

7-2-2015

Comparisons of Estimators of Small Proportion under Group Testing

Xing Wei
xwei005@fiu.edu

DOI: 10.25148/etd.FIDC000104

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>

 Part of the [Statistical Models Commons](#)

Recommended Citation

Wei, Xing, "Comparisons of Estimators of Small Proportion under Group Testing" (2015). *FIU Electronic Theses and Dissertations*. 2195.

<https://digitalcommons.fiu.edu/etd/2195>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

COMPARISONS OF ESTIMATORS OF SMALL PROPORTION
UNDER GROUP TESTING

A thesis submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE

in

STATISTICS

by

Xing Wei

2015

To: Dean Michael R. Heithaus
College of Arts and Sciences

This thesis, written by Xing Wei, and entitled Comparisons of Estimators of Small Proportion under Group Testing, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this thesis and recommend that it be approved.

Florence George

Kai Huang, Co-major Professor

Jie Mi, Co-major Professor

Date of Defense: Jul 02, 2015

The thesis of Xing Wei is approved.

Dean Michael R. Heithaus
College of Arts and Sciences

Dean Lakshmi N. Reddi
University Graduate School

Florida International University, 2015

ACKNOWLEDGMENTS

First of all, I would like to express my sincere thanks to my major professor, Dr. Mi Jie for his patient guidance, enthusiasm, encouragement and friendship throughout this whole study. I would not have completed my research without his great support.

I would also like to thank the members of my committee, Dr. Florence George and Dr. Kai Huang for their time, valuable advice and great encouragement. In addition, I would like to thank the Department of Mathematics and Statistics, all the professors who supported and encouraged me throughout my life in Florida International University.

Finally, I will express my appreciation to my parents and friends for their encouragement and support.

ABSTRACT OF THE THESIS
COMPARISONS OF ESTIMATORS OF SMALL PROPORTION
UNDER GROUP TESTING

by

Xing Wei

Florida International University, 2015

Miami, Florida

Professor Jie Mi, Co-major Professor

Professor Kai Huang, Co-major Professor

Binomial group testing has been long recognized as an efficient method of estimating proportion of subjects with a specified characteristic. The method is superior to the classic maximum likelihood estimator (MLE), particularly when the proportion is small. Under the group testing model, we assume the testing is conducted without error. In the present research, a new Bayes estimator will be proposed that utilizes an additional piece of information, the proportion to be estimated is small and within a given range. It is observed that with the appropriate choice of the hyper-parameter our new Bayes estimator has smaller mean squared error (MSE) than the classic MLE, Burrows estimator, and the existing Bayes estimator. Furthermore, on the basis of heavy Monte Carlo simulation we have determined the best hyper-parameters in the sense that the corresponding new Bayes estimator has the smallest MSE. A table of these best hyper-parameters is made for proportions within the considered range.

TABLE OF CONTENTS

CHAPTER	PAGE
I. Introduction	1
II. Model Descriptions	4
1. The Maximum Likelihood Estimate and Burrow's Estimate	4
2. Bayes Estimator and Empirical Bayes Estimator for Unrestricted p	4
3. Bayes Estimator and Empirical Bayes Estimator for Restricted p	6
III. Computation of the Posterior Mean of P and P^*	8
1. Analytical Approach	8
2. Software Approach	10
3. Monte Carlo Simulation	10
IV. Performance Evaluation of Bayesian Estimator	12
1. Definition of Bias, MSE and Bayes Risk	12
2. Relationship between Posterior Mean and Beta	13
3. Comparison of \hat{p} , \hat{p}^* and Burrows' Estimator \tilde{p}	14
4. Optimal MSE and Choice of Beta	18
5. Application and Comparison on Hepatitis C Screening	29
V. Conclusion and Possible Future Work	32
REFERENCES	33

LIST OF TABLES

TABLE	PAGE
1. Relative Bias and Relative Efficiency ($\beta \in (1, 25), n = 30, s = 10, p = 0.05$) ...	17
2. Relative Bias and Relative Efficiency ($\beta \in (1, 24), n = 50, s = 8, p = 0.05$)	19
3. Relative Efficiency and Optimal Beta for \hat{p} and \hat{p}^* ($s = 5, n = 10$ or $n = 20$) ..	21
4. Relative Efficiency and Optimal Beta for \hat{p} and \hat{p}^* ($s = 5, n = 30$ or $n = 40$) ..	22
5. Relative Efficiency and Optimal Beta for \hat{p} and \hat{p}^* ($s = 10, n = 10$ or $n = 20$) .	23
6. Relative Efficiency and Optimal Beta for \hat{p} and \hat{p}^* ($s = 10, n = 30$ or $n = 40$) .	24
7. Relative Efficiency and Optimal Beta for \hat{p} and \hat{p}^* ($s = 15, n = 10$ or $n = 20$) .	25
8. Relative Efficiency and Optimal Beta for \hat{p} and \hat{p}^* ($s = 15, n = 30$ or $n = 40$) .	26
9. Relative Efficiency and Optimal Beta for \hat{p} and \hat{p}^* ($s = 20, n = 10$ or $n = 20$) .	27
10. Relative Efficiency and Optimal Beta for \hat{p} and \hat{p}^* ($s = 20, n = 30$ or $n = 40$)	28
11. Anti-HCV data from Liu et al. (1997) with $s = 1$ and $s = 5$	30
12. Anti-HCV data from Liu et al. (1997) with $s = 5$	31

I. Introduction

Dorfman (1943) put forward an application for group testing which related to the United States Public Health Service and the Selective Service System's project to eliminate all syphilitic men called up for induction. From then on, group testing has become a subject of continuing research. Though group testing was first utilized for blood testing, because of the wide applications of group testing, it has been applied to many other fields, such as plant disease assessment, fisheries and vector transmission of viruses (Romanow et al., 1986; Murrall et al., 1996; Ornaghi et al., 1999). The present thesis will focus on the estimation problem. Suppose that we have drawn experimental units from a population and pooled them into n groups with group size s . Also suppose that we are interested in a certain characteristic of the experimental units, and it can be found that each experimental unit either possesses the characteristic or not. We denote the appearance of this specific characteristic of the experimental unit as "success" and the unrevealed as "failure". What we want to estimate is the proportion of the individuals with "success" in the population. In the process of conducting group testing, we test groups of experimental units and observe the outcome of each group instead of testing each individual. In general, if one of the individuals in a group has the specific characteristic, then the characteristic will be observed for this group. In the present case we will simply say that a group is defective, or positive for brevity. The reason for calling the group is defective is because most of time the above mentioned specific characteristic is defectiveness, sickness, infected etc. If all of the individuals in one group do not have such characteristic, then the result of

the group testing will be non-defective or negative. Further, we need to assume the sensitivity and specificity of group testing is perfect.

Our goal is to estimate the proportion $p \in (0, 1)$ of the individuals that possesses the specified characteristic in the target population. Chen and Swallow (1990) have shown that the retesting of individuals in the defective groups is not worthwhile. Hung and Swallow (1999) have shown that group testing is often more preferable to individual testing for estimating proportions of defectives when the proportion of defective individuals is low.

Under the group testing models, there are many approaches to the estimation of proportion of defective individuals, such as MLE, Burrows bias corrected MLE, Bayes estimators, empirical Bayes estimators, and various interval estimators. Some research on the comparison of these methods have been done by Bilder and Tebbs (2005). They utilized both the Bayes evaluation and frequentist evaluation. A method that uses on the score statistic with a correction for skewness and another method using the logit function is applied to the MLE was recommended by Hepworth (2005). Usually, the object of selecting an estimator in group testing is to find the one that minimizes the MSE of the estimator of p as mentioned by Hung and Swallow (1999). The classic estimator derived from likelihood is not accurate for small proportions p , and particularly in practice often our interest is in the occurrence of rare events, so we will restrict our study on the case when the defect rate p is in the interval of $(0, 0.2]$, and then we will compare the different approaches of estimating the proportion of defectives in the population.

