Annals of
Reviews and Research

Juniper Publishers
key to the Researchers

Review Article
Volume 2 Issue 3 - June 2018

Ann Rev Resear

Multiple-Test Pool-Testing Strategy for Estimating HIV/AIDS-Prevalence and Its Extension to Multi-Stage

Nyongesa LK*

Department of Mathematics, Masinde Muliro University of Science and Technology, Africa

Submission: April 06, 2018; **Published:** June 27, 2018

***Corresponding author:** Nyongesa LK, Associate Professor, Masinde Muliro University of Science and Technology, Africa, Tel: +254-723816558; Email: knyongesa@mmust.ac.ke; knyongesa@hotmail.com

Abstract

Testing units of a population one-at-a-time for the presence of a trait is expensive and tedious especially when the population is large. The remedy is to pool the population samples into pools and test each pool for the presence of a trait. There is loss of sensitivity in pooling strategies especially in the presence of inspection errors, and to recover some lost sensitivity is through re-testing pools classified as positive as proposed by Monzon et al. [1]. Pool testing with retesting entails testing the pools, and pools that are classified as positive are retested be-fore being classified as either positive or negative. This study develops a statistical pool-testing model with retesting based on Monzon et al. [1]. Retesting strategy and generalizes it to multi-stage retesting model. The multistage model improves the efficiency of the estimators as evident via the computation of asymptotic relative efficiency. Also studied are moment estimators of prevalence which have been shown to be correlated. Comparison with previous studies have been illustrated through application of the model on an example on estimating HIV/AIDS prevalence of which it was demonstrated that the proposed estimators are superior to previously studied estimators.

Keywords: Asymptotic; Estimator; Likelihood; Pool; Sub pool; Truncated Multinomial

Introduction

Developing countries wishing to set up programs to detect HIV/AIDS infection have major problems of financing the costs of expensive surveys [1]. The use of sera pooled in specific batch size is a cost-effective approach [2]. Pool testing involves pooling samples into pools, testing the pools, and classifying each pool as defective or non-defective. A pool being non-defective is taken to mean that none of the samples constituting the pool possesses the characteristic of interest and a pool testing positive is taken to mean that at least one of the individuals that possess the characteristic is present. It is important however that the sensitivity of the assay in use be maintained, especially in low prevalence countries where identification and counseling of an infected patient is a major preventive approach to limit the spread of infection Monzon et al. [1]. Testing of pools made from the prospectively collected sera started way back during the Second World War by Dorfman [2]. The procedure has been shown to be technically feasible, cost effective, and accurate for estimating sera prevalence in large population surveys Kline et al. [3].

The Dorfman [2] procedure has been extended and generalized. Greater savings can be obtained by hierarchical testing schemes [4-6]. With the stigma associated with HIV/AIDS, the procedure is also applicable where the identity of the subject is not revealed Gastwirth & Hammick [7]. Pool testing is twofold: the rest being the identification of positive individuals in a large population Dorfman [2], Nyongesa [4,5]. This is the area that has received most attention. The second objective is estimating the rate of characteristic of importance. This objective was championed by Thompson [8] and also studied by Sobel & Elashoff [9]. This paper focuses on the second objective but uses a modified design as suggested by Monzon et al. [1].

More studies have focused on the second objective and mainly generalized Thompson [8-13] used pool testing to estimate HIV/AIDS prevalence cost-effectively. Xie et al. [14] demonstrated how pool testing can reduce costs in early stages of drug discovery. On the same subject Tebbs & Swallow [15] have discussed estimation in ordered binomial proportions in pool testing. It has also been observed in studies that benefited from pool-testing that pool-testing depends on the size of the pools [16]. Pool-testing problems where the probability of response is a composite or depends on other variables have been studied by Hung & Swallow [17] Tu et al. [13]. Recently, more efforts have been put in determining optimal group sizes for instance see Ding [18,19] have developed multi-stage adaptive pool-testing strategy of which it has been shown to be more efficient than the ordinary pool-testing scheme. A combination of experiments can yield better estimators in the presence of test errors, for more discussion on the subject see Matiri et al. [20].

For simplicity, throughout our study we shall assume that samples being pooled are independent and identically distributed. In addition, the tests are also independent of one another [21]. Furthermore, the sensitivity and specificity of the test kits will be assumed to remain constant throughout the discussion. The

rest of the paper is arranged as follows: Section 2 discusses the design proposed by Monzon et al. [1] and others in their field of study. Derivation of moment estimators of prevalence is provided in Section 3. Maximum likelihood estimator of the prevalence is presented in Section 4. Extension of the model to multistage is presented in Section 5. Section 6 provides the discussion of the results while Section 7 provides the conclusion to the present study.

Design of Estimating HIV/AIDS-Prevalence in Pooled Sera

The design that is discussed here is of pooled sera as suggested by Dorfman [2]. However, the design differ from the classical pool testing in that pools that test positive are given a second test (duplicate test) as suggested by Monzon et al. [1] in their experiment as shown in Figure 1. In the pooled sera testing design, a population under investigation is of size f say pooled into n pools each of equal sizes k. The n constructed pools are then subjected to testing. Pools that test negative are dropped from further investigation. Pools that test positive are given a second screen test (duplicate test), and pools that test positive on duplicate test, constituent components are subjected to screening as shown diagrammatically in Figure 1. For purpose of estimating HIV/AIDS prevalence in a population pooled into n pools, we need not subject pools that test positive on the duplicate test to individual testing in order not to reveal the identity of the subject [7]. One would wish to subject pools that test positive initially to Western Blot (WB) or a Gold Standard test [22] but due to high cost associated with it, it is too expensive for most developing countries to mount surveys where the Gold Standard Tests are employed. Hence, more savings can be attained by just using the duplicate tests as re-ported by Monzon et al. [1] and in fact the duplicate test will minimize the error of misclassification [4].

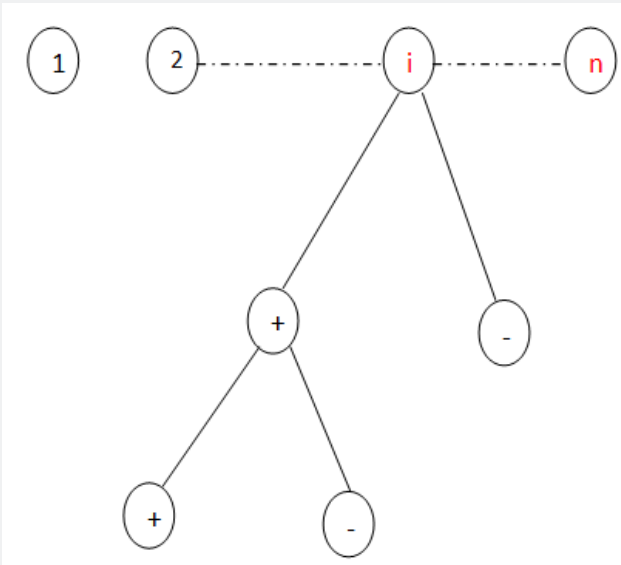

Figure 1: Anti-HIV testing strategy of pooled sera using enzyme immuno assay [1].

