Estimation of Proportion of a Trait by Batch Testing with Errors in Inspection in a Quality Control Process

Ronald W. Wanyonyi[1,*], Kennedy L. Nyongesa[2], Adu Wasike[2]

[1]Department of Mathematics, Egerton University, Nakuru, Kenya
[2]Department of Mathematics, Masinde Muliro University of Science and Technology, Kakamega, Kenya

Abstract   Batch testing has long been recognized as an efficient method of classifying all the experimental units into two mutually exclusive categories: defective or non-defective. In recent years, more focus has been on estimation of the population proportion p of a trait of interest using batch testing. However, in most of the applications, the tests used are prone to errors leading to misclassifications even though misclassifications are kept at a minimal. In this paper, an estimator of the unknown proportion of a trait using batch testing model based on a quality control process in the presence of inspection errors is constructed and its properties discussed. In quality control, a batch is rejected if the constituent defective members are greater than l, the cut off value. The proposed model is superior to the existing models when the proportion of a trait is relatively high with fixed sensitivity and specificity.

Keywords   Quality control, Batching testing, Misclassifications, Cut off value, Proportion, Imperfect tests, Sensitivity, Specificity

1. Introduction

In many applications, items can be classified as defective or non-defective. Batch testing, also known as group testing, involves pooling such items into batches and then testing the batches rather than individual items. A batch is declared non-defective if all its items do not possess the trait of interest. Batch testing has been shown to be cost and time effective when the trait of interest is rare. Batch testing has a rich history going back to Dorfman (1943) and since his seminal work; batch testing has been applied to problems in blood bank screening, genetics, drug discovery, epidemiology and quality control (see e.g Gastwirth, 2000; Remlinger et al., 2006; Hughes-Oliver, 2006, Biggerstaff, 2008; Bilder et al., 2010). In all these applications, batch testing has two main objectives; first is the identification of positive individuals in a large population (Dorfman, 1943, Johnson et. al., 1991, and Nyongesa, 2004). Second is estimation of the proportion of character or trait of interest (Thomson, 1962, Sobel and Elashoff, 1975, Garstwirth and Johnson, 1994, Brookmeyer, 1999, Hughes-Oliver and Swallow, 1994, Tu et. al., 1995 and Xie et. al., 2001).

Recent studies in batch testing have concentrated on the second objective by generalizing Thomson (1962) and Sobel and Elashoff (1975) studies. The various scholars have done this by considering; unequal pool sizes (Brookmeyer, 1999, Hepworth, 1999 and Biggerstaff, 2008), multistage pooling schemes (Brookmeyer, 1999, Hughes-Oliver and Swallow, 1994 and Nyongesa, 2004) and imperfect tests (Tu et. al., 1995 and Nyongesa, 2012).

But, in quality control processes a batch is classified positive or negative according to whether d, the number of constituent members with quality characteristic is greater or less than a fixed cut off value or threshold (Montegomery, 2009). This is similar to threshold batch testing introduced by Damaschke (2006) where he considered two thresholds, $l < u$, called the lower and upper threshold respectively. Several others authors have also considered the threshold batch testing among them are Chen and De Bonis (2011). However, they were mainly concerned with coming up with algorithms that can be used to determine the minimum number of tests required to determine the defective items. Tu et. al. (1995) described the use of batch testing in HIV surveillance and implemented the possibility of testing errors in the process. Two terms are commonly used when discussing the perfection or imperfection of a test: sensitivity and specificity. Sensitivity is the probability of a positive result given that the unit tested is truly positive. Specificity is the probability of a negative result given that the unit tested is actually negative. In this paper we present the maximum likelihood estimator for the proportion of detective, p and its properties when using batch testing in the presence of the test errors in a quality control process. That is, there is possibility of misclassification of defective or non-defective item being declared non-defective or defective respectively.

2. Estimator of Proportion

Suppose we have a set of items that can be characterized as defective or non defective and that the probability, p of being defective is same for all items in the population. Further, assume that the items can be subdivided into n batches each of size, k. Then the batches are tested with kits that are not 100% perfect. The objective is to estimate the proportion of defectiveness that characterizes the population under investigation by testing a batch. If a batch is labeled defective, then the number of constituent items with the quality character of interest, d is greater than a predetermined cut off value l. Given that $\eta(k, d > l)$ and $\phi(k, d \leq l)$ are sensitivity and specificity of the test, then the probability of classifying a batch as positive;

$$\pi_1(p) = \eta(k, d > l)(1 - F(l)) + \{1 - \phi(k, d \leq l)\}(F(l)) \quad (1)$$

where

$$F(l) = \sum_{d=0}^{l} \binom{k}{d} p^d (1-p)^{k-d} \quad (2)$$

We investigate the relationship between $\pi_1(p)$ and p by plotting $\pi_1(p)$ against p for different k and l. The value of sensitivity and specificity is fixed and assumed equal in the course of testing. Figure 1 shows the relationship between $\pi_1(p)$ versus p given that the sensitivity and specificity are equal to 99%. It can be noted that the probability increases rapidly for low p converging to the value of sensitivity.
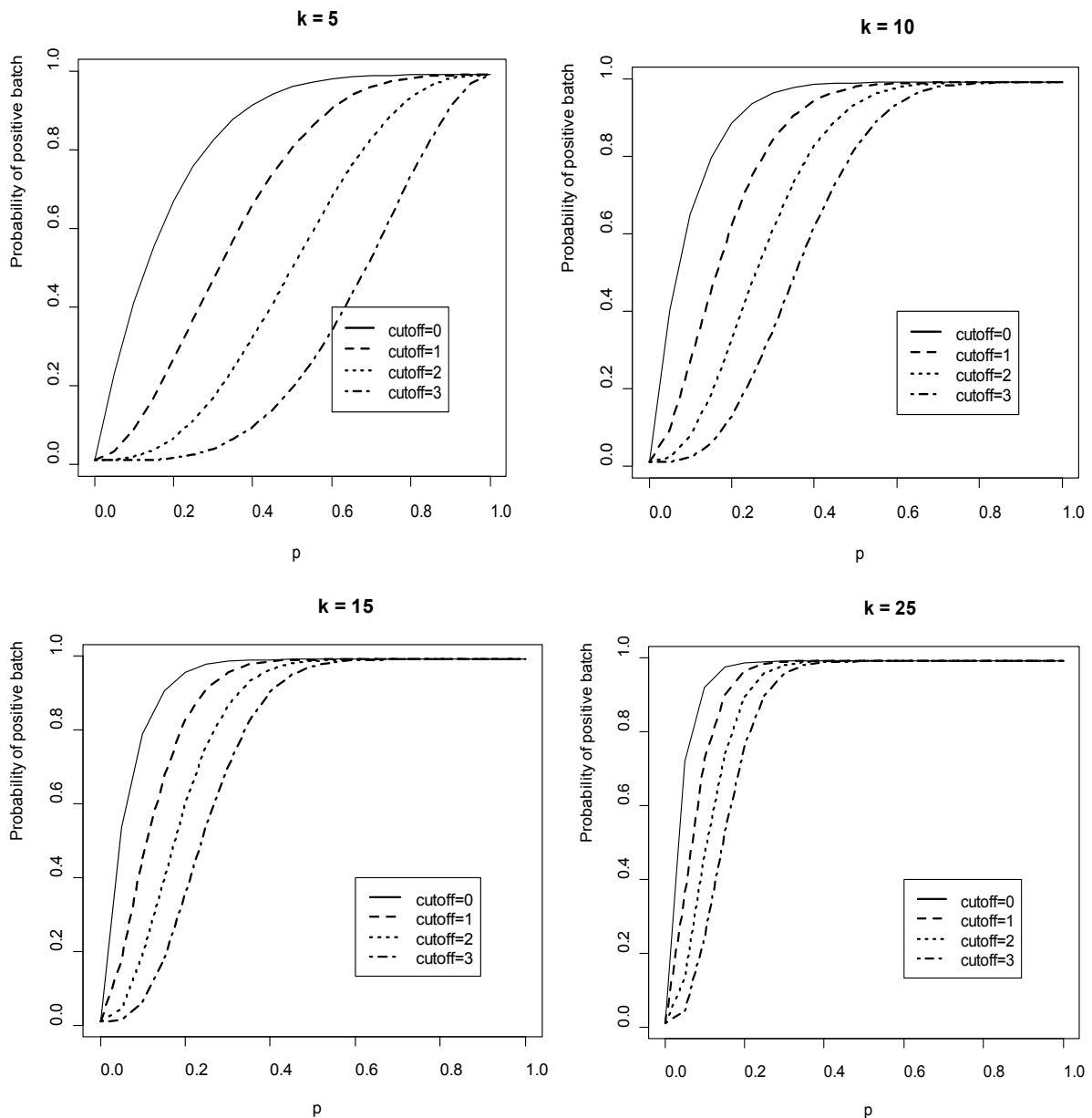