In the following section, the model used in the present study will be introduced in detail and MLE, Burrows' estimate and Bayes estimation of p will be reviewed (Burrows 1986; Fang et al., 2007). Further, a new Bayes estimator will be introduced. Section 3 mainly shows how to calculate the proposed estimator. Section 4 gives the performance evaluation of the proposed estimator and application of proposed estimator. Section 5 provides a discussion of the results.

II. Model Descriptions

1. The Maximum Likelihood Estimate and Burrow's Estimate

Suppose that a sample of size N is randomly selected from the target population and divided equally into n groups, and each group contains s experimental units ($N = ns$). For the discussion of unequal group sizes, the reader can refer to Chen and Swallow (1995). One observes Y_1, Y_2, \dots, Y_n which are independently and identically distributed Bernoulli (θ) random variables, where $\theta = 1 - (1 - p)^s$ (s is group size, p is proportions of defectives in the target population, and θ is the probability that a group is defective), and $Y_i=1$ if the i -th group is defective and $Y_i=0$ if the i -th group is non-defective. Let $T = \sum_{i=1}^n Y_i$ denote the number of the defective groups and assuming independence among the groups, then T follows binomial distribution with parameters n and θ . It is also assumed that the testing is conducted without error, in other words, the sensitivity and specificity are perfect. The MLE for p under the group testing model is $\hat{p}_{MLE} = 1 - (1 - T/n)^{1/s}$. An alternative estimate, $\tilde{p} = 1 - ((2s(n - T) + s - 1)/(2ns + s - 1))^{1/s}$, which is proposed by Burrows(1986). Burrows has shown that this estimator's bias and mean square error properties are uniformly superior to \hat{p}_{MLE} except for $s = 1$ when these two estimates are identical to the minimum variance unbiased estimate.

2. Bayes Estimator and Empirical Bayes Estimator for Unrestricted p

As a prior distribution for p , we choose the one-parameter beta family for the random variable P as in Tebbs et al. (2003). It will give us great simplifications when

we continue our computation. The prior distribution (Beta(1, β)) will have the form:

$$f_P(p|\beta) = \beta(1-p)^{\beta-1}I(0 < p < 1) \quad (1)$$

The number of the defect groups T follows the binomial distribution with parameter θ . We can derive the joint probability density function of T and P as

$$\begin{aligned} f_{T,P}(t, p) &= f(t|p)f_p(p|\beta) \\ &= \binom{n}{t} \beta [(1-p)^{s(n-t+\beta/s)-1}] [1 - (1-p)^s]^t \end{aligned} \quad (2)$$

The marginal PDF of T then have the form:

$$\begin{aligned} f_T(t) &= \int_0^1 f_{T,P}(t, p) dp \\ &= \binom{n}{t} \frac{\beta \Gamma(n-t+\beta/s) \Gamma(t+1)}{s \Gamma(n+\beta/s+1)} \end{aligned} \quad (3)$$

Using the one-parameter beta prior distribution with a given β Tebbs et al. (2003) showed that the posterior distribution of p given $T = t$ will have the form:

$$\begin{aligned} f_{p|T}(p|t) &= \frac{f_{T,P}(t, p)}{f_T(t)} \\ &= \frac{s \Gamma(n+\beta/s+1)}{\Gamma(n-t+\beta/s) \Gamma(t+1)} (1-p)^{s(n-t)+\beta-1} [1 - (1-p)^s]^t \end{aligned} \quad (4)$$

For $0 < p < 1$, the posterior mean of p is:

$$\begin{aligned} E_{P|T}[p|t] &= \int_0^1 p f_{P|T}(p|t) dp \\ &= 1 - \frac{\Gamma(n+\beta/s+1) \Gamma(n-t+\beta/s+1/s)}{\Gamma(n-t+\beta/s) \Gamma(n+\beta/s+1/s+1)} \end{aligned} \quad (5)$$

In the present study, we consider the squared-error loss function $L(p, a) = (p - a)^2$ though the more general loss function $L(p, a) = w(p)(p-a)^2$ could be used. Obviously, the square loss function is the special case of the general loss function when $w(p) = 1$.

The Bayes estimator of p corresponding to this loss function is given by:

$$\begin{aligned} a &= \frac{\int_0^1 p f_{P|T}(p|t) dp}{\int_0^1 f_{P|T}(p|t) dp} \\ &= 1 - \frac{\Gamma(n + \beta/s + 1)\Gamma(n - t + \beta/s + 1/s)}{\Gamma(n - t + \beta/s)\Gamma(n + \beta/s + 1/s + 1)} \end{aligned} \quad (6)$$

which is exactly the posterior mean of p given in (5). Empirical Bayes, represents one approach for setting hyper-parameters in the prior distribution. In our case, maximization of the marginal density of T given in Eq. (3) with respect to β provides a MLE of β . It is easy to see that when $t = 0$, Eq. (3) is strictly increasing and when $t = n$, Eq. (3) is strictly decreasing. Thus, the MLE of β exists only for $t = 1, 2, \dots, n - 1$ and denote this estimator by $\hat{\beta}$. Replacing β by $\hat{\beta}$ in (5) gives the Empirical Bayes (EB) estimators of p . Hence, for the squared error loss function the EB estimator for p is given by:

$$\hat{p}_{eb} = 1 - \frac{\Gamma(n + \hat{\beta}/s + 1)\Gamma(n - t + \hat{\beta}/s + 1/s)}{\Gamma(n - t + \hat{\beta}/s)\Gamma(n + \hat{\beta}/s + 1/s + 1)} \quad (7)$$

3. Bayes Estimator and Empirical Bayes Estimator for Restricted p

Group testing models are often used on rare events. In the case of rare events, we know that p can only be in a range much smaller than $(0, 1)$. If we want to apply this piece of information into our study, then we should restrict the range of the proportion p to be an interval $(0, c]$ where c is a constant less than 1, say $c = 0.2$ when our prior information is $p \leq 0.2$. To avoid notational confusion we will use P^* as the restricted P . With this convention we can derive the probability density function (PDF) of P^* , joint PDF of P^* and T , and the posterior mean of P^* as well.

For example, the prior distribution of P^* can be given by:

$$f_{P^*}(p^*|\beta) = \frac{\beta}{c} \left(1 - \frac{p^*}{c}\right)^{\beta-1} I(0 < p^* \leq c) \quad (8)$$

The joint PDF of P^* and T is given by:

$$\begin{aligned} f_{T,P^*}(t, p^*) &= f_{T|P^*}(t|p^*) f_{P^*}(p^*|\beta) \\ &= \frac{\beta}{c} \binom{n}{t} [(1-p^*)^{s(n-t)}] [1 - (1-p^*)^s]^t (1-p^*/c)^{\beta-1} \end{aligned} \quad (9)$$

Then, we can get the marginal PDF of T :

$$f_T(t) = \int_0^c \frac{\beta}{c} \binom{n}{t} [(1-p^*)^{s(n-t)}] [1 - (1-p^*)^s]^t (1-p^*/c)^{\beta-1} dp^* \quad (10)$$

The posterior distribution of P^* given $T = t$ will have the form:

$$\begin{aligned} f_{P^*|T}(p^*|t) &= \frac{f_{P^*,T}(p^*, t)}{f_T(t)} \\ &= \frac{[1 - (1-p^*)^s]^t (1-p^*)^{s(n-t)} (1-p^*/c)^{\beta-1}}{\int_0^c [1 - (1-p^*)^s]^t (1-p^*)^{s(n-t)} (1-p^*/c)^{\beta-1} dp^*} \end{aligned} \quad (11)$$