The Moment Estimators of Prevalence

Having introduced the model to be used in estimating HIV/AIDS-prevalence, we are now in a position to develop moment estimators. First, we establish the probability of importance in our discussion. These probabilities will be the cornerstone of our analyses. The derivation of the probability of declaring a pool as positive is

$$\pi_1(p) = (1-(1-p)^k)\eta^2 + (1-p)^k(1-\phi)^2 \quad (1)$$

Where η is the sensitivity of the test, and ϕ is the specificity of the test by sensitivity we mean, the probability of correctly classifying a positive pool or individuals whereas specificity means, the probability of correctly classifying a negative pool or individual. These two parameters will be assumed to remain constant in the entire study, and if they are not known they can be estimated from the experiment. For example, let X_1 pools be known to be defective while X_2 pools are known to be non-defective. (X_1 + X_2 = n): The pools X_1 and X_2 are subjected to testing as if their status were unknown but with the sole purpose of estimating η and ϕ. Suppose X_1^* pools out of X_1 tests positive while X_2^* pools out of X_2 tests negative when the experiment is carried out, then the estimators of η and ϕ are obtained as $\hat{\eta} = X_1^* / X_1$ and $\hat{\phi} = X_2^* / X_2$ respectively. Derivation of (1) is accomplished by the law of total probability. Notice that (1) is a composite probability, which is a function of sensitivity, specificity, pool size and the prevalence of the disease p. We also know that $p \in [0,1]$ and for given and greater than 0.5 as it is the case in practice (sensitivity and specificity are always high). So,

$$(1-\phi)^2 \le \pi_1(p) \le \eta^2 \quad (2)$$

Hence $\pi_1(p)$ is a bounded continuous function of p. Now the probability that a pool tests positive on the test instance and negative on re-test is

$$\pi_2(p) = (1-(1-p)^k)\eta(1-\eta) + (1-p)^k\phi(1-\phi) \quad (3)$$

Clearly, (3) is also a function of p. Hence it can be used to estimate the prevalence as it will be shown later. As in the discussion of (2), we also have

$$\phi(1-\phi) \le \pi_2(p) \le \eta(1-\eta) \quad (4)$$

Finally, the computation of the probability of declaring a pool as negative, denoted by $\pi_3(p)$, is given as

$$\begin{aligned}\pi_3(p) &= 1-\phi_1(p)-\phi_2(p) \\ &= [1-(1-p)^k](1-\eta)+(1-p)^k\phi\end{aligned} \quad (5)$$

It then follows that $\pi_3(p)$ is a bounded function of p

$$\phi \le \pi_3(p) \le (1-\eta) \quad (6)$$

That is to say

$$\eta \le \phi_1(p) + \pi_2(p) \le 1-\phi \quad (7)$$

In this case, the model of interest is a multinomial. The multinomial model can be put into the exponential form and the moments easily derived. This is because multinomial distribution belongs to the exponential family (Lehman & Casella, 1995: pp.

How to cite this article: Nyongesa LK. Multiple-Test Pool-Testing Strategy for Estimating HIV/AIDS-Prevalence and Its Extension to Multi-Stage. Ann Rev Resear. 2018; 2(3): 555589.

25). The likelihood function is given by

$$L(p \mid \underline{x}, \eta, \phi) \propto \pi_1^{x_1}(p)\,\pi_2^{x_2}(p)\left(1-\pi_1(p)-\pi_1(p)\right)^{n-x_1-x_2} \quad (8)$$

where x_1 are pools classified positive, and x_2 are pools declared negative by the duplicate test, $\underline{x} = (x_1, x_2)'$. Model (8) is of interest because our estimate of prevalence p will be based on it. We obtain the moment estimate of p from (8). As noted earlier, (8) belongs to exponential family. Thus, we have

$$E(X_1) = n\pi_1(p) \qquad E(X_2) = n\pi_2(p) \quad (9)$$

Equation (9) implies that $\hat{\pi}_1(p) = x_1/n$ and $\hat{\pi}_2(p) = x_2/n$ respectively. We are now in a position to provide ways of estimating the prevalence p. There are three such ways. The first moment-estimator of p is based on (1) and is given by

$$\hat{p}_1 = 1 - \left\{\frac{\eta - \hat{\pi}_1(p)}{\eta^2 - (1-\phi)^2}\right\}^{\frac{1}{k}} \quad (10)$$

The moment estimator of p provided by (10) is not valid for $\eta = 1-\phi$ or $\eta + \phi = 1$; a result also observed by Brookmeyer [6] that $\eta + \phi > 1$. In such a situation, Equation (10) is valid. The second moment-estimator of p is derived from (3) is given by

$$\hat{p}_2 = 1 - \left\{\frac{\eta(1-\eta) - \hat{\pi}_2(p)}{\eta(1-\eta) - \phi(1-\phi)}\right\}^{\frac{1}{k}} \quad (11)$$

Equation (11) is not valid for $\phi(1-\phi) = \eta(1-\eta)$, hence it will only hold in situations where $\phi(1-\phi) \neq \eta(1-\eta)$. Notice that $\hat{p}$ is a poor estimator of p and the assertion is evident in Figure 2. In most practical HIV/AIDS testing situations, $\pi_2(p)$ is roughly zero, thus making the estimator a function of sensitivity and specificity as it can be seen in (Figures 2 & 3).

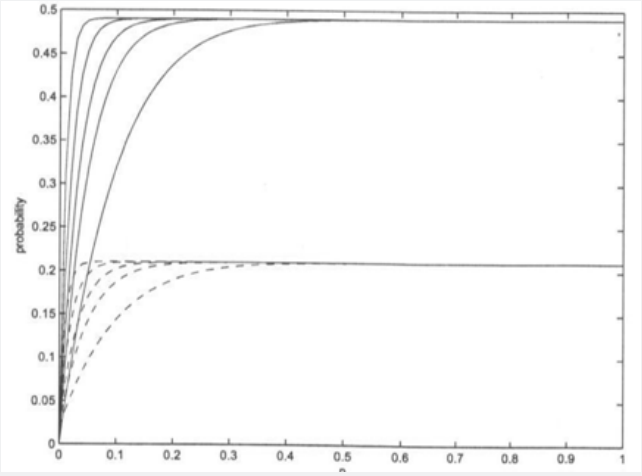

Figure 2: Plots of $_1$(p) and $_2$(p) as a function of prevalence rate for η = 0.7, ϕ = 0.98 and for pool sizes k=10, 20, 30, 50 and 100. ($\pi_1(p) = \_\_, \pi_2(p) = ...$)

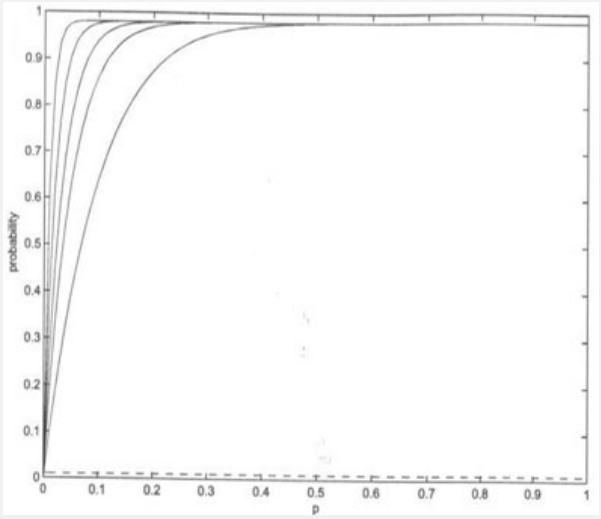

Figure 3: Plots of $\pi_1(p)$ and $\pi_2(p)$ as a function of prevalence rate for η = ϕ =0.99 and for pool sizes k = 10, 20, 30, 50 and 100.($\pi_1(p) = \_\pi_2(p) = \cdots\cdots$)

How to cite this article: Nyongesa LK. Multiple-Test Pool-Testing Strategy for Estimating HIV/AIDS-Prevalence and Its Extension to Multi-Stage. Ann Rev Resear. 2018; 2(3): 555589.