Figure 1. Plot of $\pi_1(p)$ as a function of p for k = 5, 10, 15, 25 and l = 0, 1, 2, 3 with sensitivity and specificity = 99%

Next, the value of sensitivity and specificity is fixed at 90% and the relationship observed as shown in Figure 2.
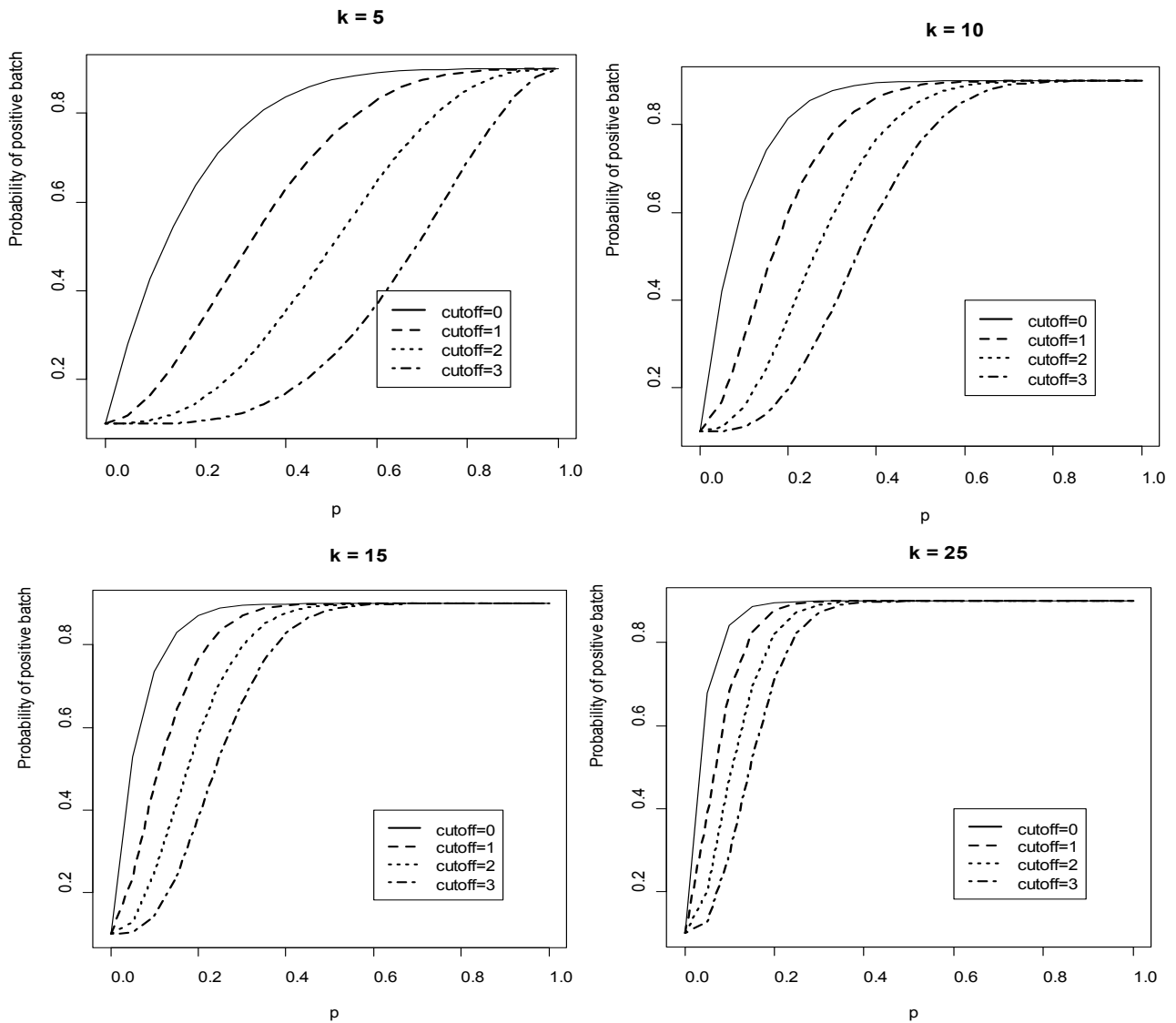

Figure 2. Plot of $\pi_1(p)$ as a function of p for $k = 5, 10, 15, 25$ and $l = 0, 1, 2, 3$ with sensitivity and specificity = 90%

The observation of Figure 2 illustrates that the relationship is monotonically increasing and the increment is steep for low values of p. The relationship was also examined for sensitivity and specificity fixed at a low value equal to 70%. This value is picked for the sole purpose of studying the relationship. This is because the tests which are normally used have higher sensitivity and specificity.

The results show that the relationship between $\pi_1(p)$ and p is least affected by the values of sensitivity and specificity.

