Biostatistics Department Technical Report

BST2009-002

Two-sided Hypothesis Testing Based on Pool Screening with Unequal Pool Sizes

Hongjiang Gao
Inmaculada Aban, PhD
Charles R. Katholi, PhD

September 2009

Department of Biostatistics
School of Public Health
University of Alabama at Birmingham
roberton@uab.edu

Two-sided Hypothesis Testing Based on Pool Screening with Unequal Pool Sizes

Hongjiang Gao, Inmaculada Aban, Charles R. Katholi

University of Alabama at Birmingham, Department of Biostatistics
Ryals Public Health Building, Suite 327, Birmingham, AL 35294

SUMMARY

Pool screening is a widely used design which provides an efficient way to estimate prevalence in vector-borne infectious disease control when the prevalence is small. Laboratory screening tests may only have the capability of handling pool sizes up some maximum value. If a pool has size larger than this maximum value, it needs to be subdivided into smaller pools so that the new pool sizes meet the requirements of the screening test. This leads to the problem of analyzing data based on unequal pool sizes. We propose and compare procedures for statistical hypothesis testing under the setting of unequal pool sizes assumed to be fixed and known. The hypothesis testing procedures considered are: (1) an exact test based on the sum of positive pools, and (2) likelihood-based test procedures. Because the asymptotic distributions of these likelihood-based tests are far from the expected Chi-squared distribution when the prevalence is small, we show that using the simulated quantiles of these likelihood-based statistics to define the new rejection region improves the performance of these tests. In the end, the exact test based on the sum of pool sizes outperforms the other tests with regard to power particularly when the prevalence is close to zero.

Key words: Pool Screening, Likelihood Ratio Test, Statistical Power, Asymptotic Distribution, Maximum Likelihood Estimate

1. Introduction

The terms “pool screening” and “group testing” are used in the literature by different authors, but they both refer to procedures that test subjects in pools or groups instead of individually. In this case subjects may be insects, virus, blood samples, chemical agents etc. The pool screening testing procedure is usually implemented when the proportion of positive subjects is very low (for instance, rare disease with prevalence less than 0.1%). The outcome of pool testing is either positive or negative. When the outcome of pool testing is negative, then all the subjects in this pool are declared negative. When the result of pool screening is positive, then one or more subjects in this pool are positive. The goal of pool screening can be the efficient classification of individuals as positive or negative or estimating the probability of individual subject being positive in whole population.

Even though earlier implementation of pool screening can be found in Marion’s(1936) research, Dorfman (1943) is often credited as the first person who discussed it in the statistics literature. The motivation of Dorfman’s work was identifying syphilitic antigen positive individuals among army man by pool screening of blood samples. Many of the statistical aspects of pool screening were widely investigated later. Retesting schemes that were explored mainly focused on improving efficiency of classifying all the positive and negative subjects (Sterrett, 1957; Milton and Groll, 1966; Chen and Swallow, 1990; Hsu, 1995). Test accuracy concerns came mostly from HIV research (Kline et al., 1989; Tu et al.,1995; Wein and Zenios, 1996). Farrington (1992) recommended generalized linear models to handle covariates. Hepworth (1996) investigated exact confidence intervals given several pool screening stages where each stage has a different pool size. Barker (2000) considered the case where the pool sizes are unequal and follow no special pattern in size.

Besides the above mentioned statistical development in pool screening, estimating the probability (denoted by p) of a subject being positive is one of the primary purposes of statistical inference. One commonly used estimator is the minimum infection rate (MIR) which is calculated as the fraction of number of positive pools over total number of subjects screened. Gu et al (2003) cautioned that this estimator will underestimate the true infection rate when positive pools contain more than one positive subject. Another estimator is the maximum likelihood estimator (MLE) which can be expressed as $\hat{p} = 1 - \left(1 - \frac{T}{M}\right)^{\frac{1}{K}}$ where K denotes the common pool size and T is the sum of positive pools. Tu et al. (1995) and Barker (2000) showed that first order approximations to the bias and variance of this estimator are

$$\text{Bias} = \frac{(K-1)\left[1-(1-p)^K\right](1-p)^{1-K}}{2MK^2}, \quad \text{Var}(\hat{p}) = \frac{1-(1-p)^K}{MK^2(1-p)^{K-2}},$$

where M denotes total number of pools. Observe that if the pool size is greater than 1, the bias is positive and so the MLE is on the average is an overestimation. Also, for a fixed pool size, bias, variance and hence mean square error (MSE) decrease as the number of pools increases. Thus the MLE converges in probability to p as the number of pools goes to infinity, i.e., the MLE is a consistent estimator of p. Finally, both bias and variance increase with p if the number of pools and the pool size are held constant.

Determining the appropriate pool size is very important in pool screening. Chiang and Reeves(1962) suggested a formula $K = \frac{\log(1/2)}{\log(1-p)}$ to compute the pool size with the aim of having half positive and half

negative pools. Thompson (1962) proposed pool size formula $K = \frac{1.5936 - p}{p}$ which minimizes the MSE. Katholi and Unnasch (2006) pointed out that for rare event where *p* is very small, the above formulae usually provide much larger pool size than can be handled in an actual lab screening test. That is, the chemistry of the test procedure places restrictions on the size of the pools. Consequentially, in practice collected subjects are subdivided into smaller samples that satisfy the requirements of the laboratory screening test. Therefore in determining pool sizes both statistical and practical requirements need to be considered. In *Wuchereria bancrofti* infection control, polymerase chain reaction (PCR) technique can be employed to detect up to 40 female mosquitoes in a pool (Helmy et al., 2004; Goodman et al., 2003;Vasuki et al., 2003;Williams et al., 2002); and in Onchocerca *volvulus* infection control program, most literature uses PCR assay method that can handle no more than 50 black flies in a pool(Yameogo et al., 1999; Guevara et al., 2003).

Hypothesis testing is another important aspect of statistical inference. Especially for disease eradication programs such as the Onchocerciasis (river blindness) Control Program in Africa. After several years' effort, hypothesis testing can be utilized to determine the progress of disease control and continuation of the program. However, as the prevalence, *p,* decreases and approaches zero, researchers must process very large number subjects because the probability of a pool being negative increases rapidly, and only a very small fraction of pools will turn out positive. Generally, it is believed that there is a level of prevalence at or below which transmission ceases. Hence testing a hypothesis of the kind $p \le p_0$ is essential. Hence it remains statistically challenging and practically crucial to investigate and compare different hypothesis testing procedures. There is a scarcity of articles in the literatures discussing statistical test and its power. Katholi (2007) summarized pool screening hypothesis testing under equal pool size situation. Tebbs and Mccann (2007) explored large sample, likelihood ratio based hypothesis tests for data stratified by categorical variable such as gender etc.

The aim of this paper is to develop and investigate two-sided exact and asymptotic tests in the unequal pool size situation. Model setting, distributional properties, and computational issues of the number of positive pools will be discussed in Section 2. Section 3 will focus on hypothesis testing procedures. Comparisons of the testing procedures in terms of statistical power will be discussed in a simulation study in Section 4. Finally, limitations and recommendations will be discussed in the Section 5.

2. Properties and Computational Method for the Distribution of Number of Positive Pools

The number of positive pools will be the basis for the exact test proposed in Section 3. In order to be able to properly use this statistic in developing inferential procedures, it is important to understand its distribution. We start by first stating the model and notations.