For $0 < p^* \leq c$, the posterior mean of p^* is given by:

$$E_{P^*|T}[p^*|t] = \frac{\int_0^c p^* [1 - (1-p^*)^s]^t (1-p^*)^{s(n-t)} (1-p^*/c)^{\beta-1} dp^*}{\int_0^c [1 - (1-p^*)^s]^t (1-p^*)^{s(n-t)} (1-p^*/c)^{\beta-1} dp^*} \quad (12)$$

Different from the case of unrestricted p , it is not easy to get the posterior mean of p^* in the restricted case. In the next chapter, different methods will be used to find the posterior mean of p^* . With respect to the same square error loss function, we can substitute $\hat{\beta}$ for β and get the Empirical Bayes estimator for p^* as follows:

$$\hat{p}_{eb}^* = \frac{\int_0^c p^* [1 - (1-p^*)^s]^t (1-p^*)^{s(n-t)} (1-p^*/c)^{\hat{\beta}-1} dp^*}{\int_0^c [1 - (1-p^*)^s]^t (1-p^*)^{s(n-t)} (1-p^*/c)^{\hat{\beta}-1} dp^*} \quad (13)$$

III. Computation of the Posterior Mean of P and P^*

1. Analytical Approach

Computation for the posterior mean of P is quite straight forward since the nice closed form given in Eq.(5). If we enter the exact values of n , β , s and t into the Eq.(5) we are able to compute the posterior mean of P .

However, computation for the posterior mean of P^* is not easy. It is difficult to find the closed form of the posterior mean of P^* . Considering the values of s , n , t in Eq. (12) are positive integers, we can use Binomial theorem to find the close form of Eq.(12).

$$E_{P^*|T}[p^*|t] = \frac{\int_0^c p^* [1 - (1 - p^*)^s]^t (1 - p^*)^{s(n-t)} (1 - p^*/c)^{\beta-1} dp^*}{\int_0^c [1 - (1 - p^*)^s]^t (1 - p^*)^{s(n-t)} (1 - p^*/c)^{\beta-1} dp^*}$$

We define $m = s(n-t)$, then we can use the binomial expansion to get the expression of numerator:

$$Num. = \int_0^c \sum_{i=0}^t \sum_{j=0}^{(t-i)s+m} \binom{t}{i} \binom{s(t-i)+m}{j} (-1)^{(s+1)(t-i)+m-j} p^{s(s-i)+m-j+1} (1-p^*/c)^{\beta-1} dp^*$$

Define $x = p/c$, then we get:

$$Num. = \int_0^1 \sum_{i=0}^t \sum_{j=0}^{(t-i)s+m} \frac{\binom{t}{i} \binom{s(t-i)+m}{j} (-1)^{(s+1)(t-i)+m-j}}{(1/c)^{s(t-i)+m-j+2}} x^{s(t-i)+m-j+1} (1-x)^{\beta-1} dx$$

Define

$$A_{ij} = \frac{\binom{t}{i} \binom{s(t-i)+m}{j} (-1)^{(s+1)(t-i)+m-j}}{(1/c)^{s(t-i)+m-j+1}}$$

$$Num. = c \sum_{i=0}^t \sum_{j=0}^{(n-i)s} A_{ij} Beta(s(n-i) - j + 2, \beta)$$

Similarly, we can get the denominator as follows:

$$Den. = \sum_{i=0}^t \sum_{j=0}^{(n-i)s} A_{ij} Beta(s(n-i) - j + 1, \beta)$$

By using the binomial expansion, we can simplify the Eq.(12) and get the close form of posterior mean for p^* :

$$\begin{aligned}
E_{P^*|T}[p^*|t] &= \frac{\int_0^c p^*[1 - (1 - p^*)^s]^t (1 - p^*)^{s(n-t)} (1 - p^*/c)^{\beta-1} dp^*}{\int_0^c [1 - (1 - p^*)^s]^t (1 - p^*)^{s(n-t)} (1 - p^*/c)^{\beta-1} dp^*} \\
&= c \frac{\sum_{i=0}^t \sum_{j=0}^{\binom{n-i}{s}} A_{ij} \text{Beta}(s(n-i) - j + 2, \beta)}{\sum_{i=0}^t \sum_{j=0}^{\binom{n-i}{s}} A_{ij} \text{Beta}(s(n-i) - j + 1, \beta)}
\end{aligned} \tag{14}$$

However, we cannot use Eq. (14) to calculate the posterior mean of P^* directly.

When sample size s and number of groups n are large, the result will be unsatisfied because the accumulation of error in computation. For example, if we define the true p equals to 0.1, sample size s equals to 10, number of groups n equals to 30, and re-samples for 500 times. For each sample, an estimator of p^* can be found. Figure 1 shows the results ranges from 0.19618 to 0.19625 approximately. The posterior mean calculated by analysis approach has huge accumulative error since the true p is only 0.1. In the following calculation, analysis approach will not be used because of the huge size of the accumulative error.

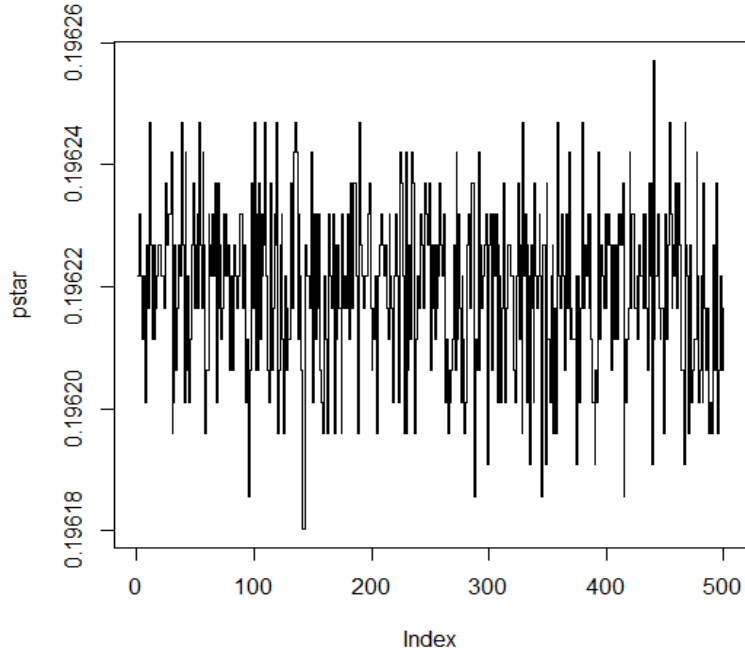


Figure 1: Plot of p^* by Analysis Approach

2. Software Approach

Using software is the most direct way of computing the posterior mean of p and p^* . We can assign values to n , β , s , t and c in Eq.(12), then we can get the posterior mean of p and p^* . Many software packages can be used, such as SAS, Minitab, R and Matlab. In this research, we use software R for calculation.

3. Monte Carlo Simulation

The Monte Carlo simulation method often used to solve questions which are too complex to get an analytical solution. Thus, we rely on repeated random sampling to obtain numerical results is an effective solution. Although the close-form expression of the posterior mean of p^* has been found, it is useless. A Monte Carlo simulation

can give another approach in solving for posterior mean of p^* . The PDF of p^* is given in Eq(8). Let $p = p^*/c$, then $p \in (0, 1)$ and follows one-parameter Beta distribution $Beta(1, \beta)$, so we can rewrite Eq.(12) as follow:

$$\begin{aligned} E_{P^*|T}[p^*|t] &= \frac{\int_0^c p^*[1 - (1 - p^*)^s]^t (1 - p^*)^{s(n-t)} (1 - p^*/c)^{\beta-1} dp^*}{\int_0^c [1 - (1 - p^*)^s]^t (1 - p^*)^{s(n-t)} (1 - p^*/c)^{\beta-1} dp^*} \\ &= c \frac{\int_0^1 p[1 - (1 - cp)^s]^t (1 - cp)^{s(n-t)} f(p) dp}{\int_0^1 [1 - (1 - cp)^s]^t (1 - cp)^{s(n-t)} f(p) dp} \end{aligned} \quad (15)$$

where $f(p)$ is the pdf of one-parameter Beta distribution.

$$f(p) = \beta(1 - p)^{\beta-1} I(0 < p < 1)$$

Define

$$g(p) = p[1 - (1 - cp)^s]^t (1 - cp)^{s(n-t)}$$

$$h(p) = [1 - (1 - cp)^s]^t (1 - cp)^{s(n-t)}$$

Then

$$E_{P^*|T}[p^*|t] = c \frac{E[g(p)]}{E[h(p)]}$$

We define another random variable ξ has a PDF $Beta(1, \beta)$ and let $p_i = \xi_i$, $1 \leq \xi \leq m$.

