

## STEP-WISE GROUP SCREENING STRATEGIES FOR ISOLATING DEFECTIVE ITEMS

M. M. MANENE and F. A. OTIENO

<sup>1</sup>Department of Mathematics,  
Catholic University of Eastern Africa,  
P. O. Box 62157, Nairobi, Kenya,  
Email: [Manene@uonbi.ac.ke](mailto:Manene@uonbi.ac.ke)

**Abstract.** In this paper, we consider the application of step-wise group screening procedure to blood screening in populations with low human immunodeficiency virus (HIV) prevalence rate. Expressions for the saving rate based on the expected number of runs (tests) and on an appropriate cost function considering both the cost of testing and the cost of pooling operations shall be derived. We shall also derive a stopping criteria based on the added cost associated with the operation of pooling. The effect of wrongly specifying the population infection rate shall also be considered.

**Key words:** Step-Wise Group Screening, Initial Step, Subsequent Steps, Blood Bank, Pooling Sera, Number of Runs, Cost Function, Stopping Criteria.

**1. Introduction.** The method of group testing was first introduced by Dorfman (1943), who proposed that instead of testing each blood sample individually for the presence of a rare disease, blood samples be pooled and analyzed together. There have been several other modifications and extensions of Dorfman procedure which has both biological and industrial applications. The notable ones among these include Sterrett (1957), Watson (1961), Patel (1962), Li (1962), Mauro and Smith (1982), Patel and Ottieno (1984), Odhiambo and Patel (1986) and Patel and Manene (1987).

Thompson (1962) used the group testing method and the method of maximum likelihood to estimate the proportion of vectors capable of transmitting auster – yellow virus in a natural population of *Macrostes fascifrons* (Stal) – the six spotted leaf hopper. W. G. Hunter and R. Mezaki (1964) used the group testing method to select the best catalyst from a list of possible catalysts for the oxidation of methane. They stated that by arranging possible catalysts for a reaction in logical groups and testing each group, the less active catalysts can be weeded out and the total number of tests (runs) reduced. Lei Zhu et al. (2001) used pooling experiments as a cost effective approach for screening chemical compounds as part of the drug discovery process in pharmaceutical companies. To complete the decoding process, they augmented the data on pooled testing with information on the chemical structure of compounds. Shou – Jen Lan et al. (1993) considered pooling strategies for screening blood in areas with low immunodeficiency virus (HIV) infection. They based their criteria to end pooling on both savings rate and the relative cost between the preparation and

## STEP-WISE GROUP SCREENING STRATEGIES FOR ISOLATING DEFECTIVE ITEMS

actual test. They restricted their method to the screening procedure suggested by Dorfman (1943) and later extended by Watson (1961) and Patel (1962).

In this paper, we shall describe a procedure for screening by pooling sample to identify seropositive units in an area with low prevalence of HIV infection. We shall base our method on the pooling procedure first describe by Sterrett(1957) and later modified by Patel and Manene (1987). We shall then compare our results with those obtained by Shou – Jen Lan et al. (1993)

**2. Assumptions.** We shall assume that;

- i) All individuals have independently the same probability 'p' of testing positive.
- ii) The sensitivity of the test is the same in various pool sizes.
- iii) The laboratory test is sensitive enough to detect a single infected case in the pooled samples.
- iv) The direction of all effects is the same.

Assumption (iv) ensures that there are no cancellation of effects.

**3.The Step-Wise group screening procedure.** Suppose that we have a population of  $f$  individuals to screen for the presence of a rare disease using a step-wise group screening procedure. The initial step is to partition the  $f$  individuals into  $g$  groups (pools) each of  $k$  individuals so that  $f = kg$ . Each of these groups is tested for its effect. The groups that are found to be negative or unimportant are set aside. In step two, we start with any positive group (pool) and test the individuals in it one by one till we find a positive individual. We set aside individuals that are found to be negative keeping the positive one separate. The remaining unclassified individuals of the group are then tested together in a pool in step three. If the test is negative, the remainder is declared good and the test procedure is complete, otherwise, the remainder forms a new positive group. Actually the test procedure carried out in the initial step and in step two is repeated in subsequent steps successfully till the analysis terminates with a test on a negative group or with a group of size one. Note however that if a negative test is performed on the  $(k-1)^{th}$  individual whilst searching for the first positive in a positive group, we can infer that the  $k^{th}$  individual is the positive one. On the other hand if a positive test is performed on the  $(k-1)^{th}$  individual, the remainders consist only of the  $k^{th}$  individual and only one further test is required.

