_4(p)(1-\pi_1(p))(1-\pi_3(p))(1-\pi_4(p))\lambda_2 k_2^2(X_1)(1-p)^{2k_2(X_1)-2} \\
&+ \pi_1(p)\pi_2(p)\pi_4(p)(1-\pi_1(p))(1-\pi_2(p))(1-\pi_4(p))\lambda_3 k_3^2(X_2)(1-p)^{2k_3(X_2)-2} \\
&+ \pi_1(p)\pi_2(p)\pi_3(p)(1-\pi_1(p))(1-\pi_2(p))(1-\pi_3(p))(1-\lambda_1-\lambda_2-\lambda_3)k_4^2(X_3)(1-p)^{2k_4(X_3)-2}\Big]
\end{aligned}
$$