2.1 Model Setting

Assume all individual subjects within the same pool and between pools are independent and identically distributed (i.i.d.). Furthermore, assume that the screening test used has perfect sensitivity and specificity. Suppose $x_1, x_2, ..., x_M$ are pool testing results of *M* pools of sizes $n_1, n_2, ..., n_M$, where

$$x_i = \begin{cases} 1, & \text{if i}^{\text{th}}\text{pool tests positive} \\ 0, & \text{if i}^{\text{th}}\text{pool tests negative} \end{cases}$$

Let p denote the probability of an individual in the population to be positive and the parameter of interest. Given i^{th} pool, the probability that the pool tests negative is $(1-p)^{n_i}$. Since one or more positive individuals in i^{th} pool will make pool positive, the probability that the i^{th} pool testing positive is $1-(1-p)^{n_i}$. In this case, the random variable X_i has *Bernoulli* distribution given by

$$f(x_i \mid p, n_i) = \left[1-(1-p)^{n_i}\right]^{x_i}\left[(1-p)^{n_i}\right]^{1-x_i}, 1 \le n_i \le N_{max}, x_i \in \{0,1\} \qquad (1)$$

2.2 Distribution and Basic Properties for Number of Positive Pools

Let $T = \sum_{i=1}^{M} X_i$ denote number of positive pools, then the probability mass function of T is

$$g(T=t \mid n_i, M) = \left[(1-p)^{\sum_{i=1}^{M} n_i}\right] \sum_{x \in \Omega_t} \prod_{i=1}^{M} \left[\frac{1-(1-p)^{n_i}}{(1-p)^{n_i}}\right]^{x_i}, 0 \le T \le M \qquad (2)$$

where Ω_t defined as $\Omega_t = \left\{ X \mid x_i \in \{0,1\}, i = 1,2,...,M, \sum_{i=1}^{M} x_i = t \right\}$ and $0 < p < 1$.

Wang (1993) provided several other expressions in a more general setting. Barker (2000) derived the above distribution in pool screening. Clearly, when all M pools have sizes equal to a constant K, pools are independent identically distributed (i.i.d.). Then T is distributed as *Binomial* $(M, 1-(1-p)^K)$ which is a special case of equation (2).

Moment and cumulant generating functions provide convenient ways to compute the moments of a distribution as well as other important properties like symmetry and kurtosis. Cumulants also play an important role in obtaining an asymptotic approximation to the distribution of a test statistic in this case, any test statistic which is a function of T. The following theorem gives us the moment and cumulant generating functions of the statistic T.

Theorem 1: Moment generating function and cumulant generating function of T are respectively given by

$$M_{\sum_{i=1}^{M} x_i}(t) = \prod_{i=1}^{M} \{[1-(1-p)^{n_i}]e^t + (1-p)^{n_i}\},$$

$$H_{\sum_{i=1}^{M} x_i}(t) = \sum_{i=1}^{M} \log\{[1-(1-p)^{n_i}]e^t + (1-p)^{n_i}\}.$$

First four cumulants, skewness and kurtosis are

1st cumulant $\quad k_1 = E(T) = \sum_{i=1}^{M}\left[1-(1-p)^{n_i}\right]$

2nd cumulant $\quad k_2 = Var(T) = \sum_{i=1}^{M}\left[1-(1-p)^{n_i}\right]\left[(1-p)^{n_i}\right]$

3rd cumulant $\quad k_3 = \sum_{i=1}^{M}\left[1-(1-p)^{n_i} - 3(1-(1-p)^{n_i})^2 + 2(1-(1-p)^{n_i})^3\right]$

4th cumulant $$k_4 = \sum_{i=1}^{M}(1-(1-p)^{n_i})(1-p)^{n_i}\left[1-6(1-p)^{n_i}+6(1-p)^{2n_i}\right]$$

Skewness $$\eta_1 = \frac{\sum_{i=1}^{M}\left\{\left[1-(1-p)^{n_i}\right]\left[2(1-p)^{n_i}-1\right]\left[(1-p)^{n_i}\right]\right\}}{\left\{\sum_{i=1}^{M}\left[(1-p)^{n_i}-(1-p)^{2n_i}\right]\right\}^{\frac{3}{2}}}$$

Kurtosis

$$\eta_2 = \frac{\sum_{i=1}^{M}\left[(1-(1-p)^{n_i})(1-p)^{n_i}\right]\left[1-6(1-p)^{n_i}+6(1-p)^{2n_i}\right]}{\left\{\sum_{i=1}^{M}\left[(1-p)^{n_i}-(1-p)^{2n_i}\right]\right\}^{2}}$$

The proof of this Theorem is straightforward, and an outline is given as follows: Since all subjects are assumed i.i.d, then pools are independently distributed, therefore the moment generating function of *T* is equal to the product of the moment generating function of each pool. The cumulant generating function can be easily found by taking the logarithm of the moment generating function. All the other results follow immediately by using their respective definitions.

First counterintuitive fact from above results is the variance of *T*. If let $\pi_i = 1-(1-p)^{n_i}$, then $\bar{\pi} = \frac{\sum_{i=1}^{M}\left[1-(1-p)^{n_i}\right]}{M}$ and $\text{var}(T) = M\bar{\pi}(1-\bar{\pi}) - \sum_{i=1}^{M}(\pi_i-\bar{\pi})^2$ .Because $\sum_{i=1}^{M}(\pi_i-\bar{\pi})^2 \geq 0$, then $\text{var}(T) \leq M\bar{\pi}(1-\bar{\pi})$. This implies that the maximum value of the variance of *T* is achieved when pool sizes are equal. Nedelman (1986) generalized this fact in other distributions.

Second noticeable fact is the sign of the skewness. Given n_i known, it is apparent that $2(1-p)^{n_i}-1$ determines the sign of each term within the summation of η_1 since $1-(1-p)^{n_i} > 0$, $(1-p)^{n_i} > 0$ and denominator greater than zero. When *p* is very small, it is possible that $(1-p)^{n_i} > \frac{1}{2}$ for all $i \in \{1,2,...,M\}$, and the distribution of *T* is right skewed. As *p* increases, more terms of $(1-p)^{n_i}$ will be less than $\frac{1}{2}$ and sign of η_1 will change from positive to zero to negative. Consequently, the distribution of *T* will change from being positively skewed to symmetrical to negatively skewed . However, there might be few positive or negative terms that dominate the other terms given very different pool sizes. Under this situation, the above observation regarding skewness could not be generalized.

When pool sizes are equal, it can be shown that *T* is a sufficient statistics of *p* and that it has the monotone likelihood ratio (MLR) property. Thus, a one-sided hypothesis test based on *T* regarding *p* is an uniformly most powerful (UMP) test by Karlin-Rubin Theorem (see for instance, Casella and Berger, 2001). In the next theorem, it will be shown that *T* is no longer sufficient for *p* when the pool sizes are unequal

although it still possesses the MLR property. Thus, it is not necessarily true that a one-sided test for *p* based on *T* is UMP.

Theorem 2: When pool sizes are unequal, the statistics *T* is not a sufficient statistics of *p*
It is not difficult to prove above theorem by using the definition of sufficient statistics or by factorization theorem.