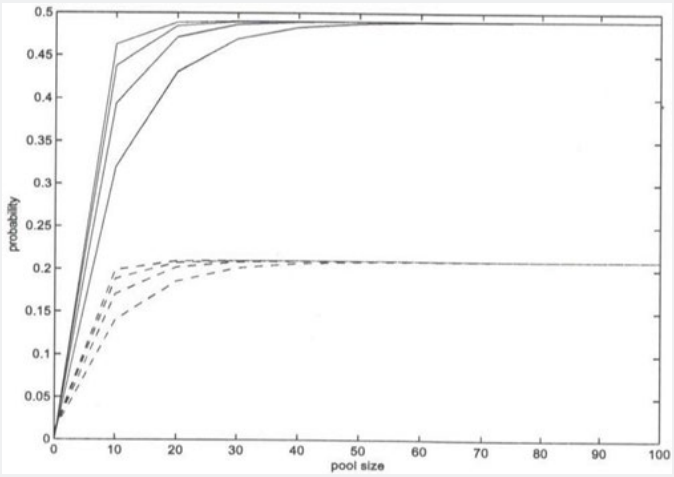

Figure 4: Plots of $\pi_1(p)$ and $\pi_2(p)$ as a function of pool size for η= 0.7,ϕ =0.99 and prevalence p = 0.1; 0.15; 0.2 and 0.25.($\pi_1(p) = \pi_2(p) = \ldots$)

Figures 2 & 3 provides the plots of $\pi_1(p)$, and $\pi_2(p)$ versus the prevalence rate p for some selected pool sizes, sensitivity (η) and specificity (ϕ). From the figures we observe that $\pi_1(p)$, and $\pi_2(p)$ increases with increase in p for $p \le 0.5$. For $\eta = 0.7$ and $\phi = 0.99$, changes in $\pi_1(p)$, and $\pi_2(p)$ remain constant for $p > 0.5$ and $p > 0.4$ respectively for all used group sizes. Also, we observe that $\pi_1(p)$, remains invariant at about $\pi_1(p) \approx 0.5$ while $\pi_2(p)$ remains invariant at about $\pi_1(p) \approx 0.2$ as shown in Figure 3. Notice that $\pi_2(p)$ vanishes with increase in $\eta \to 1$ whereas $\pi_1(p) \to 1$ for all investigated pool sizes. This illustrates that $\pi_2(p)$ is a poor estimator of p when sensitivity and specificity are high as illustrated in (Figure 3 & 4).

Plots of $\pi_1(p)$ and $\pi_2(p)$ versus pool size (k) for some selected η, ϕ, and prevalence rate p are provided in Figure 4 above. For pool size of more than 50 variation in both $\pi_1(p)$ and $\pi_2(p)$ remains constant for $\eta = 0.7$ and $\phi = 0.99$. This suggests the use of small pool sizes as is the case in practice. A third moment estimator of p is

$$\hat{p}_3 = 1 - \left\{ \frac{\eta - \hat{\pi}_1(p) - \hat{\pi}_2(p)}{\phi + \eta - 1} \right\}^{\frac{1}{k}} \tag{12}$$

With $\phi + \eta > 1$. As in the case of $\hat{p}_2$, $\hat{p}_3$ is the worst estimator of the prevalence. This can be checked by plotting $\pi_3(p)$ versus p. It is insensitive to variation in p, making it the worst estimator. Therefore, this rules out two estimators namely $\hat{p}_2$ and $\hat{p}_3$ as possible candidates for estimating the HIV/AIDS prevalence. Only one candidate $\hat{p}_1$, is the ultimate choice for moment estimator of the prevalence and its variance is

$$Var(\hat{p}_1) = \frac{(1-P)^2 \pi_1(p)(1-\pi_1(p))(1-p)^{-2k}}{nk^2(\eta^2 - (1-\phi)^2)^2} \tag{13}$$

The confidence interval (CI) for the estimator of p utilizing $\hat{p}_1$ as the possible candidate estimator is

$$\hat{p}_1 \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{(1-P)^2 \pi_1(p)(1-\pi_1(p))(1-p)^{-2k}}{nk^2(\eta^2 - (1-\phi)^2)^2}} \tag{14}$$

While the pool sizes that can minimize the variance can be computed from

$$\hat{k}_1 = \begin{pmatrix} \arg\min \\ k \end{pmatrix} \left(\frac{(1-P)^2 \pi_1(p)(1-\pi_1(p))(1-p)^{-2k}}{nk^2(\eta^2 - (1-\phi)^2)^2} \right) \tag{15}$$

Solving for k can be easily achieved by in build MATLAB functions as shown in Example 1. Now, writing the likelihood model (8) into the exponential form as

$$L(.) \propto \exp n \log(\pi_1(p)) + x_2 \log\left(\frac{\pi_2(p)}{\pi_1(p)} \right) + (n - x_1 - x_2) \log\left(\frac{1 - \pi_1(p) - \pi_2(p)}{\pi_1(p)} \right) \tag{16}$$

from which (10) can be derived. Also, it can be easily seen that

$$Cov(x_i, x_j) = \begin{cases} n\pi_i(p)(1-\pi(p)), & if \quad i = j \\ -n\pi_i(p)\pi_j & if \quad i \neq j \end{cases} \tag{17}$$

Note that we assumed tests to be independent but we have obtained estimators i.e. the moments-estimators of p that are dependent. That is to say, our estimators (10), (11) and (12) are correlated under some defined conditions. The variance covariance matrix of the moments estimators of p can be obtained by delta method [23] and it is given by

where $1 - \phi = \eta$ or $\eta + \phi > 1$. Also, note that if $1 - \phi > \eta$, we shall have negative correlation otherwise if $(1 - \phi) < \eta$ as is the case in practice. We shall have positive correlation between the probabilities $\pi_i(p)$ and $\pi_j(p), i \neq j$. For the computation of the bounds on the bias, we apply Taylor's series expansion of $\hat{p}_1, \hat{p}_2$ and $\hat{p}_3$ at about $\hat{\pi}_1, \hat{\pi}_2$ and $\hat{\pi}_3$, respectively. The inequalities are obtained as

$$E(\hat{p}_1 - p) \le \frac{(k-1)\,\mathrm{var}(\hat{\pi}_1(p))}{2k^2 n^2 [\eta^2 - (1-\phi)^2]^2} + 0(n^{-2}) \tag{18}$$

$$E(\hat{p}_2 - p) \le \frac{(k-1)\,\mathrm{var}(\hat{\pi}_2(p))}{2k^2 n^2 [\eta(1-\eta) - \phi(1-\phi)]^2} + 0(n^{-2}) \tag{19}$$

How to cite this article: Nyongesa LK. Multiple-Test Pool-Testing Strategy for Estimating HIV/AIDS-Prevalence and Its Extension to Multi-Stage. Ann Rev Resear. 2018; 2(3): 555589.

