2010

Hypothesis Testing Based On Pool Screening With Unequal Pool Sizes

Hongjiang Gao
University of Alabama at Birmingham

HYPOTHESIS TESTING BASED ON POOL SCREENING WITH UNEQUAL POOL SIZES

by

HONGJIANG GAO

INMACULADA ABAN, CHAIR
CHARLES R. KATHOLI, CO-CHAIR
YINGZI CONG
JOSHUA RICHMAN
KUI ZHANG

A DISSERTATION

Submitted to the graduate faculty of The University of Alabama at Birmingham,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

BIRMINGHAM, ALABAMA

2010

HYPOTHESES TESTING IN UNEQUAL SIZED POOL SCREENING

HONGJIANG GAO

BIOSTATISTICS

ABSTRACT

Pool screening is a widely applied technique to estimate the prevalence of a rare event. This focus of this research is on developing statistical test of hypothesis procedures under the assumption that pool sizes are unequal but known. One of the proposed test procedures is , an exact test based on the number of positive pools (denoted by *T*). Another set of proposed test procedures is a modification of the likelihood ratio, Wald's and Score tests which are commonly-used likelihood-based tests.

In paper 1, we derive the distribution of *T* which will be the basis of the exact test. Other distributional properties of T are obtained using generating functions. Due to the complexity of the form of the distribution, we propose several methods of computing probabilities using the distribution of T. It was found that in the setting being considered, the double recursion method based on the recursion relationship introduced by Marcus and Lopes is the recommended computational method.

In paper 2, we proposed an exact two-sided hypothesis test procedure based on the statistic T. We also propose modified versions of the likelihood-ratio, Wald's and Score tests where simulated quantiles are used instead of the quantiles based on the standard asymptotic distribution to obtain the rejection region for each test. Monte Carlo simulations show that the modified test procedures perform better in terms of statistical power than their original counterpart. However, the exact test based on number of

positive pools outperforms the other test when the number of pools screened is small and/or the prevalence close to zero.

The last paper focuses on the one-sided hypothesis test and the likelihood ratio (LR) test procedure. We first investigate the distributional properties and behavior of the likelihood ratio test statistic both in the finite and large sample cases. It will be shown that the distribution of one-sided LR test statistic is a mixture of distributions. We propose ways to compute the weights of the finite mixture distribution and use these weights to modify the LR test.. We also propose the use of simulated quantiles for one-sided LR test to define the rejection region. Our results show that: 1) LR test with modified weight and conventional LR test have power functions that are very similar; 2) Quantile based LR test improves the LR test but require moderate or large number of pools to be screened; and 3) When number of pools is small, exact test based on number of positive pools performs the best.

Keywords: Pool Screening, Hypothesis Testing, Number of Positive Pools, One-sided Likelihood Ratio Test, Statistical Power, Mixture Chi-square Distribution

DEDICATION

This research is dedicated to my parents, Shuxin Gao and Xiuying Li, especially to my mother who had been through so many sorrows, tragedies and sufferings. It is also dedicated to my wife, Lanfang Wang, and to my daughters, FeiFei Gao and Lily Gao. I love you all so much.

ACKNOWLEDGEMENTS

I would like to give my sincere appreciation to my dissertation chair, Dr Inmaculada Aban and co-chair, Dr Charles Katholi. Thank you for leading me into “pool screening” wonder land, for pulling me back from the dead end when I was off course, for your valuable comments and revisions, for the discussions that goes beyond the dissertation research. Thanks to my committee members: Drs Yingzi Cong, Joshua Richman, and Kui Zhang. Thanks to you for your time, patience, guidance and advice towards the completion of this dissertation. I would also like to give special thanks to Dr Gary Cutter for his support in the past several years.

I am extremely grateful to my family members, my wife, Lanfang Wang and my lovely daughters, XueFei Gao and Lily Gao. Without your support and motivation, this endeavor could not be successful.

TABLE OF CONTENTS

Page

ABSTRACT iii

DEDICATION v

ACKNOWLEDGEMENTS vi

LIST OF TABLES ix

LIST OF FIGURES x

INTRODUCTION 1

POOL SCREENING: AN EXAMPLE OF INDEPENDENT NON-IDENTICAL BERNOULLI TRIAL 14

TWO-SIDED HYPOTHESIS TESTING BASED ON POOL SCREENING WITH UNEQUAL POOL SIZES 29

ONE-SIDED LIKELIHOOD RATIO TEST IN UNEQUAL SIZED POOL SCREENING 60

SUMMARY CONCLUSIONS AND FUTURE RESEARCH 91

GENERAL LIST OF REFERENCES 97

APPENDIX

A MARCUS AND LOPES SUBROUTINE 104

B NEWTONS IDENTITY SUBROUTINE 106

C FUNCTION TO CALCULATE LIKELIHOOD 108

D FUNCTION TO CALCULATE EXPECTED FISHER INFORMATION 110

E FUNCTION TO CALCULATE VARIANCE OF SCORE STATISTICS 111

F SUBROUTINE TO FIND GIVEN QUANTILE OF ONE-DIMENSION ARRAY.....112

G PROGRAM TO CALCULATE EXACT AND SIMULATED POWERS FOR ONE-SIDED HYPOTHESIS TEST 113

LIST OF TABLES

Table *Page*

POOL SCREENING: AN EXAMPLE OF INDEPENDENT NON-IDENTICAL BERNOULLI TRIAL

1 Table 1 Example of Different Pool Sizes (Generated From Discrete Uniform Distribution [25, 50]) 22

2 Table 2 Comparing Distribution Calculated by Recursive Methods with IMSL Standard Package.(Number of pools=150, pool sizes=50, p=0.0001. Results of $t<120$ are omitted). 25

TWO-SIDED HYPOTHESIS TESTING BASED ON POOL SCREENING WITH UNEQUAL POOL SIZES

1 Table 1 Komogorov-Smironov statistics of likelihood ratio based tests given different number of pools. 43

2 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].. 49

ONE-SIDED LIKELIHOOD RATIO TEST IN UNEQUAL SIZED POOL SCREENING

1 Table 1 Exact weight c_1 in the mixture distribution under the null $p_0 = 0.0005$: pool sizes uniformly distributed in the range [25, 50]. Simulated weights are based on 100,000 simulation trials……..76

LIST OF FIGURES

Figure *Page*

POOL SCREENING: AN EXAMPLE OF INDEPENDENT NON-IDENTICAL BERNOULLI TRIAL

1 Figure 1 Distribution of T With Different p (pool sizes based on Table 1).....22

TWO-SIDED HYPOTHESIS TESTING BASED ON POOL SCREENING WITH UNEQUAL POOL SIZES

1 Figure 1 Chi-Sqaure(df=1) QQ Plot of asymptotic likelihood ratio test. Number of pools=100, p-value<0.001 for Kolmogorov-Smirnov test.....44

2 Figure 2 Chi-Square(df=1) QQ Plot of asymptotic likelihood ratio test. Number of pools=700, p-value<0.0001 for Kolmogorob-Smirnov test.....44

3 Figure 3 Simulated power curves for testing, type I error=0.05, number of pools=100. Cut off points for rejection under each alternative is defined by simulation quantile.....46

4 Figure 4 Simulated power curves for testing, type I error=0.05, number of pools=100. Cut off points for rejection under each alternative is defined by standard distribution. 47

5 Figure 5 Simulated power curves for testing, type I error=0.05, number of pools=350. Cut off points for rejection under each alternative is defined by simulation quantile.....48

6 Figure 6 Simulated power curves for testing, type I error=0.05, number of pools=350. Cut off points for rejection under each alternative is defined by standard distribution. 48

ONE-SIDED LIKELIHOOD RATIO TEST IN UNEQUAL SIZED POOL SCREENING

1 Figure 1a EDF for $-2\ln(\lambda)$ under null p=0.0005, number of pools=400, pool sizes uniformly in range [25,50] 75

2 Figure 1b $-2\ln(\lambda)$ when T=3 75

3 Figure 1c $-2\ln(\lambda)$ when T=2 75

4 Figure 2 Chi-square mixture QQ Plot. Number of pools=200, 100000 one-sided LR test statistics simulated under p=0.0005 with pool sizes in range [25,50] 80

5 Figure 3 Chi-square mixture QQ Plot. Number of pools=1000,100000 one-sided LR test statistics simulated under p=0.0005 with pool sizes in range [25,50] 80

6 Figure 4 Simulated power curves with type I error=0.05 and number of pools=100. Cut off point for rejection region for Score based on standard normal, for LRT based on 0.5+0.5*Chi-square(df=1), for LRT(M) based on0.5667+0.4333*Chi-square(df=1), for LRT(Q) based on simulation quantile 83

7 Figure 5 Simulated power curves with type I error=0.05 and number of pools=200. Cut off point for rejection region for Score based on standard normal, for LRT based on 0.5+0.5*Chi-square(df=1), for LRT(M) based on0.5168+0.4832*Chi-square(df=1), for LRT(Q) based on simulation quantile 84

8 Figure 6 Simulated power curves with type I error=0.05 and number of pools=200. Cut off point for rejection region for Score based on standard normal, for LRT based on 0.5+0.5*Chi-square(df=1), for LRT(M) based on0.4748+0.5252*Chi-square(df=1), for LRT(Q) based on simulation quantile 84

INTRODUCTION

1. Motivation Example of Pools Screening

Onchocerciasis (river blindness) is a disease caused by the microfilarial *Onchocerca volvulus*. It is transmitted by the vector black fly (genus *Simulium*) that breed along water side such as streams, rivers, and ponds. Manifestations of onchocerciasis include eye lesions, blindness, skin color changes, subcutaneous nodules etc. Therefore, serious social-economic problems such as labor loss, social economic status change and stigma of infected people have been caused by onchocerciasis. These problems are more prominent in the remote area where onchocersiasis has higher prevalence than the other places. Various means of efforts have been done to control or regionally eradicate onchocerciasis. The Onchocerciasis Control Program (OCP) was launched in 1974 in seven West African countries. The major strategy was to aerially spray insecticide to kill the larvae of black flies. Later, Merck & Co. Inc. donate ivermectin into this program in addition to vector control. Ivermectin is still the only medicine currently available in the market which can paralyze the microfilariae in the human system and stop their inseminating process. Another program, the African Program for Onchocerciasis Control (APOC) was introduced in 1995 to control onchocerciasis in the remaining endemic countries in Africa, and this program will be end in 2010. All these efforts have been successfully decreased the prevalence of disease.

One usual way to survey the prevalence of Onchocersiasis is to test the infection rate in the vector population. It is generally believed that once the infection rate below a certain level in black flies, onchocerciasis is effectively controlled in human population. However, testing a large number of black flies individually (usually done by dissection under microscope) is very time consuming and expensive. An alternative is to pool

certain number of black flies together and test the pool by polymerase chain reaction(PCR) method (Rodríguez-Pérez et al.,2004; Yamèogo et al.,1999;Goodman et al., 2003).

2. Pool Screening and Objectives

"Pool screening" and "group testing" are often used in literature by different authors, but they both refer to a procedure that tests subjects in pools or groups instead of testing individuals. In this case subjects may be insects, plants, patients, chemical agents. Pool screening testing procedure is usually implemented when the proportion of subjects testing positive is very low (for instance, rare disease with prevalence less than 0.1%), or when a screening method requires group testing in order to reach the lowest limit of getting accurate values. Therefore pool screening is more cost effective compared to individual test. Another advantage of pool screening is that the identity of subjects could be kept confidential. The outcome of pool testing is usually dichotomous, either positive or negative. When the outcome of pool testing is negative, all the subjects in this pool are declared negative. When the result of pool screening is positive, one or more subjects in this pool are positive. Note that pools do not necessarily contain the same number of subjects.

The goal of pool screening could be classifying positive and negative individuals or estimating the probability of individual being positive in whole population. In some cases, one may be interested in both classification and estimation. If it is classification problem, subjects in positive pools should be rescreened in smaller pool size or be tested individually. An example of a classification problem is applying pool screening method

to identify each human immunodeficiency virus (HIV) positive blood sample from many donors. Experiment will stop only when each positive or negative sample has been classified. On the other hand, in HIV surveillance studies, the goal is to estimate the prevalence of HIV. Investigation may just stop at the group testing level without knowing individual HIV status. Furthermore, individual subjects (e.g., mosquitoes, flies) may no longer be available for testing after the pooling process.

3. Pool Screening Assumptions

There are several commonly used assumptions in pool screening: 1) All individuals are independent, identically distributed (i.i.d) and hence pools are also independent; 2) Pool screening test has perfect sensitivity and specificity and there is no lost of accuracy of pool test compared to individual test; 3) Pool sizes are nonrandom and known; and 4) Cost of time, financial resources, manpower of pool screening is greatly reduced compared to individual screening test. Under the first assumption, we assume individuals are i.i.d. whether they are in the same pool or not. Furthermore, pooling process should be random, i.e., there is no pattern followed in grouping subjects. The second assumption requires no chance for false negative and false positive to happen in the pool screening, in other words, screening test could correctly identify each true positive and true negative without error. Furthermore, accuracy of screening test at group level should be as good as at individual level. Pool sizes are very important in pool screening; however, we assume pool sizes are nonrandom and known quantities in this research. The last assumption is a practical issue. Originally, pool screening was motivated by logistic concerns in addition

to efficiency. If there were no such major benefits then the values of pool screening should be questioned.

4. Literature Review from Statistical Perspective

Pool screening method has been widely used in many fields, such as disease vector control (Katholi and Unnasch, 2006; Durnez and Portaels, 2008; Gu et al., 2008), HIV research (Emmanuel et al., 1988; Cahoon-Young et al., 1989; Kline et al., 1989; Busch, 1991; Litvak et al., 1994; Tu et al., 1995), phytopathology (Marion, 1936; Chiang and Reeves, 1962), drug discovery (Xie et al., 2001; Zhu et al., 2001) etc. In this section, we review some common statistical issues discussed in the literature. We start by formally defining the statistical model for pool screening.

4.1 Statistical Model

Suppose $x_1, x_2, ..., x_m$ are pool testing results of M pools with sizes $n_1, n_2, ..., n_m$, where

$$x_i = \begin{cases} 1, \text{ for i}^{\text{th}} \text{ pool test positive} \\ 0, \text{ for i}^{\text{th}} \text{ pool test negative} \end{cases}$$

Let p, be 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 tests positive is $1-(1-p)^{n_i}$. In this case, the random variable X_i follows a *Bernoulli* distribution given by

$$f(x_i \mid p, n_i) = \left[1-(1-p)^{n_i}\right]^{x_i} \left[(1-p)^{n_i}\right]^{1-x_i}$$

When all M pools have sizes equal to a constant K and since pools are independent identically distributed (i.i.d.), it is well known that sum of positive pools, $T = \sum_{i=1}^{m} X_i$ is distributed as *Binomial*(M, $1-(1-p)^K$). It follows from standard theory that the maximum likelihood estimate (MLE) of $1-(1-p)^K$ is $\frac{T}{M}$. By the invariance property, the MLE of p is $\hat{p} = 1-\left(1-\frac{T}{M}\right)^{\frac{1}{K}}$. Furthermore, if we let $\pi = 1-(1-p)^k$, so that $p = 1-(1-\pi)^{\frac{1}{k}}$, then it follows from the asymptotic properties of MLEs that, as $M \to \infty$, $\sqrt{M}\,(\hat{p}-p) \xrightarrow{d} N(0, \mathrm{var}(p))$

where Var(p), by the Delta-method, is given by,

$$\mathrm{var}(p) = \left(\frac{\partial p}{\partial \pi}\right)^2 \pi(1-\pi) = \frac{1-(1-p)^K}{K^2(1-p)^{K-2}}$$

Therefore, the asymptotic distribution of $\hat{p}$ is: $\hat{p} \xrightarrow{d} N\left(p, \frac{1-(1-p)^K}{MK^2(1-p)^{K-2}}\right)$

Note that if K=1, i.e. when we are testing every subject, Var(p)=p(1-p).

4.2 Retesting Scheme of Classification Problem

Even though earlier implementation of group testing can be found in Marion's (1936) experiment where he investigated the relationship between number of viral infected aphid and plants, Dorfman (1943) is well recognized as the first person who introduced group testing into statistical field. The motivation of Dorfman's work was identifying syphilitic antigen positive individuals among army man by pool screening of blood samples. If the

screening test is positive, each individual in the positive pools will be retested until all each subject is identified as either positive or negative. After Dorfman's work, many other retesting schemes have been developed mainly focusing on improving efficiency of classifying all the positive and negative individuals. Sterrett (1957) proposed to divide positive groups into subgroups, such as half of the initial pool size, he concluded that positive individuals could be identified more efficiently than Dorfman's original method. More complicated testing/retesting schemes were developed later by others (Milton and Groll, 1966; Chen and Swallow, 1990; Hsu, 1995).

4.3 Pool Size

Since one of the major purposes of pool screening is for cost efficiency, researchers may be tempted to increase pool size. However, the accuracy of the laboratory screening test may likely decrease as pool size increases. In addition, when almost all the pools are positive, testing bias regarding p will increase because the estimator $\hat{p}$ is very close to 1 even if the actual p is low. On the other hand, when pool size is too small, it will more likely lead to a large number of negative pools and cost efficiency, which is the essence of pool screening, will be lost. Therefore, selection of a proper pool size is critical in pool screening and has been investigated by several authors. We will summarize the choice of pool size by two schools of thought: for the purpose of classification and for the purpose of estimation

Pool size for the objective of clarification. This is the classic "blood testing" problem as originally proposed by Dorfman (1943). In summary, suppose all the negative pools

are cleared, and all the subjects in positive pools will be tested individually (this design is often referred as two-stage pool screening). Based on the assumed *p*, the question is how to choose the group size to minimize the expected number of tests per subject. Some solutions to this problem were proposed by several authors (Samuels, 1978; Turner et al., 1988). The basic mathematical models are all same and can be described as follows. When the equal pool size $K \geq 2$, total number of test for a negative pool is 1 with probability $(1-p)^K$, and the total number of test for a positive pools is $K+1$ with probability $1-(1-p)^K$. Hence the expected number of test per subject is

$$f(K) = \frac{(1-p)^K + (K+1)(1-(1-p)^K)}{K} = 1-(1-p)^K + \frac{1}{K}.$$

The above question should be equivalent to the question of finding the value of *K* that can minimize $f(K)$ given a specific *p*.