Remark: However, by applying Marcus and Lopes (1957) inequality, Huynh (1994) showed that sum of independent and non-identical *Bernoulli* random variables still possess the MLR property under the condition of success probability of each trial $P_j(\theta)$ is a non decreasing function of θ. This condition is automatically satisfied in pool screening setting because $1-(1-p)^{n_i}$ is a monotone increasing function of *p*, thus statistics *T* has MLR property.

2.3 Computation of Probability Mass Function of *T*

Exact test based on *T* requires computing probabilities associated with different values of *T*. Most statistical software can easily compute the distribution of *T* when the pool sizes are equal by applying the *Binomial* distribution. However, when the pool sizes are unequal, different pools have different probabilities of being positive, and alternative methods need to be explored to compute the distribution of *T* before one can make further statistical inference. Several different computational methods will be proposed and compared in this subsection.

2.3.1 Enumeration and Saddle Point Method

The most obvious way to compute the distribution of *T* (that is, $P(T=t), t=0,1,\cdots,M$) is exhaustively enumerating all $\binom{M}{t}$ possible combinations in the second factor of equation (2) for each value of *t*. However, this method is extremely tedious because the total number of arithmetic operations in the sum requires *M*-1 multiplications. Its computational complexities increase exponentially with *M*. Practical experience shows that the enumeration method is not applicable when the total number of pools is much larger than 25. Consider the case where there are total 34 pools having different pool sizes. By equation (2), the maximum number of combination terms within the summation is $\binom{34}{17} = 2,333,606,220$ which already exceeds 2^{31}-1=2,147,483,647, the commonly used largest exact integer based on a 32 bit floating point arithmetic according to IEEE standard.

Due to the deficiency of above direct computation, Barker (2000) explored the use of the saddlepoint approximation method to calculate distribution of *T* . The saddlepoint approximation is usually applied when there is no close form for the probability density (or mass) function but the moment generating function is known or when the probability has close form but is not easy to compute. Daniels(1954) first approached this problem by using inversion of Fourier transformation. Goutis and Casella (1999) had an excellent tutorial review on this method and simplified this method into several steps. Barker (2000) combined saddle point approximation and enumeration method in unequal pool size screening using Fortran, where exact

enumeration method is used to calculate PMF at the two ends when t=0,1,2,M-2,M-1,M. And saddle point method is applied to calculate distributions when $3 \le t \le m-3$.

2.3.2 Recursive Method

From equation (2), note that distribution of T could be calculated as $g(T=t \mid p) = c(p,n_i)S_t$ where $c(p,n_i) = (1-p)^{\sum_{i=1}^{M} n_i}$, $S_t = \sum_{x \in \Omega_t} \prod_{i=1}^{M} \left[\frac{1-(1-p)^{n_i}}{(1-p)^{n_i}} \right]^{x_i}$. If let $a_i = \left[\frac{1-(1-p)^{n_i}}{(1-p)^{n_i}} \right]^{x_i}$, then S_t is the expression of coefficients of elementary symmetric polynomials. For example, $S_0 = 1$, $S_1 = a_1 + a_2 + ... + a_M$, $S_M = a_1 a_2 ... a_M$ Furthermore, let $S_t^*(a_i)$ represent S_t excluding all terms involving a_i for $i \in \{1,2,3,...,M\}$. To illustrate, $S_1^*(a_1)$ is S_1 excluding a_1 in the summation, $S_2^*(a_1)$ is S_2 excluding any terms having a_1 as shown below $S_1^*(a_1) = a_2 + a_3 + ... + a_m$, $S_2^*(a_1) = a_2 a_3 + a_2 a_4 + ... + a_2 a_m + ... + a_{m-1} a_m$. Based on double recursive relations mentioned in the proof of Theorem 1 in Marcus and Lopes (1957) paper,

$$tS_t = \sum_{i=1}^{M} a_i S_{t-1} - \sum_{i=1}^{M} a_i^2 S_{t-2}^*(a_i) \text{ , } t\text{=2…M} \qquad (3)$$

$$S_t^*(a_i) = S_t - a_i S_{t-1}^*(a_i) \qquad ,t\text{=2…M} \qquad (4)$$

The distribution of T can be calculated in the following manner:

Step 1.) Define $S_0 = 1, S_0^*(a_i) = 1,\ c(p,n_i) = (1-p)^{\sum_{i=1}^{M} n_i}$. Also define $S_1 = \sum_{i=1}^{M} a_i$, $S_1^*(a_i) = S_1 - a_i$ for $i \in \{1,2,...,M\}$. Then g(T=0)= $c(p,n_i)$, g(T=1)= $c(p,n_i)\ S_1$

Step 2.) Start loop:

S_2 will be calculated by plugging S_1 and S_0^* into equation (3)

S_2^* will be calculated by plugging S_2 and S_1^* into equation (4)

Output g(T=2)= $c(p,n_i)\ S_2$

.

.

.

S_M will be calculated by plug S_{M-1} and S_{M-2}^* into equation (3)

Output g(T=M)= $c(p,n_i)\ S_M$

End loop

Another possible recursive method to calculate the distribution of T is using Newton's identities. Newton's identities connect power sums and elementary symmetric polynomials (Mead, 1992) which can be stated in the following equation

$$tS_t = \sum_{i=1}^{M} (-1)^{i-1} S_{t-i} W_i \text{ for } t\text{=1,…}M \qquad (5)$$

where $W_i = \sum_{j=1}^{M} a_j^i$ and $S_0 = 1, S_{t-i} = 0$ for t<i. For illustration, suppose we have three pools with pool size n_1, n_2, and n_3. If we let $W_1 = a_1 + a_2 + a_3$, $W_2 = a_1^2 + a_2^2 + a_3^2, W_3 = a_1^3 + a_2^3 + a_3^3$. Then from equation (5),

$S_1 = (-1)^0 S_0 W_1 = W_1 = a_1 + a_2 + a_3$

$2S_2 = (-1)^0 S_1 W_1 + (-1)^1 S_0 W_2 = S_1 W_1 - S_0 W_2$

$3S_3 = (-1)^0 S_2 W_1 + (-1)^1 S_1 W_2 + (-1)^2 S_0 W_3 = S_2 W_1 - S_1 W_2 + S_0 W_3$.

Calculation method will be similar to the above pseudo code and omitted here.

Unfortunately, the Newton recursion is unstable and hence not very useful in practice. Among the three methods of PMF calculation of statistic *T*, the Marcus recursive method performs better than saddle point approximation in terms of precision and speed. The greater precision is not a surprise since the saddle point approach is not expected to yield more than a few digits of accuracy. However, care must be taken with the Marcus method to control underflow problem. Considering only the leading term $c(p, n_i) = (1-p)^{\sum_{i=1}^{M} n_i}$ in the PMF expression, $c(p, n_i)$ will decrease as $\sum_{i=1}^{M} n_i$ increases for a fixed *p*. Eventually, $c(p, n_i)$ will run underflow after a certain point. However, if the natural logarithm of $c(p, n_i)$ is used together with the natural logarithm of the quantities S_t the probabilities for values of *T* can be calculated successfully. There are overflow problems coming from the term $\frac{1-(1-p)^{n_i}}{(1-p)^{n_i}}$ as well because this term is an increasing function of both *p* and n_i. Therefore given large *p* and n_i, above algorithm will break down. A simple safeguard to prevent this is to set $1-(1-p)^{n_{\max}} \le \frac{1}{2}$ for maximum pool size, then $p \le 1 - \left(\frac{1}{2}\right)^{\frac{1}{n_{max}}}$. If this condition is violated, calculation should be terminated.