And

$$E(\hat{p}_3 - p) \le \frac{(k-1)[\mathrm{var}(\hat{\pi}_1(p)) + \mathrm{var}(\hat{\pi}_1(p)) + 2\,\mathrm{cov}(\hat{\pi}_1(p), \hat{\pi}_2(p))}{2k^2 n^2 [\eta + \phi - 1]^2} + 0(n^{-2}) \quad (20)$$

Applying the inequality $E(|x|) \le \{E(x^2)\}^{0.5}$ on our moment estimators of p, we have for p = 0 and $\phi < 1$

$$E(\hat{p}_1) \le \frac{(1-\phi)^2}{k^2 n^2 [\eta^2 - (1-\phi)^2]} + 0(n^{-1}) \quad (21)$$

$$E(\hat{p}_2) \le \frac{\phi(1-\phi)}{kn[\eta(1-\eta) - \phi(1-\phi)]} + 0(n^{-1}) \quad (22)$$

And

$$E(\hat{p}_3) \le \frac{\phi}{kn[\eta + \phi - 1]} + 0(n^{-1}) \quad (23)$$

To compute the confidence interval (CI) of our moment estimators $\hat{p}_1$ and $\hat{p}_2$ we use

$$\hat{p}_1 \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\pi_1(p)(1-\pi_1(p))}{\eta^2 - (1-\phi)^2}}$$

And

$$\hat{p}_2 \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\pi_2(p)(1-\pi_2(p))}{\eta(1-\eta) - \phi(1-\phi)}}$$

respectively. The computation of CI can be easily implemented in statistical software as the estimates and $\hat{p}_2$ can be easily obtained by statistical software.

Maximum Likelihood Estimator and Asymptotic Variance

Several moment estimators of prevalence have been presented in the preceding section. In this section, the objective is to compute a maximum likelihood estimator (MLE) of ρ. The MLE of p is traditionally obtained by maximizing/minimizing log L (.) as provided in (8). If it exists, it is unique. Our MLE of interest is computed from

$$\hat{p} = 1 - \left(\frac{\arg\min}{q} \right)$$

$$\left[\exp n \log(\pi_1(p)) + x_2 \log\left(\frac{\pi_2(p)}{\pi_1(p)}\right) + (n - x_1 - x_2) \log\left(\frac{1 - \pi_1(p) - \pi_2(p)}{\pi_1(p)}\right) \right] \quad (24)$$

The asymptotic variance of our estimator (24) is given by

$$\mathrm{var}(\hat{p}) = \frac{(1-p)^2 \pi_1(p) \pi_2(p) (1 - \pi_1(p) - \pi_2(p))(1-p)^{-2k}}{y} \quad (25)$$

Where

$$y = nx^2 \begin{bmatrix} \pi_2(p)1 - \pi_2(p))\left[\eta^2 - (1-\phi)^2\right] \\ +\pi_1(p)(1-\pi_1(p))\left[\eta(1-\eta) - \phi(1-\phi)\right]^2 \\ +2\pi_1(p)\pi_2(p)\left[\eta^2 - (1-\phi)^2\right][\eta(1-\eta) - \phi(1-\phi)] \end{bmatrix}$$

For individual testing, as is the case in practice, (25) becomes

$${}_1\,\mathrm{var}(\hat{p}) = \frac{[\rho\eta + (1-\phi)(1-p)][p(1-\eta) + \phi(1-p)]}{f}$$

Where f is the population size. Now, with (24) and (25) at hand, the CI for is given by

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{(1-p)^2 \pi_1(p)(1 - \pi_1(p) - \pi_2(p))(1-p)^{-2k}}{y}}$$

With z ~ normal (0.1).

For large sample size, i.e., n→ ∞, we have

$$\sqrt{n(\hat{p} - p)} \to normal(0, A), \quad (26)$$

Where $A = \mathrm{var}(\hat{p})$ as provided by (25). For the MLE of ρ obtained in (24), we have

$$L(\hat{p} \setminus \chi, \eta, \phi, n) \ge L(p \setminus \chi, \eta, \phi, n) \quad (27)$$

For large samples, under the assumption that the distribution tends to normality as given in (26), the estimator is the most efficient for p. Similarly, it can be easily seen that it is a sufficient estimator of p. The bias in an estimation problem is a measure of how much error we have on average. In our estimation, when we use the computed in (24) to estimate the prevalence p, the bias is given by

$$bias(\hat{p}) = E(\hat{P}) - P$$

Notice that the computation of $E(\hat{P})$ from (24) is not that easy. If our estimator was unbiased, then the expected value of the estimator would be $E(\hat{P})$ = p; implying that bias $\hat{p}$ = 0: On the other hand, the mean squared error (MSE) of a given estimator is defined by

$$MSE(\hat{\rho}) = E\left[(\hat{\rho} - \rho)\right]^2 \quad (28)$$

Computation of the bias and MSE in this problem can be accomplished by Monte Carlo simulations.

Example 1: In this example we provide a brief history of HIV and AIDS in Kenya. Between 1983 and 1985, 26 cases of AIDS were reported in Kenya. Sex workers were the first group affected. A study from 1985 reported an HIV prevalence of 59% amongst a group of sex workers in the capital city Nairobi. Towards the end of 1986 there was an average of four new AIDS reported to the World Health Organization each month. This, totaled 286 cases by the beginning of 1987, 38 of which had been fatal. One of the Kenya Government's first responses was to publish informative articles in the press and launch a poster campaign urging people to use condoms and avoid indiscriminate sex. A year later in 1988, the minister of Health announced a year-long health and education programmers, funded by development partners to the tune of sterling pounds 2 million. By 1988 HIV appeared to be spreading rapidly among the population. An estimated 1.2% of adults in the capital city of Nairobi were infected with the virus, and HIV prevalence among pregnant women in the city had increased from 6.5% to staggering 13% between 1989 and 1990. By 2000 an estimated 100,000 people had already died from AIDS. And around 100 in 1000 (1 in 100) people were infected with HIV and AIDS. There have been deterring efforts in the fight of HIV and AIDS in the country and in 1999 it was declared a national disaster and the formation of National AIDS control council that

How to cite this article: Nyongesa LK. Multiple-Test Pool-Testing Strategy for Estimating HIV/AIDS-Prevalence and Its Extension to Multi-Stage. Ann Rev Resear. 2018; 2(3): 555589.

was formed with the sole purpose of mobilizing resources in combating HIV/AIDS [24]. To estimate the prevalence of HIV/AIDS in the Kenyan population using the proposed testing scheme, we assume that the test kits in use have sensitivity and specificity of 95%. The population is pooled into 100 pools each of size 10 via simple random sampling mechanism. From the computed results, we have the corresponding estimators of the HIV/AIDS prevalence as = 0.0869 and = 0.0973. Implying that. Note that the true prevalence rate is 0.1. Hence is a better estimator than, the moment estimator. In this example, we have applied our model in estimating the HIV/AIDS prevalence rate and obtained results within the CI of the true value.