**4. Saving rate based on the total number of runs.** For Step-Wise group screening with equal prior probabilities and no errors in observations, Patel and Manene (1987) gave the expected number of tests as

$$E(R) = 1 + fp + \frac{2fq}{k} + f - \frac{f}{kp} \{1 - q^{k+1}\} \quad (4.1)$$

Where  $f$  is the number of individuals to be investigated  $p$  is the a-prior probability of an individual testing positive ( $q = 1 - p$ ), and  $k$  is the size of the group (pool) at the initial step. Using the method of finite differences, they obtained the approximate value of  $k$  that minimizes  $E(R)$  as



$$k \approx \left(\frac{2}{p}\right)^{\frac{1}{2}} + \frac{1}{2} + \left(\frac{9}{8}\right)\left(\frac{p}{2}\right)^{\frac{1}{2}} \quad (4.2)$$

upto order  $p$  and the corresponding approximate value of minimum  $E(R)$  as

$$\text{Min } E(R) \approx 1 + f(2p)^{\frac{1}{2}} + \frac{5}{12}fp \quad (4.3)$$

upto order  $p$ .

Let us now define the saving rate function as

$$S_r = \frac{f+1 - \text{min } E(R)}{f+1} \quad (4.4)$$

Where  $f+1$  is the total number of tests (runs) when there is no pooling and  $E(R)$  is the minimum expected total number of runs when pooling is done and the step-wise method used. The one extra run in each case is the control run.

Using the equation (4.3) in (4.4) we obtain

$$S_r = \frac{(f+1) - \left[1 + f(2p)^{\frac{1}{2}} + \frac{5}{12}fp\right]}{f+1} = \frac{f \left[1 - (2p)^{\frac{1}{2}} - \frac{5}{12}p\right]}{f+1} \quad (4.5)$$

For the step-wise design to be more economical than individual testing in terms of number of tests (runs),  $S_r$  must be greater than zero. That is

$$1 - (2p)^{\frac{1}{2}} - \frac{5}{12}p > 0$$

$$25p^2 - 408p + 144 > 0 \quad (4.6)$$

Equation (4.6) simplifies to

$$p < 0.3609 \quad (4.7)$$

Thus we can conclude that the step-wise group screening strategy will be cost effective in terms of number of runs only if the a-priori probability of an individual testing positive 'p' is less than 0.3609. If  $p \geq 0.3609$ , then we shall resort to individual testing.

**5. Added number of preparations.** Before any testing can be done, blood samples from all individuals will have to be prepared. Thus we shall have a total of  $f+1$  preparations where the 1 extra preparation is the control. The  $f$  preparations will then be divided into  $g$  groups each of size  $k$ . A pooled sample is then obtained for each group and tested for its effect. For each positive group, blood samples for individuals in the group will have to be prepared again for retesting using the step-wise screening procedure. The preparation of the samples leads to extra cost on top of the actual cost of testing the sample. We shall therefore consider the added number of preparations when the step-wise group screening procedure is performed.

## STEP-WISE GROUP SCREENING STRATEGIES FOR ISOLATING DEFECTIVE ITEMS

Let  $g'$  be the number of defective groups (pools) isolated in the initial step. Then

$$E(g') = g(1 - q^k) = \frac{f}{k}(1 - q^k) \quad (\text{since } g = \frac{f}{k}). \quad (5.1)$$

Patel and Manene (1987) gave the expected number of tests required to analyze each defective group (pool) as

$$E(R_s^o) = (1 - q^k)^{-1} \left[ k + 1 + kp - 2p - \frac{1}{p}(1 - q^{k+1}) \right] \quad (5.2)$$

Let  $A$  be the added number of preparations then

$$E(A) = E[g'E(R_s^o)] = \frac{f}{k} \left[ k + 1 + kp - 2p - \frac{1}{p}(1 - q^{k+1}) \right] \quad (5.3)$$

using (5.1) and (5.2)

Now  $\frac{1}{p}(1 - q^{k+1}) \approx k + 1 - \frac{k(k+1)}{2}p$  up to order  $p$ . Substituting this in

(5.3) and simplifying we obtain.  $E(A) \approx \frac{fp(k-1)(k+4)}{2k}$ .

**6. Stopping criteria based on the added cost associated with the operation of pooling.** The criteria to stop pooling at  $p \geq 0.3609$  based on the saving on number of tests is appropriate only if the cost on added number of preparations is negligible compared to the test of making an individual test. Testing of individual blood samples from defective pooled samples and re-pooling again leads to seeking new serum of these individuals and mixing the sera in new test tubes before testing. This results to added costs. Due to these added costs, we may need to stop using the stepwise screening procedure when  $p = 0.3609$  has not been reached.