3. Two-Sided Hypothesis Testing Based on Sum of Positive Pools and Asymptotic Results

3.1 Exact Test

Using the exact distribution of *T*, its properties and the computational methods discussed in the preceding section, an exact test using *T* as the test statistic will be proposed. It was shown in the preceding section that *T* possesses a monotone likelihood ratio property. Consider a two-sided size α hypothesis test for $H_0 : p = p_0$ versus $H_a : p \ne p_0$ based on sum of positive pools. Let γ_1 and γ_2 be two constants taking values between 0 and 1. Because *T* is a discrete random variable, a randomized test (see for instance, Lehmann and Romano, 2005) will be utilized to test this set of hypotheses. The left and right critical values of the test and the constants, γ_1 and γ_2, can be respectively solved using following equations

$$\frac{\alpha}{2} = \sum_{T=0}^{T_l - 1} g(T \mid p_0) + \gamma_1 g(T = T_l \mid p_0) \qquad (6)$$

$$\frac{\alpha}{2} = \sum_{T_r+1}^{M} g(T \mid p_0) + \gamma_2 g(T = T_r \mid p_0) \qquad (7)$$

Given an alternative p_a, the formula for the statistical power, β, is given by

$$\beta = \sum_{T=0}^{T_l-1} g(T \mid p_a) + \gamma_1 g(T = T_l \mid p_a) + \gamma_2 g(T = T_r \mid p_a) + \sum_{T=T_r+1}^{M} g(T \mid p_a) \qquad (8)$$

3.2 Asymptotic Tests

Asymptotic test procedures are commonly used in practice because, in most cases, the asymptotic distribution is either normal or chi-square distribution. These tests are typically based on the likelihood function. Three of the standard likelihood-based test procedures are the likelihood ratio (LR) test, Wald's test and the Score test. In the most general case where pool screening applies, the pools do not necessarily have the same size which makes the sample independent but not identically distributed. This being the case, the usually quoted results concerning the asymptotic properties of the MLE parameter estimate do not apply. Bradley and Gart (1962) defined a special situation called "associated population" where observations come from different (sub)populations but have some parameters in common. In their paper, they proved that MLE is a consistent estimator, it is asymptotically normally distributed, and asymptotic Chi-square distribution still followed for the asymptotic likelihood ratio test under certain regularity conditions.

Suppose there are total M pools, m_i pools have pool size n_i and $M = \sum_{i=1}^{k} m_i$ (k is the total number of distinct sized pools). Then the log of the likelihood is

$$\log L(p, x) = \sum_{i=1}^{k} \sum_{j=1}^{m_i} \left\{ x_{ij} \log\left[1 - (1-p)^{n_i}\right] + n_i (1 - x_{ij}) \log(1-p) \right\} \qquad (9)$$

And first derivative of log likelihood is

$$\frac{\partial \log L(p, x)}{\partial p} = \sum_{i=1}^{k} \sum_{j=1}^{m_i} \left\{ \frac{x_{ij} n_i (1-p)^{n_i - 1}}{1 - (1-p)^{n_i}} - \frac{n_i (1 - x_{ij})}{1-p} \right\} \qquad (10)$$

The MLE for p solves equation (10) when it is set equal to 0. Unlike the equal pool size case, there is no explicit expression for the MLE ($\hat{p}$) but it can be obtained using numerical methods. In this research, the inverse quadratic interpolation was utilized to find the root of above partial derivative equation.
The next theorem summarizes the asymptotic likelihood-based test procedures being considered.

Theorem 3: When $0 < p < 1$ and assuming $\frac{m_i}{M}$ is constant as $M \to \infty$. For the hypothesis $H_o : p = p_o$ versus $H_a : p \neq p_o$, an approximate level α test rejects for the likelihood based methods when:

I) Likelihood ratio test: $\chi_L^2 = -2\left[\log L(p_0, x) - \log L(\hat{p}, x)\right] > \chi_\alpha^2(1)$

II) Wald's test: $\chi_w^2 = \left(\frac{\hat{p} - p_0}{\sqrt{I_{\hat{p}}^{-1}}} \right)^2 > \chi_1^2(\alpha)$ where $I = \left(\sum_{i=1}^{k} m_i \frac{n_i^{\,2} (1-p)^{n_i - 2}}{1 - (1-p)^{n_i}} \right)$

III) Score test: $\chi_s^2 = \left(\frac{S(p_o)}{\sqrt{I_{p_o}}} \right)^2 > \chi_1^2(\alpha)$ where $S(p) = \frac{\partial \log L(p \mid x)}{\partial p}$ as given in equation (10).

Proof of Theorem 3 is given in theAppendix.

Remark: One can easily modify the above Wald's and score tests for one sided hypothesis. In addition,test statistics $Z_W = \left(\frac{\hat{p} - p_0}{\sqrt{I_{\hat{p}}^{-1}}} \right)$ and $Z_S = \left(\frac{S(p_o)}{\sqrt{I_{p_o}}} \right)$ can be compared with standard normal critical values instead of their chi-square counterparts.

4. Simulation Study and Results

To compare the exact and likelihood ratio based test procedures in terms of statistical power, a series of Monte Carlo simulations were conducted using Fortran (Absoft Pro Fortran 10.1). Consider testing H_o: $p_o = 0.0005$ versus $p_o \neq 0.0005$ as an example for a very low prevalence of certain infectious disease. Although this may seem extreme, there are applications where such a prevalence rate is of interest such as in Tropical Medicine research (see for instance, Guevara et al., 2003; Yamèogo et al.,1999). Thus, it is important to be able to test if the prevalence is less than 5 in 10,000 or even 1 in 10,000.

It is worth noting that although Theorem 3 states that the asymptotic distribution of the likelihood-based tests being considered follows a chi-square distribution, it will require extremely large number of pools for the asymptotic results to provide a good approximation when *p* is near 0. In practice, the typical number of pools used is between 100 and 250. To illustrate, consider the following simulation studies where pool sizes were randomly drawn from a discrete uniform [25,50]. In this case, the probability of a pool testing positive ranged from 0.012 (when pool size is 25) to 0.025 (when pool size is 50). Varying the number of pools from 50 to 700, the Kolmogorov-Smironov (KS) goodness of fit test statistic values for the LR, Wald's and Score statistics compared to a chi-square with 1 degree of freedom (df) are displayed in Table 1. As expected, the values of the KS statistic decrease as the number of pools increases indicating that the chi-square 1 df is a reasonable fit. However, Figure 1 and Figure 2 show that the speed of convergence is unsatisfactory. These graphs are quantile-quantile plots of the LR statistic compared to a chi-square with df=1 when the number of pools is 100 and 700. Similar observations were obtained for Wald's and Score statistics as well as other cases. Therefore, test procedures using tabulated chi-square values to define the critical points of the rejection region may be inaccurate in cases where *p* is near zero possibly leading researchers to erroneous conclusions.

To address this issue, an alternative method to define the rejection regions for these tests is proposed. Simulated quantiles will determine the cut off point instead of the tabulated values based on a chi-square distribution. The power function based on the quantile method will be compared with the exact test and the standard asymptotic test as defined in Theorem 3. It is hoped that using simulated quantiles will improve the performance of the likelihood-based tests.