Multistage Estimation Model

In this section, a discussion of a modified Monzon et al. [1] design as shown in Figure 1 is given. The proposed design is a generalization of the model proposed by Monzon et al. [1] this testing strategy is presented in Figure 5 diagrammatically. In the proposed design, we pool the population into n pools and each pool is subjected to testing. If a pool tests negative (-), we drop it from further investigation. If it tests positive (+) on the initial test, it is subjected to a duplicate test. If it tests negative on the duplicate test, it is dropped from further investigation. Otherwise, if it tests positive on the duplicate test, it is split into two smaller pools and the procedure is repeated on the sub-pools as shown in (Figure 5). Clearly, this is a generalization of the preceding discussion. In fact, the design discussed earlier is merely a special case as will be shown. The probability that a sub-pool is declared as positive at stage i is given

$$\frac{[1-(1-p)^{k_i}]\eta^2+(1-p)^{k_i}(1-\varphi)^2}{[1-(1-p)^{k_{i-1}}]\eta^2+(1-p)^{k_{i-1}}(1-\varphi)^2}$$

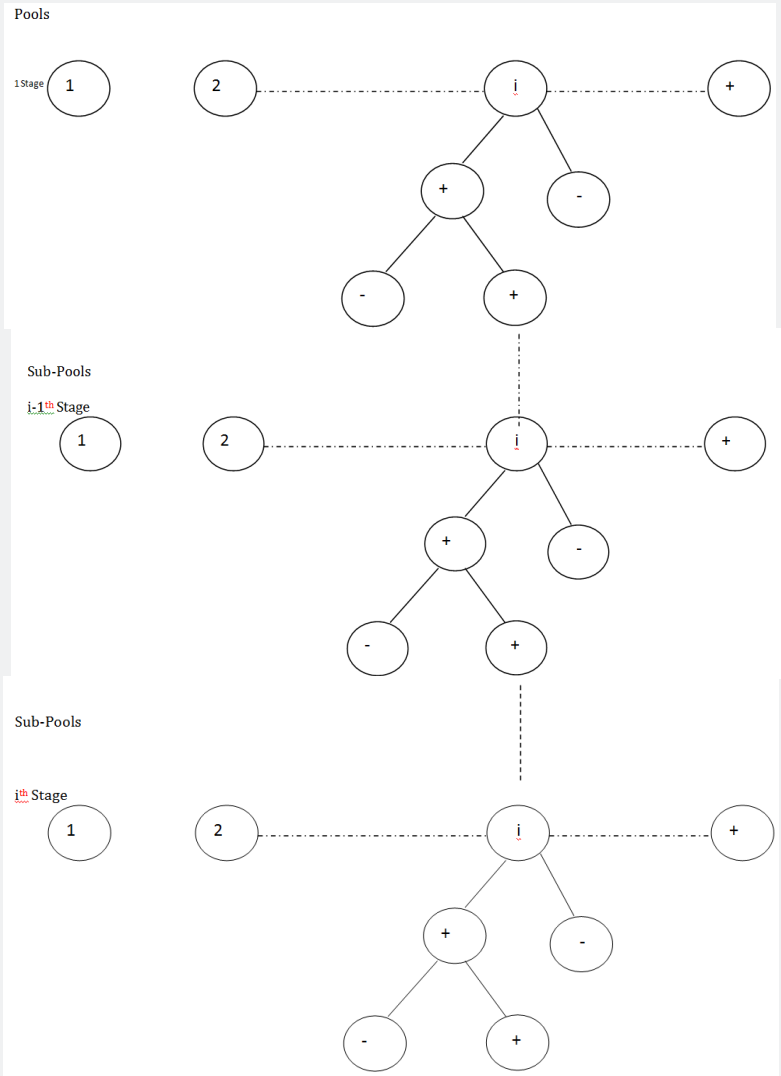

Figure 5: Multi-stage anti-HIV testing strategy of pooled sera using enzyme immune.

Clearly, (29) is a truncated binomial model. The model will play a major role in the formation of our MLE of p at the i^{th} stage. The probability that a sub-pool tests positive on the initial test and tests negative on the duplicate test at the i^{th} stage is

$$\frac{[1-(1-p)^{k_i}]\eta(1-\eta)+(1-p)^{k_i}\varphi(1-\varphi)}{[1-(1-p)^{k_{i-1}}]\eta^2+(1-p)^{k_{i-1}}(1-\varphi)^2} \quad (30)$$

How to cite this article: Nyongesa LK. Multiple-Test Pool-Testing Strategy for Estimating HIV/AIDS-Prevalence and Its Extension to Multi-Stage. Ann Rev Resear. 2018; 2(3): 555589.

another truncated binomial model. Finally, the probability that a sub-pool tests negative at the i[th] stage is given as

$$\frac{[1-(1-p)^{k_i}](1-\eta)+(1-p)^{k_i}\varphi(1-\varphi)}{[1-(1-p)^{k_{i-1}}]\eta^2+(1-p)^{k_{i-1}}(1-\varphi)^2} \quad (31)$$

which is also a truncated binomial distribution. Therefore, the three truncatedbinomial models (29), (30) and (31) are the tools required to develop the likelihood function. It is obvious that the likelihood function will be a truncated multinomial distribution. The a likelihood function at the i[th] stage is given as

$$L_i(p|\underline{x},\eta,\phi) \propto \left[\frac{[1-(1-p)^{k_i}]\eta^2+(1-p)^{k_i}(1-\varphi)^2}{[1-(1-p)^{k_{i-1}}]\eta^2+(1-p)^{k_{i-1}}(1-\varphi)^2}\right]^{x_{1i}} \times \left[\frac{[1-(1-p)^{k_i}]\eta(1-\eta)+(1-p)^{k_i}\varphi(1-\varphi)}{[1-(1-p)^{k_{i-1}}]\eta^2+(1-p)^{k_{i-1}}(1-\varphi)^2}\right]^{x_{2i}} \times \left[\frac{[1-(1-p)^{k_i}](1-\eta)+(1-p)^{k_i}\varphi}{[1-(1-p)^{k_{i-1}}]\eta^2+(1-p)^{k_{i-1}}(1-\varphi)^2}\right]^{x_{1i-1}-x_{1i}-x_{2i}} \quad (32)$$

where $\underline{x}=(x_{1i}, x_{2i}, x_{1i-1}-x_{1i}-x_{2i})'$. It is easy to see that (32) simplifies to

$$L_i(.) \propto \left[[1-(1-p)^{k_i}]\eta^2+(1-p)^{k_i}(1-\varphi)^2\right]^{x_{1i}} \times \left[[1-(1-p)^{k_i}]\eta(1-\eta)+(1-p)^{k_i}\varphi(1-\varphi)\right]^{x_{2i}} \times \frac{[[1-(1-p)^{k_i}](1-\eta)+(1-p)^{k_i}\varphi]^{x_{1i-1}-x_{1i}-x_{2i}}}{[[1-(1-p)^{k_i}-1]\eta^2+(1-p)^{k_i}-(1-\varphi)^2]^{x_{1i-1}}} \quad (33)$$

Thus, for the s-stage multi-stage testing scheme, the likelihood function is

$$L_i(p|\underline{x},\eta,\phi) \propto \prod_{i=1}^{S} L_i(.) \quad (34)$$

with

$$L_1(.) \propto \left[[1-(1-p)^{k_i}]\eta^2+(1-p)^{k_i}(1-\varphi)^2\right]^{x_{1i}} \times \left[[1-(1-p)^{k_i}]\eta(1-\eta)+(1-p)^{k_i}\varphi(1-\varphi)\right]^{x_{2i}} \times \left[[1-(1-p)^{k_i}](1-\eta)+(1-p)^{k_i}\varphi\right]^{x_1-x_{11}-x_{21}} \quad (35)$$