Pool size for the objective of estimation. Chiang and Reeves(1962) suggested to choose pool size that there should be half positive pools and half negative pools. They solved the equation $1-(1-p)^K = (1-p)^K = 1/2$ with pool size $K = \log(1/2)/\log(1-p)$ and pointed out that pool size is a decreasing function of *p*. Thompson (1962) chose the pool size by minimizing mean square error (MSE) , he noticed that when *p* is small and number of subjects is large, the square of the bias term in MSE decreased rapidly, and hence, it is reasonable to minimize asymptotic variance term by minimizing the MSE. In his result, he found that the appropriate pool size is computed as $K = \frac{1.5936-p}{p}$.

Besides above theoretical pool size constraints, there are also application constraints such as capability of screening test which is well explained by Katholi and Unnasch (2006). An instructive example in their paper assumes testing p=0.0025. Chiang and Reeves formula recommends a pool size of 277. Thompson suggests a pool size of 635. But all these pool sizes are far beyond current laboratory testing capability. The maximum pool size is 100 for PCR and 50 for anti-body. Therefore authors suggested that the upper limit for pool size is a function of both statistical property and testing applicability, and it is more likely to be defined by the latter.

4.4 Sensitivity, Specificity, Dilution Effect of Pool Screening Test

Test accuracy concerns came mostly from HIV research literatures within the past two decades (Kline et al., 1989; Litvak et al., 1994; Tu et al., 1995; Wein and Zenios, 1996). Tu et al.(1995) pointed out that pool screening actually improved the test accuracy than individual test (ELISA test in HIV) when the condition $\phi + \psi > 1$ satisfied, where ϕ, ψ denote sensitivity and specificity respectively. In addition to this main conclusion, there are several other interesting statistical results in this paper as shown below

a) MLE of p is $\hat{p} = 1 - \left(\frac{\phi - \frac{T}{M}}{\phi + \psi - 1} \right)^{\frac{1}{K}}$, and $\hat{p}$ asymptotically normal with mean p and variance $\mathrm{var}(\hat{p}) = \frac{(1-p)^2 \left[1 - f_c(p)\right](1-p)^{-2K} f_c(p)}{MK^2(\phi + \psi - 1)^2}$ for $0 < p < 1$

,where $f_c(p) = \left[1 - (1-p)^K\right]\phi + (1-p)^K(1-\psi)$.

b) Bias $E(\hat{p}-p)=\frac{K-1}{2(1-p)}\text{var}(\hat{p})+O(M^{-3/2})\geq 0 \text{ for } 0<p<1$

Dilution effect happens when *p* decreases, i.e., the probability of finding true negative subjects decrease. When *p* is small, there might be individuals in this pool who are positive but the pool tested negative. Hung and Swallow (1999) recommended using smaller pool sizes in pool screening over individual tests.

4.5 Pool Screening Estimators and their Properties

Other than maximum likelihood estimate, there is another widely used estimator called the minimum infection rate (MIR) defined as the fraction of number of positive pools over total number of pools. Apparently this estimator will severely underestimate true *p* when studying none rare disease since positive pools could contain one or more than one individuals. This problem has been cautioned by several researchers (Gu et al., 2003; Katholi and Unnasch, 2006; Gu et al., 2008).

Barker(2000) took the Taylor expansion about the expectation of sum of positive pools in her dissertation research on MLE $\hat{p}=1-\left(1-\frac{T}{M}\right)^{\frac{1}{K}}$. Her results are summarized by Katholi and Unnasch (2006). However, if we let $\phi=1$ and $\psi=1$ (perfect test assumption) in above results of Tu et al.(1995), exactly the same results could be reached regarding bias and asymptotic variance ,

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

Swallow(1985) had more detailed discussion regarding variance, bias, total number of subjects screened, and pool size. Similar points could be noticed from above two

expressions here. First, MLE has upward bias given $K > 1$ and $0 < p < 1$, if we fixed the constants *p* and *K*, both bias and variance decrease as the number of pools (*M*) increases. Hence mean square error (MSE) decreases. Secondly, if *K* and *M* are constant, then bias, variance and consequently, the MSE, will increase with *p* increases.

Burrows(1987) proposed another point estimator $\tilde{p} = 1 - \left[\frac{2KT + K - 1}{2KM + K - 1} \right]^{\frac{1}{K}}$. He suggested that $\tilde{p}$ performs uniformly better than $\hat{p}$ in terms of bias and MSE when *K*>1.

5. Hypothesis Testing in Pool Screening and Dissertation Outline

From preceding literature review, note that most statistical inference problems have been well studied and some results are commonly accepted under equal pool size situation. However, pools do not necessarily contain the same number of subjects either due to the study design or by the nature of sample collection. For example, a project needs to collect specific insects in an area and then apply pool screening method to estimate the prevalence of infection among the vector population. It makes more sense that field workers will stop the collecting process in a certain time interval rather than counting the number of individual insects to reach a specific constant number. For large scale study across different regions, laboratories will most likely use different PCR machines, and each PCR machine has its own sample size range. Furthermore, it is more likely that field workers bring back all collected samples and divide them into different pool sizes to apply laboratory screening test.

Hypothesis testing is usually applied to assist decision making of whether to continue or stop the disease control program. Choosing the most appropriate test procedure is

extremely important not only just for the statistical properties but also for the well being of the residents in the endemic area. In order to achieve the above objective, we ask the following questions: 1) How many test procedures are available? 2) Which one is better than the others and how to compare them? 3) Is there any uniformly most powerful (UMP) test procedure, and how to prove or disprove it? 4) Can we use a sufficient statistic to develop a testing procedure? 5) What is the distribution of the sufficient statistics and how can we use this distribution to compute probabilities and quantiles of to define a rejection region for the test? 6) Suppose there is an exact test, how does it compare with the asymptotic tests?

This dissertation will try to address the above problems under unequal pool size situation. We will also focus on the settings that when p is extremely low such as in the above onchocerciasis example. After several decades of disease control efforts, the probability that a black fly being positive may, for instance, be between 1/1000 and 1/10000. For the case of equal pool size, the number of positive pools is a natural test statistic to choose in hypothesis testing. In Chapter 2, we will find the sufficient statistic for p and then define the distribution and characteristics of number of positive pools, T. Recursive methods to calculate the distribution of T are explored. Chapter 3 focuses on two-sided hypothesis tests. We propose an exact test based on the number of positive pools and modifications of the likelihood based tests where simulation quantiles used instead of critical points from standard distribution to define the rejection region. We compare these proposed tests and the asymptotic likelihood based tests with each other in terms of statistical powerChapter 4 investigates the one-sided likelihood ratio test in unequal pool size. In particular, we examine the distribution and behavior of the one-

sided likelihood ratio (LR) statistic. We derive the connection between the distribution of LR statistic with the distribution of number of positive pools. We propose modifications of the one-sided LR test and compared these tests with the exact and asymptotic tests in terms of statistical power via Monte Carlo simulations . Finally, in Chapter 5, we summarize all the results from previous chapters and discuss future research topics.

POOL SCREENING: AN EXAMPLE OF INDEPENDENT NON-IDENTICAL BERNOULLI TRIAL

by

HONGJIANG GAO, INMACULADA ABAN, CHARLES R. KATHOLI

In preparation for submission

Format adapted for dissertation

SUMMARY

Pool screening is widely used design for studies which aims to make inferences about the probability of a subject being positive when the probability is extremely low. When pool sizes are unequal, the outcomes of pool screening are assumed to be independent non-identical *Bernoulli* distribution. Of interest in this paper is the distribution of, *T*, the number of positive pools. It is shown that this distribution is the sum of the Cartesian products of *W* different *Binomial* distributions where *W* denotes the number of distinct pool sizes. Characteristics of the distribution of *T* are obtained using the generating functions. Recursive methods of computing the distribution are discussed.

KEY WORDS: Pool Screening, Number of Positive Pools, Recursive Method.

1. INTRODUCTION

When investigating a very rare dichotomous event, pool screening technique is widely applied for the purpose of efficiency. In this technique, a number of subjects are usually pooled together to form groups. Each group will be tested as one unit instead of testing individual subjects. 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.

Estimating the probability (denoted by *p*) of a subject being positive is one of the primary purposes of statistical inference in pool screening. Tu et al. (1995) and Barker (2000) showed that first order approximations to the bias and variance of the maximum likelihood estimate (MLE) are

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

$$\mathrm{Var}(\hat{p}) = \frac{1-(1-p)^K}{MK^2(1-p)^{K-2}} ,$$

where M denotes total number of pools and K denotes pool size. 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 $K = \log(1/2)/\log(1-p)$ to compute the pool size with the aim of having half positive and half negative pools. Thompson (1962) proposed pool size formula $K = (1.5936 - p)/p$ which minimizes the MSE.

Further statistical inference such as hypothesis testing and confidence interval construction should based on appropriate test statistics. One natural choice would be number of positive pools. Katholi (2009) summarized pool screening hypothesis testing based on this statistic under equal pool size situation. In this study note, we will investigate the same test statistics under unequal pool sizes setting with respect to its distribution, statistical properties, and computational methods.

2. DISTRIBUTION FOR NUMBER OF POSITIVE POOLS

Let us first assume that all individual subjects in pool screening are independent, identically distributed (i.i.d.), and the pools are independent of each other. Secondly, we

assume that the screening test has perfect sensitivity and specificity. Suppose that there are *W* distinct values of the pool sizes denoted by $n_1, n_2, .., n_w$, and, corresponding to each distinct pool size, there are $m_1, m_2, .., m_w$ pools so that $M = \sum_{i=1}^{W} m_i$. Given a pool of size n_i, the probability that the pool tests negative is $(1-p)^{n_i}$. Let $x_{ij} = 1$ if the ijth pool is positive, and $x_{ij} = 0$ otherwise. The random variable X_{ij} has *Bernoulli* distribution given by

$$f(x_{ij} \mid p, n_i) = \left[1-(1-p)^{n_i}\right]^{x_{ij}} \left[(1-p)^{n_i}\right]^{1-x_{ij}}, \; i = 1, 2, ..., W \text{ and } j = 1, 2, ..., m_i \;,$$

$$1 \le n_i \le N_{max}, x_{ij} \in \{0,1\} \qquad (1)$$

Define $t_i = \sum_{j=1}^{m_i} x_{ij}$, for *i=1,2, …, W,* as the number of positive pools among the pools with the same size n_i. The following theorem summarizes the joint distribution of (T_1 $T_{2, \ldots,}$ T_w).

Theorem 1: Let each element of vector $\langle T_1, T_2, ..., T_w \rangle'$ denote the sum of positive pools corresponding to pool size $n_1, n_2, ..., n_w$, then this vector

1) has probability mass function: $\prod_{i=1}^{W} \binom{m_i}{t_i} \left[1-(1-p)^{n_i}\right]^{t_i} [(1-p)^{n_i}]^{m_i - t_i}$ for $t_i = 0, 1, ..., m_i$ and $0 < p < 1$

2) is jointly sufficient for *p*.

Proof: Apparently, each T_i has the *Binomial* distribution

$P(T_i = t_i \mid m_i, p) = \binom{m_i}{t_i} \left[1-(1-p)^{n_i}\right]^{t_i} \left[(1-p)^{n_i}\right]^{m_i - t_i}$ for $i = 1, 2, ..., W$. Since all subjects are independent, pools are independent as well. Thus the joint distribution of vector

$\langle T_1, T_2, ..., T_w \rangle'$ is equal to the product of the above *Binomial* distributions, which is

$P(T_1 = t_1, T_2 = t_2, ..., T_W = t_w \mid p, m_i) = \prod_{i=1}^{W} \binom{m_i}{t_i} \left[1-(1-p)^{n_i}\right]^{t_i} [(1-p)^{n_i}]^{m_i - t_i}$ .The likelihood for a given sample is $L(x_{ij}, n_i, p) = \prod_{i=1}^{W} \left[1-(1-p)^{n_i}\right]^{t_i} \left[(1-p)^{n_i}\right]^{m_i - t_i}$. Since

$\frac{L(x_{ij}, n_i, p)}{P(T_1 = t_1, T_2 = t_2, ..., T_W = t_w \mid p, m_i)} = \left(\prod_{i=1}^{W} \binom{m_i}{t_i} \right)^{-1}$ is independent of *p*, we get the sufficiency result using the definition of sufficient statistics. □

Next we investigate the distribution of the statistic denoting the number of positive pools across all pool size categories.

Theorem 2: Let $T = \sum_{i=1}^{W} t_i$ be the total number of positive pools. The probability mass function of *T* is

$$g(T = t \mid m_i, p) = \sum_{t \in Q} \prod_{i=1}^{W} \binom{m_i}{t_i} \left[1-(1-p)^{n_i}\right]^{t_i} \left[(1-p)^{n_i}\right]^{m_i - t_i}, T = 0,1,2,...,M \qquad (2)$$

Where Q is defined as

$Q_t = \left\{ X \mid x_{ij} \in \{0,1\}, i = 1,2,...,W,\ j = 1,2,...,M_i, \sum_{j=1}^{M_i} x_{ij} = t_i \text{ and } t = \sum_{i=1}^{W} t_i \right\}$ and

$0 < p < 1$.

Proof: From Theorem 1, {T_i, *i=1* ,.., *W*} are independent $Binomial(M_i, 1-(1-p)^{n_i})$ random variables so that their joint distribution is simply the product of their marginal distributions. Finally, the distribution of T is equal to the sum of all possible permutations to get $\sum_{i=1}^{W} t_i = t$. □

Clearly, when all *M* pools have sizes, say *K*, pools are independent identically distributed (i.i.d.). Thus, *T* has *Binomial* $(M, 1-(1-p)^K)$ distribution which is a special case of equation (2).

Example: Suppose 3 pools are screened: 1 pool has size n_1, the other 2 pools have size n_2. From the first part of Theorem 1, $\langle T_1, T_2 \rangle'$ has distribution

$$P(T_1 = t_1, T_2 = t_2 \mid p, m_1, m_2) = \binom{1}{t_1}\left[1-(1-p)^{n_1}\right]^{t_1}\left[(1-p)^{n_1}\right]^{1-t_1}\binom{2}{t_2}\left[1-(1-p)^{n_2}\right]^{t_2}\left[(1-p)^{n_2}\right]^{2-t_2}$$

where $T_1 = \{0,1\}$ *and* $T_2 = \{0,1,2\}$. By the second part of Theorem 1, vector $\langle T_1, T_2 \rangle'$ is a sufficient statistic for *p* . According to Theorem 2, the total number of positive pools, $t \in \{0,1,2,3\}$. For instance, the distribution of $T = 2$ is

$$\begin{aligned} P(T = 2) = &\binom{1}{0}\left[1-(1-p)^{n_1}\right]^{0}\left[(1-p)^{n_1}\right]^{1}\binom{2}{2}\left[1-(1-p)^{n_2}\right]^{2}\left[(1-p)^{n_2}\right]^{0} \\ &+\binom{1}{1}\left[1-(1-p)^{n_1}\right]^{1}\left[(1-p)^{n_1}\right]^{0}\binom{2}{1}\left[1-(1-p)^{n_2}\right]^{1}\left[(1-p)^{n_2}\right]^{1} \end{aligned}$$

Barker (2000) described the following expressions for pool screening

$$g(T = t \mid n_i, M) = \left[(1-p)^{\sum_{i=1}^{M} n_i}\right]\sum_{x \in \Omega_t}\prod_{i=1}^{M}\left[\frac{1-(1-p)^{n_i}}{(1-p)^{n_i}}\right]^{x_i} \qquad (3)$$

where Ω_t is defined as $\Omega_t = \left\{ X \mid x_i \in \{0,1\}, i = 1,2,...,M, t = \sum_{i=1}^{M} x_i \right\}$. Comparing equations (2) and (3), the only difference is that equation (2) further grouped samples based on known distinct pool size. Otherwise, they are essentially equivalent.

Theorem 3: The moment generating function and cumulant generating function of *T* are respectively given by

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

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

Therefore, the first four cumulants are given by

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

Furthermore, skewness and kurtosis are given by

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

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

Proof: 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 Theorem 3 is the variance of *T*. If let $\pi_i = 1-(1-p)^{n_i}$, then $\bar{\pi} = \sum_{i=1}^{M}\left[1-(1-p)^{n_i}\right]/M$ 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 (i.e., Table 1), 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 (Figure. 1). 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.

Table 1 Example of Different Pool Sizes (Generated From Discrete Uniform distribution [25, 50]).