According to the Law of Large Numbers, the average of the results obtained from a large number of trials will converge to the expected value as the number of trials increased to infinity.

$$\begin{aligned} \lim_{m \rightarrow \infty} \frac{\sum_{i=1}^m g(p_i)}{m} &= E[g(p)] \\ \lim_{m \rightarrow \infty} \frac{\sum_{i=1}^m h(p_i)}{m} &= E[h(p)] \end{aligned}$$

IV. Performance Evaluation of Bayesian Estimator

1. Definition of Bias, MSE and Bayes Risk

From a frequentist's point of view, estimators are evaluated by using bias and MSE. In the present study. We will use the relative bias and relative efficiency comparing with the Burrows' estimator (\tilde{p}) because Burrows' estimator performs better than the classic MLE(\hat{p}_{MLE}) regardless of whether bias or MSE is concerned. Suppose that \hat{p} is an estimator of p , the definition of bias and MSE are given by:

$$Bias(\hat{p}) = E(\hat{p} - p) = E(\hat{p}) - p$$

$$MSE(\hat{p}) = E(\hat{p} - p)^2$$

The relative bias and relative efficiency with respect to the Burrows' estimator of \tilde{p} are defined as:

$$RB(\hat{p}) = \frac{Bias(\hat{p})}{Bias(\tilde{p})}$$
$$RE(\hat{p}) = \frac{MSE(\tilde{p})}{MSE(\hat{p})}$$

respectively. If the absolute value of $RB(\hat{p})$ is less than 1, the result indicates that the bias of \hat{p} has a smaller absolute value than that of the Burrows' estimator; likewise, if $RE(\hat{p})$ is less than 1, then the MSE of \hat{p} is less than the MSE of the Burrows' estimator.

From a Bayesian point of view, Bayes estimators can be compared by using the Bayes risk which is defined as follows: let \hat{p} be an Bayes estimator of p , then its Bayes risk associated with the squared error loss function is given as:

$$r(\hat{p}) = E_P[E_{T|P}(\hat{p} - p)^2]$$

In general, it is very difficult to get the close form of Bayes risk. Bilder and Tebbs(2004) have derived the Bayes risk when p has prior on $(0, 1)$ under squared error loss function. In our research, we focus on comparing the relative bias and relative efficiency.

2. Relationship between Posterior Mean and Beta

For clarity of statement the posterior mean of P is denoted as \hat{p}_B when the prior distribution of P is defined on $(0, 1)$, and when the prior distribution of P is restricted to $(0, c]$ the posterior mean of P^* is denoted as \hat{p}^* , where we use P^* to emphasize that P is restricted to $(0, c]$.

Figure 2 shows the relationship between posterior mean of P and parameter β . We set the β ranges from 1 to 20000, $n = 30$, $s = 10$ and the true value of p is 0.10. In Figure 2, the curve shows the changing of posterior mean when β changes and the horizontal line shows the true value of p . It can be observed that when β is small, the posterior mean of P overestimates the true p . As β surpasses a specific value, the posterior mean of P then underestimates the true p .

Let $c = 0.2$, then the prior distribution is restricted to $(0, 0.2]$ and keep the other assumptions the same as in Figure 2 then we can get the posterior mean of P^* . Figure 3 shows the relationship between posterior mean P^* and parameter β . When β increases, the posterior mean P^* decreases which displays the same pattern as in Figure 2. When we compare the curves in Figure 2 and Figure 3, we find the curve in Figure 3 is much more closer to the origin $(0, 0)$ than the curve in Figure 2, indicating that they have different change rate. Again, when β is small, the posterior mean of

P^* overestimates the true p . As β surpasses a specific value, the posterior mean of P will underestimate the true p .

It is true that the Bayes estimate of p decreases as the parameter β increases no matter what the prior distribution is defined on $(0, 1)$ or $(0, c]$.

3. Comparison of \hat{p} , \hat{p}^* and Burrows' Estimator \tilde{p}

The performance of \hat{p} and \hat{p}^* can be compared further to see the influence of prior information on the posterior mean. Suppose $0 < a < b < 1$, a and b are constants. To avoid confusion of notation when the prior distribution is defined on $(0, a]$, the posterior mean is denoted as \hat{p}_1^* ; and when the prior distribution is defined on $(0, b]$, the posterior mean is denoted as \hat{p}_2^* . The performance of \hat{p}_1^* and \hat{p}_2^* can be measured by the contribution of prior information. For example, let $a = 0.10$, $b = 0.20$, $\beta \in \{1, 2, \dots, 20\}$, $n = 30$, $s = 10$, and the true value of p be 0.05. Let us generate 500 independent samples of T to obtain the classic MLE and Bayes estimates for p . And then we are able to compute relative bias and relative efficiency. The results are shown in Table 1. In Table 1, relative bias and relative efficiency with respect to MLE are used as a criteria for comparing these estimators. If the absolute value of relative bias less than 1, it indicates this estimator has smaller absolute bias than that of the classic MLE. On the other hand if the relative efficiency is less than 1, it indicates that the estimator has a smaller MSE than the classic MLE. For both relative bias and relative efficiency, the smaller absolute value, the better performance of corresponding estimator has. The relative bias and relative efficiency respect to MLE are defined as below.

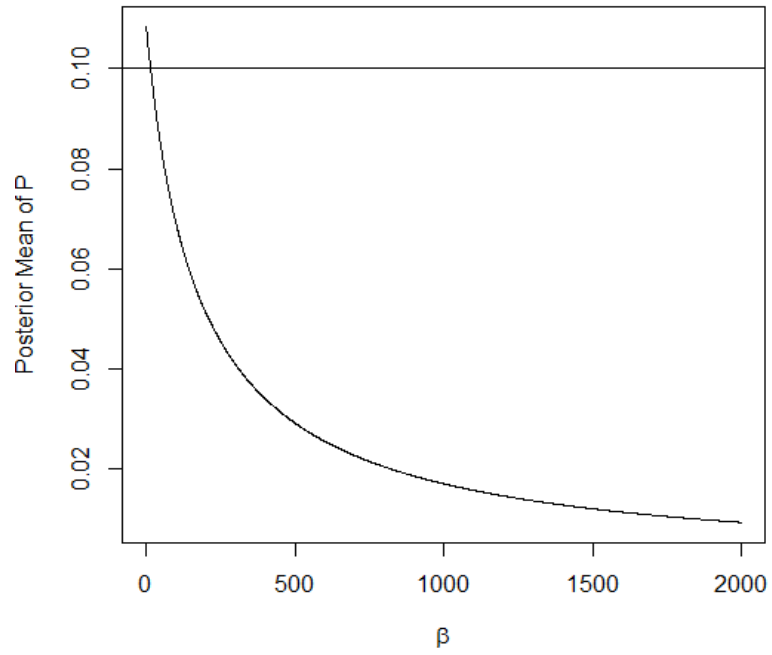


Figure 2: Posterior Mean of P and β ($\beta \in (1, 2000)$, $n = 30$, $s = 10$, $p = 0.10$)