Therefore,

$$L_i(p|\underline{x},\eta,\varphi) \propto L_1(.) \prod_{i=2}^{S} L_i(.) \quad (36)$$

That is

$$L(.) \propto \left[[1-(1-p)^{k_s}]\eta^2+(1-p)^{k_{si}}(1-\varphi)^2\right]^{x_{1s}} \times \prod_{i=1}^{S}\{\left[[1-(1-p)^{k_1}]\eta(1-\eta)+(1-p)^{k_1}\varphi(1-\varphi)\right]^{x_{2i}} \times \left[[1-(1-p)^{k_i}](1-\eta)+(1-p)^{k_i}\varphi\right]^{x_{1i-1}-x_{11}-x_{21}}\} \quad (37)$$

In order to obtain our estimator of interest, we must find a q = 1 − p that maximizes or minimizes (37) or its log likelihood given by

$$\log L(.) \propto x_{1s}\log\left[[1-(1-p)^{k_s}]\eta^2+(1-p)^{k_s}(1-\varphi)^2\right] + \sum_{i=1}^{S}\{\left[x_{2i}\log[1-(1-p)^{k_1}]\eta(1-\eta)+(1-p)^{k_1}\varphi(1-\varphi)\right] + (x_{1i-1}-x_{1i}-x_{2i})\log\left[[1-(1-p)^{k_i}](1-\eta)+(1-p)^{k_i}\varphi\right]\} \quad (38)$$

To obtain the MLE of our estimator in the proposed design, we follow similar argument as in Section 4, the MLE is accomplished by

$$\hat{p} = 1\left(\arg\min_{q}\right) - \{x_{1s}\log\left[[1-(1-p)^{k_s}]\eta^2+(1-p)^{k_s}(1-\varphi)^2\right] + \sum_{i=1}^{S}\{\left[x_{2i}\log[1-(1-p)^{k_1}]\eta(1-\eta)+(1-p)^{k_1}\varphi(1-\varphi)\right] + (x_{1i-1}-x_{1i}-x_{2i})\log\left[[1-(1-p)^{k_i}](1-\eta)+(1-p)^{k_i}\varphi\right]\} \quad (39)$$

Setting s = 1 in (37) and after intensive computation, we obtain the results as in (24). The variance obtained in (25) is the variance of the estimator when s = 1 testing scheme. Next we consider the derivation of the variance in the multistage testing strategy whose technical details are omitted. The asymptotic variance of the multistage estimator obtained in (39) is derived by applying the Crammer Rao lower bound i.e.

$$\text{Var}(\hat{p}) = \left[-E\left(\frac{\partial^2 \log L(.)}{\partial p^2}\right)\right]^{-1} \quad (40)$$

Where

$$\text{Var}(\hat{p}) = -E\left[\frac{\partial^2 \log L(.)}{\partial p^2}\right] = n_s k_s^2(1-p)2k_s - 2\left\{\frac{[\eta^2-(1-\varphi)^2]^2}{\hat{\pi}_{1s}} + (1-\frac{1}{k_s})[\eta^2-(1-\varphi)^2]^2(1-p)^{-k_s}\right\} + \sum_{i=1}^{S} n_i k_i^2(1-p)^{2k_i-2}[\frac{[\eta(1-\eta)-\varphi(1-\varphi)]^2}{\hat{\pi}_{2i}} + \frac{(\eta+\varphi-1)^2}{1-\hat{\pi}_{1i}-\hat{\pi}_{2i}} - (1-\frac{1}{k_i})[\eta^2-(1-\varphi)^2](1-p)^{-k_i}] \quad (41)$$

And η_i is the total number of sub-pools that are tested at the i[th] stage.

Where

$$\text{var}(\hat{p}) = -E\left[\frac{\partial^2 \log L(.)}{\partial \rho^2}\right]$$

and n_i is the total number of sub-pools that are tested at the i[th] stage. In particular, if we set s = 1 in (39) for simplicity, (25) is easily deduced. Thus, we have generalized the design proposed by Monzon et al. [1] and, at the same time, generalized both our estimator and its asymptotic variance. The large-sample property of our estimator can be studied without reference to the loss function in the multi-stage design. Therefore, approximation to or limit of performance measures as the number of pool sizes increases, i.e., n →∞ for the multi-stage model by central limit theorem (CLT) is

$$\sqrt{n}(\hat{p}-p) \to normal\left(0, \text{var}(\hat{p})\right). \quad (42)$$

Where can be derived from (39) and Var $(\hat{p})$ is obtained from (41). Also, the asymptotic relative efficiency (ARE) of the proposed multi-stage testing scheme is

$$ARE_1 = \frac{\text{var}(\hat{p})}{1\,\text{var}(\hat{p})},$$

Asymptotic relative efficiency (ARE) helps in assessing the efficiency of our estimator as compared to one-at-a-time testing procedure. Plots for various values of Var ($\hat{p}$ $_1$Var $(\hat{p})$ versus k are provided in (Figures 6 & 7).

How to cite this article: Nyongesa LK. Multiple-Test Pool-Testing Strategy for Estimating HIV/AIDS-Prevalence and Its Extension to Multi-Stage. Ann Rev Resear. 2018; 2(3): 555589.

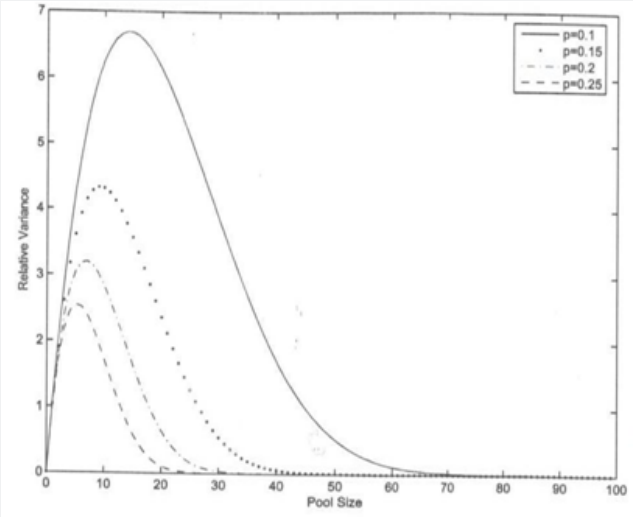

Figure 6: Plots of $Var(\hat{p})$ / var(p) versus pool sizes for various values of p and $\eta = 0.7$ and $\phi = 0.99$.

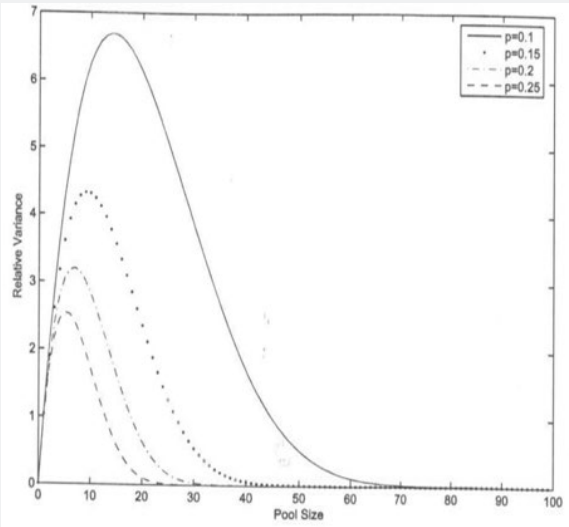

Figure 7: Plots sizes for various values of p and η = 0:7 and ϕ = 0:99.