26,29,25,26,47,38,40,29,42,28,41,32,27,50,29,47,33,39,47,48,50,26,49,46,32,33,49,40, 31,34,43,41,50,25,44,36,27,37,41,28,42,37,36,40,47,27,43,37,40,27

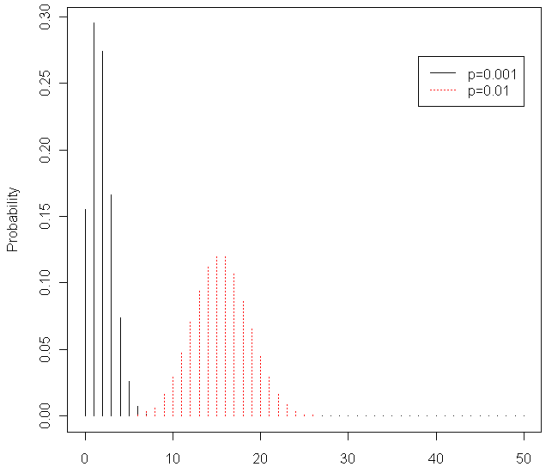

Figure 1 Distribution of T With Different p (pool sizes based on Table 1)

3. COMPUTING THE DISTRIBUTION OF *T*

One direct approach to compute distribution of T is to derive all W binomial distributions based on equation (2), and then calculate Cartesian products of all these *Binomial* distributions. The sum of the terms of the above Cartesian product with $\sum_{i=1}^{W} t_i = t$ is the probability of T. For a given value of T and W, the sum, $\sum_{i=1}^{W} t_i = t$, would be over the set of *weak compositions* of t defined by $\Omega_t = \left\{ (t_1, t_2, ..., t_W) \mid \forall i, 0 \le t_i \le m_i \text{ and } t = \sum_{i=1}^{W} t_i \right\}$. Note that when $t \le \min(m_1, ..., m_W)$ then

there are exactly $\binom{t+W-1}{t}$ such compositions and these are easily generated and do not require use of the additional constrains. However, when $t \geq \min(m_1,...,m_W)$ the constrains apply, consequently, applications become more difficult.

Another option to calculate the distribution of T is to apply recursive methods by equation (3). First way 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

$$tS_t = \sum_{i=1}^{M} (-1)^{i-1} S_{t-i} R_i \quad \text{for } t = 1,...,M \qquad (4)$$

Where $R_i = \sum_{j=1}^{M} a_j^i$, $S_0 = 1, S_{t-i} = 0$ for $t < i$, and $a_i = \left[\frac{1-(1-p)^{n_i}}{(1-p)^{n_i}} \right]^{x_i}$. For illustration, suppose we have three pools with pool size n_1, n_2, and n_3. If we let

$R_1 = a_1 + a_2 + a_3, R_2 = a_1^{\,2} + a_2^{\,2} + a_3^{\,2}, R_3 = a_1^{\,3} + a_2^{\,3} + a_3^{\,3}$. Then from equation (4),

$S_1 = (-1)^0 S_0 R_1 = R_1 = a_1 + a_2 + a_3, 2S_2 = (-1)^0 S_1 R_1 + (-1)^1 S_0 R_2 = S_1 R_1 - S_0 R_2$

$3S_3 = (-1)^0 S_2 R_1 + (-1)^1 S_1 R_2 + (-1)^2 S_0 R_3 = S_2 R_1 - S_1 R_2 + S_0 R_3$.

If we define $c(p, n_i) = (1-p)^{\sum_{i=1}^{M} n_i}$, then distribution of T is $g(T = i \mid n_i, M) = c(p, n_i) S_i$ for $i = 0,...,3$. Unfortunately, the Newton recursion is unstable and essentially worthless for practical computation.

The second recursive method we have explored is applying double recursive relations in the proof of Marcus and Lopes (1957) inequality. If we let $D_t = \sum_{x \in \Omega_t} \prod_{i=1}^{M} \left[\frac{1-(1-p)^{n_i}}{(1-p)^{n_i}} \right]^{x_i}$,

and $D_t^*(a_i)$ represent D_t excluding all terms involving a_i for $i \in \{1,2,3,...,M\}$. Then the recursive relations can be express as

$$tD_t = \sum_{i=1}^{M} a_i D_{t-1} - \sum_{i=1}^{M} a_i^2 D_{t-2}^*(a_i) \ , t=2\ldots M,$$

$$D_t^*(a_i) = D_t - a_i D_{t-1}^*(a_i) \qquad ,t=2\ldots M$$

Detailed discussions of this method including numerical issues can be found in Gao et al.(2009).

In order to illustrate Marcus and Lopes recursive method, we consider the equal pool size where *T* has a *Binomial* distribution. Because *T* has a *Binomial* distribution, its probabilities may be also obtained using standard statistical or mathematical programs. We can then compare the probabilities given by these two methods. In this illustration, we use FORTRAN 95 (Absoft) and IMSL to compute the *Binomial* probabilities. Suppose there are a total of 150 (*M*=150) pools screened of equal pool sizes n=50. Let the probability of each subject being positive be 0.0001 (*p*=0.0001).Under this setting, the number of positive pools *T* has $Binomial(150,(1-0.0001)^{50})$ distribution. Results of these computations are displayed in Table 2 which shows that Marcus and Lopes recursive method and Fortran IMSL give practically identical results considering the significant digits provided.

Table 2 Comparing Distribution Calculated by Recursive Methods with IMSL Standard Package.(Number of pools=150, pool sizes=50, p=0.0001. Results of $t<120$ are omitted).

T	IMSL	Marcus and Lopes Method
120	2.88695198042033E-127	2.88695198042079E-127
121	3.67173022365900E-129	3.67173022365982E-129
122	4.47717686978659E-131	4.47717686978748E-131
123	5.22820359182048E-133	5.22820359182128E-133
124	5.83969130068907E-135	5.83969130069022E-135
125	6.23086791115188E-137	6.23086791115334E-137
126	6.34181148345855E-139	6.34181148345940E-139
127	6.14774957726811E-141	6.14774957726872E-141
128	5.66668869341169E-143	5.66668869341158E-143
129	4.95744203737101E-145	4.95744203737094E-145
130	4.10798555064724E-147	4.10798555064752E-147
131	3.21723605530638E-149	3.21723605530665E-149
132	2.37551576375674E-151	2.37551576375756E-151
133	1.64920294119818E-153	1.64920294119865E-153
134	1.07328125266586E-155	1.07328125266612E-155
135	6.52521877989170E-158	6.52521877989419E-158
136	3.69183895420507E-160	3.69183895420559E-160
137	1.93528769933870E-162	1.93528769933822E-162
138	9.35201343297653E-165	9.35201343297439E-165
139	4.14158794208808E-167	4.14158794208867E-167
140	1.66927103202685E-169	1.66927103202611E-169
141	6.07299696376894E-172	6.07299696376974E-172
142	1.97447913544355E-174	1.97447913544368E-174
143	5.66632940630066E-177	5.66632940629919E-177
144	1.41296916987072E-179	1.41296916987051E-179
145	2.99924023722591E-182	2.99924023722179E-182
146	5.26894569172252E-185	5.26894569173594E-185
147	7.35464472759044E-188	7.35464472758972E-188
148	7.64744844230237E-191	7.64744844232092E-191
149	5.26569384244740E-194	5.26569384240432E-194
150	1.80077588225930E-197	1.80077588229788E-197

4. CLOSING REMARKS

We have presented the distribution and properties of number of positive pools in pool screening which a special case of independent non-identical *Bernoulli* trial. The recursive computational methods for this distribution are also explored. Further statistical inference such as hypothesis testing, confidence interval construction regarding this test statistic can be achieved based on the results in this article.

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,AL.

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

GAO,H., ABAN,I., and KATHOLI,C.R.,2009, "Two-sided Hypothesis Testing Based on Pool Screening with Unequal Pool Sizes", Biostatistics Department Technical Report, BST2009-002, Available at https://www.soph.uab.edu/bst/technical.

Katholi, C. R., "The Role of Polymerase Chain Reaction Techniques for Assessing Lymphatic Filariasis Transmission", Appendix 5 and Annexes A – E, "Hypothesis Testing With the Pool Screening Model", World Health Organization, Report Number WHO/HTM/NTD/PCT/2009.1, 2009.

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

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.

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

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.

TWO-SIDED HYPOTHESIS TESTING BASED ON POOL SCREENING WITH UNEQUAL POOL SIZES

by

HONGJIANG GAO, INMACULADA ABAN, CHARLES R. KATHOLI

In preparation for submission

Format adapted for dissertation

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 num of positive pools 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

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

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 laboratory 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(Yamèogo 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 (2009) 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. Limitations and recommendations will be discussed in the Section 5.

2. Distribution and Computation Method for 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 and Distribution for Number of Positive Pools

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 $n_1, n_2, ..., n_M$,

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

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

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

Further more, 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. Detailed statistical properties of statistic T will be discussed later by authors in another study note. 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).

Since $1-(1-p)^{n_i}$ is a monotone increasing function of p, Marcus and Lopes (1957) inequality condition is satisfied. It can be shown that number of positive pools possesses

the monotone likelihood ratio (MLR) property with respect to the parameter *p* when pool sizes are unequal. Detailed proof of above property in a more general setting was done by Huynh (1994).

2.2 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.

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 $3 \le t \le m-3$.

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

$S_1^*(a_1) = a_2 + a_3 + ... + a_m$, $S_2^*(a_1) = a_2a_3 + a_2a_4 + ... + a_2a_m + ... + a_{m-1}a_m$. Based on double recursive relations mentioned in the proof of Theorem 1 in Marcus and Lopes (1957) paper,

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

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

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

Above 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 is overflow problem 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

$H_a : p \neq p_0$ 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 1: 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 is given in the Appendix.

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 1 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.

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

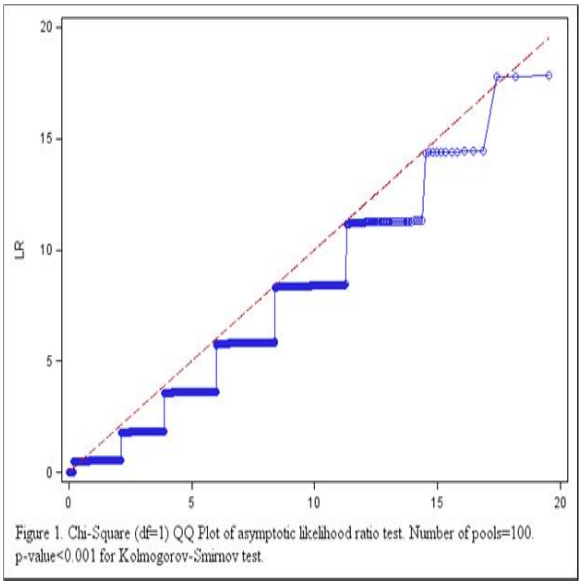

Figure 1. Chi-Square (df=1) QQ Plot of asymptotic likelihood ratio test. Number of pools=100. p-value<0.001 for Kolmogorov-Smirnov test.

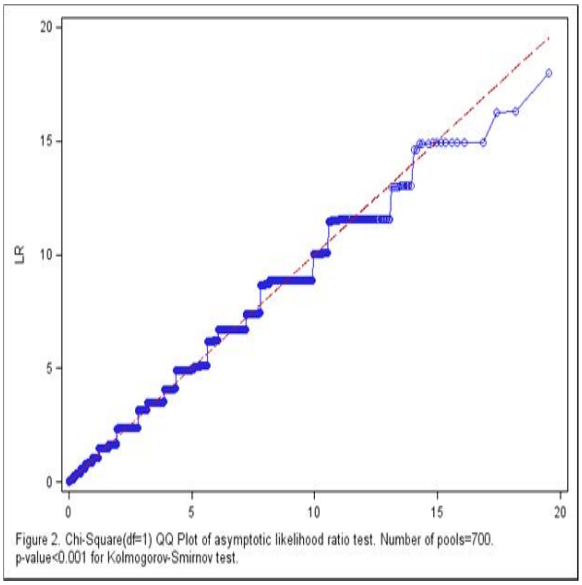

Figure 2. Chi-Square(df=1) QQ Plot of asymptotic likelihood ratio test. Number of pools=700. p-value<0.001 for Kolmogorov-Smirnov test.

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 1. 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.

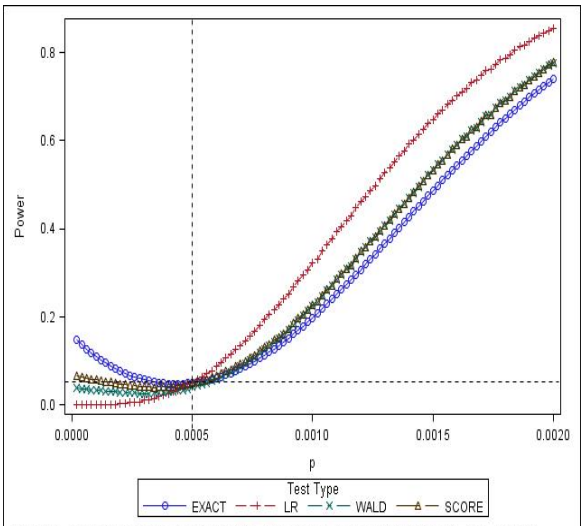

Figure 3. Simulated power curves for testing , type I error=0.05, number of pools=100. Cut off points for rejection under each alternative is defined by simulation quantile.

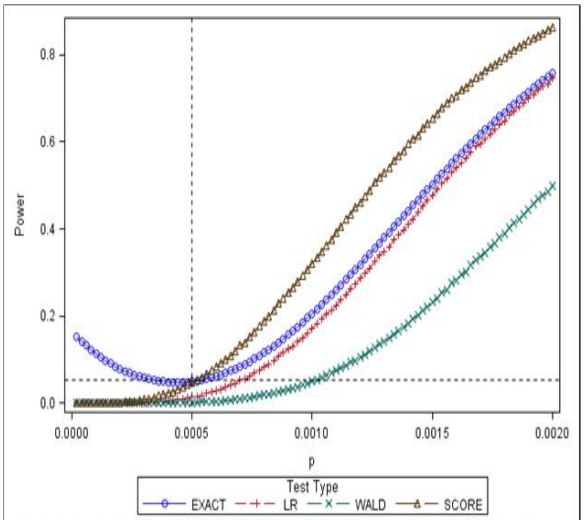

Figure 4. Simulated power curves for testing , type I error=0.05, number of pools=100. Cut off points for rejection under each alternative is defined by standard distribution.

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.

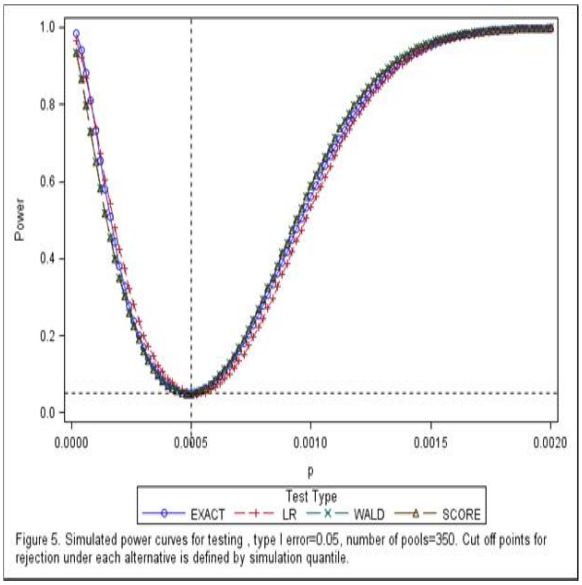

Figure 5. Simulated power curves for testing , type I error=0.05, number of pools=350. Cut off points for rejection under each alternative is defined by simulation quantile.

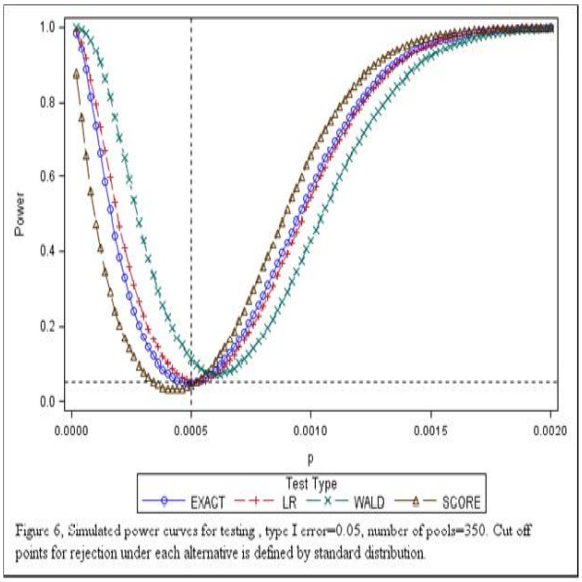

Figure 6, Simulated power curves for testing , type I error=0.05, number of pools=350. Cut off points for rejection under each alternative is defined by standard distribution.

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.

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	0.786	0.882	0.916	0.948	0.974	0.987	0.992	0.995	0.997	0.999

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 1

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 $n_i = 2$, $\frac{\partial ln f_i}{\partial p}$, $\frac{\partial ln^2 f_i}{\partial p^2}$, 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 $n_i = 1$,

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

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

$E(g_i(x_{ij} \mid n_i, p)) = \frac{2}{(1-p)^2} + \frac{2}{p^2}$,let $W_i = E(g_i(x_{ij} \mid n_i, p)) + \varepsilon$, where $\varepsilon > 0$, so

$$\left|\frac{\partial \log^3 f_{ij}}{\partial^3 p}\right| < g_i(x_{ij} \mid n_i, p) \text{ and } E(g_i(x_{ij} \mid n_i, p)) < W_i$$

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

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

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

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

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

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

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

$$J(p) = \sum_{i=1}^{k} m_i \sum_{x_{ij}=0,1} \frac{\partial \log f_{ij}}{\partial p} \frac{\partial \log f_{ij}}{\partial p} f_{ij}$$

$$= \sum_{i=1}^{k} m_i \left[\left(\frac{-n_i}{1-p} \right) \left(\frac{-n_i}{1-p} \right) (1-p)^{n_i} + \left(\frac{n_i (1-p)^{n_i - 1}}{1-(1-p)^{n_i}} \right) \left(\frac{n_i (1-p)^{n_i - 1}}{1-(1-p)^{n_i}} \right) (1-(1-p)^{n_i}) \right]$$

$$= \sum_{i=1}^{k} m_i \frac{n_i^2 (1-p)^{n_i - 2}}{1-(1-p)^{n_i}}$$

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.