Below is a summary of the simulation steps taken to obtain results in this section:
Step 1.) For a certain number of pools such as *k*, generate pool sizes from a discrete uniform distribution over the range [25, 50]. Note that this range of pool sizes is typically required by PCR laboratory screening test.
Step 2.) Given Type I error set at 0.05, find γ_1, γ_2 and critical values T_l, T_r satisfying equations (6) and (7). Given a value of *p* and the computed values of γ_1, γ_2, T_l, T_r, calculate exact power associated with the test statistic *T* using equation (8).

Step 3.) Do first simulation: Generate 100,000 samples under the null. Calculate LR, Wald's and Score test statistics under each sample. Find 97.5^{th} and 2.5^{th} quantiles of each test statistic. These quantiles will be used to define the rejection region as an alternative to the rejection region based on the chi-squared distribution. Thus, H_0 is rejected when the test statistic value is either less than its corresponding 2.5^{th} quantile or greater than its corresponding 97.5^{th} quantile.
Step 4.) Do simulation two: Generate 100,000 samples under alternative. Calculate LR, Wald's and Score test statistics under each sample. Compute the percent of times a test rejects the null hypothesis – either using the simulated quantiles or the tabulated chi-square values. The resulting percentage is the respective simulated power for LR, Wald's, and Score for that particular value of *p*.

Remark: Samples where all pools are either positive or negative were excluded from the simulation because in these cases the MLE is either 1 or 0. Consequently, the test statistics associated with the likelihood ratio, Wald's, and Score tests cannot be computed. When this happens, another set of sample is simulated in order to reach the total of 100,000.

Figure 3 and Figure 4 display the power functions of the exact test and the likelihood-based tests when the number of pools is 100. For the likelihood-based tests, the power curves in Figure 3 are based on the simulated quantiles while the power curves in Figure 4 were based on the chi-square distribution. Because the test based on the number of positive pools is exact, the power at the null hypothesized value is around the set significance level of 5%. A striking feature in these power curves is that only the exact test is unbiased, i.e., the power under the alternative at least that of the power under the null. The power curves of the likelihood-based tests were significantly improved by using simulated quantiles but these modified tests are still biased, in particular, when the alternative value is less than the null hypothesized value. Although the power function is higher for the modified LR test in Figure 3 relative to the power for the exact test, this is only true when *p* is greater than the null hypothesized value. Modified LR performs poorly when p is small. The bias problem of the likelihood-based tests is more likely due to fact that most pools are negative. Therefore, in cases like this, it is recommended that the likelihood-based tests not be used.

When number of pools is increased to 350, the exact test procedure and the likelihood-based procedures using simulated quantiles have power curves that are very similar (see Figure 5). At the null value, the simulated levels are at around 5% and at any of the alternative values, the power increases as the alternative value gets farther away from the null. Finally, all tests are unbiased. However, these observations do not hold for the standard likelihood-based (see Figure 6). Score test is biased for values less than the null while Wald's test has an inflated type I error rate.

Of major interest for applied researchers is determining the number of subjects needed if they desire to perform a hypothesis test based on the number of positive pools. To illustrate how this can be done using the exact test, consider the case where all pool sizes are known uniformly ranging from 25 to 50. Let the null hypothesis be p=0.0005 and the significance level be set at 5%. The estimated power for varying number of subjects is summarized in Table 2 for different alternatives. Based on this table, if the true prevalence is less than or equal to p=0.0001, then obtaining 200 pools (7493 total subjects) is estimated to provide a power of about 86% while increasing the number of pools to 250 (9335 total subjects) increased the estimated power to about 88%.

Remark: The computed power values given in Table 2 are sensitive to the specific pool sizes being considered

5. Conclusion

Although the distribution of the number of positive pools is complex when pool sizes are unequal, it is no longer difficult to compute given the recursive methods explored in this research. In addition to this, exact test performs very well in terms of statistical power compared to all the other tests considered in this paper. The standard asymptotic likelihood-based tests need to be modified to address the issue of slow convergence when the prevalence is near 0 by using simulated quantiles to define the critical values of the rejection region. In spite of the improvements due to this modification, the exact test still performed better than these likelihood-based tests especially when the number of pools is not large Furthermore, calculating the MLE and obtaining simulated quantiles are computationally demanding to researchers.Thus, test procedure based on the number of positive pools is more appealing. Therefore, the exact test based on the number of positive pools is recommended regardless of the number of pools. This manuscript focused on two sided tests. In practice, one sided hypothesis tests are more often of interest particularly in disease elimination programs. This will be the focus of future research.

APPENDIX

Proof of Theorem 3

Results of this theorem follow immediately by applying Bradley and Gart's theorems which states that the maximum likelihood estimate ($\hat{p}$) is a consistent estimator of p and $\sqrt{M}(\hat{p}-p_0)$ has asymptotic normal distribution with mean 0 and variance $\left(\sum_{i=1}^{k}\frac{m_i}{M}\frac{{n_i}^2(1-p)^{n_i-2}}{1-(1-p)^{n_i}}\right)^{-1}_{p=p_0}=\frac{1}{I}$ where I is the expected Fisher information. And also from results of section 2.4 in their original paper, they showed that $-2\left[\log L(p_0,x)-\log L(\hat{p},x)\right]$ has an asymptotic Chi-square distribution with 1 degree of freedom. Therefore, what remains to be done is to show that in this particular case, the conditions required in the Theorems of Bradley and Gart are satisfied.

Condition I(i) $\frac{\partial lnf_i}{\partial p}$**,** $\frac{\partial ln^2 f_i}{\partial p^2}$**,** $\frac{\partial ln^3 f_i}{\partial p^3}$ **exist.**

$lnf_i = x_i ln\left[1-(1-p)^{n_i}\right]+n_i(1-x_i)ln(1-p)$, first derivative of lnf_i exist and equal to

$$\frac{\partial lnf_i}{\partial p}=\frac{n_i x_i(1-p)^{n_i-1}}{1-(1-p)^{n_i}}-\frac{n_i(1-x_i)}{1-p}\ .$$

$$\frac{\partial ln^2 f_i}{\partial p^2}=\frac{-x_{ij}n_i(n_i-1)(1-p)^{n_i-2}\left[1-(1-p)^{n_i}\right]-x_{ij}{n^2}_i(1-p)^{2(n_i-1)}}{\left[1-(1-p)^{n_i}\right]^2}-\frac{n_i}{(1-p)^2}+\frac{n_i x_{ij}}{(1-p)^2}.$$

$$\frac{\partial ln^3 f_i}{\partial p^3}=\frac{x_{ij}n_i(n_i-1)(n_i-2)(1-p)^{n_i-3}}{\left[1-(1-p)^{n_i}\right]}+\frac{3x_{ij}n_i^2\,(n_i-1)(1-p)^{2n_i-3}}{\left[1-(1-p)^{n_i}\right]^2}+\frac{2x_{ij}{n_i}^3(1-p)^{3(n_i-1)}}{\left[1-(1-p)^{n_i}\right]^3}-\frac{2n_i}{(1-p)^3}+\frac{2n_i x_{ij}}{(1-p)^3}$$

From above equations, $\frac{\partial lnf_i}{\partial p}$, $\frac{\partial ln^2 f_i}{\partial p^2}$,and $\frac{\partial ln^3 f_i}{\partial p^3}$ exist only when $p\neq 0$ and $p\neq 1$, which could be automatically satisfied by the pool screening problem. Please also note when $n_i=1$, result is simply binomial and is well known. When $n_i=2$, $\frac{\partial lnf_i}{\partial p}$**,** $\frac{\partial ln^2 f_i}{\partial p^2}$**,** and $\frac{\partial ln^3 f_i}{\partial p^3}$ are still exist.