2.1. Maximum Likelihood Estimator

Suppose there are n batches, each of size k available for testing and X batches test positive. Then according to Dorfman (1943), X is a random variable that follows binomial distribution $(n, \pi_1(p))$. Thus the likelihood function is;

$$L(p/n, x, \eta, \phi) = \binom{n}{x}[\pi_1(p)]^x[1 - \pi_1(p)]^{n-x} \quad (3)$$

To obtain the MLE of p, we solve the equation

$$\frac{x}{\pi_1(p)}\pi_1'(p) - \frac{(n-x)}{1 - \pi_1(p)}\pi_1'(p) = 0 \quad (4)$$

where $\pi_1'(p) = \dfrac{\partial \pi_1(p)}{\partial q}$ and $q = 1 - p$.

In solving Equation (4), there are two possible solutions;

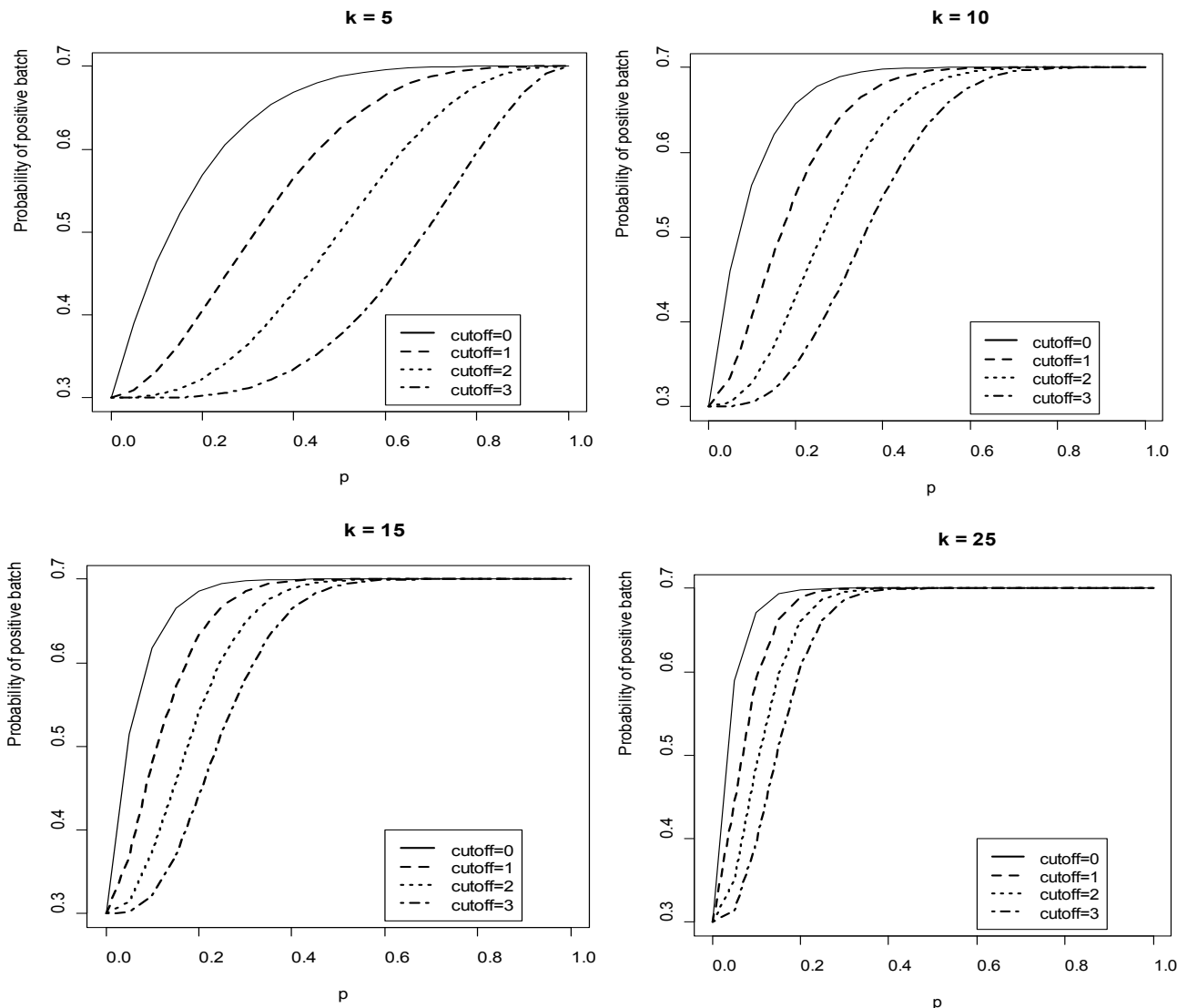$$\pi_1'(p) = 0 \quad \text{or} \quad \pi_1(p) = \frac{x}{n}$$

Figure 3. Plot of $\pi_1(p)$ as a function of p for $k = 5$, 10, 15, 25 and $l = 0$, 1, 2, 3 when sensitivity and specificity = 70%

The candidate $\pi'(p) = 0$ is dropped because it gives only two extreme solutions $\hat{q} = 0$ or $\hat{q} = 1$ and thus $\hat{p} = 0$ or $\hat{p} = 1$. The estimate $\hat{p} = 0$ is most unlikely to happen since in practice it is not possible to avoid defectives and have only good items. On the other hand $\hat{p} = 1$ means that all the batches or items test positive which is certainly an overestimate of p as it is most unusual for every item in the population to be positive (Hepworth and Watson, 2009). With the candidate $\pi_1(p) = \dfrac{x}{n}$ we get;

$$\sum_{d=0}^{l} \binom{k}{d} p^d (1-p)^{k-d} = \frac{\eta(k, d > l) - \dfrac{x}{n}}{\eta(k, d > l) + \varphi(k, d \leq l) - 1} \quad (5)$$

For $l = 0$, Equation (5) reduces to

$$\hat{p}_1 = 1 - \left\{ \frac{\eta - \dfrac{x}{n}}{\eta + \phi - 1} \right\}^{\frac{1}{k}} \quad (6)$$

This is a result obtained by Nyongesa (2012). However, Equation (5) has no solution in closed form when $l > 0$. Therefore, the equation is solved iteratively using an R code that we developed which is presented in Appendix A.

3. Properties of the Estimator

In this section, the properties of the maximum likelihood estimator of proportion are discussed.