GOODMAN DS, O.J., ROBERTS JM, LAMMIE PJ, STREIT TG., 2003, PCR and Mosquito dissection as tools to monitor filarial infection levels following mass treatment.: Filaria J., v. Jul 7;2(1): 11.

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.

HELMY H, F.P., FARID HA, BRADLEY MH, RAMZY RM., 2004, Test strip detection of Wuchereria bancrofti amplified DNA in wild-caught Culex

pipiens and estimation of infection rate by a PoolScreen algorithm.: Trop Med Int Health., v. 9(1):158-63.

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., “The Role of Polymerase Chain Reaction Techniques for Assessing Lymphatic Filariasis Transmission”, Appendix 5 and Annexes A – E, “Hypothesis Testing With the Pool Screening Model”, World Health Organization, Report Number WHO/HTM/NTD/PCT/2009.1, 2009.

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.

VASUKI V, H.S., SADANANDANE C, JAMBULINGAM P., 2003, A simple and rapid DNA extraction method for the detection of Wuchereria bancrofti infection in the vector mosquito, Culex quinquefasciatus by Ssp I PCR assay.: Acta Trop., v. 86(1):109-14.

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.

WILLIAMS SA, L.S., BIERWERT LA, SAUNDERS LJ, BOAKYE DA, FISCHER P, GOODMAN D, HELMY H, HOTI SL, VASUKI V, LAMMIE PJ, PLICHART C, RAMZY RM, OTTESEN EA., 2002, Development and standardization of a rapid, PCR-based method for the detection of Wuchereria bancrofti in mosquitoes, for xenomonitoring the human prevalence of bancroftian filariasis.: Ann Trop Med Parasitol., v. 96, p. Suppl 2:S41-6

YAMÈOGO, 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.

ONE-SIDED LIKELIHOOD RATIO TEST IN UNEQUAL SIZED POOL SCREENING

by

HONGJIANG GAO, INMACULADA ABAN, CHARLES R. KATHOLI

In preparation for submission

Format adapted for dissertation

SUMMARY

The use of the one-sided likelihood ratio (LR) test to make inferences about parameters is often based on the asymptotic properties of the test statistic which assume some regularity conditions and large sample size. Some users of this test blindly assume these conditions hold. In this paper, we are interested in applying the one-sided LR method to test whether the prevalence of a certain rare disease has decreased based on pool screening data where the pool sizes are unequal. This paper examines the finite sample properties as well as the large sample properties of the one-sided LR statistic under this setting. We will also propose modified LR tests and compare them with the Score test and an exact test based on the number of positive pools using Monte Carlo simulations. These simulations show that the conventional LR test has inflated simulated type I error rates resulting in an artificially high power values under the alternative hypothesis. Modifying the LR test based on simulated quantiles greatly improves the performance of the LR test. However, this method requires large enough sample size to ensure at least one pool is positive. In the simulation setting considered where the prevalence rate in the null hypothesis is 0.05%, 200 pools (with pool sizes ranging between 25 and 50) are sufficient for the quantile-based LR test to work. The exact test based on the number of positive pools performs the best among all tests considered when the sample size is small.

Key words: Pool Screening, One-sided Likelihood Ratio Test, Simulated Power

1. MOTIVATION AND BACKGROUND

Pool screening is a widely used technique to estimate prevalence (denoted by *p*) of a rare disease. A researcher would often pool individual subjects or units together to test them as a group instead of testing each individual unit to be more cost effective. All subjects in a pool will be declared negative if the pool tests negative. If a pool tests positive, then at least one subject in this particular pool is positive for the disease. A common application of pool screening is in tropical infectious disease control, for instance, the Onchocerciasis Control Program (OCP) which was launched in 1974 in seven West African countries. Onchocerciasis (river blindness) is a disease caused by the microfilarial *Onchocerca volvulus*. It is transmitted by the black flies. Manifestations of onchocerciasis include blindness, skin color change, subcutaneous nodules etc.. One way to survey the infection rate is to dissect a large number of black flies individually under microscope. This is usually very time consuming and expensive. An alternative is to pool certain number of black flies together and test the pool by polymerase chain reaction (PCR) method (Rodríguez-Pérez et al.,2004; Yamèogo et al.,1999;Goodman et al., 2003). The primary objective in OCP is to eradicate this disease. An indicator of the effectiveness of this eradication program is when the transmission potential goes below a certain target level. Transmission potential is typically measured by the infection rate in vector population, i.e., carrier of the disease such as black flies in this case. From a statistical perspective, the main interest is testing the hypotheses about the parameter, *p*, defined by $H_0 : p = p_0$ versus $H_A : p < p_0$ where p_0 is a hypothesized target value of *p*.

The likelihood ratio (LR) procedure is a popular method used to test such hypotheses because of the intuitive idea behind the procedure as will be discussed in the next section. In application, the LR test is commonly used for testing two-sided hypothesis. Furthermore, the LR test procedure is often based on asymptotic results which rely heavily on the validity of the assumed regularity conditions as well as a large sample size. Some researchers use the asymptotic LR test procedures without checking the regularity conditions and assume that the sample size they have is large enough for the test to be valid. Gao et al. (2009) showed that the likelihood based test procedures (Wald, Score and LR tests) for testing two-sided hypothesis about the parameter p based on pool screening when p is very small do not perform very well. Moreover, the sample size needed for the asymptotic results to provide good approximations is too large to be practical.

A major difference between the one-sided test and two-sided LR test is that the statistic for the one-sided has a mixture distribution which makes it more difficult to obtain its distributional property. To add to this complexity is the case that the parameter of interest, p, is assumed to have values near 0. These issues associated with the one-sided LR test will be the main focus of this paper. In addition, we will propose modifications to the LR test and investigate the power functions of these tests in relation to other tests.

The outline of this paper is as follows. In the next section, we start with a review of the basic concepts of the LR test and pool screening. It will be followed in Section 3 by the presentation of the results about the exact distributional properties and the asymptotic properties of the one-sided LR test. In Section 4, we use Monte Carlo simulations to

evaluate and compare the power of the one-sided LR tests (conventional and modified) with an exact test based on the sum of positive pools and Score tests. Finally in Section 5, we summarize our conclusions and provide recommendations.

2. BACKGROUND

2.1 Likelihood Ratio Test

R. A. Fisher (1922) first gave the definitions of "*likelihood*", "*sufficiency*" and introduced "*Method of Maximum Likelihood*" for estimation in his seminal paper "*On the mathematical foundations of theoretical statistics*". Several years after, Neyman and Pearson (1928) formulated the likelihood ratio test by considering hypothesis both under the null and alternative. Wilks (1938) obtained the limiting distribution of the two-sided LR test using characteristic functions. At present, LR test is now one of the commonly used procedures for statistical inference.

The rational behind LR test is as follows: Let $L(\theta \mid X)$ denote the likelihood function from a random sample **X** with discrete probability mass function (pmf) $f(x \mid \theta)$ where θ is the scalar parameter in the simplest case. The LR test statistic for testing the hypotheses $\mathrm{H}_0 : \theta \in \Theta_0$ versus $\mathrm{H}_\mathrm{A} : \theta \in \Theta_0^c$ is defined as

$$\lambda(X) = \frac{\sup_{\Theta_0} L(\theta \mid X)}{\sup_{\Theta} L(\theta \mid X)}$$

where Θ is the unrestricted parameter space, Θ_0 is the parameter space restricted by the null hypothesis and Θ_0^c is the complement of Θ_0, i.e., $\Theta - \Theta_0$. Note that $0 \le \lambda(X) \le 1$. Given the observed sample, the null hypothesis is rejected when $\lambda(X)$ is small because this implies that the most likely parameter value that generated the data belongs to the

alternative parameter space. Therefore, a critical point *C,* $0 \le C \le 1$, is chosen based on a significance level, α, such that $P(\lambda(X) < C) = \alpha$. In this case, whenever, $\lambda(X) < C$, we reject H_0. Typically, the numerator and denominator of the LR test statistic achieve their suprema when the parameter is replaced by the maximum likelihood estimator (MLE) over the corresponding parameter space. Thus, if we let $\hat{\theta}_0$ and $\hat{\theta}$ be the MLE under restricted parameter space (null hypothesis) and unrestricted parameter space (alternative hypothesis), respectively, then the LR test statistic simplifies to

$$\lambda(X) = \frac{L(\hat{\theta}_0 \mid X)}{L(\hat{\theta} \mid X)}.$$

A desirable property of a statistical inference method is that the statistic being used is a function of the sufficient statistic. Using a sufficient statistic is a way to reduce the dimension of the data without losing information about the parameter of interest (for more details on the concept of sufficiency see for instance Casella and Berger (2001)). By invoking the Factorization Theorem, it can be shown that the likelihood function is always a function of the sufficient statistic, hence, making likelihood-based procedures desirable. A drawback in using the LR test is that oftentimes the form of the test statistic is complex, and its distribution is difficult to obtain. In cases where a one-dimensional minimal sufficient statistic exists and the distribution of this test statistic is more manageable, one may be able to construct a test equivalent to the LR test but based on the sufficient statistic. However, in the case of distribution of sufficient statistic is neither easily obtained nor calculated, one would instead use the asymptotic results. For a two-sided test of a single parameter, it is a well-known result that, under some regularity

conditions, $-2\ln(\lambda(X))$ converges in law to distribution χ^2 with 1 degree of freedom (see Lehmann,1998).

2.2 Pool Screening

In pool screening, *N* independent and identically distributed units are collected and divided into *m* pools. Given that each unit can either be positive or negative with probability, p, of being positive, then the probability that the j^{th} pool yields positive result is $1-(1-p)^{n_j}$, where n_j is the size of the j^{th} pool. . Now, if we let X_j be the indicator function that takes on a value of 1 if the j^{th} pool is positive, and 0 if the j^{th} pool is negative, then X_j is a *Bernoulli* random variable with probability mass function

$$f_{x_j}(x \mid p, n_j) = \left[1-(1-p)^{n_j}\right]^{x_j}\left[(1-p)^{n_j}\right]^{1-x_j},\ x_j \in \{0,1\}, 1 \le n_j \le n_{max},\ j = 1,..,M\ .$$

When pool sizes are equal, X_js are independent and identically distributed random variables. Consequently, the distribution of the number of positive pools, denoted by *T*, is $Binomial(M, 1-(1-p)^k)$ where *k* is the pool size. In this case, it can easily be shown using the definition of sufficiency or the Factorization Theorem, that the number of positive pools, denoted by *T*, is a sufficient statistic for *p* and has a monotone likelihood ratio property (Huynh,1994). The resulting one-sided hypothesis test for *p* will be uniformly most powerful among the tests with the same significance level.

In practice, the size of the pools is limited by laboratory requirements of a particular screening method (Katholi and Unnasch, 2006). For instance, in *Wuchereria bancrofti* (a filaria can cause lymphatic filariasis) 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;Williams et al., 2002); and in *Onchocerca volvulus* infection

control, most literature uses PCR assay method which can handle no more than 50 female black flies in a pool (Yamèogo et al., 1999; Guevara et al., 2003). Similar issues are also encountered in large-scale multi-site studies where laboratories may have varying requirements on the pool size to obtain the best readings. Thus, it is very likely that the collected data will be based on unequal pool sizes. In such cases, pools will no longer be identically distributed (because the probability of a positive pool depends on the pool size), although they are still independent, and the number of positive pools, *T*, is no longer a sufficient statistics for *p* (Gao et al., 2009). However, as shown by Gao et. al., two-sided tests based on *T* has desirable properties.

3. One-sided LR Test

Consider testing the hypotheses $H_0 : p = p_0$ versus $H_A : p < p_0$. Given the pair of data $(n_j, X_j), j = 1, \cdots, m$ on m pools where n_j, $j=1,\ldots m$, are known constants, let $\hat{p}$ be the MLE of *p*. It can be shown that the likelihood ratio statistic λ is defined as,

$$\lambda = \begin{cases} 1, \text{ when the MLE } \hat{p} \geq p_0 \\ \dfrac{\prod_{j=1}^{m} \left\{ \left[1-(1-p_0)^{n_j} \right]^{X_j} \left[(1-p_0)^{n_j} \right]^{1-X_j} \right\}}{\prod_{j=1}^{m} \left\{ \left[1-(1-\hat{p})^{n_j} \right]^{X_j} \left[(1-\hat{p})^{n_j} \right]^{1-X_j} \right\}}, \text{ when } \hat{p} < p_0 \end{cases} \qquad (1.0)$$

and that the cumulative distribution function of $W = -2\ln(\lambda)$ for the one sided test has the form

$$F_W(w) = c_1 H_1(w) + c_2 H_2(w), c_1, c_2 > 0, c_1 + c_2 = 1 \qquad (2.0)$$

where $H_1(w)$ is the indicator function defined as $H_1(w) = \begin{cases} 1, w = 0 \\ 0, w < 0 \end{cases}$ and $H_2(w)$ is a step function .

3.1 Finite Sample Case

Based on equation (2.0), the exact distribution of the LR test statistic depends on 2 unknown components: c_1 and $H_2(w)$. The other constant, c_2, can be obtained from c_1 while $H_1(w)$ is simply an indicator function. In the case we are considering, it is difficult to obtain the form of $H_2(w)$. However, we are still able to investigate and characterize the weights, c_1 and c_2, for the finite sample case. This will be the focus of this subsection.

We shall first show that under the case of unequal pool size, the MLE corresponding to having observed T positive pools, is bounded in an interval $\left[p_L(T), p_U(T) \right]$. For the special case of equal pool size, each of these intervals satisfy $p_L(T) = p_U(T) = \hat{p}(T)$, and since T takes on values in the set $\{0,1,2,\cdots,m\}$, it follows that the values of $\hat{p}(T)$ are "well separated"; that is, $\hat{p}(T) << \hat{p}(T+1)$. Katholi and Aban (2009) found that the values of the Score Statistic for the two-sided and one-sided hypotheses fall into disjoint intervals both for small or large number of pools. Given $T = \tilde{t}$, it is very likely that $\hat{p}(\tilde{t}) < p_0 < \hat{p}(\tilde{t}+1)$. In this case, it follows that $c_1 = P(T \geq \tilde{t}+1)$ and $c_2 = P(T \leq \tilde{t})$. Using the algorithm in Gao et al. (2009) for computing probabilities based on the exact distribution of T, the values of the weights can be determined.

For the more general case of unequal pool sizes, it is possible that there is a $\tilde{t}$ such that when p_0 is small, $\hat{p}_U(\tilde{t}) < p_0 < \hat{p}_l(\tilde{t}+1)$, in which case the weights may again be

calculated. In the unfortunate case where p_0 does not satisfy this condition, the exact weights cannot be calculated but may still be approximated via simulation.

The following theorem provides the basis for these observations.

Theorem 1: Let (n_j, X_j), $j = 1, \cdots, m$ be ordered pairs associated with testing m pools where the j-th pool is of size n_j and X_j, the result of the test, is a Bernoulli random variable with probability mass function,

$$f_{X_j}(x \mid n_j, p) = [1-(1-p)^{n_j}]^{X_j}[(1-p)^{n_j}]^{(1-X_j)},\ X_j \in \{0,1\},\ n_j \geq 1$$

Let $T = \sum_{j=1}^{m} X_j$, and let the likelihood function be $L(p \mid X_1, \cdots, X_m) = \prod_{j=1}^{m} f_{X_j}(X_j \mid n_j, p)$, then for $T = t$, the MLE of p, $\hat{p}$ can take on any one of $\binom{m}{t}$ values depending on which vector from the set

$$\Omega_t = \left\{ \underline{X} \mid \underline{X} = (X_1, \cdots, X_m),\ X_j \in \{0,1\}, \forall j = 1, \cdots, m \text{ and } t = \sum_{j=1}^{m} X_j \right\}$$

is observed. The set of all values has cardinality no larger than $\binom{m}{t}$ nor less than one and is such that we can associate with the set a smallest element $\hat{p}_L(t)$ and a largest element $\hat{p}_U(t)$, i.e., $\forall \underline{X} \in \Omega_t$, $\hat{p}_L(t) \leq \hat{p}(\underline{X}) \leq \hat{p}_U(t)$. Without loss of generality we may assume that the data are ordered so that $n_1 \leq n_2 \leq \cdots \leq n_m$. Then $\hat{p}_L(t)$ is equal to the MLE when $X_1 = X_2 = \cdots X_t = 1$ while $X_{t+1} = X_m = 0$; similarly, $\hat{p}_U(t)$ is equal to the MLE when $X_m = X_{m-1} = \cdots = X_{m+1-t} = 1$ while $X_{m-t} = \cdots = X_1 = 0$.

Proof: To begin**,** we note that Katholi and Aban (2009) showed that the function

$$h(\varsigma) = \frac{\varsigma}{1-(1-p)^{\varsigma}}, 0 < p < 1, \varsigma \geq 1$$

is strictly monotonically increasing as a function of ς for any fixed *p*. It is also easily shown that for fixed ς, the function is strictly monotonically decreasing for $p \in (0,1)$. Next note that the derivative of the natural log of the likelihood function in this case is the function $\frac{1}{(1-p)} f(p \mid \underline{X}, n_1, \cdots, n_m)$ where

$$f(p \mid \underline{X}, n_1, \cdots, n_m) = f(p) = \left(\sum_{j=1}^{m} X_j \left[\frac{n_j}{1-(1-p)^{n_j}} \right] \right) - \sum_{j=1}^{m} n_j \qquad (3.0)$$

The derivative of the log likelihood is zero at the MLE and this requires that $f(p)\big|_{p=\hat{p}} = 0$.