Condition I(ii) $\sum_{x_i\in R_i}\frac{\partial f_i}{\partial p}$ **and** $\sum_{x_i\in R_i}\frac{\partial^2 f_i}{\partial^2 p}$ **converges uniformly for all** $p\in Q$**, and** $\left|\frac{\partial^3 \log f_{ij}}{\partial p^3}\right|<H_i(x_i)$ **where** $\sum_{x_i\in R_i}H_i(x_i)f_i<W_i$ **for all** $p\in Q$ **and** W_i **are finite positive numbers.**

$$\frac{\partial f_i}{\partial p} = x_{ij}\left[1-(1-p)^{n_i}\right]^{x_{ij}-1} n_i(1-p)^{2n_i-n_i x_{ij}-1} - \left[1-(1-p)^{n_i}\right]^{x_{ij}} (n_i - n_i x_{ij})(1-p)^{n_i-n_i x_{ij}-1}$$

$$\sum_{x_i \in R_i} \frac{\partial f_i}{\partial p} = n_i(1-p)^{n_i-1} - n_i(1-p)^{n_i-1} = \mathbf{0}$$

$$\frac{\partial^2 f_i}{\partial^2 p} = \frac{\partial}{\partial p} x_{ij}\left[1-(1-p)^{n_i}\right]^{x_{ij}-1} n_i(1-p)^{2n_i-n_i x_{ij}-1} - \left[1-(1-p)^{n_i}\right]^{x_{ij}} (n_i - n_i x_{ij})(1-p)^{n_i-n_i x_{ij}-1}$$

$$= n_i^2 x_i(x_i-1)\left[1-(1-p)^{n_i}\right]^{x_i-2}(1-p)^{3n_i-n_i x_i-2} - n_i x_i\left[1-(1-p)^{n_i}\right]^{x_i-1}(2n_i-n_i x_i-1)(1-p)^{2n_i-n_i x_i-2}$$

$$-n_i^2(1-x_i)x_i\left[1-(1-p)^{n_i}\right]^{x_i-1}(1-p)^{2n_i-n_1 x_i-2} + n_i(1-x_i)\left[1-(1-p)^{n_i}\right]^{x_i}(n_i-n_i x_i-1)(1-p)^{n_i-n_i x_i-2}$$

$$\sum_{x_i \in R_i} \frac{\partial^2 f_i}{\partial^2 p} = n_i(n_i-1)(1-p)^{n_i-2} - n_i(n_i-1)(1-p)^{n_i-2} = 0$$

$$\log f_{ij} = x_{ij} log(1-(1-p)^{n_i}) + n_i(1-x_{ij}) log(1-p)$$

$$\frac{\partial \log f_{ij}}{\partial p} = \frac{x_{ij} n_i (1-p)^{n_i-1}}{1-(1-p)^{n_i-1}} - \frac{n_i(1-x_{ij})}{1-p}$$

$$\frac{\partial \log^2 f_{ij}}{\partial^2 p} = \frac{-x_{ij} n_i (n_i-1)(1-p)^{n_i-2}}{\left[1-(1-p)_i^n\right]^2} + \frac{x_{ij} n_i (n_i-1)(1-p)^{2n_i-2}}{\left[1-(1-p)^{n_i}\right]^2} - \frac{x_{ij} n_i^2 (1-p)^{2n_i-2}}{\left[1-(1-p)^{n_i}\right]^2} - \frac{n_i(1-x_{ij})}{(1-p)^2}$$

$$= -\frac{x_{ij} n_i (n_i-1)(1-p)^{n_i-2}}{\left[1-(1-p)_i^n\right]^2} - \frac{x_{ij} n_i (1-p)^{2n_i-2}}{\left[1-(1-p)^{n_i}\right]^2} - \frac{n_i(1-x_{ij})}{(1-p)^2}$$

$$\frac{\partial \log^3 f_{ij}}{\partial^3 p} = \frac{x_{ij} n_i (n_i-1)(n_i-2)(1-p)^{n_i-3}}{\left[1-(1-p)^{n_i}\right]^2} + \frac{2 x_{ij} n_i^2 (n_i-1)(1-p)^{2n_i-3}}{\left[1-(1-p)^{n_i}\right]^3} +$$

$$+ \frac{2 x_{ij} n_i (n_i-1)(1-p)^{2n_i-3}}{\left[1-(1-p)^{n_i}\right]^2} + \frac{2 x_{ij} n_i^2 (1-p)^{3n_i-3}}{\left[1-(1-p)^{n_i}\right]^3} - \frac{2 n_i (1-x_{ij})}{(1-p)^3}$$

When all $n_i = 1$,

$$\frac{\partial \log^3 f_{ij}}{\partial^3 p} = \frac{2x_{ij}}{p^3} - \frac{2(1-x_{ij})}{(1-p)^3}$$

$$\left|\frac{\partial \log^3 f_{ij}}{\partial^3 p}\right| \le g_i(x_{ij} \mid n_i, p) = \frac{2x_{ij}}{p^3} + \frac{2(1-x_{ij})}{(1-p)^3}$$

$E(g_i(x_{ij} \mid n_i, p)) = \dfrac{2}{(1-p)^2} + \dfrac{2}{p^2}$,let $W_i = E(g_i(x_{ij} \mid n_i, p)) + \varepsilon$, where $\varepsilon > 0$, so $\left|\dfrac{\partial \log^3 f_{ij}}{\partial^3 p}\right| < g_i(x_{ij} \mid n_i, p)$ and $E(g_i(x_{ij} \mid n_i, p)) < W_i$

When $n_i \geq 2$ and n_i is constant,

$$\left|\frac{\partial \log^3 f_{ij}}{\partial^3 p}\right| \leq g_i(x_{ij} \mid n_i, p) = \frac{x_{ij} n_i (n_i - 1)(n_i - 2)(1-p)^{n_i - 3}}{\left[1-(1-p)^{n_i}\right]^2} + \frac{2x_{ij} n_i^2 (n_i - 1)(1-p)^{2n_i - 3}}{\left[1-(1-p)^{n_i}\right]^3} + \frac{2x_{ij} n_i (n_i - 1)(1-p)^{2n_i - 3}}{\left[1-(1-p)^{n_i}\right]^2} + \frac{2x_{ij} n_i^2 (1-p)^{3n_i - 3}}{\left[1-(1-p)^{n_i}\right]^3} + \frac{2n_i (1 - x_{ij})}{(1-p)^3}$$

let

$$E(g_i(x_{ij} \mid n_i, p)) = 2n_i (1-p)^{n_i - 3} + \frac{n_i (n_i - 1)(n_i - 2)(1-p)^{n_i - 3}}{1-(1-p)^{n_i}} + \frac{2n_i^2 (n_i - 1)(1-p)^{2n_i - 3}}{\left[1-(1-p)^{n_i}\right]^2} + \frac{2n_i (n_i - 1)(1-p)^{2n_i - 3}}{1-(1-p)^{n_i}} + \frac{2n_i^2 (1-p)^{3n_i - 3}}{\left[1-(1-p)^{n_i}\right]^2}$$