The cost of testing for HIV can be roughly divided into two categories;

- i) The cost of actual testing including test kit and personnel, and
- ii) The cost of preparing samples before testing, including laboratory supplies and personnel. These two types of cost can vary from place to place and from period to period. In this section, we shall take into consideration the relative magnitude of the two types of cost.

Denoted by  $c_1$  the cost of testing per unit and by  $c_2$  the cost of preparation per unit. Suppose that  $c_2 = c_1 W$ , where  $0 \leq W \leq 1$ . When testing is done without pooling, the total cost  $C$  is given by

$$C = (c_1 + c_2)(f + 1) \quad (6.1)$$

When the step-wise group screening procedure is used, the total cost is given approximately by

$$\begin{aligned}
 C^* &\approx c_1[E(R)] + c_2[(f+1) + E(A)] \\
 &\approx c_1\left[1 + f(2p)^{\frac{1}{2}} + \frac{5}{12}fp\right] + c_2\left[f + 1 + \frac{fp(k-1)(k+4)}{2k}\right] \quad (6.2)
 \end{aligned}$$

using (4.3) and (5.4). The step-wise group screening procedure will be preferred to individual testing when  $C - C_1^* > 0$ .

That is when

$$c_1f\left[1 - (2p)^{\frac{1}{2}} - \frac{5}{12}p\right] - c_2f\left[\frac{p(k-1)(k+4)}{2k}\right] > 0 \quad (6.3)$$

Putting  $c_2 = c_1W$ , inequality (6.3) reduces to

$$\begin{aligned}
 &\left[1 - (2p)^{\frac{1}{2}} - \frac{5}{12}p\right] - W\left[\frac{p(k-1)(k+4)}{2k}\right] > 0. \\
 \text{That is } W &< \frac{k\left[2 - (8p)^{\frac{1}{2}} - \frac{5}{6}p\right]}{p(k^2 + 3k - 4)} \quad (6.4)
 \end{aligned}$$

An approximate upper bound for  $W$  can be obtained by using the approximate optimum group size given by

$$k \approx \left(\frac{2}{p}\right)^{\frac{1}{2}} + \frac{1}{2} + \left(\frac{9}{8}\right)\left(\frac{p}{2}\right)^{\frac{1}{2}} \quad (\text{c.f. (4.2)})$$

The cost adjusted percentage saving rate

$$S_r(CA) = \frac{f\left[(2p)^{\frac{1}{2}}Wf[p(k-1)(k+4)]\right]}{(1+W)(f+1)} \times 100 \quad (6.5)$$

If  $W$  does not satisfy inequality (6.4), then we shall resort to individual testing.



**STEP-WISE GROUP SCREENING STRATEGIES FOR ISOLATING DEFECTIVE ITEMS**

7. **Effects of mis-specifying population infection rate.** Suppose the true infection rate is  $p$ . The number of groups (pools) in the initial step will then be

$$g = \frac{f}{k} \approx \frac{f}{\left(\frac{2}{p}\right)^{\frac{1}{2}} + \frac{1}{2} + \left(\frac{9}{8}\right)\left(\frac{p}{2}\right)^{\frac{1}{2}}} \quad (7.1)$$

using (4.2).

In practice the exact true infection rate is not known. If we erroneously assumed the infection rate to be  $p'$ , the number of groups (pools) in the initial step will then be

$$g_e = \frac{f}{k'} \approx \frac{f}{\left(\frac{2}{p'}\right)^{\frac{1}{2}} + \frac{1}{2} + \frac{9}{8}\left(\frac{p'}{2}\right)^{\frac{1}{2}}} \quad (7.2)$$

It should be noted that after the initial step laboratory tests, the infection rate can be estimated exactly and over-pooling or under pooling corrected. The estimator  $\hat{p}$  of  $p$  is obtained by maximizing the function.

$$L(g') = \frac{g!}{g!(g-g')!} p^{*g'} (1-p^*)^{g-g'} \quad (7.3)$$

Where  $p^* = 1 - q^k$ ,  $g$  is the total number of pools at the initial step and  $g'$  is the observed number of positive pools based on the laboratory test results. Alternatively we can obtain  $\hat{p}$  through the following reasoning. The probability that an initial step group (pool) is defective is  $p^* = 1 - q^k$