Again, by assumption $n_1 \leq n_2 \leq \cdots \leq n_m$ and using the inequalities previously developed it follows that,

$$g_U(p) = \sum_{j=1}^{t} \left(\frac{n_{m-t+j}}{1-(1-p)^{n_{m-t+j}}} \right) - N \geq \sum_{\substack{j=1 \\ \underline{X} \in \Omega_t}}^{m} X_j \left(\frac{n_j}{1-(1-p)^{n_j}} \right) - N \geq \sum_{j=1}^{t} \left(\frac{n_j}{1-(1-p)^{n_j}} \right) - N = g_L(p) \quad (4.0)$$

where *t* is the observed number of positive X_i. Next note that as $p \to 1$, $g_U(p) \to \sum_{j=1}^{t} n_{(m+1-j)} - N < 0$ and that as $p \to 0$, $g_U(p) \to \infty$. But $g_U(p)$ is continuous and strictly increasing as $p \to 0$ and so there a unique point $\hat{p}_U$ such that $g_U(\hat{p}_U) = 0$. At this point it is clear from equation (4.0) then $g(\hat{p}_U) < 0$ and $g_L(\hat{p}_U) < 0$ where

$$g(p \mid \underline{X}) = \sum_{j=1}^{m} \frac{X_j n_j}{1-(1-p)^{n_j}} - N \text{, for any } \underline{X} \in \Omega_t$$

For $g_U(p)$ to be greater than or equal to zero, *p* must be less than $\hat{p}_U$. Similarly, for $g_L(p)$ to be zero, it is necessary for p to be closer to zero than $\hat{p}_U$ and so $\hat{p}_L < \hat{p}_U$.

Finally we note that for any of the other $\binom{m}{t}$-2 elements in Ω_t the zero of the function $g(p)$ must be in the closed interval $[\hat{p}_L, \hat{p}_U]$.

End of proof.

From equation (1), when $\hat{p} \geq p_0$, then $\lambda = 1$ and $H_1(w) = 1$.Therefore, c_1 is equal to the probability of $\hat{p} \geq p_0$ under the null. When all pool sizes are equal, its exact value can be calculated applying the following theorem.

Theorem 2: Suppose all *m* pools have equal pool sizes *k*, then c_1 in equation (2.0) can be expressed as

$$c_1 = \sum_{i=\lceil M(1-(1-p_0)^k) \rceil}^{m} \binom{m}{i} (1-(1-p_0)^k)^i (1-p_0)^{k(m-i)}$$

and

$$c_2 = \sum_{i=0}^{\lfloor M(1-(1-p_0)^k \rfloor} \binom{m}{i} (1-(1-p_0)^k)^i (1-p_0)^{k(m-i)} \ ,$$

where $\lceil M(1-(1-p_0)^k \rceil$ represents the smallest integer not less than $M(1-(1-p_0)^k$ and $\lfloor M(1-(1-p_0)^k \rfloor$ represents largest integer not greater than $M(1-(1-p_0)^k$.

Proof

From equation 1.0, $c_1 = \mathrm{P}(\hat{p} \geq \mathrm{p}_0)$, by invariant property of MLE, $\hat{p} = 1-(1-T/m)^{1/k}$.

Hence $c_1 = \mathrm{P}(\hat{p} \geq \mathrm{p}_0) = \mathrm{P}((1-T/m)^{1/k} \geq \mathrm{p}_0) = \mathrm{P}(T \geq m(1-(1-p_0)^k)$. Under the null, T follows the distribution of $Binomial(m, 1-(1-p_0)^k)$, therefore, c_1 and c_2 can be calculated exactly.

End of proof.

When pool sizes are unequal, the following Corollary defines an alternative way to obtain the situations where the exact weight can be calculated.

Corollary 1: Under the assumptions of Theorem 1, let $T = \tilde{t}$, n_1 denote the smallest pool size, n_m denote the largest pool size, and N the total number of subjects screened. Define $\hat{p}_U^*$ and $\hat{p}_L^*$ as

$$\hat{p}_U{}^*(\tilde{t}) = 1 - (1 - \frac{n_m \times \tilde{t}}{N})^{\frac{1}{n_m}}$$

and

$$\hat{p}_L{}^*(\tilde{t}+1) = 1 - (1 - \frac{n_1 \times (\tilde{t}+1)}{N})^{\frac{1}{n_1}}.$$

In the case where $\hat{p}_U^*(\tilde{t}) \le \hat{p}_L^*(\tilde{t}+1)$ and p_0 lies in the interval $(\hat{p}_U^*(\tilde{t}),\ \hat{p}_L^*(\tilde{t}+1))$, by equation (2.0),

$$c_1 = \sum_{i=\tilde{t}+1}^{m} \Pr(T = i \mid p_0).$$

Proof

By equation (4.0) in Theorem 1, $\sum_{j=1}^{\tilde{t}} \left(\frac{n_m}{1-(1-p)^{n_m}} \right) \ge \sum_{j=1}^{\tilde{t}} \left(\frac{n_{m-\tilde{t}+j}}{1-(1-p)^{n_{m-\tilde{t}+j}}} \right)$, hence

$\hat{p}_U{}^*(\tilde{t}) \ge \hat{p}_U(\tilde{t})$. Note that the left side of this inequality assumes the extreme case where pool sizes are all equal to the maximum pool size. Since the pmf of X_j can be written as

$$f_{X_j}(x \mid n_j, p) = [(1-p)^{n_j}] \left[\frac{1-(1-p)^{n_j}}{(1-p)^{n_j}} \right]^{X_j}$$

then

$$L(p, n_j, \tilde{t} \mid x) = [(1-p)^{\sum_{j=1}^{m} n_j}]\left[\frac{1-(1-p)^{n_m}}{(1-p)^{n_m}}\right]^{\tilde{t}}.$$

The MLE of this likelihood, denoted by $\hat{p}_U^{\,*}(\tilde{t})$, is given by

$$\hat{p}_U^{\,*}(\tilde{t}) = 1-(1-\frac{n_m \times \tilde{t}}{N})^{\frac{1}{n_m}}.$$

Replacing $\tilde{t}$ by $\tilde{t}+1$, we find $\hat{p}_L^{\,*}(\tilde{t}+1)$, and hence, proved that $\hat{p}_L^{\,*}(\tilde{t}+1) \le p_L(t)$ by directly applying Theorem 1.

Therefore, if p_0 falls in the interval $(\hat{p}_U^{\,*}(\tilde{t}), \hat{p}_L^{\,*}(\tilde{t}+1))$, it must also fall in the interval $(\hat{p}_U(\tilde{t}), \hat{p}_L(\tilde{t}+1))$ because $\hat{p}_U^{\,*}(\tilde{t}) \ge \hat{p}_U(\tilde{t})$ and $\hat{p}_L^{\,*}(\tilde{t}+1) \le p_L(t)$ as shown above. Consequently, the probability that $\hat{p} \ge p_0$ should be equal to the probability that $T \ge \tilde{t}+1$ which can be calculated exactly using the methods discussed in Gao et. al. (2009) to compute the distribution of *T*.

End of proof.

The relevance of the result in Corollary 1 is as follows: If p_0 lies in the shorter interval $(\hat{p}_U^{\,*}(\tilde{t}), \hat{p}_L^{\,*}(\tilde{t}+1))$, it must also fall in the wider interval $(\hat{p}_U(\tilde{t}), \hat{p}_L(\tilde{t}+1))$ that contains this smaller interval. Under this assumption, it will be easier to compute $\hat{p}_U^{\,*}(\tilde{t})$ and $\hat{p}_L^{\,*}(\tilde{t}+1)$ than $\hat{p}_U(\tilde{t})$ and $\hat{p}_L(\tilde{t}+1)$. On the other hand, if p_0 does not lie in the interval, p_0 may still be contained in $(\hat{p}_U(\tilde{t}), \hat{p}_L(\tilde{t}+1))$, and the results of Theorem 1 still applies.

To illustrate the above results, we simulate 100,000 data from a distribution with $p_o = 0.0005$, number of pools *m*=400, and pool sizes uniformly distributed in the range [25, 50]. We then calculated the one-sided LR test statistics using equation (1.0). Figure

1a displays the empirical distribution function of $-2\ln(\lambda)$. It is evident in this graph that $-2\ln(\lambda)$ has a discrete distribution. Each "jump" in this graph corresponds to a cluster of values of $-2\ln(\lambda)$ associated with a specific value of *T*. Consider 2 cases: when *T*=2 and when *T*=3. The values of the statistic associated with these values of *T* are the circled areas in Figure 1a. The flat area between these two jumps indicates there is no overlap with these two clusters. Figures 1b and 1c show in more detail the values of the LR test statistics. We observe that each cluster has a series of values that fall into interval $[5.7722, 5.7906]$ when *T*=2 and $[3.5356, 3.5577]$ when *T*=3 in the simulated data. Theoretically, suppose all *m* pools have different pool sizes and calculated $\hat{p}$ is less than p_0 given a specific observed $\tilde{t}$ value, then the maximum number of different one-sided $-2\ln(\lambda)$ in the cluster corresponding to $\tilde{t}$ is $\binom{m}{\tilde{t}}$. From above theorem, these $\binom{m}{\tilde{t}}$ values are bounded. When the pool sizes are equal, the probability of any $\tilde{t}$ pools out of *m* being positive is a constant, therefore, all these $\binom{m}{\tilde{t}}$ values will collapse into one single $-2\ln(\lambda)$ point. Unlike what we currently have in Figure 1b and Figure 1c, there is only one histogram bar in equal sized pool screening.

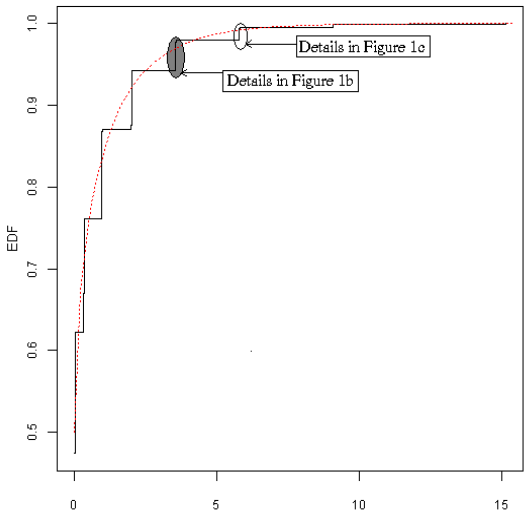

Figure 1a: EDF for -2ln(λ) under null p=0.0005,number of pools=400, pool sizes uniformly in range[25,50]

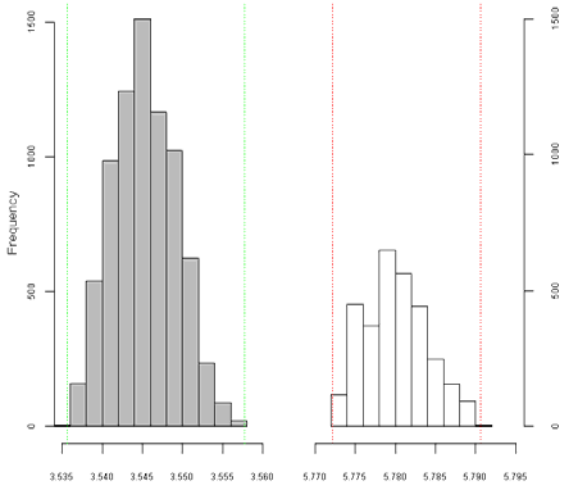

Figure 1b: -2ln(λ) when T=3 Figure 1c: -2ln(λ) when T=2

Next we illustrate Corollary 1. Table 1 displays the results of computing the exact weights for varying number of pools: 100, at 200 and at 400. The exact weight is the probability of number of positive pools, *T*, greater than or equal to 1, 3, and 7 for the number of pools equal to 100, 200 and 400, respectively. We also obtained simulated weights from 100,000 simulated data. One can see that the exact and simulated weights are very close to each other. Note that exact weight depends on the set of pool sizes in a sample.

Table 1 Exact weight c_1 in the mixture distribution under the null $p_0 = 0.0005$: pool sizes uniformly distributed in the range [25, 50]. Simulated weights are based on 100,000 simulation trials.

Number of Pools (m)	Total Subjects (N)	$\tilde{t}$	$\hat{p}_U{}^*(\tilde{t})$	$\hat{p}_L{}^*(\tilde{t}+1)$	Exact Weight	Simulated Weight
100	3822	1	0.000263	0.000527	0.5667	0.5680
200	7563	3	0.000401	0.000532	0.5168	0.5188
400	15217	7	0.000465	0.000529	0.4748	0.4741

3.2 Large Sample Case

Suppose **X** is a vector of random sample from a population with distribution $f(X \mid \theta)$. Let $L(\theta, x) = \prod_{i=1}^{n} f(x_i \mid \theta)$ be the likelihood function and $\hat{\theta}_n$ denote the MLE of θ. Assume that $\hat{\theta}_n$ exists and has a limiting normal distribution $N(\theta, I^{-1}(\theta))$, where $I(\theta) = E_\theta\left[\left(\frac{\partial}{\partial\theta}\log L\right)^2\right]$ is the Fisher Information matrix. Chernoff (1954) first showed that, for testing $H_0: \theta = \theta_0$ vs. $H_A: \theta < \theta_0$, the test statistic $-2\ln(\lambda(X))$ has cumulative distribution function defined in equation (2.0) satisfying

$$F_{\theta_0}(-2\ln(X) < C) \xrightarrow{L} 1/2 + F(\chi^2 < C, df = 1)/2 \qquad (5.0)$$

under the null as $n \to \infty$. [Note: Other researchers such as Kudô (1963), Shapiro (1985, 1988), and Silvapulle and Sen (2005) extended chi-square mixture results to the multiparameter case.] If a Chi-squared distribution with zero degree of freedom (χ_0^2) will be used to denote the constant $1/2$, then the cumulative distribution function of $-2\ln(\lambda(X))$ is a mixture distribution of two chi-square distributions, χ_0^2 and χ_1^2, under the null. Given an observed value of $-2\ln(\lambda(X))$, say C^*, the asymptotic p-value is equal to p=(1/2)*P(χ_1^2 >C*).

The above result due to Chernoff assumed independent and identically distributed random variables. In unequal sized pool screening, screening results are still assumed to be independent but no longer non-identically distributed because the probability of a pool being positive is an increasing function of pool size. However, using the results due to Bradley and Gart (1962), Chernoff's results may be extended to the case of non-identically distributed random variables. In their paper, Bradley and Gart defined a special case called "*associated population*" where observations come from different (sub) populations but have some parameters in common. In such cases, they proved that the MLE still possesses the property of being a consistent estimator and of having an asymptotically normal distribution. Unequal sized pool screening is one example of above defined situation since pools are independent but non-identically distributed, but they share the same parameter *p*. All the required regularity conditions for above asymptotic results to be true in pool screening with unequal pool sizes were verified by Gao et al., 2009. Therefore, the likelihood ratio test statistic for one-sided test in unequal pool screening still satisfies the asymptotic property as defined in (3).

Wald's and Score test are other likelihood-based procedures commonly used in practice. If we let m_i denote the number of pools having pool size n_i, $m = \sum_{i=1}^{W} m_i$ where W is the total number of distinct pool size and $\hat{p}$ denote the MLE of p, then for testing the hypothesis $\mathrm{H_o} : p = p_o$ versus $\mathrm{H_A} : p < p_o$, the asymptotic level α tests based on Wald's and Score test statistics are defined as follows:

Wald's test: Reject the null if $Z_w = \dfrac{\hat{p} - p_0}{\sqrt{I_{\hat{p}}^{-1}}} < Z(\alpha)$

Score Test: Reject the null if $Z_s = \frac{S(p_o)}{\sqrt{I_{p_o}}} < Z(\alpha)$,

where S(p) is given by

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

$$I = \left(\sum_{i=1}^{W} m_i \frac{{n_i}^2(1-p)^{n_i-2}}{1-(1-p)^{n_i}}\right) \quad \text{(Fisher information number)},$$

and $Z(\alpha)$ is the critical value from standard normal distribution with an α area to the right.

Remark: When all pools are negative, so that the MLE is 0, Wald's statistic cannot be computed because the Fisher information number is undefined at $\hat{p} = 0$. This is a major drawback in using Wald's test for rare events because the probability of observing no positive results is relatively high.

Asymptotic results in cases of rare events require a very large number of pools for the approximation to perform well. Consider the case when $p = 0.0005$ and pool sizes range from 25 to 50 inclusively. The probability of a pool testing positive is between 0.0124 and 0.0247. Using the Negative Binomial distribution, the expected total number of pools needed to find just one positive pool ranges from 41(1025 subjects when pool size is 25) to 81(4050 subjects when pool size is 50). Therefore, in order to find more than one positive pool, researchers need to have large number of subjects which may no longer be practical.