3.1. Relationship between $\hat{p}_1$ and p

The relationship between $\hat{p}_1$ and p when the value of sensitivity and specificity of the test is fixed at 99% is presented in Figure 4. It can be noted that there is significant increase in the maximum likelihood estimate as the cut off value increases. But the increase diminishes with the increase in the batch size.
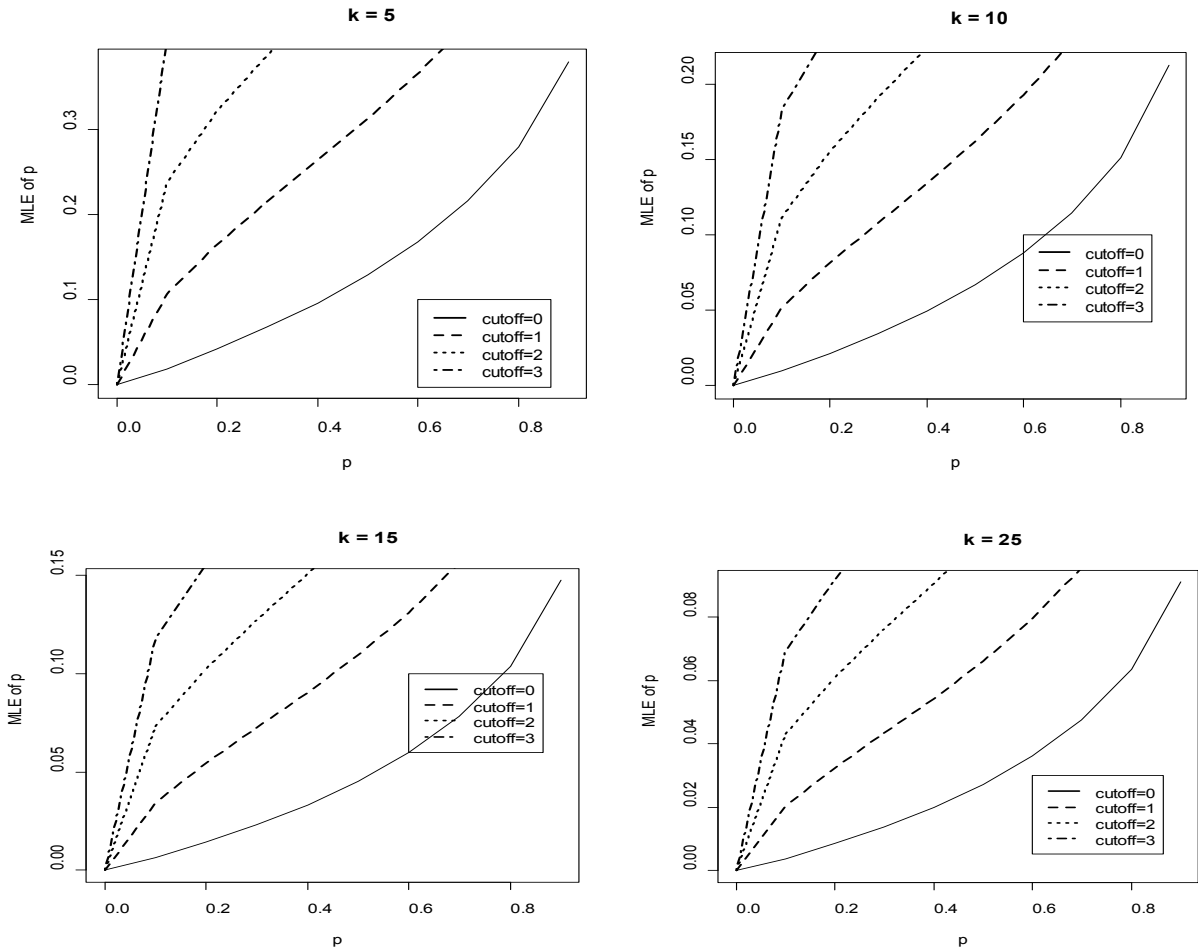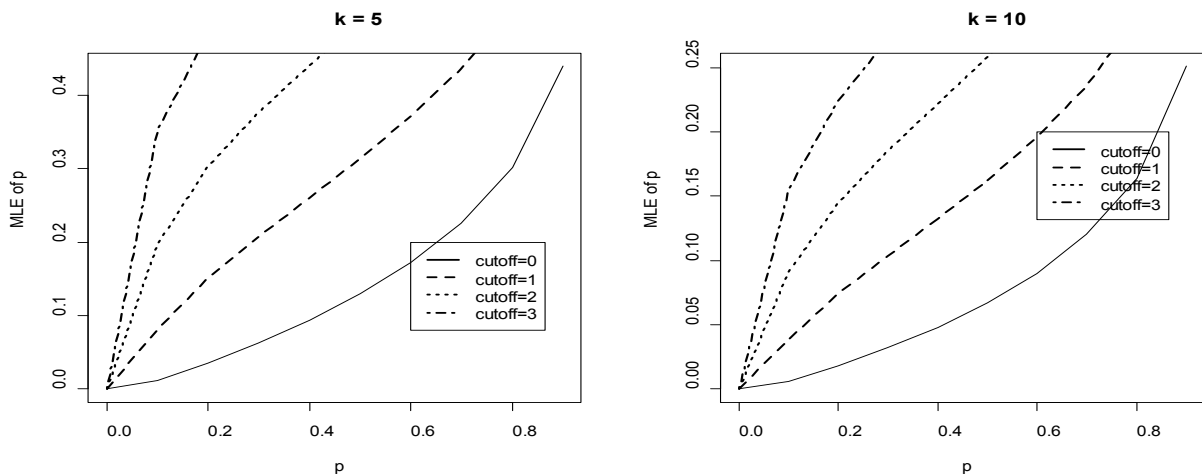

Figure 4. Plot of $\hat{p}_1$ against p for $k = 5, 10, 15, 25$ and $l = 0, 1, 2, 3$ with sensitivity and specificity = 99%

The relationship is studied when the value of sensitivity and specificity is set at 95%. The results are shown in Figure 5 below.
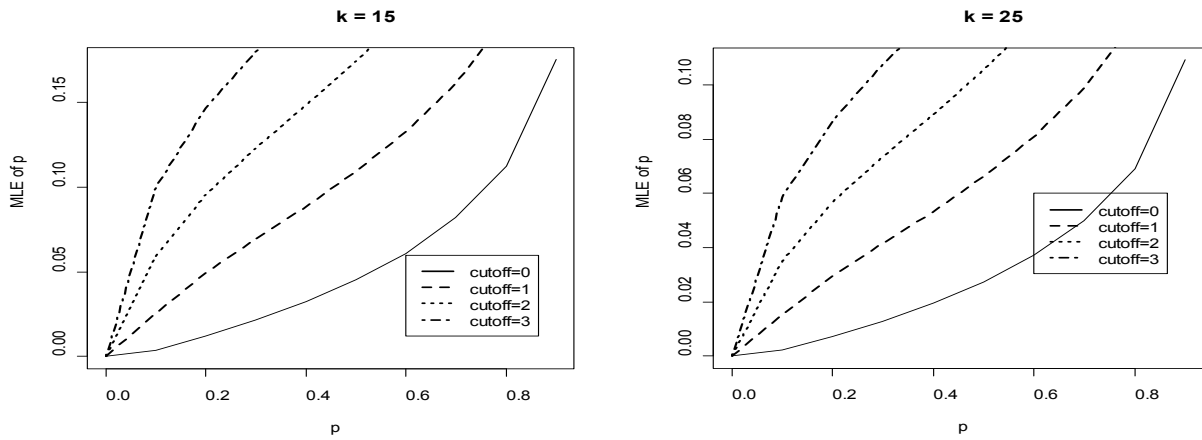
Figure 5. Plot of $\hat{p}_1$ against p for k = 5, 10, 15, 25 and l = 0, 1, 2, 3 with sensitivity and specificity = 95%