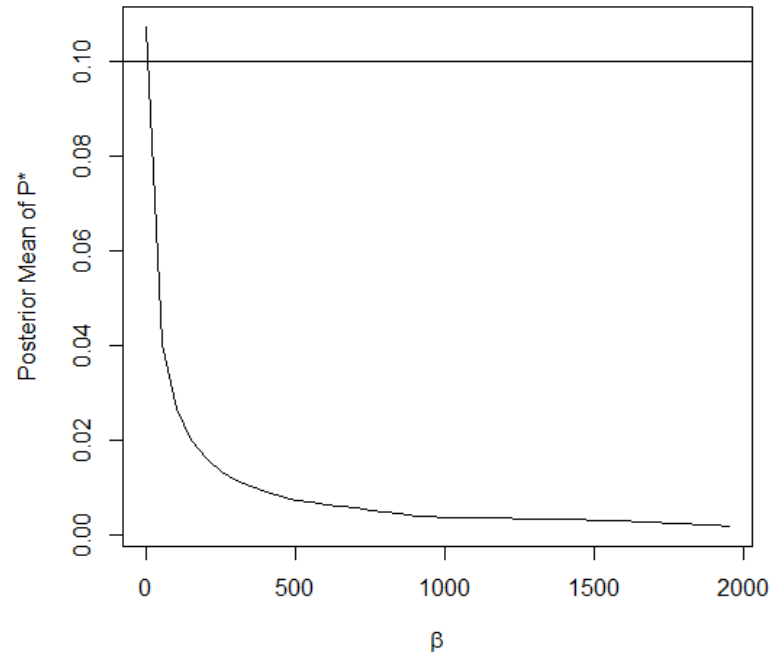


Figure 3: Posterior Mean of \hat{P}^* and β ($\beta \in (1, 2000)$, $n = 30$, $s = 10$, $p = 0.10$)

$$RB(\hat{p}) = \frac{Bias(\hat{p})}{Bias(\hat{p}_{MLE})}$$

$$RE(\hat{p}) = \frac{MSE(\hat{p})}{MSE(\hat{p}_{MLE})}$$

On the basis of the data in Table 1, the absolute value of relative bias and relative efficiency have the same pattern that as β increases at first these values decrease and then increase. Clearly, these values change with different rates. But regardless of the different rates these values will achieve their minimum for a certain value of the hyper-parameter β . It implies then at those specific values of β the Bayes estimators \hat{p}_1^* or \hat{p}_2^* will have minimum absolute bias or minimum MSE. Focused on relative bias, when β is small ($\beta \in [1, 10]$), $RB(\hat{p}_1^*)$ achieved the smallest absolute value of relative bias while when β is big ($\beta \in [11, 25]$), $RB(\hat{p})$ achieved the smallest absolute value. The bias of these estimates are highly related to the value of β . Considering the relative efficiency, when β is small ($\beta \in [1, 10]$), $RE(\hat{p}_2^*)$ achieved the smallest absolute value of relative efficiency. Thus both bias and mean square errors of the estimates are highly related to the value of β . Then the bias and mean square error are very sensitive to the choice of β when conducting group testing.

Burrows (1989) has provided an alternative method of estimation of defect rate to maximum likelihood estimator with superior bias and mean squared error properties. The reduced biased estimator is given by $\tilde{p} = 1 - ((2s(n - T) + s - 1)/(2ns + s - 1))^{1/s}$. Burrows' estimator improves the efficiency of group testing and extends the range of conditions where group testing is more efficient than individual

Beta	$RB(\hat{p})$	$RB(\hat{p}_1^*)$	$RB(\hat{p}_2^*)$	$RE(\hat{p})$	$RE(\hat{p}_1^*)$	$RE(\hat{p}_2^*)$
1	4.359	4.358	3.243	1.143	1.142	0.818
2	4.153	2.953	-0.644	1.12	0.974	0.559
3	3.949	1.676	-3.378	1.099	0.855	0.523
4	3.748	0.501	-5.549	1.078	0.771	0.575
5	3.547	-0.588	-7.366	1.058	0.714	0.669
6	3.349	-1.605	-8.934	1.038	0.678	0.788
7	3.153	-2.557	-10.313	1.019	0.658	0.92
8	2.958	-3.454	-11.546	1.001	0.653	1.059
9	2.765	-4.3	-12.659	0.984	0.658	1.202
10	2.573	-5.101	-13.672	0.968	0.674	1.347
11	2.384	-5.862	-14.602	0.952	0.697	1.492
12	2.196	-6.585	-15.459	0.936	0.726	1.635
13	2.009	-7.275	-16.255	0.922	0.761	1.777
14	1.824	-7.933	-16.995	0.908	0.801	1.916
15	1.641	-8.563	-17.688	0.894	0.845	2.052
16	1.459	-9.166	-18.337	0.882	0.892	2.186
17	1.279	-9.744	-18.948	0.869	0.943	2.317
18	1.1	-10.3	-19.524	0.858	0.995	2.444
19	0.923	-10.834	-20.069	0.847	1.049	2.569
20	0.748	-11.349	-20.585	0.836	1.105	2.69
21	0.573	-11.844	-21.075	0.826	1.163	2.809
22	0.401	-12.322	-21.541	0.817	1.221	2.924
23	0.229	-12.784	-21.985	0.808	1.281	3.037
24	0.059	-13.229	-22.408	0.799	1.341	3.147
25	-0.109	-13.66	-22.813	0.791	1.402	3.254

Table 1: Relative Bias and Relative Efficiency ($\beta \in (1, 25), n = 30, s = 10, p = 0.05$)

testing. Because of these nice properties, comparisons among \hat{p} , \hat{p}^* will be respected to the Burrows' estimator instead of MLE. Set $c = 0.2$, then \hat{p}^* is the estimate when the prior distribution is defined on $(0, 0.2)$. Table 2 shows the relative bias and relative efficiency respected to Burrows' estimator.

When comparing MLE to Burrows' estimator, the absolute value of relative bias and relative efficiency are constant and greater than 1 which means Burrows' estimates has lower bias and mean square error. For the performance of \hat{p} , \hat{p}^* , they have the same pattern that decreasing as β increasing, and then increasing. But when β is small, \hat{p}^* has smaller MSE. If focused on bias, the Burrows' estimator tends to have smaller bias than other estimates. Only under a few choice, \hat{p} and \hat{p}^* will have smaller bias than Burrows' estimate, for example, when $\beta = 21$ and $\beta = 22$, absolute value of $RB(\hat{p})$ is less than 1. Considering the MSE, it is much more complicate. The Bayesian estimator seem to have a smaller MSE than the Burrows' estimator, and under a certain range of β , \hat{p}^* even have smaller MSE than \hat{p} . These results naturally lead us to consider choosing the best hyper-parameter to get the smaller MSE.