$W_i = E(g_i(x_{ij} \mid n_i, p)) + \varepsilon$, where $\varepsilon > 0$, so $\left|\dfrac{\partial \log^3 f_{ij}}{\partial^3 p}\right| < g_i(x_{ij} \mid n_i, p)$ and $E(g_i(x_{ij} \mid n_i, p)) < W_i$

Condition I(iii) $J(p) = \sum_{i=1}^{k} \sum_{j=1}^{M_i} \sum_{x_{ij}=0,1} \dfrac{\partial \log f_{ij}}{\partial p} \dfrac{\partial \log f_{ij}}{\partial p} f_{ij}$ **is positive definite (real positive number)**

$$\begin{aligned} J(p) &= \sum_{i=1}^{k} m_i \sum_{x_{ij}=0,1} \frac{\partial \log f_{ij}}{\partial p} \frac{\partial \log f_{ij}}{\partial p} f_{ij} \\ &= \sum_{i=1}^{k} m_i \left[\left(\frac{-n_i}{1-p} \right) \left(\frac{-n_i}{1-p} \right) (1-p)^{n_i} + \left(\frac{n_i (1-p)^{n_i - 1}}{1-(1-p)^{n_i}} \right) \left(\frac{n_i (1-p)^{n_i - 1}}{1-(1-p)^{n_i}} \right) (1-(1-p)^{n_i}) \right] \\ &= \sum_{i=1}^{k} m_i \frac{n_i^2 (1-p)^{n_i - 2}}{1-(1-p)^{n_i}} \end{aligned}$$

Hence, for $0 < p < 1$, $J(p) > 0$.

This ends the proof.

REFERENCES

Absoft Coporation, Fortan 95,Absoft Pro Fortran 10.1 32bit versions

BARKER, J.T., 2000, Statistical Estimates of Infection Potential Based on PCR Pool Screening with Unequal Pool Sizes, PhD thesis, Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL

BRADLEY, R.A., and GART, J.J., 1962, The Asymptotic Properties of ML Estimators when Sampling from Associated Populations: Biometrika, v. 49, p. 205-214.

BUSCH MP, E.B., KHAYAM-BASHI H, HEILBRON D, MURPHY EL, KWOK S, SNINSKY J, PERKINS HA, VYAS GN., 1991, Evaluation of Screened Blood Donations for Human Immunodeficiency Virus Type 1 Infection by Culture and DNA Amplification of Pooled Cells.: N Engl J Med., v. 325, p. 1-5.

CASELLA, G., and BERGER, R., 2001, Statistical Inference: Duxbury, Pacific Grove,CA.

CHEN, C.L., and SWALLOW, W.H., 1990, Using Group Testing to Estimate a Proportion, and to Test the Binomial Model: Biometrics, v. 46, p. 1035-1046.

CHIANG, C., and REEVES, W., 1962, Statistical Estimation of Virus Infection Rates in Mosquito Vector Populations Am J Hyg, v. 75, p. 377-391.

DANIELS, H.E., 1954, Saddlepoint Approximations in Statistics: The Annals of Mathematical Statistics, v. 25, p. 631-650.

DORFMAN, R., 1943, The Detection of Defective Members of Large Populations: The Annals of Mathematical Statistics, v. 14, p. 436-440.

FARRINGTON, C.P., 1992, Estimating Prevalence by Group Testing Using Generalized Linear Models: Statistics in Medicine, v. 11, p. 1591-1597.

GOUTIS, C., and CASELLA, G., 1999, Explaining the Saddlepoint Approximation: The American Statistician, v. 53, p. 216-224.

GU, W., LAMPMAN, R., and NOVAK, R.J., 2003, Problems in Estimating Mosquito Infection Rates Using Minimum Infection Rate: Journal of Medical Entomology, v. 40, p. 595-596.

GUEVARA AG, V.J., LILLEY BG, LÓPEZ A, VIEIRA N, RUMBEA J, COLLINS R, KATHOLI, and CR, U.T., 2003, Entomological evaluation by pool screen polymerase chain reaction of Onchocerca volvulus transmission in Ecuador following mass Mectizan distribution.: Am J Trop Med Hyg., v. 68(2):222-7.

HEPWORTH, G., 1996, Exact Confidence Intervals for Proportions Estimated by Group Testing: Biometrics, v. 52, p. 1134-1146.

HSU, L., 1995, New Procedures for Group-Testing Based on the Huffman Lower Bound and Shannon Entropy Criteria: Lecture Notes-Monograph Series, v. 25, p. 249-262.

HUYNH, H., 1994, A New Proof For Monotone Likelihood Ratio For the Sum of Independent Bernoulli Random Variables: Psychometrika, v. 59, p. 77-79.

KATHOLI, C.R., 2007, Hypothesis Testing with the Pool Screening Model: In Press.

KATHOLI, C., TOÉ, L., MERRIWEATHER, A., and UNNASCH, T., 1995, Determining the Prevalence of Onchocerca *Volvulus* Infection in Vector Populations by Polymerase Chain Reaction Screening of Pools of Black Flies.: J Infect Dis., v. 172, p. 1414-1417.

KATHOLI, C.R., and UNNASCH, T.R., 2006, Important Experimental Parameters for Determining Infection Rates in Arthropod Vectors Using Pool Screening Approaches: Am J Trop Med Hyg, v. 74, p. 779-785.

KLINE, R.L., BROTHERS, T.A., BROOKMEYER, R., ZEGER, S., and QUINN, T.C., 1989, Evaluation of Human Immunodeficiency Virus Seroprevalence in Population Surveys Using Pooled Sera: J. Clin. Microbiol., v. 27, p. 1449-1452.

LEHMANN, E.L., and ROMANO, J.P., 2005, Testing Statistical Hypotheses: Springer.

LITVAK, E., TU, X.M., and PAGANO, M., 1994, Screening for the Presence of a Disease by Pooling Sera Samples: Journal of the American Statistical Association, v. 89, p. 424-434.

MARCUS, M., and LOPES, L., 1957, Inequalities for symmetric functions and Hermitian matrices: Canadian Journal of Mathmatics, v. 9, p. 305-312.

MARION, A.W., 1936, Factors Affecting the Amount of Infection Obtained by Aphis Transmission of the Virus Hy. III: Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, v. 226, p. 457-489.

MEAD, D.G., 1992, Newton's Identities: The American Mathematical Monthly, v. 99, p. 749-751.

NEDELMAN, J., and WALLENIUS, T., 1986, Bernoulli Trials, Poisson Trials, Surprising Variances, and Jensen's Inequality: The American Statistician, v. 40, p. 286-289.

STERRETT, A., 1957, On the Detection of Defective Members of Large Populations: The Annals of Mathematical Statistics, v. 28, p. 1033-1036.

MILTON, S., and GROLL, P.A., 1966, Binomial Group-Testing with an Unknown Proportion of Defectives: Technometrics, v. 8, p. 631-656.

THOMPSON, K.H., 1962, Estimation of the Proportion of Vectors in a Natural Population of Insects: Biometrics, v. 18, p. 568-578.

TEBBS, J.M., and MCCANN, M.H., 2007, Large-Sample Hypothesis Tests for Stratified Group-Testing Data: Journal of Agriculture, Biological, and Environmental Statistics, v. 12, p. 534-551.

TU, X.M., LITVAK, E., and PAGANO, M., 1995, On the Informativeness and Accuracy of Pooled Testing in Estimating Prevalence of a Rare Disease: Application to HIV Screening: Biometrika, v. 82, p. 287-297.