Figure 6. Plot of $\mathrm{var}(\hat{p}_1)$ as a function of p for k = 5, 10, 15, 25 and l = 0, 1, 2, 3 with sensitivity and specificity = 99%

The relationship is observed to be monotonically increasing and the rate of increment depends on k and l. Comparing Figures 4 and 5 it is found that the relationship is slightly affected by the values of sensitivity and specificity. Also to note is the fact that the Brookmeyer (1999) model gives lower estimates than the proposed model. This is indicated by the solid line lying below other lines.

3.2. Asymptotic Variance of Estimator

In this section, we consider the computation of the

asymptotic variance of our estimator. The asymptotic variance is obtained from the Fisher's information given by;

$$\left\{ -E\left(\frac{\partial^2 \log L(.)}{\partial p^2} \right) \right\}^{-1} \quad (7)$$

And we obtain;

$$\text{var}(\hat{p}_1) = \frac{\pi_1(p)(1-\pi_1(p))p^2(1-p)^2}{n\left[\{\eta(k,d>l)+\phi(k,d\leq l)-1\}pk(1-A)-E \right]^2} \quad (8)$$

where

$$E = \sum_{d=0}^{l} d \binom{k}{d} p^d (1-p)^{k-d}$$

$$A = 1 - \sum_{d=0}^{l} \binom{k}{d} p^d (1-p)^{k-d}$$

The derivation of Equation (8) is presented in Appendix B.

The figures that follow below indicate the relationship between $\text{var}(\hat{p}_1)$ and p. The point of interest in the figures is to locate the region where the curves of the model with $l > 0$ are below the curve with $l = 0$ (solid line). That is, identify the region where the model is more efficient than the Brookmeyer (1999) model.

As indicated in Figure 6, the Brookmeyer (1999) model is more efficient than the proposed model for small p. But the p for which the proposed model is more efficient than the Brookmeyer (1999) model reduces as k increases. For example, with $k = 5$, the proposed model is more efficient than Brookmeyer (1999) model for $p > 0.3$ while if $k = 25$, that value is $p > 0.06$. In Figure 7 the sensitivity and specificity was set at 95% and the relationship of $\text{var}(\hat{p}_1)$ with respect to p observed. The plot indicates that the p for which the proposed model is more efficient than the Brookmeyer (1999) model is relatively higher for small batch size as opposed to large batch size.
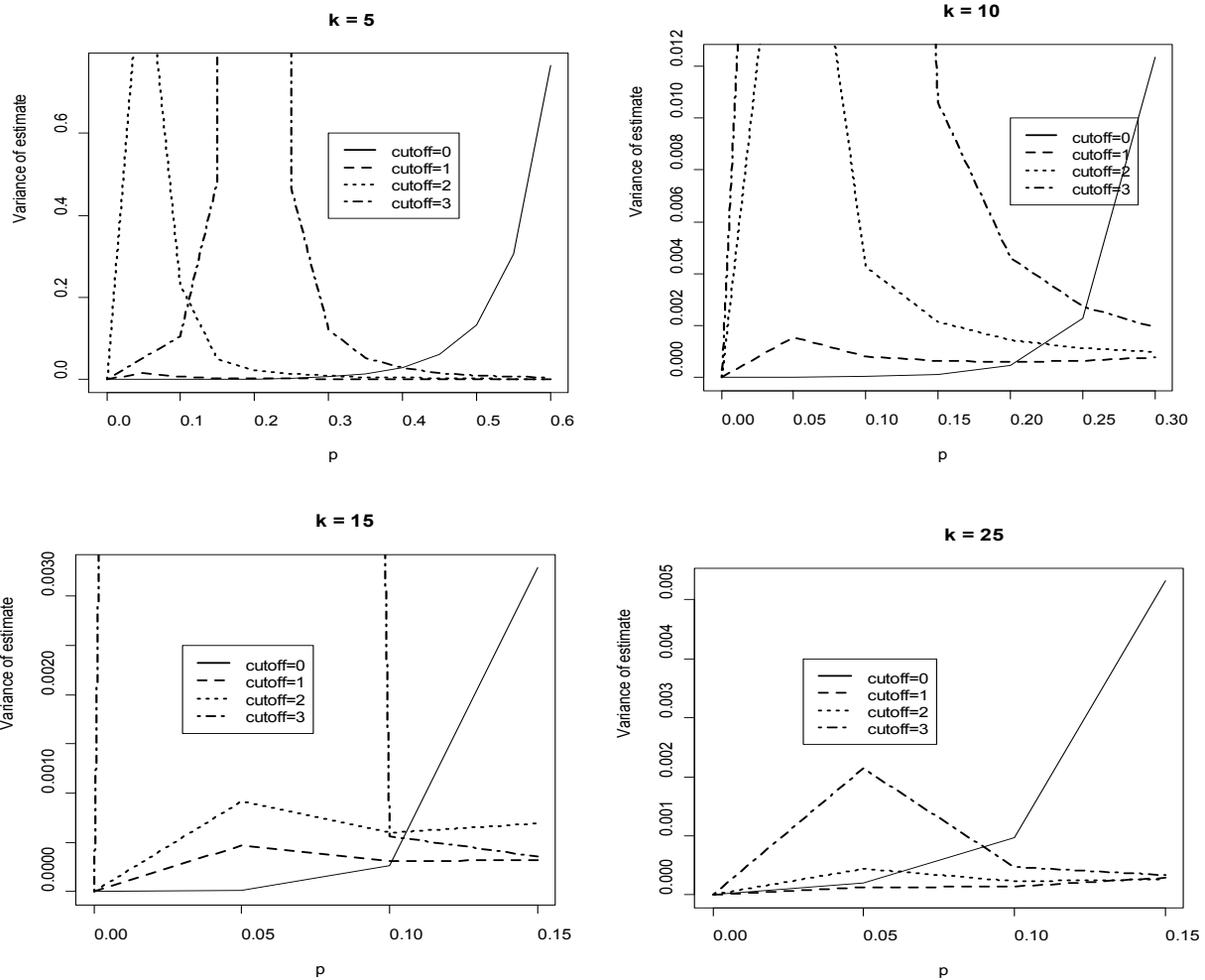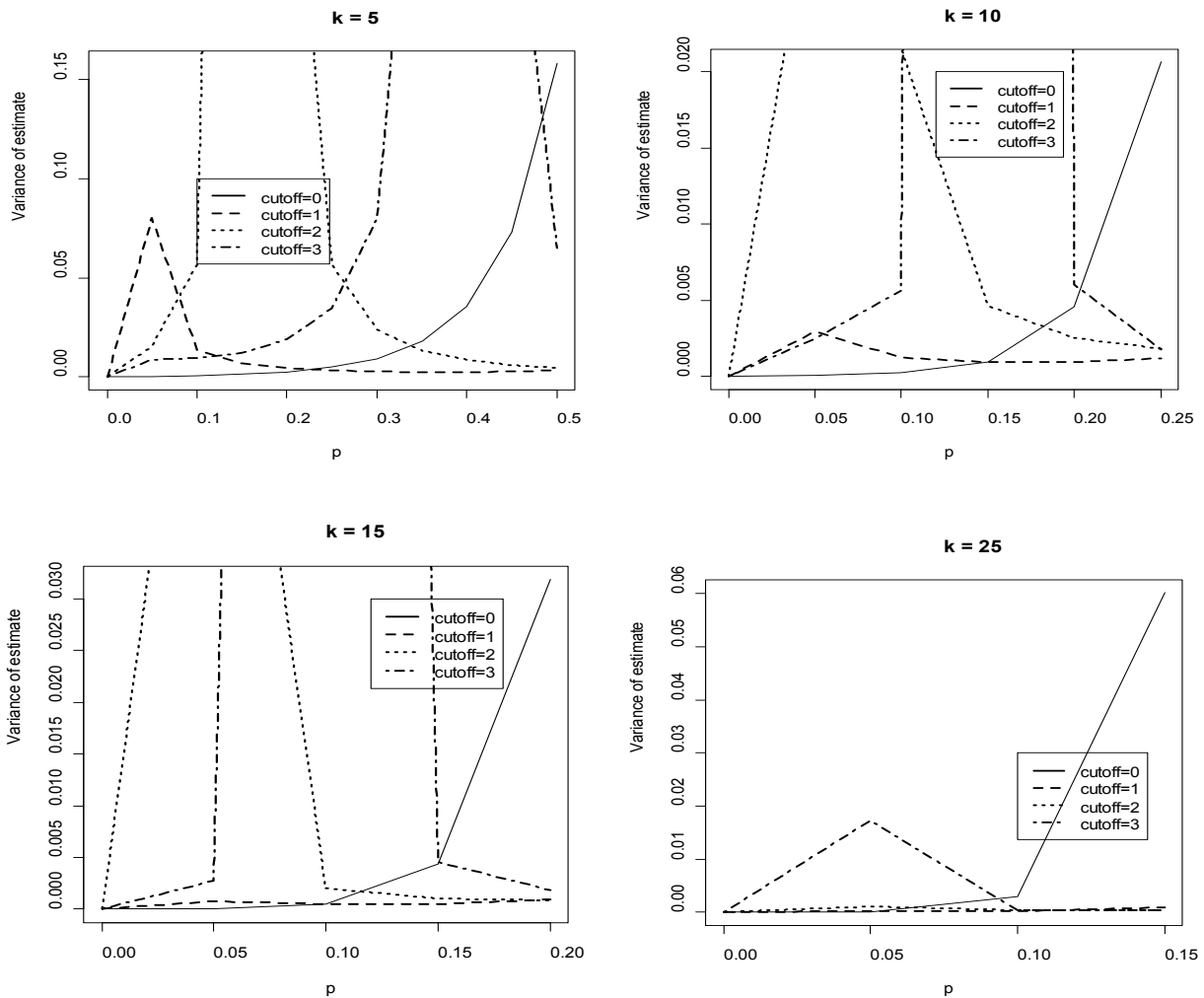