4. Optimal MSE and Choice of Beta

The bias and MSE of the estimators highly dependent on the choice of hyper-parameter β . In practice, if we can choose an appropriate β on the basis of the prior information about p , then smaller MSE of the estimator would be able to be achieved. On the basis of the simulated data, tables of the optimal β have been provided. In the following tables, the relative efficiency is the ratio of MSE of a proposed estimator

Beta	$RB(\hat{p}_{MLE})$	$RB(\hat{p})$	$RB(\hat{p}^*)$	$RE(\hat{p}_{MLE})$	$RE(\hat{p})$	$RE(\hat{p}^*)$
1	-2.434	-20.113	-20.113	1.027	1.108	1.108
2	-2.434	-19.074	-13.229	1.027	1.093	1.004
3	-2.434	-18.041	-6.734	1.027	1.079	0.925
4	-2.434	-17.015	-0.582	1.027	1.066	0.867
5	-2.434	-15.995	5.262	1.027	1.052	0.827
6	-2.434	-14.981	10.827	1.027	1.040	0.801
7	-2.434	-13.973	16.139	1.027	1.028	0.788
8	-2.434	-12.971	21.219	1.027	1.016	0.786
9	-2.434	-11.976	26.086	1.027	1.004	0.793
10	-2.434	-10.987	30.756	1.027	0.994	0.809
11	-2.434	-10.003	35.245	1.027	0.983	0.833
12	-2.434	-9.026	39.563	1.027	0.973	0.862
13	-2.434	-8.054	43.724	1.027	0.963	0.898
14	-2.434	-7.088	47.737	1.027	0.954	0.938
15	-2.434	-6.128	51.611	1.027	0.945	0.983
16	-2.434	-5.173	55.355	1.027	0.937	1.031
17	-2.434	-4.225	58.977	1.027	0.929	1.083
18	-2.434	-3.281	62.483	1.027	0.921	1.138
19	-2.434	-2.344	65.881	1.027	0.914	1.196
20	-2.434	-1.412	69.174	1.027	0.907	1.256
21	-2.434	-0.485	72.370	1.027	0.900	1.318
22	-2.434	0.436	75.472	1.027	0.894	1.381
23	-2.434	1.352	78.487	1.027	0.888	1.447
24	-2.434	2.263	81.417	1.027	0.883	1.513

Table 2: Relative Bias and Relative Efficiency ($\beta \in (1, 24)$, $n = 50$, $s = 8$, $p = 0.05$)

respects to the MSE of Burrows' estimator given by the following equation.

$$RE(\text{Proposed Estimator}) = \frac{MSE(\text{Proposed Estimator})}{MSE(\tilde{p})}$$

If we choose the optimal beta, Bayesian estimators tend to have smaller MSE than Burrows' estimator. In the mean while, the performances of unrestricted p and the restricted p are different. These tables can be a reference for practitioners to select appropriate hyper-parameter β when the binomial group testing method is applied in their research. With these tables, practitioners may select the restricted prior of p to achieve smaller MSE when the unrestricted prior of p produces large MSE.

There are infinity combinations of sample size and number of groups when practitioners to design their tests. However we only consider several combination of sample size and number of groups and assume the true p is between 0 and 0.2. Dorfman (1943) has given a table shows the optimum group size and relative testing costs for selected prevalence rates. Swallow (1985) and Hughes-Oliver and Swallow (1994) have investigated the choice of s . There are also some biological considerations associated with test assays solely determined the choice of s . For example, Kline et al. (1989) and Monzon et al. (1992) have reported that the sensitivity and specificity of tests used in HIV screening nearly perfect when $s \leq 2$. Neil and Conradie (1992, 1994) have found that tests used in hepatitis C screening are reliable when $s \leq 8$. Fang et al. (2007) have discussed these combinations: sample size of 5, 10, 20 and number of groups of 10, 30, 50, 100, 150, 200. In our study we will only consider sample sizes can be 5, 10, 15, 20 and the numbers of groups 10, 20, 30, 40. In total, 16 combination of sample size and number of groups would be given.

p	s=5, n=10				s=5, n=20			
	\hat{p}		\hat{p}^*		\hat{p}		\hat{p}^*	
	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.
0.01	131	0.0914	25	0.0860	153	0.1915	29	0.1835
0.02	74	0.1957	13	0.1795	83	0.3515	15	0.3290
0.03	53	0.2771	9	0.2464	58	0.4573	10	0.4194
0.04	42	0.3312	7	0.2840	46	0.5153	8	0.4599
0.05	34	0.3897	5	0.3236	37	0.5775	6	0.5043
0.06	29	0.4270	4	0.3403	30	0.6265	5	0.5315
0.07	25	0.4641	3	0.3609	28	0.6299	4	0.5157
0.08	24	0.4752	3	0.3437	24	0.6705	3	0.5406
0.09	21	0.4934	2	0.3456	23	0.6681	3	0.4984
0.10	19	0.5200	2	0.3284	20	0.7016	2	0.5230
0.11	17	0.5377	2	0.3261	18	0.7175	2	0.4842
0.12	17	0.5376	1	0.3006	18	0.6962	2	0.4265
0.13	15	0.5548	1	0.2538	17	0.7153	1	0.4289
0.14	14	0.5651	1	0.2434	16	0.7076	1	0.3466
0.15	14	0.5504	1	0.2281	15	0.7301	1	0.3132
0.16	12	0.5447	1	0.2363	14	0.7447	1	0.3172
0.17	12	0.5718	1	0.2953	13	0.7403	1	0.3373
0.18	12	0.5481	1	0.3234	13	0.7396	1	0.3965
0.19	11	0.5622	1	0.4042	12	0.7389	1	0.4887
0.20	11	0.5560	1	0.4808	12	0.7421	1	0.6054

Table 3: Relative Efficiency and Optimal Beta for \hat{p} and \hat{p}^* ($s = 5, n = 10$ or $n = 20$)

p	s=5, n=30				s=5, n=40			
	\hat{p}		\hat{p}^*		\hat{p}		\hat{p}^*	
	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.
0.01	157	0.2888	30	0.2792	167	0.3602	32	0.3499
0.02	88	0.4589	16	0.4360	88	0.5479	16	0.5243
0.03	62	0.5601	11	0.5222	60	0.6494	11	0.6124
0.04	48	0.6231	8	0.5706	48	0.6962	9	0.6459
0.05	39	0.6686	7	0.6003	40	0.7342	7	0.6724
0.06	33	0.7042	5	0.6198	34	0.7603	6	0.6837
0.07	29	0.7254	5	0.6210	29	0.7874	5	0.6966
0.08	26	0.7483	4	0.6228	26	0.7994	4	0.6853
0.09	23	0.7669	3	0.6199	23	0.8234	3	0.6990
0.10	21	0.7780	3	0.6009	21	0.8246	3	0.6614
0.11	20	0.7709	2	0.5687	19	0.8426	3	0.6671
0.12	18	0.7857	2	0.5318	19	0.8312	2	0.6037
0.13	18	0.7806	2	0.4959	17	0.8401	2	0.5735
0.14	16	0.8076	1	0.4726	15	0.8573	1	0.5677
0.15	15	0.8113	1	0.3985	15	0.8477	1	0.4566
0.16	14	0.8202	1	0.3646	15	0.8500	1	0.3919
0.17	14	0.8066	1	0.3985	14	0.8513	1	0.3744
0.18	13	0.8162	1	0.4141	13	0.8630	1	0.4303
0.19	12	0.8231	1	0.5330	13	0.8448	1	0.5183
0.20	12	0.8056	1	0.6561	13	0.8476	1	0.6911

Table 4: Relative Efficiency and Optimal Beta for \hat{p} and \hat{p}^* ($s = 5, n = 30$ or $n = 40$)

p	s=10, n=10				s=10, n=20			
	\hat{p}		\hat{p}^*		\hat{p}		\hat{p}^*	
	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.
0.01	151	0.1888	29	0.1807	167	0.3520	32	0.3414
0.02	83	0.3400	15	0.3181	93	0.5172	17	0.4929
0.03	61	0.4170	11	0.3791	65	0.6097	12	0.5701
0.04	48	0.4672	8	0.4106	53	0.6476	9	0.5922
0.05	42	0.4819	6	0.4100	43	0.6852	7	0.6109
0.06	36	0.5160	5	0.4234	38	0.6944	6	0.6044
0.07	31	0.5291	4	0.4010	33	0.7227	5	0.6087
0.08	28	0.5351	4	0.4010	30	0.7220	4	0.5870
0.09	25	0.5629	3	0.3962	27	0.7315	4	0.5738
0.10	23	0.5626	2	0.3736	26	0.7134	3	0.5218
0.11	22	0.5346	2	0.3189	25	0.6923	3	0.4783
0.12	20	0.5397	2	0.3052	23	0.7155	2	0.4508
0.13	19	0.5402	1	0.2683	22	0.6949	2	0.4047
0.14	18	0.5548	1	0.2331	21	0.6679	1	0.3469
0.15	17	0.5542	1	0.2093	19	0.6913	1	0.2940
0.16	16	0.5679	1	0.2275	19	0.6383	1	0.2355
0.17	15	0.6050	1	0.2881	18	0.6256	1	0.2471
0.18	14	0.6285	1	0.3740	17	0.6476	1	0.3128
0.19	13	0.6665	1	0.5061	16	0.6423	1	0.4004
0.20	12	0.6890	1	0.6639	16	0.6326	1	0.5181