Thus from the observed positive rate

$$\frac{g'}{g} = 1 - q^k.$$

This implies

$$q^k = 1 - \frac{g'}{g} = \frac{g''}{g}. \text{ That is } (1-p)^k = \left(\frac{g''}{g}\right).$$

Which imply  $p = 1 - \left(\frac{g''}{g}\right)^{\frac{1}{k}}.$  (7.4)

Note that  $g''$  is the number of pooled blood samples found to be non-defective at the initial step. The percentage total saving rate is given by

$$S_r\left(\frac{o}{o}\right) = \frac{f+1-E(R)}{f+1} \times 100. \quad (7.5)$$

The difference  $d$  in the total saving rate due to mis-specifying  $p$  the proportion defective is obtained as

$$d = S_{rm}(\circ/\circ) - S_r(\circ/\circ). \quad (7.6)$$

Where  $S_{rm}(\circ/\circ)$  is the percentage saving rate when  $p$  is mis-specified and  $S_r(\circ/\circ)$  is percentage saving rate with the correct  $p$ .

Table 1 below shows the cost adjusted saving rate when the cost of preparation is included. The value of  $k$  is as given in (4.2) and the value of  $S_r(CA)$  percentage is as given in (6.5) for  $W=0.1$ . The upper bound for  $W$ ,  $W_U$  is also given. From this table it is easy to see that for  $p > 0.3$ , the upper bound for  $W$ ,  $W_U$  is less than 0.1. Thus the use of step-wise screening procedure should be stopped and we resort to individual testing.

**STEP-WISE GROUP SCREENING STRATEGIES FOR ISOLATING DEFECTIVE ITEMS**

**Table 1**

P	k	S <sub>r</sub> (CA)%	W <sub>c</sub>
0.001	45.2	87.02	39.66
0.01	14.7	78.46	9.79
0.10	5.2	49.85	1.37
0.11	5.0	47.71	1.22
0.12	4.9	45.66	1.09
0.13	4.7	43.68	0.98
0.14	4.6	41.76	0.88
0.15	4.5	39.90	0.79
0.16	4.4	38.10	0.71
0.17	4.3	36.34	0.64
0.18	4.2	34.62	0.58
0.19	4.1	32.95	0.52
0.20	4.0	31.31	0.47
0.22	3.9	28.13	0.38
0.24	3.8	25.07	0.30
0.26	3.7	22.11	0.23
0.28	3.6	19.25	0.17
0.30	3.5	16.46	0.12
0.32	3.5	13.75	0.079
0.34	3.4	11.11	0.04
0.35	3.4	9.82	0.02
0.36	3.3	8.54	0.002
0.361	3.3	8.41	-0.0001
0.362	3.3	8.28	-0.002

Table II (a) below shows for the proposed procedure the pool size, saving rate on the total number of tests, added number of preparations, and the total number of preparations needed for different estimated infection rates  $\hat{p}$  in a population with a true infection rate  $p = 0.08$ . The adjusted cost saving rate is based on



$W = \frac{c_2}{c_1} = 0.1$ . Table II (b) is the equivalent of Table II (a) for the procedure proposed by Shou - Jen Lan et al. (1993).

**Table II (a)** The number of individuals to be tested  $f = 10,000$ .

$\hat{p}$	k	E(R)	$S_r$ (%)	E(A)	$S_r$ (CA) (%)
	1	10,001	0	0	0
0.25	3	5117	48.83	1783	49.81
0.125	4	4747	52.53	2246	49.32
0.08	4	4640	53.60	2639	48.89
0.05	5	4737	52.63	3307	48.08
0.04	8	4854	51.47	3603	47.69
0.025	10	5136	48.64	4135	46.93
0.02	11	5288	47.12	4378	46.55
0.01	15	5889	41.11	5222	45.06
0.005	21	6675	33.25	6198	42.85
0.0025	29	7478	25.22	7133	39.93
0.0004	71	9304	6.97	9162	24.62

**Table II (b)**  $f=10,000$  using the procedure proposed by Shou – Jen Lan et al. (1993)

$\hat{p}$	k	E(R)	$S_r$ (%)	E(A)	$S_r$ (CA) (%)	Total Stages needed
	1	10001	0	0	0	1
0.25	2	6537	34.33	1536	33.10	2
0.125	3	5548	44.52	2214	42.32	2
0.08	4	5537	46.63	2836	43.80	2
0.05	5	5019	49.81	5291	44.53	3
0.025	7	4897	51.03	6416	44.62	3
0.01	10	4964	50.36	7736	42.63	3
0.0025	20	5291	47.09	10880	36.22	3
0.0004	50	5496	45.04	12682	32.37	3