Figure 7. Plot of $\text{var}(\hat{p}_1)$ as a function of p for k = 5, 10, 15, 25 and l = 0, 1, 2, 3 with sensitivity and specificity = 95%

Table 1. ARE of proposed model relative to Brookmeyer (1999) model for k = 5, 10, 15, 25 with sensitivity and specificity = 99%

		l	
p	1	2	3
k = 5			
0.02	0.025	0.004	0.005
0.06	0.273	0.000	0.014
0.08	0.404	0.004	0.018
0.15	0.808	0.151	0.011
0.20	1.073	0.373	0.000
0.25	1.345	0.675	0.032
0.30	1.643	1.070	0.155
k = 10			
0.02	0.162	0.001	0.009
0.06	0.693	0.114	0.005
0.08	0.909	0.284	0.001
0.15	1.595	1.197	0.405
0.20	2.145	2.179	1.251
0.25	2.879	3.699	2.884
0.30	4.029	6.412	6.254
k = 15			
0.02	0.321	0.002	0.012
0.06	1.023	0.422	0.025
0.08	1.306	0.793	0.178
0.15	2.369	2.731	1.971
0.20	3.573	5.499	5.469
0.25	6.016	12.216	15.464
0.30	12.089	33.852	53.659
k = 25			
0.02	0.603	0.075	0.006
0.06	1.575	1.253	0.534
0.08	2.039	2.121	1.386
0.15	5.095	9.776	12.211
0.20	13.595	41.472	71.373
0.25	40.679	247.332	620.728
0.30	92.492	1380.420	6104.280

Comparing Figures 6 and 7 indicate that the proposed model is more efficient than Brookmeyer (1999) at lower p when sensitivity and specificity is low given small batches. For instance, if $k = 5$ and sensitivity and specificity = 99%, the proposed model is more efficient than Brookmeyer (1999) when $p > 0.25$. But when sensitivity and specificity = 95%, the model is more efficient with $p > 0.21$. However, if k is large, then p for which the proposed model is more efficient than the Brookmeyer (1999) model is relatively lower than one when both sensitivity and specificity are equal to 99% as opposed to when sensitivity and specificity are equal to 95%. For example, when $k = 25$ the proposed model is efficient compared to Brookmeyer (1999) when $p > 0.01$ if sensitivity and specificity is 99%. However, for the same batch size, the model is more efficient than Brookmeyer (1999) for $p > 0.05$

given the test has sensitivity and specificity equal to 95%.