Table 5: Relative Efficiency and Optimal Beta for \hat{p} and \hat{p}^* ($s = 10, n = 10$ or $n = 20$)

p	s=10, n=30				s=10, n=40			
	\hat{p}		\hat{p}^*		\hat{p}		\hat{p}^*	
	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.
0.01	173	0.4682	33	0.4567	182	0.5396	35	0.5282
0.02	92	0.6379	17	0.6141	100	0.6841	19	0.6610
0.03	71	0.6878	13	0.6509	71	0.7504	13	0.7179
0.04	54	0.7404	10	0.6923	55	0.7914	10	0.7487
0.05	44	0.7727	8	0.7117	46	0.8169	8	0.7630
0.06	39	0.7783	7	0.7036	39	0.8309	7	0.7642
0.07	35	0.7946	5	0.7008	34	0.8477	6	0.7685
0.08	32	0.7881	5	0.6676	31	0.8484	5	0.7482
0.09	29	0.7962	4	0.6515	29	0.8435	4	0.7220
0.10	27	0.7994	3	0.6393	26	0.8563	4	0.7133
0.11	26	0.7881	3	0.5777	25	0.8493	3	0.6662
0.12	24	0.7911	2	0.5588	25	0.8335	3	0.6159
0.13	22	0.7944	2	0.4976	23	0.8419	2	0.5722
0.14	22	0.7716	2	0.4550	22	0.8295	2	0.5142
0.15	21	0.7601	1	0.3726	21	0.8134	1	0.4442
0.16	20	0.7443	1	0.2919	21	0.7892	1	0.3395
0.17	20	0.7081	1	0.2686	20	0.7908	1	0.3037
0.18	18	0.7148	1	0.3172	19	0.7718	1	0.3343
0.19	18	0.6853	1	0.3957	19	0.7289	1	0.3948
0.20	17	0.6794	1	0.5350	19	0.7101	1	0.5420

Table 6: Relative Efficiency and Optimal Beta for \hat{p} and \hat{p}^* ($s = 10, n = 30$ or $n = 40$)

p	s=15, n=10				s=15, n=20			
	\hat{p}		\hat{p}^*		\hat{p}		\hat{p}^*	
	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.
0.01	166	0.2614	32	0.2520	183	0.4420	35	0.4302
0.02	96	0.4031	17	0.3797	99	0.6063	19	0.5820
0.03	67	0.4855	12	0.4450	75	0.6531	13	0.6142
0.04	52	0.5314	9	0.4726	57	0.7015	10	0.6473
0.05	46	0.5202	7	0.4455	50	0.6994	8	0.6295
0.06	39	0.5451	6	0.4501	42	0.7279	7	0.6381
0.07	35	0.5288	5	0.4137	39	0.7063	6	0.5956
0.08	32	0.5404	4	0.4007	36	0.6999	5	0.5646
0.09	29	0.5231	3	0.3666	34	0.6494	4	0.4882
0.10	26	0.5544	3	0.3658	32	0.6396	3	0.4620
0.11	24	0.5794	2	0.3493	29	0.6438	3	0.4303
0.12	22	0.6071	2	0.3374	27	0.6341	2	0.3886
0.13	20	0.6491	1	0.3532	26	0.6083	2	0.3317
0.14	19	0.7007	1	0.3028	23	0.6407	1	0.3295
0.15	17	0.7618	1	0.3057	22	0.6545	1	0.2657
0.16	16	0.8136	1	0.3649	20	0.6805	1	0.2499
0.17	14	0.8460	1	0.4849	19	0.7300	1	0.3037
0.18	13	0.8455	1	0.6697	17	0.7867	1	0.4314
0.19	12	0.8084	1	0.8878	16	0.8446	1	0.6508
0.20	11	0.7321	1	1.1091	14	0.8895	1	0.9464

Table 7: Relative Efficiency and Optimal Beta for \hat{p} and \hat{p}^* ($s = 15, n = 10$ or $n = 20$)

p	s=15, n=30				s=15, n=40			
	\hat{p}		\hat{p}^*		\hat{p}		\hat{p}^*	
	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.
0.01	187	0.5598	36	0.5483	196	0.6233	38	0.6114
0.02	107	0.6870	20	0.6637	105	0.7609	20	0.7404
0.03	76	0.7451	14	0.7105	77	0.7930	14	0.7637
0.04	59	0.7825	11	0.7366	64	0.8087	12	0.7672
0.05	50	0.7975	9	0.7385	50	0.8429	9	0.7927
0.06	44	0.8062	7	0.7315	45	0.8453	8	0.7830
0.07	40	0.7953	6	0.6991	41	0.8430	6	0.7610
0.08	38	0.7763	5	0.6529	37	0.8393	6	0.7340
0.09	35	0.7672	5	0.6242	35	0.8288	5	0.6999
0.10	33	0.7490	4	0.5706	33	0.8148	4	0.6535
0.11	32	0.7123	3	0.5057	32	0.7810	4	0.5933
0.12	30	0.6848	3	0.4527	31	0.7571	3	0.5242
0.13	28	0.6667	2	0.3929	29	0.7387	2	0.4776
0.14	26	0.6587	2	0.3549	28	0.6909	2	0.3838
0.15	24	0.6730	1	0.3077	26	0.6850	1	0.3560
0.16	23	0.6721	1	0.2505	25	0.6678	1	0.2628
0.17	21	0.6933	1	0.2610	23	0.6891	1	0.2505
0.18	20	0.6994	1	0.3213	22	0.6793	1	0.2891
0.19	19	0.7472	1	0.4849	21	0.7071	1	0.4202
0.20	17	0.7928	1	0.7383	19	0.7400	1	0.6373

Table 8: Relative Efficiency and Optimal Beta for \hat{p} and \hat{p}^* ($s = 15, n = 30$ or $n = 40$)

p	s=20, n=10				s=20, n=20			
	\hat{p}		\hat{p}^*		\hat{p}		\hat{p}^*	
	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.
0.01	177	0.3215	34	0.3114	182	0.5297	35	0.5175
0.02	101	0.4535	19	0.4285	104	0.6577	20	0.6331
0.03	73	0.5030	13	0.4614	77	0.7040	14	0.6655
0.04	57	0.5404	9	0.4816	65	0.6953	11	0.6409
0.05	50	0.5237	8	0.4490	54	0.7164	9	0.6441
0.06	44	0.5139	6	0.4204	48	0.7047	7	0.6118
0.07	39	0.5332	5	0.4169	45	0.6410	6	0.5296
0.08	34	0.5622	4	0.4160	40	0.6426	5	0.5085
0.09	31	0.5931	3	0.4265	37	0.6223	4	0.4665
0.10	27	0.6595	3	0.4434	34	0.6239	4	0.4450
0.11	25	0.7201	2	0.4581	31	0.6317	3	0.4110
0.12	22	0.8022	2	0.4748	28	0.6727	2	0.4118
0.13	19	0.8488	1	0.5114	25	0.7399	2	0.4081
0.14	17	0.8397	1	0.3983	23	0.8009	1	0.4273
0.15	15	0.7692	1	0.3662	21	0.8668	1	0.3709
0.16	14	0.6507	1	0.3950	18	0.9024	1	0.3720
0.17	12	0.5087	1	0.4557	16	0.8762	1	0.4855
0.18	11	0.3758	1	0.5336	15	0.7821	1	0.6392
0.19	10	0.2808	1	0.6077	13	0.6427	1	0.8113
0.20	9	0.2054	1	0.6694	12	0.4975	1	0.9526