In Figure 6, we notice that the relative variances increase with increase in k up to some optimal group-size, and then drops. It also reduces with increase in p. Figure 7 shows the effect of a slight change in. In this figure, we observe that the relative variance increases with decrease in relative accuracy of the tests in use. Therefore, repeated testing model of Monzon et al. [1] should be applied in situations where the efficiency of the test kits is low for better results in practice. The multi-stage model outperforms one-at-a-time testing procedure as the AREs > 1: For instance, for only a simple case s = 1; they ARE increases with increase in pool size as it can be easily seen in Figures 6 & 7. It is therefore anticipated that as the number of stages increases the ARE will increase and MSE reduce as demonstrated by Brookmeyer [6] who investigated the problem of multi-stage estimation with perfect tests. Our study generalizes Brookmeyer [6] by introducing the error component as well as the duplicate test as it is the case in practice Monzon et

$$\hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\text{var}\left(\hat{p}\right)}$$

Where Var $(\hat{p})$ is provided in (41) and for simple case where s = 1 is discussed in Section 4.

Results

We have developed estimators based on the screening design proposed by Monzon et al. [1] and extended it to the situation with sub-pooling algorithms. Although the estimator of standard screening problem holds well when the prevalence of the population is low, our approach performs better in slightly higher prevalence population. Our approach is even more robust in situations where the efficiency of the test kits are low, i.e., re-testing is preferred when the efficiency of the test kits are low as observed in the discussion. This result concurs with observations made in practice as alluded to by Sterret [25]. It is in order therefore to conclude that the procedure is viable for developing countries as retesting with the same type of kit is relatively cheap compared to the more expensive test kits like the Western Blot.

How to cite this article: Nyongesa LK. Multiple-Test Pool-Testing Strategy for Estimating HIV/AIDS-Prevalence and Its Extension to Multi-Stage. Ann Rev Resear. 2018; 2(3): 555589.

Our approach is easy to implement. At the same time it is more efficient than individual testing as observed in Figures 6 & 7. Also, the procedure is more efficient in relatively higher prevalence population than the classical pool- screening algorithms as repeated testing improves the efficiency of the testing strategy. Notice that we ruled out two possible candidates for estimating the prevalence namely $\hat{p}_2$ and $\hat{p}_3$. Thus we had only two candidates for consideration for estimating the prevalence, these are $\hat{p}_1$ and the $\hat{p}$. This leads to the question: which one of the two is recommended for this study? To help the discussion we compute the AREs

$$ARE_1 = \frac{(1-\pi_1(p))y}{nk^2(\eta+\phi-1)\pi_2(p)(1-\pi_1(p)-\pi_2(p))} \quad (45)$$

The variable y is provided in (23). Also, it is reported in Tu et al. [13] that the classical MLE for estimating prevalence with imperfect tests in the ordinary Dorfman [2] testing procedure is

$$\hat{P}_G = 1-\left\{\frac{\eta-\hat{\pi}(p)}{\eta+\phi-1}\right\}^{\frac{1}{k}} \quad (46) \quad (46)$$

and its corresponding variance given by

$$Var(\hat{p}_G) = \frac{(1-p)^2\pi(p)(1-\pi(p))(1-p)^{-2k}}{nk^2(\eta+\phi-1)^2} \quad (47)$$

where $\hat{\pi}(p) = x/n, x$ is the pools classified as positive by a single test. Note that $\pi(p)$ is the probability of classifying a pool as positive and is $\pi(p) = \eta(1-(1-p)^k + (1-\phi)(1-p)^k$. It can easily be seen from (12) that if $\pi_1(p) \approx \pi(p)$, then $Var(\hat{p}_1) = Var(\hat{p}_G)$, for $\eta = \phi$, because $(\eta^2-(1-\phi)^2)^2 = (\eta+\phi-1)^2$ for all η and $\phi(0 \le \eta \le 1, 0 \le \phi \le 1)$. If $\eta > \phi$, then $(\eta^2-(1-\phi)^2 > (\eta+\phi-1)^2$ implying that $Var(\hat{p}_1) < Var(\hat{p}_G)$. Hence, in such situations, the estimator $\hat{p}_1$ is superior to $\hat{p}_G$ of Tu et al. [13]. Similarly, if $\eta < \phi$, then $(\eta^2-(1-\phi)^2 < (\eta+\phi-1)^2$, implying that $Var(\hat{p}_1) > Var(\hat{p}_G)$. Hence, in this case, the estimator $\hat{p}_G$ is superior to $\hat{p}_1$ and call for no retesting in such situations. They ARE off $\hat{P}_G$ to $\hat{p}_1$ is

$$ARE_2 = \left(\frac{\eta^2-(1-\phi)^2}{\eta+\phi-1}\right)^2 \frac{\pi(p)(1-\pi(p))}{\pi_1(p)(1-\pi_1(p))} \quad (48)$$

Further, comparing the asymptotic variance of $\hat{p}$ with the classical variance of $\hat{p}_G$, we obtain ARE as

$$ARE_3 = \frac{\pi(p)(1-\pi(p))y}{nk^2(\eta+\phi-1)^2\pi_1(p)\pi_2(p)(1-\pi_1(p)-\pi_2(p))} \quad (49)$$

Simulations of the AREs for various values of $p, k, \eta,$ and ϕ are provided in Table 1 below. For illustration purposes, we have used MATLAB package to compute the AREs in the table (Table 1).

Table 1: Asymptotic relative efficiency for the estimators and of prevalence.

P	0.001	0.01	0.05	0.1	0.15	0.2
ARE_1	1	1	1.01	1.03	1.05	1.08
ARE_2	7.42	1.92	1.16	1.04	0.99	0.95
ARE_3	8.25	2.14	1.3	1.19	1.16	1.14

$k = 5, \eta = \phi = 0.95$

P	0.001	0.01	0.05	0.1	0.15	0.2
ARE_1	1	1	1.03	1.08	1.15	1.26
ARE_2	4.9	1.46	1.05	0.96	0.88	0.8
ARE_3	5.45	1.63	1.2	1.06	1.16	1.12

$k = 10, \eta = \phi = 0.95$

p	0.001	0.01	0.05	0.1	0.15	0.2
ARE_1	1	1	1.01	1.02	1.04	1.08
ARE_2	1.98	1.09	1.01	0.99	0.97	0.93
ARE_3	2.03	1.12	1.04	1.03	1.02	1.02

$k = 10, \eta = \phi = 0.95$

p	0.001	0.01	0.05	0.1	0.15	0.2
ARE_1	1	1.01	1.02	1.05	1.09	1.14
ARE_2	6.75	2.56	1.31	1.09	0.99	0.93
ARE_3	8.45	3.22	1.67	1.44	1.16	1.32

$k = 5, \eta = \phi = 0.90$

p	0.001	0.01	0.05	0.1	0.15	0.2
ARE_1	1	1.01	1.05	1.13	1.23	1.35
ARE_2	5.46	1.84	1.1	0.95	0.84	0.76
ARE_3	6.84	2.32	1.45	1.33	1.29	1.28

$k = 5, \eta = \phi = 0.95$

How to cite this article: Nyongesa LK. Multiple-Test Pool-Testing Strategy for Estimating HIV/AIDS-Prevalence and Its Extension to Multi-Stage. Ann Rev Resear. 2018; 2(3): 555589.