We assess how close the asymptotic distribution approximates the finite distribution of the LRT test statistic by conducting goodness of fit tests. In particular, we use the Kolmogorov-Smirnov test and quantile-quantile plots. We simulated 100,000 data with

p=0.0005 and computed the LR test statistic. The number of pools considered was 200 and 1000. The resulting quantile-quantile plots are displayed in Figure 2 and Figure 3. The values of the Komogorov-Smironov (KS) statistic decreased from 0.108 to 0.059 when the number of pools was increased from 200 to 1000 and the graphs show improvement in the fit for the case where m=1000 relative to m=200 but still unsatisfactory. The asymptotic p-values of KS test are all less than 0.0001. However, because we simulated 100,000 data values, the p-value associated with the KS test is expected to be very small regardless of the fit of the model, and hence will not be very meaningful. We note that the empirical distribution of the LR statistic is very discrete although we expect this distribution to approximate a continuous distribution as the number of pools increases. However, the convergence of the LR statistic to a continuous chi-square distribution is very slow in the particular case considered. Katholi and Aban (2009) showed that the r^{th} cumulant of the Score test statistics is $K_r = O\left[\left(\frac{1-p}{mp}\right)^{\frac{r}{2}-1}\right]$ for $r \geq 3$. In order for Score test statistics to converge to a $Normal(0,1)$, K_r must converge to zero. Given that p is near 0, m needs to be very large for this to happen. As a consequence of slow convergence, test procedures using standard critical values for rejection may be inaccurate and possibly misleading.

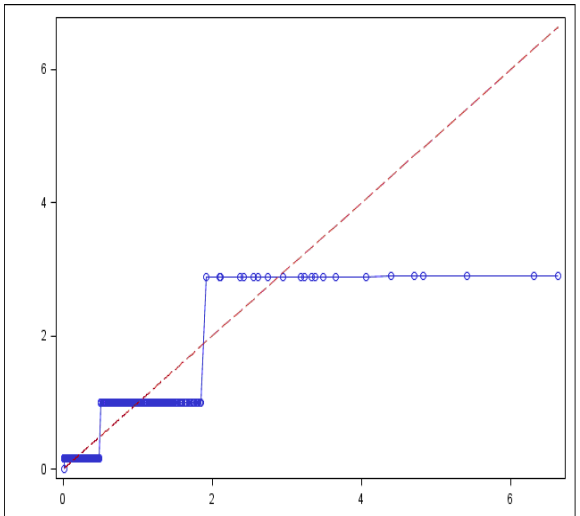

Figure 2.Chi-square mixture QQ plot. Number of pools=200, 100000 one-sided LR test statistics simulated under p=0.0005 with pool sizes in range [25,50].

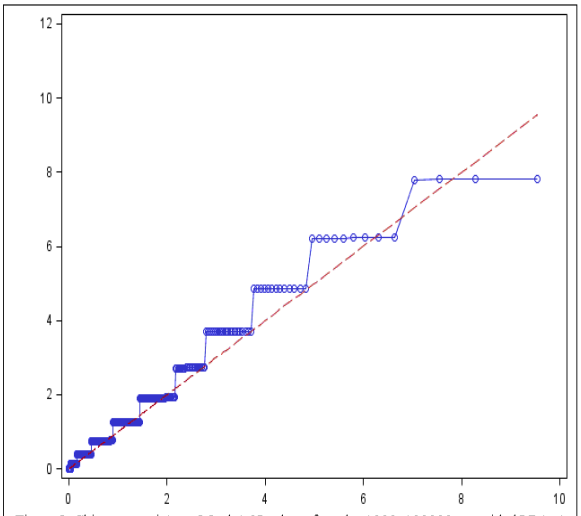

Figure 3. Chi-square mixture QQ plot. Number of pools=1000, 100000 one-sided LR test statistics simulated under p=0.0005 with pool sizes in range [25,50].

Because we cannot rely on the asymptotic distribution, we investigate two alternative methods to see if any of these methods will improve the performance of the LR test. The first method is to obtain simulated quantiles to determine the cut off for the rejection region. The second method is to incorporate the exact weight into the mixture Chi-square distribution instead of using 1/2. For instance, from Table 1, the exact weight is 0.4748 for this specific simulation when number of pools screened is 400. If we use this weight, the resulting mixture of Chi-square distributions is

$0.4748 + 0.5252 \times \chi^2(df = 1)$ under the null instead of $0.5 + 0.5 \times \chi^2(df = 1)$. Consequently, for $\alpha = 0.05$ test, the critical point for the rejection region is 2.784. Following the same argument, the critical point for the rejection region is 2.654 and 2.479 when number of pools screened is 200 and 100, respectively. These modified tests will be compared with the conventional LR test in the next section to see how their power functions differ.

4. Comparison of Power Functions

We compare the performance of the LR tests with other test procedures with respect to their power functions when testing $\mathrm{H_o}$: $p_o = 0.0005$ versus $\mathrm{H_A}$: $p_o < 0.0005$ through a series of Monte Carlo simulations. In particular, we would like to compare LR test with the Score test and the exact test based on the *T* statistics (number of positive pools). As noted in the previous section, the Wald's test is not applicable when one observes a sample with no positive pools, i.e., *T*=0. Because of this, we excluded the Wald's test in the comparison.

Due to the discreteness of *T*, we utilized an ancillary randomized test (see for instance, Lehmann and Romano, 2005) in order to conduct a size α hypothesis test. Let γ be a constant that satisfy $0 < \gamma < 1$. In a randomized test for one-sided, the critical value T_c and constant γ are determined such that the following equation is satisfied:

$$\alpha = \sum_{i=0}^{T_c - 1} \Pr(T = i \mid p_0) + \gamma \Pr(T = T_c \mid p_0) \qquad (6.0)$$

Given an alternative p_a , statistical power will be computed by

$$\beta = \sum_{i=0}^{T_c-1} \Pr(T = i \mid p_a) + \gamma \Pr(T = T_c \mid p_a) \qquad (7.0)$$

Below is a summary of the steps in the simulation. All calculations and simulations were performed using FORTRAN (Absoft Pro FORTRAN 10.1).

Step 1: Set the significance level at 5%.

Step 2: Given the number of pools, say *m*=100,200,400, generate pool sizes from a discrete uniform distribution over the range [25, 50].

Step 3: Compute the power for each test:

- Exact test based on *T*: Given 0.05 significance level, find γ and critical value T_c using equation (6.0). Given a value of *p* under the alternative and using the computed values of γ and T_c , calculate exact power associated with the test statistic *T* by equation (7.0).
- LRT: LR test based on asymptotic distribution: Generate 100,000 samples under the different values of the parameter and calculate LR test statistics under each sample. Given a value of *p* under the alternative, the simulated power is the percent of times out of 100,000 for which LR statistic is greater than the critical point C such that $P(\chi_1^2 \le C) = 0.90$ where χ_1^2 is the chi-square random variable with one degree of freedom.
- LRT(M): LR test based on asymptotic distribution but using exact weight as shown in Table 1: Generate 100,000 samples under the different values of the parameter and calculate LR test statistics under each sample. Given a value of *p* under the alternative, the simulated power is the percent of times out of 100,000 for which LR statistic is greater than the critical point C_M such that $P(\chi_1^2 \le C_M) = \frac{0.95 - c_1}{1 - c_1}$ where χ_1^2 is the chi-square random variable with one degree of freedom.
- LRT(Q): LR test where rejection region is based on simulated quantiles: Generate 100,000 samples under the null, calculate LR test statistic and find the 95th quantile. Given a value of *p* under the alternative, compute the simulated power by generating another set of 100,000 samples using p, and calculate LR test statistics under each sample. The simulated power is the percent of times out of 100,000 for which LR statistic is greater than the 95th quantile.

Figures 4, 5, and 6 display the simulated power for each test when m=100, 200, and 400, respectively. Simulated type I error rate for each test is the value of the power function associated with p=0.0005. As expected, the power function of a given test improves as m increases, i.e., the type I error rates get closer to the set 5% level and the power function is higher for a given value of p under the alternative. The exact test based on the number of positive pools performs the best among all tests with regard to the type I error. This is no surprise because there are no approximations involved in this test and the critical point was obtained to attain a 5% size test.

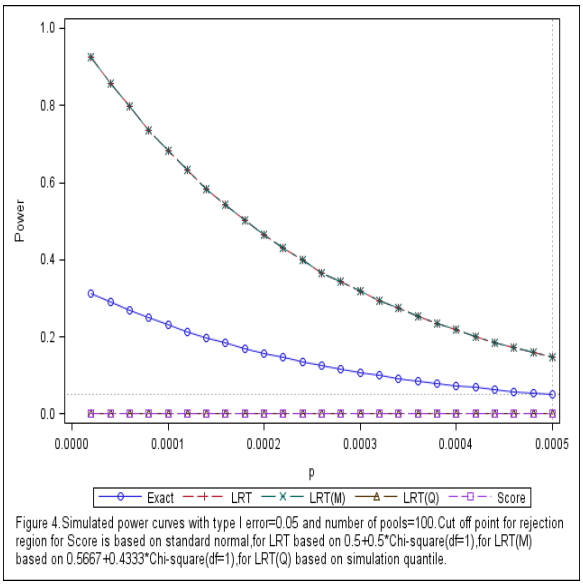

Figure 4.Simulated power curves with type I error=0.05 and number of pools=100.Cut off point for rejection region for Score is based on standard normal,for LRT based on 0.5+0.5*Chi-square(df=1),for LRT(M) based on 0.5667+0.4333*Chi-square(df=1),for LRT(Q) based on simulation quantile.

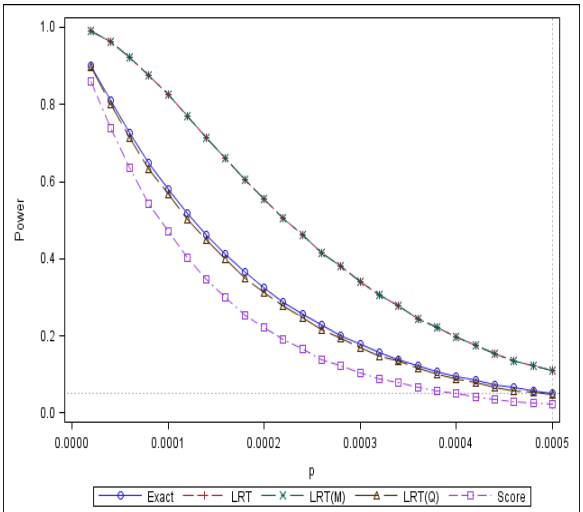

Figure 5.Simulated power curves with type I error=0.05 and number of pools=200.Cut off point for rejection region for Score is based on standard normal,for LRT based on 0.5+0.5*Chi-square(df=1),for LRT(M) based on 0.5168+0.4832*Chi-square(df=1),for LRT(Q) based on simulation quantile.

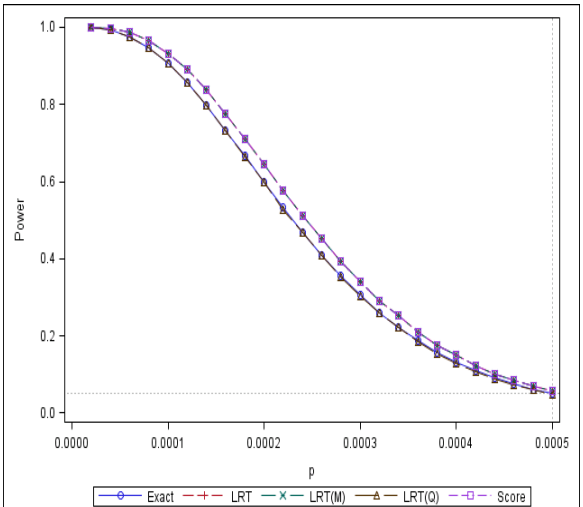

Figure 6.Simulated power curves with type I error=0.05 and number of pools=400.Cut off point for rejection region for Score is based on standard normal,for LRT based on 0.5+0.5*Chi-square(df=1),for LRT(M) based on 0.4748+0.5252*Chi-square(df=1),for LRT(Q) based on simulation quantile.

LRT(M) and LRT power values are not very different from each other which implies that replacing the weights based on asymptotic results by the exact weight did not significantly change the performance of the conventional LR test. Among all the tests, both of these tests have inflated type I error rates but improves as the number of pools increases which supports our claim that the finite sample distribution under the null hypothesis converges slowly to the asymptotic distribution. A consequence of inflated type I error is that the power function is artificially high under the alternative values of *p*.

The LRT(Q) and the Score tests perform very poorly when m=100 where their power functions are the zero function. This implies that these two tests never reject the null hypothesis regardless of the true value of *p*. The main reason for this behavior of LRT(Q) is that when m=100, the number of pools that are positive is near 0. Hence, the critical value corresponding to the 95^{th} quantile by equation (1.0) is the value of the test statistic associated with T=0, i.e.,

$$-2\ln(\lambda) = -2 \times 3822 \times \ln(1 - p_o) = 3.823$$

Similar argument can be made to explain the behavior of the Score function.

When m=200, power functions of LRT(Q) and Score tests significantly improved. In fact LRT(Q) power function is practically mimics the power function of the exact test. When number of pools screened increased to 400 (Figure 6), the type I error rates for all tests are at about the target 5% level. This implies that the asymptotic distributions under the null of the LR and Score tests are good approximations. Furthermore, the Score, LRT and LRT(M) exceed the power function of LRT(Q) and exact test.

5. Conclusions

One-sided LR test statistics in unequal sized pool screening has a mixture distribution. The exact weight may be computed but the form of the exact distribution is difficult to obtain. Large sample distribution of the LR statistic is known and often used in practice. However, for rare diseases where the prevalence is near 0, the asymptotic results require large sample sizes to be applicable. In the simulation studies performed, 100 pools are not sufficient when the prevalence rate is 0.05%. In this case, the distribution of the LR statistic is discrete and clustered around corresponding number of positive pools.

In cases where the exact weight can be calculated, one may modify the LR test based on the asymptotic distribution of the test statistic by replacing the asymptotic weights by exact weights. However, this did not improve the performance of the LR test with regard to power function for small samples. Another modification of the LR test is to obtain the critical value based on simulated quantiles. This method improved the performance of the LR test but this method fails when the sample size is not large enough to observe at least one positive pools. In our simulations, if the number of pools is greater than or equal to 200, this method works well but fails when the number of pools is 100. Exact test based on the number of positive pools is still the best choice for small samples. For moderately large samples (at least 400 in the case considered), LR test and Score test are recommended because they provide higher power values under the alternative than the exact test.

REFERENCES

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.

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

CHERNOFF, H., 1954, On the Distribution of the Likelihood Ratio: The Annals of Mathematical Statistics, v. 25, p. 573-578.

FISHER, R.A., 1922, On the Mathematical Foundations of Theoretical Statistics: Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, v. 222, p. 309-368.

GAO,H., ABAN, I., and KATHOLI,C.R., Two-sided Hypothesis Testing Based on Pool Screening with Unequal Pool Sizes: Biostatistics Department Technical Report, BST2009-002, Available at https://www.soph.uab.edu/bst/technical.

GAO,H., ABAN, I., and KATHOLI,C.R., 2009, Pool Screening: An Example of Independent Non-identical Bernoulli Trial, manuscript under review

GOODMAN DS, O.J., ROBERTS JM, LAMMIE PJ, STREIT TG., 2003, PCR and Mosquito dissection as tools to monitor filarial infection levels following mass treatment.: Filaria J., v. Jul 7;2(1): 11.

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.

HELMY H, F.P., FARID HA, BRADLEY MH, RAMZY RM., 2004, Test strip detection of Wuchereria bancrofti amplified DNA in wild-caught Culexpipiens and estimation of infection rate by a PoolScreen algorithm.: Trop Med Int Health., v. 9(1):158-63.

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., The Role of Polymerase Chain Reaction Techniques for Assessing Lymphatic Filariasis Transmission: Appendix 5 and Annexes A – E, Hypothesis Testing With the Pool Screening Model: World Health Organization, Report Number WHO/HTM/NTD/PCT/2009.1, 2009.

KATHOLI,C.R. and ABAN,I.,2009, Investigation of the Distribution of the Score Statistics for a Simple Hypothesis in Pool Screening, Biostatistics Department Technical Report, BST2009-003, Available at https://www.soph.uab.edu/bst/technical.

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.

KUDÔ, A., 1963, A Multivariate Analogue of the One-Sided Test: Biometrika, v. 50, p. 403-418.

LEHMANN, E.L., 1998, Elements of Large-Sample Theory: Springer. Third Printing(2004).

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

NEYMAN, J., and PEARSON, E.S., 1928, On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part I: Biometrika, v. 20A, p. 175-240.

RODRÍGUEZ-PÉREZ MA, L.B., DOMÍNGUEZ-VÁZQUEZ A, SEGURA-ARENAS R,, and LIZARAZO-ORTEGA C, M.-H.A., REYES-VILLANUEVA F, UNNASCH TR., 2004, Polymerase chain reaction monitoring of transmission of Onchocerca volvulus in two endemic states in Mexico.: Am J Trop Med Hyg., v. 70(1):38-45.

SHAPIRO, A., 1985, Asymptotic Distribution of Test Statistics in the Analysis of Moment Structures Under Inequality Constraints: Biometrika, v. 72, p. 133-144.

SHAPIRO, A., 1988, Towards a Unified Theory of Inequality Constrained Testing in Multivariate Analysis: International Statistical Review / Revue Internationale de Statistique, v. 56, p. 49-62.

SILVAPULLE, M.J., and SEN, P.K., 2005, Constrained Statistical Inference:Inequality, Order, and Shape Restrictions: John Wiley & Sons, Inc., Hoboken, New Jersey.

WILKS, S.S., 1938, The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses: The Annals of Mathematical Statistics, v. 9, p. 60-62.

WILLIAMS SA, L.S., BIERWERT LA, SAUNDERS LJ, BOAKYE DA, FISCHER P, GOODMAN D, HELMY H, HOTI SL, VASUKI V, LAMMIE PJ, PLICHART C, RAMZY RM, OTTESEN EA., 2002, Development and standardization of a rapid, PCR-based method for the detection of Wuchereria bancrofti in mosquitoes, for xenomonitoring the human prevalence of bancroftian filariasis.: Ann Trop Med Parasitol., v. 96, p. Suppl 2:S41-6

YAMÊOGO, 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.