## STEP-WISE GROUP SCREENING STRATEGIES FOR ISOLATING DEFECTIVE ITEMS

From the two tables II(a) and II(b) it is evident that over estimating the pool size increases the number of runs more rapidly in the step-wise than in the procedure proposed by Shou – Jen Lan et al. However the step-wise procedure requires fewer added number of preparations when the value of  $p$  is underestimated than the procedure proposed by Shou- Jen Lan et al. the cost adjusted saving rate with  $W= 0.1$  is higher for the step-wise procedure for  $p \geq 0.0025$  than the procedure proposed by Shou-Jen Lan et al. It should be noted that tables II (a) and II (b) are only for illustrative purpose and are not exhaustive.

**8. Application.** To illustrate how well the saving costs are achieved by the proposed step-wise procedure, we apply the procedure to two selected practical examples.

**Example 1.** An HIV infection rate of about 8% has been reported among children in Romania, Rudin, C., et al. (1990). Suppose that blood samples of 10,000 children in such an area are to be tested and that the true infection rate is 8%. To identify the children who are HIV positive, we first note that  $p = 0.08 < 0.3609$  and from table I cannot be affected by  $W$ . With  $p=0.08$ , the group (pool) size is  $k=5$ . The number of pooled samples ( $g$ ) to be formed at the initial step is

$$g = \frac{f}{k} = \frac{10,000}{5} = 2,000$$

The total number of tests (runs) to be performed will be

$$\begin{aligned} E(R) &= 1 + fp + \frac{2fq}{k} + f - \frac{f}{kp} \{1 - q^{k+1}\} \\ &= 1 + 10,000 \times 0.08 + 2(2000)0.92 + 10,000 - \frac{2000}{0.08} \{1 - 0.92^{5+1}\} \\ &= 4639.875 \approx 4640 \end{aligned}$$

The percentage total saving rate from pooling is

$$S_r(\%) = \frac{10,001 - 4640}{10,001} \times 100\% = 53.6\%$$

**Example 2.** The Central Bureau of Statistics Ministry of Planning and National Development “Kenya Demographic Health Survey (2003)” reports that the HIV infection prevalence rate in general population in Kenya is approximately 0.067.

Suppose that blood banks in such an area have together 1000000 units of blood to be examined and that there are 67,000 HIV positive units. The true HIV positive prevalence rate being 0.067. We illustrate how to screen this blood by pooling using the proposed procedure.

First let us assume incorrectly that  $\hat{p} = 0.10$ . Since  $\hat{p} = 0.10 < 0.3609$ , we shall have some saving in the number of tests (runs). For  $\hat{p} = 0.10$ , the



optimum group size is  $k=5$  (c.f. Patel and Manene 1987). The number of groups (pools) formed at the initial step is

$$g = \frac{100000}{5} = 200,000$$

The expected total number of test will be

$$E(R) = g + 1 + g \left[ k + 1 + kp - 2p - \frac{1}{p} \{1 - q^{k+1}\} \right]$$

$$= 200001 + 200000 \left[ 6 + 0.067 \times 5 - 2 \times 0.067 - \frac{1}{0.067} \{1 - 0.933^6\} \right]$$

$$= 424123.4$$

The percentage total saving rate is

$$S_r(\%) = \frac{1000001 - 424123.4}{1000001} \times 100\% = 57.59\%$$

If the correct prevalence rate  $p = 0.067$  was used then the optimum pool size  $k = 6$  and the expected total number of tests is given by

$$E(R) = 1 + fp + \frac{2fq}{k} + f - \frac{f}{kp} \{1 - q^{k+1}\}$$

$$= 1 + 0.067 \times 1,000,000 + \frac{2 + 100000 \times 0.93}{6} + 1000000$$

$$- \frac{100000}{0.067 \times 6} \{1 - 0.0933^7\} = 421334.0$$

The percentage total saving rate is

$$S_r(\%) = \frac{1000001 - 421334}{1000001} \times 100\% = 57.87\%$$

The difference in saving rate due to mis-specifying  $p$  is  
 $(57.87 - 57.59)\% = 0.28\%$

The saving rate decreases by 0.28% when we use  $\hat{p} = 0.10$  to partition the blood into pools of size 5 in the initial step instead of using the correct value  $p=0.067$  leading to pools of size 6.

It should be noted that even when we approximate the prevalence rate, there is a substantial saving if a group testing procedure is used.