Clearly, the simulated $ARE_1 > 1$; implying that the proposed MLE $\hat{p}$ is more efficient than the moment-estimator $\hat{p}_1$.Thus, $\hat{p}$ is more superior to $\hat{p}_1$ in estimating HIV/AIDS prevalence in an African population where it is believed the prevalence is high. Also, note that $ARE_3 > 1$ and the estimator $\hat{p}_G$ is provided as the MLE of the prevalence in the past literature-i.e., Thompson [8] and Brookmeyer [6] derived $\hat{p}_G$ for perfect tests and Tu et al. [13] derived $\hat{p}_G$ for imperfect test. Therefore, the proposed MLE estimator is superior to the past estimators and can be recommended for screening purposes as opposed to the $\hat{p}_G$ estimator common in pool-testing literature. It can be easily noticed from the tabulated results that the proposed MLE estimator becomes superior to $\hat{p}_G$ when the prevalence of the population is small, calling for re-testing when the prevalence of the population is low. That is, retesting is recommended for high risk population Nyongesa [21]. Also, it is easily seen that the proposed MLE is more efficient in situations where the sensitivity and specificity are low. For $ARE_2 > 1$; meaning that $\hat{p}_1$ is more efficient than $\hat{p}_G$ and this complicates the applicability of $\hat{p}_G$ in estimating HIV/AIDS in low prevalence populations confounded with the fact that there is loss of sensitivity when pooling strategies are applied in screening HIV/AIDS Kline et al. [3] and the loss in sensitivity can be recovered by retesting. The $ARE_2 < 1$ for high prevalence rates such as 0.2 makes $\hat{p}_1$ a poor estimator. For example, if the prevalence of the population is 0.15, the ARE_2 is 0.88 when $k = 10, \eta = \phi = 0.95$, thus for $ARE_2 < 1$ makes $\hat{p}_1$ a poor estimator of prevalence in high prevalence population [26].

Conclusion

This study proposed moment estimators for prevalence estimation and MLE. It has been demonstrated that the proposed ME $\hat{p}_1$ is superior to the studied $\hat{p}_G$ in some situations whereas the MLE $\hat{p}$ outperform both $\hat{p}_1$ and $\hat{p}_G$. More efficient estimators can be realized if multi-stage models are applied. However, multi-stage models are only possible if there enough samples to allow creation of sub-pools. The study only discussed halving multistage model but a generalized multi-stage model that allows creation of any number of sub-pools at any stage can be studied and results compared with the one proposed in this study.

References

1. Mitchel S, Monzon OT, Paladin E, Fem Julia P, Dimaandal E, et al. (1992) Relevance of antibody content and test format in HIV/AIDS testing of pooled sera. AIDS 6(1): 43-48.
2. Dorfman R (1943) The detection of defective members of large population. Annals of Mathematical Statistics 14(4): 436-440.
3. Kline RL, Bothus TA, Brookmeyer R, Zeyer S, Quinn TC (1989) Evaluation of human Immunodeficiency virus sero prevalence in population surveys using pooled sera. J Clin Microbiol 27(7): 1449-1452.
4. Nyongesa LK (2004) Multistage group testing procedure (Group screening) Communication in Statistics-Simulation and computation. 33(3): 621-637.
5. Johnson Nl, Kotz S, Wu X (1991) Inspection errors for attributes in quality control. London; Chapman and Hall, UK.
6. Brookmeyer R (1999) Analysis of multistage pooling studies of Biological specimens for Estimating Disease Incidence and prevalence. Biometric 55(2): 608-612.
7. Gastwirth JL, Hammick PA (1989) Estimation of the prevalence of a rare disease preserving the anonymity of the subject by Group-testing; Application to estimating the prevalence of AIDS antibodies in blood donors. Journal of statistical planning and inference 22(1): 15-27.
8. Thompson KH (1962) Estimation of the population of vectors in a natural population of insects. Biometrics 18(40029): 568-578.
9. Sobel M, Elasho RM (1975) Group-testing with a new goal estimation. Biometrika 62(1): 181-193.
10. Behets F, Bertezzi S, Kasali M, Kashamuka M, Atikala L, et al. (1990) Successful use of pooled sera to determine HIV/AIDS-1 seroprevalence in Zaire with development of cost-effective models. AIDS 4(8): 737-741.
11. Gastwirth JL, Johnson WO (1994) Screening with cost-effective quality control: Potential applications to HIV/AIDS and drug testing. Journal of the American Statistical Association 89(427): 972-981.
12. Hammick PA, Gastwirth JL (1994) Extending the applicability of estimation of prevalence of sensitive characteristics by group testing to moderate prevalence populations. International Statistical Review 62: 319-331.
13. Tu MX, Litrak E, Pagano M (1995) On the in formativeness and accuracy of pooled testing in estimating prevalence of a rare disease: Application to HIV/AIDS screening. Biometrika 82(2): 287-297.
14. Xie M, Tatsuoka K, Sacks J, Young S (2001) Group testing with Blockers and synergism. Journal of American Statistical Association 96(453): 92-102.
15. Tebbs MJ, Swallow HW (2003) Estimating ordered binomial proportions with the use of group testing. Biometrika 90(2): 471-477.
16. Swallow WH (1985) Group testing for estimating infection rates and probability of disease transmission. Phytopathology.
17. Hung M, Swallow (2000) Use of binomial group testing in tests of hypothesis for classification or quantitative covariates. Biometrics 56(1): 319-331.
18. Ding J, Xiong W (2015) Robust group testing for multiple traits with miscalculation. Journal of Applied Statistics.
19. Okoth AW, Nyongesa LK, Kwatch BO (2017) Multi-stage adaptive pool-testing model with Test errors; Improved efficiently. Journal of Mathematics 13(1): 43-55.
20. Matiri G, Nyongesa K, Ali I (2017) Sequentially Selecting Between Two Experiment for Optimal Estimation of a Trait with Misclassification. American Journal of Theoretical and Applied Statistics 6(2): 79-89.
21. Nyongesa LK (2004b) Testing for the Presence of Disease by Pooling Samples. Australian and New Zealand Journal of Statistics 46(3): 383-390.
22. Johnson WO, Pearson LM (1999) Dual Screening. Biometrics 55(3): 867-873.
23. Billingsley P (1995) Probability and measure (3rd edn) John Wiley and Sons Inc, New Jersey, United States.
24. (2005) National AIDS Control Council Kenya National HIV/AIDS Strategic Plan (KNASP).
25. Monzon OT, Palalin FJ, Dimaal E, Balis AM, Samson C, et al. (1957) On the detection of defective members of large population. Annals of Mathematical Statistics 28(4): 1033-1036.
26. Lehmann El, Casella G (1998) Theory of point Estimation (2nd edn) Springer Verlag, New-York Inc, USA.

How to cite this article: Nyongesa LK. Multiple-Test Pool-Testing Strategy for Estimating HIV/AIDS-Prevalence and Its Extension to Multi-Stage. Ann Rev Resear. 2018; 2(3): 555589.

How to cite this article: Nyongesa LK. Multiple-Test Pool-Testing Strategy for Estimating HIV/AIDS-Prevalence and Its Extension to Multi-Stage. Ann Rev Resear. 2018; 2(3): 555589.