SUMMARY, CONCLUSIONS AND FUTURE RESEARCH

1. Summary and Conclusions

We first review the simulation settings and model assumptions considered in this dissertation which defines the limitations of our results. We assume that pool sizes are known constants, and the maximum possible pool size is bounded to reflect the PCR technique requirement corresponding to the lowest and highest sample size that a PCR machine can handle to have the best readings. We also assume that the laboratory testing procedure has perfect sensitivity and specificity. The prevalence, p, in the simulations is set to be equal to 5 out of 10000 to address the rareness of the event of interest. This number might be too high for some studies or too low for other studies. However, in general, pool screening may not longer be recommended when p is relatively high, for example, near or greater than 0.1, because pool screening will result in serious concerns about the bias of the MLE. Furthermore, if the pool size is greatly decreased when p is high, cost-efficiency which is the essence of pool screening is lost as well. Lastly, we should recognize the difficulties to collect large number of positive subjects due to the extremely low p. At least one pool should be positive in pool screening for estimation problem, otherwise, MLE estimate will be 0 and Wald's test will not be defined.

Next we summarize our general conclusions of this dissertation research. 1) Exact test has close performance with all the other hypothesis testing procedures. When the number of pools screened is small (100) and p is extremely low (0.0005), exact test outperforms all the other test procedures. 2) For asymptotic test procedures such as Wald, Score and LRT, the modified tests based on quantiles should be preferred over the standard procedures when number of pools screened is moderate(200). Please note in order to

apply Wald's test, at least one positive pool should be observed. These conclusions are mainly attributed to the unsatisfactory speed of convergence of the standard asymptotic tests under this setting. 3) FORTRAN codes (provided in Appendix A-G) were developed as part of this dissertation to implement the proposed tests. To conclude, we recommend the use of either the exact test based on the number of positive pools or the modified likelihood based tests using simulated quantiles.

2. Future Research

In pool screening applications, experiment setting will never be as perfect as simulations. Additional statistical complexities due to different designs, diseases, individual subjects, and pooling processes need to be considered. Therefore, there are still many unanswered statistical questions in this field. We will discuss future research interests and divide this section into two parts. In the first part, we introduce pool size and number of pools combination issues in pool screening. We have a separate subsection for this because we present some preliminary results that we have obtained. The second part contains some ideas and thoughts on other open problems.

2.1 Combination of Number of Pools and Pool Sizes

Most often, total number of subjects (denoted by N) collected from fields will be brought back to a laboratory and then divided into a number of pools for screening test. The question would be given N fixed, how to choose proper pool size to attain the maximum statistical power? A simple simulation has been done as follows. Suppose there are 2000 (N=2000) subjects collected, and the pool sizes (denoted by K) are in the

range 10 to 100 corresponding to usual laboratory screening test can handle . We are interested in the following combinations of number of pools and pool size : 200 pools with pool size 10; 100 pools with pool size 20; 50 pools with pool sizes 40; 25 pools with pool sizes 80; 20 pools with pool sizes 100. We also want to test one-side hypothesis in the form of $H_o : p = p_o$ *versus* $H_a : p < p_o$. In order to see the trend under the different null values, we considered $p_o = 0.0005$, 0.001, 0.003, 0.006, 0.009, 0.012. Given equal pool sizes, the number of positive pools has distribution of $Binomial(N/K, 1-(1-p)^K)$, and test based on this test statistics is an UMP test.

Preliminary results show that maximum statistical power is attainable at pool size=1 and number of pools=2000. Otherwise, the statistical power slightly decreases as pool size increases. Another interesting fact is that when p_o is very low, the power curves of different combinations are almost non-distinguishable with each other.

Our conclusion is that when *p* is extremely low (for instance, $p_o = 0.001$), statistical power is mainly determined by the total number of subjects N. In other words, for a fixed and small p_0 and fixed N, the power functions associated with the different pool size combinations do not differ much. As p_o increases, the differences in the simulated power become more obvious: the lower the pool sizes, the higher the statistical power. However, decreasing pool sizes will result in an increase in number of screening tests and the cost efficiency of pool screening will be gradually lost. Based on these preliminary results, we recommend that pool size be determined by financial resources available and laboratory requirements when for p_o <0.001 in order to maximize the power. Future research needs

to focus on how to define a utility function that will help provide balance between statistical power and research budget in determining the pool size.

2.2 Other Future Research

There is one critical assumption we have been making in this research, i.e., pool sizes are known. In many situations, pool sizes could be random. How to statistically address this problem will be very interesting. Suppose we want to use Bayesian's method, which prior do we want to pick for pool sizes? And why pick this prior instead of others? How about p, if we think it is random and having Beta distribution, then what kind of Beta distribution do we need to choose to reflect the rareness of the dichotomous event?

Another assumption we have been using is perfect test assumption, in reality, sensitivity and specificity are never perfect. Their values may vary with the sample sizes handled by the PCR machine. Furthermore, different PCR machine has different sensitivity and specificity because manufacturers are not same.

Finally, it may be of interest to consider the case where more than one vector species can carry the same disease. These species are caught and pooled separately. The question of interest is: How do we compare the prevalence across different vectors? One may also be interested in the problem of analyzing data from pool screening where the same vector carries two or more diseases. In this case, the screening results follow a multinomial model. How do you perform a statistical test of hypothesis about the prevalence of each diseases considering the possibility of correlated outcomes?

GENERAL LIST OF 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.

BURROWS, P., 1987, Improved estimation of pathogen transmission rates by group testing: Phytopathology, v. 77, p. 363-365.

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.

CAHOON-YOUNG, B., CHANDLER, A., LIVERMORE, T., GAUDINO, J., and BENJAMIN, R., 1989, Sensitivity and specificity of pooled versus individual sera in a human immunodeficiency virus antibody prevalence study: J. Clin. Microbiol., v. 27, p. 1893-1895.

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.

CHERNOFF, H., 1954, On the Distribution of the Likelihood Ratio: The Annals of Mathematical Statistics, v. 25, p. 573-578.

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.

DURNEZ L, E.M., MGODE GF, KATAKWEBA A, KATHOLI CR, MACHANG'U RR, KAZWALA RR,, and PORTAELS F, L.H., 2008, First detection of mycobacteria in African rodents and insectivores, using stratified pool screening.: Appl Environ Microbiol, v. 74, p. 768-773.

EMMANUEL, J.C., BASSETT, M.T., SMITH, H.J., and JACOBS, J.A., 1988, Pooling of sera for human immunodeficiency virus (HIV) testing: an economical method for use in developing countries: J Clin Pathol, v. 41, p. 582-585, doi: 10.1136/jcp.41.5.582.

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

FISHER, R.A., 1922, On the Mathematical Foundations of Theoretical Statistics: Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, v. 222, p. 309-368.

GAO,H., ABAN,I., and KATHOLI,C.R.,2009, "Two-sided Hypothesis Testing Based on Pool Screening with Unequal Pool Sizes", Biostatistics Department Technical Report, BST2009-002, Available at https://www.soph.uab.edu/bst/technical.

GAO,H., ABAN, I., and KATHOLI,C.R., 2009, Pool Screening: An Example of Independent Non-identical Bernoulli Trial, manuscript under review

GOODMAN DS, O.J., ROBERTS JM, LAMMIE PJ, STREIT TG., 2003, PCR and Mosquito dissection as tools to monitor filarial infection levels following mass treatment.: Filaria J., v. Jul 7;2(1): 11.

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.

GU, W., UNNASCH, T.R., KATHOLI, C.R., LAMPMAN, R., and NOVAK, R.J., 2008, Fundamental issues in mosquito surveillance for arboviral transmission: Transactions of the Royal Society of Tropical Medicine and Hygiene, v. 102, p. 817-822.

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.

HELMY H, F.P., FARID HA, BRADLEY MH, RAMZY RM., 2004, Test strip detection of Wuchereria bancrofti amplified DNA in wild-caught Culex pipiens and estimation of infection rate by a PoolScreen algorithm.: Trop Med Int Health., v. 9(1):158-63.

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.

HUNG, M., and SWALLOW, W.H., 1999, Robustness of Group Testing in the Estimation of Proportions: Biometrics, v. 55, p. 231-237.

Katholi, C. R., "The Role of Polymerase Chain Reaction Techniques for Assessing Lymphatic Filariasis Transmission", Appendix 5 and Annexes A – E, "Hypothesis Testing With the Pool Screening Model", World Health Organization, Report Number WHO/HTM/NTD/PCT/2009.1, 2009.

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.

KATHOLI,C.R. and ABAN,I.,2009, Investigation of the Distribution of the Score Statistics for a Simple Hypothesis in Pool Screening, Biostatistics Department Technical Report, BST2009-003, Available at https://www.soph.uab.edu/bst/technical.

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.

KUDÔ, A., 1963, A Multivariate Analogue of the One-Sided Test: Biometrika, v. 50, p. 403-418.

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.

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

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

NEYMAN, J., and PEARSON, E.S., 1928, On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part I: Biometrika, v. 20A, p. 175-240.

RODRÍGUEZ-PÉREZ MA, L.B., DOMÍNGUEZ-VÁZQUEZ A, SEGURA-ARENAS R,, and LIZARAZO-ORTEGA C, M.-H.A., REYES-VILLANUEVA F, UNNASCH TR., 2004, Polymerase chain reaction monitoring of transmission of Onchocerca volvulus in two endemic states in Mexico.: Am J Trop Med Hyg., v. 70(1):38-45.

SAMUELS, S.M., 1978, The Exact Solution to the Two-Stage Group-Testing Problem: Technometrics, v. 20, p. 497-500.

SHAPIRO, A., 1985, Asymptotic Distribution of Test Statistics in the Analysis of Moment Structures Under Inequality Constraints: Biometrika, v. 72, p. 133-144.

SHAPIRO, A., 1988, Towards a Unified Theory of Inequality Constrained Testing in Multivariate Analysis: International Statistical Review / Revue Internationale de Statistique, v. 56, p. 49-62.

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

SILVAPULLE, M.J., and SEN, P.K., 2005, Constrained Statistical Inference:Inequality, Order, and Shape Restrictions: John Wiley & Sons, Inc., Hoboken, New Jersey.

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.

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

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.

TURNER, D.W., TIDMORE, F.E., and YOUNG, D.M., 1988, A Calculus Based Approach to the Blood Testing Problem: SIAM Review, v. 30, p. 119-122.

VASUKI V, H.S., SADANANDANE C, JAMBULINGAM P., 2003, A simple and rapid DNA extraction method for the detection of Wuchereria bancrofti infection in the vector mosquito, Culex quinquefasciatus by Ssp I PCR assay.: Acta Trop., v. 86(1):109-14.

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.

WILKS, S.S., 1938, The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses: The Annals of Mathematical Statistics, v. 9, p. 60-62.

WILLIAMS SA, L.S., BIERWERT LA, SAUNDERS LJ, BOAKYE DA, FISCHER P, GOODMAN D, HELMY H, HOTI SL, VASUKI V, LAMMIE PJ, PLICHART C, RAMZY RM, OTTESEN EA., 2002, Development and standardization of a rapid, PCR-based method for the detection of Wuchereria bancrofti in mosquitoes, for xenomonitoring the human prevalence of bancroftian filariasis.: Ann Trop Med Parasitol., v. 96, p. Suppl 2:S41-6

XIE, M., TATSUOKA, K., SACKS, J., and YOUNG, S.S., 2001, Group Testing with Blockers and Synergism: Journal of the American Statistical Association, v. 96, p. 92-102.

YAMÈOGO L, T.L., HOUGARD JM, BOATIN BA, UNNASCH TR., 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(1):124-8.

ZHU, L., HUGHES-OLIVER, J.M., and YOUNG, S.S., 2001, Statistical Decoding of Potent Pools Based on Chemical Structure: Biometrics, v. 57, p. 922-930.

APPENDIX A MARCUS AND LOPES SUBROUTINE

```
!Last change: Hongjiang Gao, Dec 2008

!-----------------------------------------------------------------------------------------------------
! This routine calculates the distribution of the number of
! successes in m independent trial with changing probability of
! success on each trial using Marcus and Lopes (Marcus and Lopes, 1957) double
! recursions.
!-----------------------------------------------------------------------------------------------------
! The arguments of the routine have the following meanings:
!
!k(integer)=total number of pools
!p(real, double precision)=probability of individual subject being positive
!PoolSize(array with dimension k, double precision )=Pool Size
!probability(array with dimension k+1,double precision)=output PMF of the distribution
!
!Author/Implementor: Hongjiang Gao
!
!Latest revision: Dec 2008
!
!-----------------------------------------------------------------------------------------------------
!
!-----------------------------------------------------------------------------------------------------
!                              C A V I A T R E C E P T O R
!
! This is a research program and is not warranted by the authors
! to be free from "bugs". Having a copy of this program should in no
! way be considered as a warranty on the part of the authors to
! provide technical support or changes to the program. Any changes
! you might make to the program are your responsibility. In short,
! we make no claims with respect to the quality or correctness of
! any results you might obtain using this program.
!-----------------------------------------------------------------------------------------------------
!
!-----------------------------------------------------------------------------------------------------
subroutine Marcus_pmf(k,p,PSZ,probability)
implicit none
integer(4):: i,j,m,k
real(8),dimension(k)::a,b,PSZ
real(8),dimension(k+1)::probability
real(8),allocatable::s(:),Sminus_a(:,:), prob(:)
real(8)::LT,sum1,sum2,cdf,v,dlgomx,sumpsz,ptrial,p
```

```
allocate(s(k+1),Sminus_a(k+1,k),prob(k+1))

do i=1,k
        a(i)=(1.0d0-(1.0d0-p)**psz(i))/((1.0d0-p)**psz(i))
        Sminus_a(0,i)=1.0d0
end do
LT=(1.0D0-P)**SUM(PSZ)
prob(0)=LT
S(0)=1.0d0                              ! Initiate S(0)
S(1)=sum(a)                             ! Initiate S(1)
Prob(1)=S(1)*LT
do i=1, k
        Sminus_a(1,i)=S(1)-a(i)
end do

do i=2,k
        sum1=0.0d0
        sum2=0.0d0
        do j=1,k
                sum1=sum1+a(j)*s(i-1)
                sum2=sum2+(a(j)**2.0D0)*Sminus_a(i-2,j)
        end do
        S(i)=(sum1-sum2)/real(i,8)
        prob(i)=LT*S(i)
        do j=1,k
                Sminus_a(i,j)=S(i)-a(j)*Sminus_a(i-1,j)
        end do
end do
do i=0,k
        probability(i)=prob(i)
end do
return
end subroutine Marcus_pmf
```

APPENDIX B NEWTONS IDENTITY SUBROUTINE

```
!-----------------------------------------------------------------------------------------------------
! This routine calculates the distribution of the number of
! successes in m independent trial with changing probability of
! success on each trial using Newton's Identities (Mead, 1992)
!-----------------------------------------------------------------------------------------------------
! The arguments of the routine have the following meanings:
!
!k(integer)=total number of pools
!p(real, double precision)=probability of individual subject being positive
!PSZ(array with dimension k, double precision )=Pool Size
!pmf(array with dimension k+1,double precision)=output PMF of the distribution
!
!Author/Implementor: Hongjiang Gao
!
!Latest revision: Dec 2008
!
!-----------------------------------------------------------------------------------------------------
!
!-----------------------------------------------------------------------------------------------------
!                              C A V I A T R E C E P T O R
!
! This is a research program and is not warranted by the authors
! to be free from "bugs". Having a copy of this program should in no
! way be considered as a warranty on the part of the authors to
! provide technical support or changes to the program. Any changes
! you might make to the program are your responsibility. In short,
! we make no claims with respect to the quality or correctness of
! any results you might obtain using this program.
!-----------------------------------------------------------------------------------------------------
!
!-----------------------------------------------------------------------------------------------------

Subroutine Newtonpmf(k,p,psz,pmf)
implicit none
integer(4):: k,i,j
real(8),dimension(k):: psz,a,prob,esum2
real(8),dimension(k+1):: pmf,esum1
real(8)::p,sum,lt,esum,probsum
lt=(1.0d0-p)**sum(psz)


do j=1,k
```

```
        Probsum=0.0d0
        do i=1,k
                Probsum=probsum+a(i)**real(j,8)
        end do
        prob(j)=probsum
end do

esum1(0)=1.0d0
do i=1,k
        do j=1,I
                if((mod(j-1,2)==0)) then
                        esum=esum+esum1(i-j)*prob(j)
                else
                        esum=esum-esum1(i-j)*prob(j)
                endif
        end do
        esum(i)=esum/real(I,8)
end do

do i=0,k
        pmf(i)=lt*esum1(i)
end do

return

end subroutine Newtonpmf
```