3.3. Asymptotic Relative Efficiency

We compare our results with those of Brookmeyer (1999) among others by computing Asymptotic Relative Efficiency (ARE). If the estimator of Brookmeyer (1999) among others is denoted by $\hat{p}_B$ and our estimator $\hat{p}_l$ by then

$$ARE = \frac{\text{var}(\hat{p}_B)}{\text{var}(\hat{p}_l)} \qquad (9)$$

Table 2. ARE of proposed model relative to Brookmeyer (1999) model for k = 5, 10, 15, 25 with sensitivity and specificity = 95%

		l	
p	1	2	3
k = 5			
0.02	0.002	0.036	0.037
0.06	0.053	0.064	0.098
0.08	0.133	0.062	0.132
0.15	0.515	0.005	0.255
0.20	0.819	0.023	0.305
0.25	1.157	0.163	0.284
0.30	1.559	0.458	0.189
k = 10			
0.02	0.014	0.049	0.065
0.06	0.394	0.006	0.167
0.08	0.631	0.010	0.176
0.15	1.501	0.644	0.001
0.20	2.348	1.804	0.282
0.25	3.739	4.214	1.783
0.30	6.343	9.961	6.905
k = 15			
0.02	0.081	0.047	0.092
0.06	0.766	0.057	0.129
0.08	1.117	0.292	0.038
0.15	2.747	2.633	0.936
0.20	5.287	7.956	5.896
0.25	11.077	25.835	29.042
0.30	23.332	91.078	146.807
k = 25			
0.02	0.303	0.012	0.135
0.06	1.478	0.731	0.017
0.08	2.177	1.775	0.476
0.15	8.897	19.063	21.297
0.20	25.138	112.318	209.465
0.25	55.805	569.611	1947.016
0.30	100.113	2081.690	14468.410

Therefore, ARE > 1 implies that the proposed model is more efficient than the Brookmeyer (1999) model. The tables that follow show ARE for some values of p with sensitivity and specificity equal to 99% and 95%.

From Table 1, it can be seen that the proposed model is more efficient than Brookmeyer (1999) model for large values of p. However, the value of p for which our model is more efficient reduces with increase in k. Next sensitivity and specificity is reduced to 95% and the results are shown in the table below;

The results in Table 2 indicate that the proposed model is more efficient than Brookmeyer (1999) model for relatively higher values of p as indicated by ARE > 1.

3.4. Relative Mean Squared Error

If the estimator of Brookmeyer (1999) among others is denoted by $\hat{p}_B$ and our estimator $\hat{p}_l$ by then the Relative Mean Squared Error (RMSE) is given by;

$$RMSE = \frac{MSE(\hat{p}_B)}{MSE(\hat{p}_l)} \qquad (9)$$

Table 3. MSE of the estimate for the proposed model relative to the Brookmeyer (1999) model with sensitivity and specificity = 99%

	l		
p	1	2	3
k = 5			
0.005	0.308	-	-
0.01	8.349	-	-
0.05	0.343	-	-
0.10	3.484	-	-
0.15	0.274	-	-
0.20	0.372	-	-
k = 10			
0.005	251.094	0.581	0.005
0.01	0.177	0.000	0.000
0.05	1.139	5.948	2.899
0.10	0.009	6.322	8.148
0.15	0.131	0.001	0.001
0.20	0.531	0.439	0.186
k = 15			
0.005	0.652	0.636	0.023
0.01	0.002	0.023	0.000
0.05	2.701	0.315	302.103
0.10	9.143	9.062	0.009
0.15	0.001	0.001	0.173
0.20	6.044	0.213	3132.134
k = 25			
0.005	9.316	23.982	0.450
0.01	1.542	0.379	0.342
0.05	0.003	0.002	0.015
0.10	11.362	0.694	1.042
0.15	0.110	61.562	16.829
0.20	167.710	14.973	1.095

Table 3 gives the RMSE for a range of proportion values when the sensitivity and specificity is fixed at 99%.

It can be seen that the efficiency of the proposed model depends on the p, k and l. For example, if $p = 0.005$ and $l = 1$, the proposed model is more efficient than Brookmeyer (1999) model when $k = 10$ or 25. But it is less efficient when k = 5 or 15.

Next, the sensitivity and specificity is reduced to 95% and the RMSE calculated for a range of proportion and cut off values. The results are presented in Table 4.

Table 4. The MSE of the estimate for the proposed model relative to the Brookmeyer (1999) model with sensitivity and specificity = 95%

	l		
p	1	2	3
k = 5			
0.005	0.226	-	-
0.01	8.228	-	-
0.05	0.343	-	-
0.10	3.473	-	-
0.15	0.322	-	-
0.20	0.370	-	-
k = 10			
0.005	224.652	0.624	0.009
0.01	0.166	0.000	0.000
0.05	1.138	5.971	2.174
0.10	0.009	6.087	6.069
0.15	0.138	0.001	0.001
0.20	0.532	0.441	0.187
k = 15			
0.005	0.000	0.000	0.000
0.01	0.287	1.548	0.002
0.05	21.699	2.524	1054.744
0.10	29.810	29.639	0.033
0.15	0.028	0.029	7.527
0.20	26.175	0.921	12253.150
k = 25			
0.005	9.320	22.684	0.216
0.01	1.542	0.379	0.289
0.05	0.003	0.002	0.015
0.10	11.379	0.694	1.043
0.15	0.111	62.444	16.951
0.20	165.304	14.817	1.085

The results indicate that the efficiency of the proposed model is reliant on the p, k and l. The proposed model performs better than the Brookmeyer (1999) model in several settings of p, l and k. But overall, the proposed model is more efficient when batch size increases.

Lastly, the sensitivity and specificity was set to 90% and the models compared using the RMSE.

From Tables 3, 4 and 5, it can be noted that the efficiency of the proposed model relative to the Brookmeyer (1999)

model is affected by the value of sensitivity and specificity. However, the relationship between the relative efficiency of the proposed model and sensitivity and specificity is not a simple one. For instance, for large $k = 15$ and 25 with $l = 1$ and $p = 0.05$, the relative efficiency of the model is 2.701 increasing to 21.699 and then reducing to 21.660 for sensitivity and specificity equal to 99%, 95% and 90% respectively.

Table 5. The MSE of the estimate for the proposed model relative to the Brookmeyer (1999) model with sensitivity and specificity = 90%

		l	
p	1	2	3
$k = 5$			
0.005	0.028	-	-
0.01	3.814	-	-
0.05	0.341	-	-
0.10	3.464	-	-
0.15	0.663	-	-
0.20	0.393	-	-
$k = 10$			
0.005	55.922	0.040	0.043
0.01	0.130	0.000	0.000
0.05	1.139	5.911	1.543
0.10	0.010	5.915	3.889
0.15	0.131	0.001	0.001
0.20	0.531	0.439	0.186
$k = 15$			
0.005	0.000	0.000	0.000
0.01	0.286	0.031	0.001
0.05	21.660	47.415	518.652
0.10	29.809	19.726	0.033
0.15	0.028	0.048	7.582
0.20	25.727	1.885	9807.048
$k = 25$			
0.005	9.307	18.615	0.133
0.01	1.541	0.378	0.228
0.05	0.003	0.002	0.015
0.10	11.332	0.693	1.041
0.15	0.111222	62.62097	16.96471
0.20	164.3565	14.98504	1.098260

On the other hand if $k = 10$, the relative efficiency of the proposed model increases as the sensitivity and specificity increases. Therefore, one would need such tables to be able to set up a more efficient model provided that there is information about the proportion p, cut off value, l and sensitivity and specificity of tests.