WANG, Y.H., 1993, On the Number of Success in Independent Trials: Statistica Sinica, v. 3, p. 295-312.

WEIN, L.M., and ZENIOS, S.A., 1996, Pooled Testing for HIV Screening: Capturing the Dilution Effect: Operations Research, v. 44, p. 543-569.

YAMEOGO, L., TOE, L., HOUGARD, J.M., BOATIN, B.A., and UNNASCH, T.R., 1999, Pool screen polymerase chain reaction for estimating the prevalence of Onchocerca volvulus infection in Simulium damnosum sensu lato: results of a field trial in an area subject to successful vector control: Am J Trop Med Hyg, v. 60, p. 124-128.

Table 1: Komogorov-Smironov statistics of likelihood ratio based tests given different number of pools.

Number of Pools	50	100	250	400	550	700
Wald Statistics	0.3762	0.2224	0.1633	0.1312	0.1124	0.0932
Score Statistics	0.3584	0.2261	0.1264	0.0941	0.0830	0.0700
LR Statistics	0.5542	0.3053	0.1318	0.1268	0.0844	0.0785

Table 2: Examples of statistical power and number of pools (subjects) for exact test under different alternatives against null p=0.0005, significance level=0.05, assuming pool size has discrete uniform distribution[25,50].

Alternative	**Number of Pools (Total Number of Subjects)**												
	50 (1906)	100 (3709)	150 (5665)	200 (7493)	250 (9335)	300 (11368)	350 (12954)	400 (14916)	450 (16887)	500 (19082)	550 (20660)	600 (22506)	650 (24366)
Pa=0.00002	0.063	0.148	0.380	0.863	0.884	0.979	0.982	0.997	0.998	1.000	1.000	1.000	1.000
Pa=0.0001	0.055	0.110	0.241	0.478	0.523	0.694	0.723	0.824	0.863	0.907	0.943	0.953	0.973
Pa=0.001	0.142	0.199	0.327	0.356	0.450	0.489	0.546	0.623	0.692	0.715	0.749	0.790	0.825
Pa=0.0015	0.320	0.489	0.726	c0.786	0.882	0.916	0.948	0.974	0.987	0.992	0.995	0.997	0.999

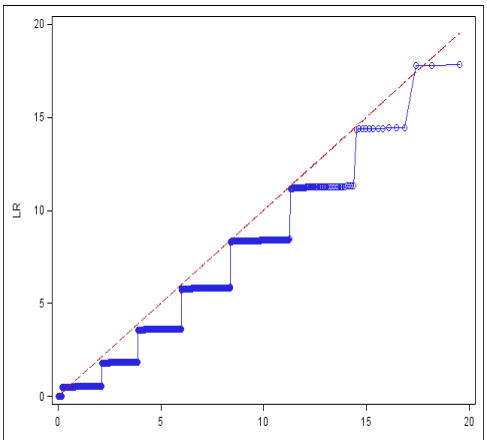

Figure 1. Chi-Square (df=1) QQ Plot of asymptotic likelihood ratio test. Number of pools=100. p-value<0.001 for Kolmogorov-Smirnov test.

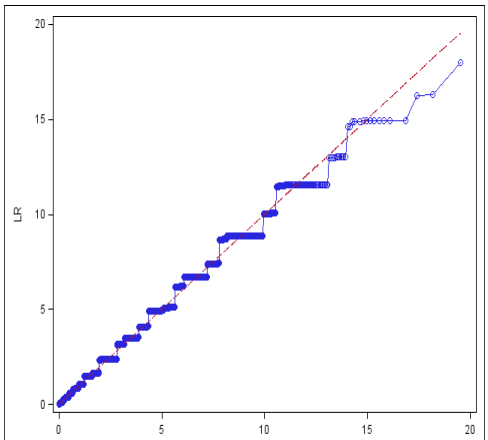

Figure 2. Chi-Square(df=1) QQ Plot of asymptotic likelihood ratio test. Number of pools=700. p-value<0.001 for Kolmogorov-Smirnov test.

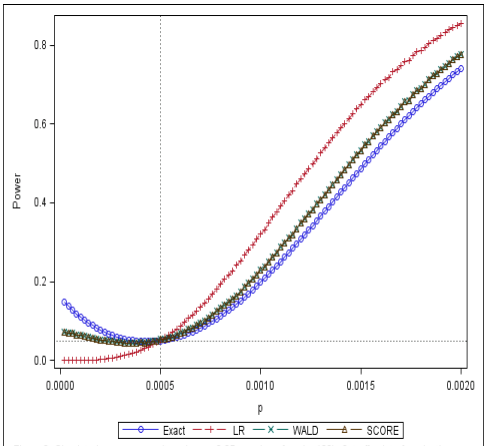

Figure 3. Simulated power curves:(type I error=0.05, number of pools=100). Cut off points for rejection are based on simulated quantiles.

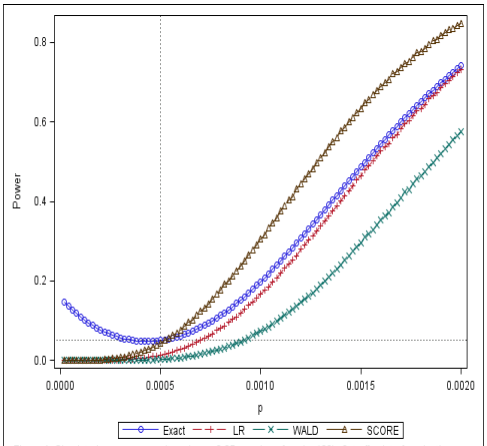

Figure 4. Simulated power curves: (type I erro=0.05, number of pools=100). Cut off points for rejection are based on standard distributions.

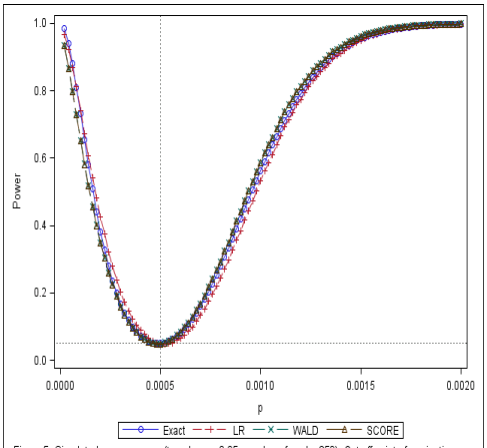

Figure 5. Simulated power curves: (type I error=0.05, number of pools=350). Cut off points for rejection are based on simulated quantiles.

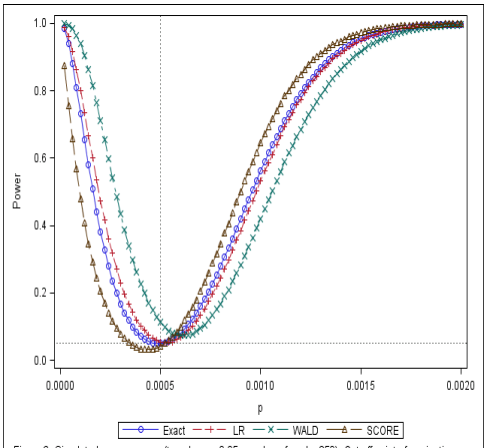

Figure 6. Simulated power curves: (type I error=0.05, number of pools=350). Cut off points for rejection are based on standard distributions.