APPENDIX C FUNCTION TO CALCULATE LIKELIHOOD

```
!------------------------------------------------------------------------------------------------
!The purpose of this function is to calculate likelihood
!------------------------------------------------------------------------------------------------
!The arguments of the function have the following meanings
!K=How many pools
!A=Pool size, one-dimensional array with size k
!B=Test result, screening test result of each pool, 0.0 or 1.0, one-dimensional array with
! size k
!P=probability of an individual to be positive
!
!------------------------------------------------------------------------------------------------
!Author/implementer: Hongjiang Gao
!Date last update: Aug 13,2008
!------------------------------------------------------------------------------------------------
!                          C A V I A T R E C E P T O R
!
! This is a research program and is not warranted by the authors
! to be free from "bugs". Having a copy of this program should in no
! way be considered as a warranty on the part of the authors to
! provide technical support or changes to the program. Any changes
! you might make to the program are your responsibility. In short,
! we make no claims with respect to the quality or correctness of
! any results you might obtain using this program.
!------------------------------------------------------------------------------------------------
!
!------------------------------------------------------------------------------------------------

REAL(8) FUNCTION LL(K,A,B,P)
INTEGER(4),INTENT(IN)::K
INTEGER(4)::I
REAL(8),INTENT(IN)::A(K),B(K),P
REAL(8)::LT,PROD
LT=(1.0D0-P)**SUM(A)
!IF(LT .LE. EPSILON(LT)) THEN
! WRITE(*,*)'WARNING! WARNING!!'
! WRITE(*,*)'LEADING TERM IS LESS THAN MINIMUM DOUBLE
VALUE=(0.5)^52 '
! WRITE(*,*)'PLEASE USE OTHER PROGRAM TO CACULATE THE PDF AND
CDF'
! STOP
!ELSE
PROD=1.0D0
```

```
DO I=1,K
PROD=PROD*((1.0D0-(1.0D0-P)**A(I))/(1.0D0-P)**A(I))**B(I)
END DO
!END IF
LL=LT*PROD
RETURN
END FUNCTION LL
```

APPENDIX D FUNCTION TO CALCULATE EXPECTED FISHER INFORMATION

```
!-----------------------------------------------------------------------------------------------
! The purpose of this function is to calculate the EXPECTED Fisher Information
!-----------------------------------------------------------------------------------------------
!The argument of the function have the following measnings
! A=Total Number of Pools
! B(A)=The Vector of Pool Size
! C= p_null or p_hat
!
! Author/Implementer: Hong J Gao
! Last Revision: Oct 3, 2008
!
!-----------------------------------------------------------------------------------------------
!
! C A V I A T R E C E P T O R
! This is a research program and is not warranted by the authors
! to be free from "bugs". Having a copy of this program should in no
! way be considered as a warranty on the part of the authors to
! provide technical support or changes to the program. Any changes
! you might make to the program are your responsibility. In short,
! we make no claims with respect to the quality or correctness of
! any results you might obtain using this program.
!-----------------------------------------------------------------------------------------------
!
!-----------------------------------------------------------------------------------------------
REAL(8) FUNCTION VAR(A,B,C)
INTEGER(4),INTENT(IN)::A
REAL(8),INTENT(IN)::B(A),C
INTEGER(4)::I,J
REAL(8)::VARSUM
VARSUM=0.0D0
DO I=1,A
VARSUM=VARSUM+(B(I)**2.0D0*(1.0D0-C)**(B(I)-2.0D0))/(1.0D0-(1.0D0-
C)**B(I))
END DO
VAR=1.0D0/(VARSUM/A)
END FUNCTION
```

APPENDIX E FUNCTION TO CALCULATE VARIANCE OF SCORE STATISTICS

```
!-----------------------------------------------------------------------------------------------
! The purpose of this function is to calculate the second derivative of log likelihood
! under the null (variance for Score)
!-----------------------------------------------------------------------------------------------
!The argument of the function have the following meanings
! A=Total Number of Pools
! B(A)=The Vector of Pool Size
! C(A)=Pool Test Result
! C= p_null
! Author/Implementor: Hong J Gao
!
!Last Revision: Oct 15, 2008
!
!-----------------------------------------------------------------------------------------------
!
! C A V I A T R E C E P T O R
! This is a research program and is not warranted by the authors
! to be free from "bugs". Having a copy of this program should in no
! way be considered as a warranty on the part of the authors to
! provide technical support or changes to the program. Any changes
! you might make to the program are your responsibility. In short,
! we make no claims with respect to the quality or correctness of
! any results you might obtain using this program.
!-----------------------------------------------------------------------------------------------
!
!-----------------------------------------------------------------------------------------------

REAL(8) FUNCTION SCORE_Ftn(A,B,C,D)
IMPLICIT NONE
INTEGER(4),INTENT(IN)::A
REAL(8),INTENT(IN)::B(A),C(A),D
INTEGER(4)::I,J
REAL(8)::SCORESUM
SCORESUM=0.0D0
DO I=1,A
SCORESUM=SCORESUM+C(I)*B(I)*(1.0D0-D)**(B(I)-1.0D0)/(1.0D0-(1.0D0-
D)**B(I))-B(I)*(1.0D0-C(I))/(1.0D0-D)
END DO
SCORE_Ftn=SCORESUM
END FUNCTION
```

APPENDIX F SUBROUTINE TO FIND GIVEN QUANTILE OF ONE-DIMENSION ARRAY

```
!-----------------------------------------------------------------------------------------------
! The purpose of this subroutine is to find the quantile of one-dimensional array
!-----------------------------------------------------------------------------------------------
!
!-----------------------------------------------------------------------------------------------
!The argument of the function have the following meanings
!N=Size of the array(Input)
! P=P^th Quantile(Input)
! X=X array (Input)
! Q=Output
!-----------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------------------
!
! C A V I A T R E C E P T O R
! This is a research program and is not warranted by the authors
! to be free from "bugs". Having a copy of this program should in no
! way be considered as a warranty on the part of the authors to
! provide technical support or changes to the program. Any changes
! you might make to the program are your responsibility. In short,
! we make no claims with respect to the quality or correctness of
! any results you might obtain using this program.
!-----------------------------------------------------------------------------------------------
!
!-----------------------------------------------------------------------------------------------

SUBROUTINE QUANTILE(NOBS,P,X,Q)
USE EQTIL_int
INTEGER(4),PARAMETER:: NQPROP=1
INTEGER(4)::NOBS,NMISS
REAL(8):: P,X(NOBS),Q(1),XEMP(NQPROP),
XHI(NQPROP),XLO(NQPROP),QPROP(NQPROP)
QPROP=P
CALL D_EQTIL (X, NQPROP, QPROP, XEMP, XLO, XHI, NMISS)
Q=XEMP
RETURN
END SUBROUTINE
```

APPENDIX G PROGRAM TO CALCULATE EXACT AND SIMULATED POWERS FOR ONE-SIDED HYPOTHESIS TEST

```
!-----------------------------------------------------------------------------------------------
! The purpose of this program is to calculate exact and simulated powers of different
! testing methods for one-sided hypothesis test(Details in Chapter 4 )
!-----------------------------------------------------------------------------------------------
!
!-----------------------------------------------------------------------------------------------
!The argument of the function have the following meanings
!N=Size of the array(Input)
! P=P^th Quantile(Input)
! X=X array (Input)
! Q=Output
!-----------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------------------
!
! C A V I A T R E C E P T O R
! This is a research program and is not warranted by the authors
! to be free from "bugs". Having a copy of this program should in no
! way be considered as a warranty on the part of the authors to
! provide technical support or changes to the program. Any changes
! you might make to the program are your responsibility. In short,
! we make no claims with respect to the quality or correctness of
! any results you might obtain using this program.
!-----------------------------------------------------------------------------------------------
!
!-----------------------------------------------------------------------------------------------


PROGRAM MAIN
IMPLICIT NONE
INTEGER(4)::I,J,CRITICAL_L,CRITICAL_R,K,SIMU1,SIMU2,LR_COUNT,LR_CO
UNT_M,LR_COUNT_C,WALD_COUNT,WALD_COUNT_C,P_COUNT
INTEGER(4)::SCORE_NEW_COUNT,SCORE_NEW_COUNT_C
INTEGER(4),ALLOCATABLE::B(:)
REAL(8)::Q,P,CDF,V,POWER_L,POWER_R,POWER,POWER1,POWER2,TYPE1_L,
TYPE1_R,SIG,LBD,UBD,GAMMA_L,GAMMA_R,LL
REAL(8)::P_NULL,LR_POWER,LR_POWER_M,LR_POWER_C,WALD_POWER,W
ALD_POWER_C,SCORE_NEW_POWER,SCORE_NEW_POWER_C,ML_EST,&
NUMRATR,DENOM,LAMDA
REAL(8)::LR_L,LR_R,SCORE_R,WALD_R,VAR,SCORE_FTN,WALD,SCORE_NE
W
REAL(8),ALLOCATABLE::C(:),PROBABILITY(:),PSZ(:),TMP(:),TMP2(:),RSLT(:),L
AMDA_C(:),LRT(:),WALD_C(:),SCORE_NEW_C(:)
```

```
EXTERNAL::DRNUN,RNSET
WRITE(*,*) 'HOW MANY SIMULATIONS FOR ASYMPTOTIC TESTS CRITICAL
VALUES(INTEGER)?'
READ(*,*) SIMU1
WRITE(1,*)'SIMU1=',SIMU1
WRITE(*,*) 'HOW MANY SIMULATIONS FOR SIMULATED POWER(INTEGER)?'
READ(*,*) SIMU2
WRITE(1,*)'SIMU2=',SIMU2
WRITE(*,*) 'PREVERLENCE?(DOUBLE)'
READ(*,*) Q
WRITE(1,*) 'PREVERLENCE=',Q
WRITE(*,*) 'SIGNIIFICANCE LEVEL(DOUBLE)'
READ(*,*) SIG
WRITE(1,*) 'SIGNIFICANCE(TYPE I ERROR)=',SIG
! WRITE(*,*) 'HOW MANY POOLS?(INTEGER)'
! READ(*,*) K
WRITE(*,*) 'LOWER BOUD OF POOL SIZE?(DOUBLE)'
READ(*,*) LBD
WRITE(1,*) 'LOWER BOUND=',LBD
WRITE(*,*) 'UPPER BOUD OF POOL SIZE?(DOUBLE)'
READ(*,*) UBD
WRITE(1,*) 'UPPER BOUND=',UBD
WRITE(4,*) "P ","K ","Exact_POWER ","LR_POWER_M ","LR_POWER
","LR_POWER_C ","WALD_POWER ","WALD_POWER_C ",&
"SCORE_NEW_POWER ", "SCORE_NEW_POWER_C "
do k=100,100,100
CALL RNSET(19720612)
!k=500
WRITE(*,*) 'Num of Pools=',K
WRITE(1,*) '-----------------------------------------------------------------'
WRITE(1,*) 'Num of Pools=',K
ALLOCATE(B(K),C(K),PROBABILITY(K+1),PSZ(K),TMP(K),TMP2(K),RSLT(K),L
AMDA_C(SIMU1),LRT(SIMU1),WALD_C(SIMU1),&
SCORE_NEW_C(SIMU1))
CALL DRNUN(K,TMP)
PSZ=DNINT(LBD+(UBD-LBD)*TMP)
CALL MARCUS_PMF(K,Q,PSZ,PROBABILITY)
!-------------------------------------------------------------------------------------------------
! CRITICAL VALUE ON THE LEFT EXACT DISTRIBUTION
!-------------------------------------------------------------------------------------------------
CDF=0.0D0
DO I=0,K
CDF=CDF+PROBABILITY(I)
IF (CDF .GE. SIG) EXIT
END DO
CRITICAL_L=I-1
```

```
TYPE1_L=CDF-PROBABILITY(I)
GAMMA_L=(SIG-TYPE1_L)/PROBABILITY(I)
!-----------------------------------------------------------------------------------------------------
! CRITICAL VALUES OF LR TEST
!-----------------------------------------------------------------------------------------------------
SIMULATION1:DO J=1,SIMU1
P_COUNT=0
DO WHILE(P_COUNT .EQ. 0)
CALL DRNUN(K,TMP2)
DO I=1,K
IF (TMP2(I) .LE. 1.0D0-(1.0D0-Q)**PSZ(I)) THEN
RSLT(I)=1.0D0
P_COUNT=P_COUNT+1
ELSE
RSLT(I)=0.0D0
END IF
END DO
END DO
!-----------------------------------------------------------------------------------------------------

NUMRATR=LL(K,PSZ,RSLT,Q)
CALL MLE(PSZ,RSLT,ML_EST,K)
DENOM=LL(K,PSZ,RSLT,ML_EST)
LAMDA_C(J)=NUMRATR/DENOM
IF(Q .LE. ML_EST) THEN
LRT(J)=0.0D0
ELSE
LRT(J)=-2.0D0*DLOG(NUMRATR/DENOM)
END IF
!-----------------------------------------------------------------------------------------------------

WALD_C(J)=(ML_EST-Q)/sqrt(VAR(K,PSZ,ML_EST)/real(K,8))
SCORE_NEW_C(J)=SCORE_FTN(K,PSZ,RSLT,Q)/sqrt(1.0D0/VAR(K,PSZ,Q)*real(K
,8))
!-----------------------------------------------------------------------------------------------------

END DO SIMULATION1
CALL QUANTILE(SIMU1,0.95D0,LRT,LR_R)
CALL QUANTILE(SIMU1,0.05D0,Wald_C,WALD_R)
CALL QUANTILE(SIMU1,0.05D0,SCORE_NEW_C,SCORE_R)
WRITE(*,*) 'LR_R=',LR_R,'***','Wald_R=',Wald_R,'Score_R=',Score_R
!-----------------------------------------------------------------------------------------------------
! SIMULATED POWER
!-----------------------------------------------------------------------------------------------------
DO P=0.00002D0,0.00050D0,0.00002D0
CALL MARCUS_PMF(K,P,PSZ,PROBABILITY)
```

```
POWER_L=0.0D0
DO I=0,CRITICAL_L
POWER_L=POWER_L+PROBABILITY(I)
END DO
POWER1=POWER_L+GAMMA_L*PROBABILITY(CRITICAL_L+1)
POWER=POWER1
WRITE(*,"(F8.6,5X,F8.6,5X,I,5X,A10)")P, POWER,K,'EXACT'
WRITE(3,"(F8.6,5X,F8.6,5X,I,5X,A10)")P, POWER,K,'EXACT'
LR_COUNT_M=0
LR_COUNT=0
LR_COUNT_C=0
WALD_COUNT=0
WALD_COUNT_C=0
SCORE_NEW_COUNT=0
SCORE_NEW_COUNT_C=0
SIMULATION2:DO J=1,SIMU2
P_COUNT=0
DO WHILE(P_COUNT .EQ. 0)
CALL DRNUN(K,TMP2)
DO I=1,K
IF (TMP2(I) .LE. 1.0D0-(1.0D0-P)**PSZ(I)) THEN
RSLT(I)=1.0D0
P_COUNT=P_COUNT+1
ELSE
RSLT(I)=0.0D0
END IF
END DO
END DO
!-------------------------------------------------------------------------------------------------

NUMRATR=LL(K,PSZ,RSLT,Q)
CALL MLE(PSZ,RSLT,ML_EST,K)
DENOM=LL(K,PSZ,RSLT,ML_EST)
IF(Q .LE. ML_EST) THEN
LRT(J)=0.0D0
ELSE
LRT(J)=-2.0D0*DLOG(NUMRATR/DENOM)
END IF
IF(LRT(J) .GT.LR_R) THEN
LR_COUNT=LR_COUNT+1
END IF
IF(LRT(J) .GT. 2.70554345d0) THEN
LR_COUNT_C=LR_COUNT_C+1
END IF
IF(LRT(J) .GT. 2.478755967d0) THEN
LR_COUNT_M=LR_COUNT_M+1
```

```
END IF
!-----------------------------------------------------------------------------------------------

WALD=(ML_EST-Q)/sqrt(VAR(K,PSZ,ML_EST)/real(K,8))
SCORE_NEW=SCORE_FTN(K,PSZ,RSLT,Q)/sqrt(1.0D0/VAR(K,PSZ,Q)*real(K,8))
IF((WALD .LE. WALD_R)) THEN
WALD_COUNT=WALD_COUNT+1
END IF
IF(WALD .LE. -1.644853627D0) THEN
WALD_COUNT_C=WALD_COUNT_C+1
END IF
IF(SCORE_NEW .LE. SCORE_R) THEN
SCORE_NEW_COUNT=SCORE_NEW_COUNT+1
END IF
IF(SCORE_NEW .LE. -1.644853627D0) THEN
SCORE_NEW_COUNT_C=SCORE_NEW_COUNT_C+1
END IF
!-----------------------------------------------------------------------------------------------

END DO SIMULATION2
LR_POWER_M=REAL(LR_COUNT_M,8)/REAL(SIMU2,8)
LR_POWER=REAL(LR_COUNT,8)/REAL(SIMU2,8)
LR_POWER_C=REAL(LR_COUNT_C,8)/REAL(SIMU2,8)
WALD_POWER=REAL(WALD_COUNT,8)/REAL(SIMU2,8)
WALD_POWER_C=REAL(WALD_COUNT_C,8)/REAL(SIMU2,8)
SCORE_NEW_POWER=REAL(SCORE_NEW_COUNT,8)/REAL(SIMU2,8)
SCORE_NEW_POWER_C=REAL(SCORE_NEW_COUNT_C,8)/REAL(SIMU2,8)
WRITE(*,'(F8.6,5X,F8.6,5X,I,5X,A10)')P,LR_POWER,K,'LR'
WRITE(*,'(F8.6,5X,F8.6,5X,I,5X,A10)')P,LR_POWER_C,K,'LR_C'
WRITE(*,'(F8.6,5X,F8.6,5X,I,5X,A10)')P,LR_POWER,K,'LR'
WRITE(*,'(F8.6,5X,F8.6,5X,I,5X,A10)')P,LR_POWER_C,K,'LR_C'
WRITE(4,'(F8.6,5X,I,8(5X,F22.15))')
P,K,POWER,LR_POWER_M,LR_POWER,LR_POWER_C,WALD_POWER,WALD_P
OWER_C,SCORE_NEW_POWER, &
SCORE_NEW_POWER_C
END DO
DEALLOCATE(B,C,PROBABILITY,PSZ,TMP,TMP2,RSLT,LAMDA_C,LRT,WALD
_C,SCORE_NEW_C)
END DO
END PROGRAM MAIN
```