Note that RMSE could not be determined for $k = 5$ with $l = 2$ and 3 as indicated by missing values in Tables 3, 4 and 5. Hence for small batch size, $k = 5$, we should avoid large cut off values, $l > 1$.

4. Discussion and Conclusions

We have constructed an estimator for batch testing based on the quality control procedure in the presence of test errors. The probability of detecting a positive batch is much affected by the cut off value and batch size but least affected by sensitivity and specificity. The plots show significant increase in the probability of detecting a positive batch with increase in batch size for the given value of sensitivity and specificity. But the rate of increase diminishes with increase in the cut off value. The MLE of p derived is a generalization of one obtained by Nyongesa (2012) and has no closed form solution when $l > 0$. However, a computer code using an in-built R function, uniroot, is developed to determine the estimate of p. It is found out that the MLE of p increases with increase in sensitivity and specificity. It is also noted that the proposed model gives higher values for MLE of p than Brookmeyer (1999) model for fixed batch size, cut off value and sensitivity and specificity. If the sensitivity and specificity is assumed to be constant in the entire testing time, the proposed model improves the efficiency of the estimator over other models previously developed especially for higher values of p. The problem of errors in inspection can be mitigated by re-testing. It has been shown that re-testing of batches reduces misclassification errors (cf. Nyongesa, 2004). In fact, lost sensitivity of a test is recovered by re-testing batches classified as negative in the initial stage if the tests are imperfect. A batch testing model as applied in quality control where re-testing is allowed can be developed and investigated.

Appendix A: R Code to Determine MLE

```
#Program to find the estimate of p when test is imperfect
#Specify the parameter
k = 15
x = 2
l = 1
n = 20
d = c (0: l)
Sensitivity = 0.99
Specificity = 0.99
#Define the function whose single root is to be determine
in interval (0, 1)
fun<-function(p)
(sensitivity+specificity-1)*(sum(choose(k,d)*p^d*(1-p)^(k-d
)))+x/n-sensitivity
p.estimate<-uniroot(fun,c(0,1))$root
p.estimate
```

Appendix B: Derivation of Asymptotic Variance

We consider the computation of the asymptotic variance of our estimator as presented in Equation (9). The variance is

computed by solving for

$$\left\{ - E\left(\frac{\partial^2 \log L(.)}{\partial p^2} \right) \right\}^{-1}$$

But

$$\frac{\partial \log L(.)}{\partial q} = \frac{x}{\pi_1(p)} \pi_1'(p) - \frac{(n-x)}{1 - \pi_1(p)} \pi_1'(p) \quad (10)$$

where $\pi_1'(p) = \dfrac{\partial \pi_1(p)}{\partial p}$ and

$$\frac{\partial^2 \log L(.)}{\partial p^2} = x \frac{\partial}{\partial p}\left[\frac{\pi_1'(p)}{\pi_1(p)} \right] - (n-x)\frac{\partial}{\partial p}\left[\frac{\pi_1'(p)}{1 - \pi_1(p)} \right]$$

$$= x\left[\frac{\pi_1(p)\pi_1''(p) - (\pi_1'(p))^2}{\pi_1^2(p)} \right]$$

$$- (n-x)\left[\frac{(1 - \pi_1(p))\pi_1''(p) + (\pi_1'(p))^2}{(1 - \pi_1(p))^2} \right] \quad (11)$$

where $\pi_1''(p) = \dfrac{\partial^2 \pi_1(p)}{\partial p^2}$

Thus

$$- E\left(\frac{\partial^2 \log L(.)}{\partial p^2} \right) = \frac{n(\pi_1'(p))^2}{\pi_1(p)(1 - \pi_1(p))} \quad (12)$$

Hence

$$\operatorname{var}(p) = \frac{[\eta(k, d > l) + \phi(k, d \le l) - 1][pk(1 - A) - E]}{p(1 - p)} \quad (13)$$

where

$$E = \sum_{d=0}^{l} d \binom{k}{d} p^d (1 - p)^{k-d}$$

and

$$A = 1 - F(l)$$

REFERENCES

[1] Biggerstaff, B. (2008). Confidence Intervals for the Difference of two proportions Estimated from Pooled samples. *Journal of Agricultural, Biological and Environmental Statistics*, 13, 478-496.

[2] Brookmeyer, R. (1999). Analysis of Multistage Pooling Studies of Biological Specimens for Estimating Disease Incidence and Prevalence. *Biometrics,* 55, 608-612.

[3] Chen, H. and Bonis, A. D. (2011). An almost Optimal Algorithm for Generalized Threshold Group Testing with Inhibitors. *Journal of Computational Biology*, 18, 851-864.

[4] Damaschke, P. (2006). Threshold Group Testing. *Information Transfer and Combinatorics, Lecture notes in Computer Science*, 4123, 707-718.

[5] Dorfman, R. (1943). The detection of defective members of large population. *Annals of Matematical Statistics,* 14,436-440.

[6] Hepworth, G. (1999). Estimation of Proportions by Group Testing. PhD thesis, University of Melbourne.

[7] Hepworth, G. and Watson, R. (2009). Debiased Estimation of Proportions in Group Testing. Applied Statistics, 58, 105-121.

[8] Hughes-Oliver, M. J. (2006). Pooling Experiments for Blood Screening and Drug Discovery. *Screening: Methods for Experimentation in Industry, Drug Discovery and Genetics*, New York, Springer.

[9] Hughes-Oliver, M. J. and Swallow, H. W. (1994). A Two-Stage Adaptive Procedure for Estimating Small Proportions. *Journal of the American Statistical Association*, 89, 982-993.

[10] Montegomery, D. C. (2009). Introduction to Statistical Quality Control (6th edition). New York, John Wiley and Sons.

[11] Nyongesa, L. K., (2004). Multistage Group testing procedure (Batch screening). *Communication in Statistics-Simulation and Computation*, 33, 621-637.

[12] Nyongesa, L. K., (2012). Dual Estimation of Prevalence and Disease Incidence in Pool-Testing Strategy. *Communication in Statistics Theory and Method.* 40, 1-12.

[13] Sobel, M., and Elashoff, R.M., (1975). Group -testing with a new goal, Estimation. *Biometrika*, 62, 181-193.

[14] Thomson, K. H. (1962). Estimation of the Proportion of Vectors in a Natural Population of Insects, *Biometrics,* 18, 568-578.

[15] Tu, M. X., Litvak, E. and Pagano, M. (1995). On the Informative and Accuracy of Pooled Testing in Estimating Prevalence of a Rare Disease: Application to HIV Screening. *Biometrika,* 82, 287-297.