Table 9: Relative Efficiency and Optimal Beta for \hat{p} and \hat{p}^* ($s = 20, n = 10$ or $n = 20$)

p	s=20, n=30				s=20, n=40			
	\hat{p}		\hat{p}^*		\hat{p}		\hat{p}^*	
	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.	OPT. Beta	REL. Eff.
0.01	200	0.6139	40	0.6018	195	0.6953	38	0.6845
0.02	107	0.7466	20	0.7251	112	0.7905	21	0.7717
0.03	75	0.7983	14	0.7662	80	0.8321	15	0.8045
0.04	65	0.7940	12	0.7494	67	0.8381	12	0.7985
0.05	56	0.7921	10	0.7319	58	0.8375	10	0.7859
0.06	51	0.7756	8	0.6950	52	0.8303	8	0.7617
0.07	46	0.7608	7	0.6575	48	0.8129	7	0.7273
0.08	42	0.7452	6	0.6202	45	0.7872	6	0.6737
0.09	39	0.7165	5	0.5656	41	0.7796	5	0.6392
0.10	37	0.6661	4	0.4882	40	0.7215	4	0.5573
0.11	34	0.6670	3	0.4614	36	0.7112	4	0.5173
0.12	32	0.6657	3	0.4284	35	0.6555	3	0.4286
0.13	29	0.6832	2	0.3921	32	0.6486	2	0.3944
0.14	27	0.7054	2	0.3777	30	0.6575	2	0.3475
0.15	24	0.7848	1	0.3508	27	0.7044	1	0.3478
0.16	22	0.8491	1	0.3228	24	0.7859	1	0.2990
0.17	19	0.9061	1	0.4007	22	0.8402	1	0.3244
0.18	17	0.9199	1	0.5881	20	0.9107	1	0.4937
0.19	15	0.8615	1	0.8352	18	0.9312	1	0.7738
0.20	14	0.7407	1	1.0998	16	0.8816	1	1.1304

Table 10: Relative Efficiency and Optimal Beta for \hat{p} and \hat{p}^* ($s = 20, n = 30$ or $n = 40$)

Based on Table 3 to Table 10, we can make the following conclusions:

- (1) When the true value of p increases, the optimal β for \hat{p} and \hat{p}^* will be non-increasing;
- (2) For a fixed value of p between 0 and 0.2, the optimal β for \hat{p} will be greater than the optimal β for \hat{p}^* . Thus, when the practitioners want to choose smaller β , the restricted p would be preferred;
- (3) For a fixed value of p between 0 and 0.2, the relative efficiency are smaller than one, except for the combination that $s = 20, n = 30$ and $s = 20, n = 40$ when true value of p is 0.2. The MSE of the \hat{p} and \hat{p}^* are smaller than that of Burrows' estimator for the other combinations;
- (4) Most of the cases, \hat{p}^* has better performance than \hat{p} when comparing the MSE. The relative efficiency of unrestricted p is greater than that of the restricted p .

5. Application and Comparison on Hepatitis C Screening

Hepatitis C (HCV) is a viral infection that affects liver and it can begin as an acute infection or a short-term illness that occurs within six months after someone is infected. Not everyone with the disease exhibits symptoms. In most cases, hepatitis C becomes a long-term or chronic infection. Chronic infection occurs in about 75%-80% of those infected. An estimated 3.2 million people in the United States are living with chronic hepatitis C infection, and most do not feel ill or know they are infected, according to the Centers for Disease Control and Prevention (CDC). There are approximately 17,000 new hepatitis C cases each year in the United States, many of them go unreported since people often do not have symptoms and are not diagnosed.

Mohd et al.(2013) have concluded that most recent estimates of disease burden show an increase in seroprevalence over the last 15 years to 2.8%. Since the costs associated with testing for HCV can be very high, group testing will be a better choice for HCV screening experiments, especially in developing nations. Neill and Conradie (1992, 1994) proposed the use of group testing to screen for the prevalence of HCV and conduct sensitivity and specificity analyses for standard ELISA tests. The first published HCV group-testing application was provided by Liu et al. (1997). Joshua et al. (2006) re-used their data and got the MLE (\hat{p}_{MLE}) and Empirical Bayesian (\hat{p}_{EB}) estimates summarized in Table 11.

	Individual Sera (s=1)	Pooled Sera (s=5)
Number of Pools (n)	1875	375
Positive Pools	42	37
Estimate of p	0.0224	$\hat{p}_{MLE}=0.020562$ $\hat{p}_{EB}=0.020557$

Table 11: Anti-HCV data from Liu et al. (1997) with $s = 1$ and $s = 5$

Before we apply the new method to get the estimate when p is restricted to $(0, 0.2]$, we need to find the optimal β for unrestricted p and restricted p . In the past few years the seroprevalence in the world is around 2.8%, we can consider the true p is between 0.02 and 0.03. Since $s = 5$, $n = 375$ is out of the range of the table provided in the previous section of our study, the same method as before is tried to find the optimal β for cases of $p = 0.02$ and 0.03 . Then we can obtain the

Bayesian estimates for both the unrestricted regular p and restricted p . According to our numerical results in the case of p defined on $(0, 1)$ the minimum MSE can be achieved when $\beta = 100$, whereas the best $\beta = 23$ in the case of p restricted on $(0, 0.2]$ if the true p is 0.02. On the other hand if true p is 0.03, then in the case of p defined on $(0, 1)$ the minimum MSE can be achieved when $\beta = 77$, and the best $\beta = 14$ in the case of p restricted on $(0, 0.2]$. On the basis of the choices of the best β we can get the estimates for both the unrestricted regular p and restricted p shown in Table 12.

	p=0.02	p=0.03
Regular p	0.01999509 ($\beta = 100$)	0.02024025 ($\beta = 77$)
Restricted p	0.01981248 ($\beta = 23$)	0.02035439 ($\beta = 14$)
pMLE	0.02056169	0.02056169
pBurrows	0.02053884	0.02053884

Table 12: Anti-HCV data from Liu et al. (1997) with $s = 5$

V. Conclusion and Possible Future Work

In our research, a new Bayesian estimator has been proposed to deal with the binomial group testing estimation. Usually, in group testing application, experimenters lack the freedom to choose the sample size and the number of groups because of cost or other restrictions. Thus, the choice of hyper-parameter β would be very important. On the basis of heavy Monte Carlo simulation we have determined the best hyper-parameters in the sense that the corresponding new Bayes estimator has the smallest MSE. A table of these best hyper-parameters is made for proportions within the considered range $((0, 0.2])$. The table is useful for practitioners to select appropriate hyper-parameters when the binomial group testing method is applied in their research work.

One important possible future work is to consider the Bayes risk of the proposed estimator. In this research, we only focused on minimizing the MSE of the estimator, the comparison of Bayes risk of \hat{p} and \hat{p}^* has not been discussed. Another possible future work would be interval estimation of p base on the posterior distribution of \hat{p}^* . Finally, in this research, we considered only the square-error loss function, we may use other loss functions in the future study.

REFERENCES

- Bilder, C. R., Tebbs, J. M., (2005). Empirical Bayesian Estimation of the Disease Transmission Probability in Multiple-Vector- Transfer Designs. *Biometrical Journal*, v.47, no. 4, p. 502-516.
- Burrows, P. M., (1987). Improved Estimation of Pathogen Transmission Rates by Group Testing. *Phytopathology*, v.77, p, 363-365.
- Chaubey, Y., Li, W., (1995). Comparison Between Maximum Likelihood and Bayes Methods for Estimation of Binomial Probability with Sample Compositing. *Journal of Official Statistics*, v.11, p.379-390.
- Chen, C., Swallow, W., (1990). Using Group Testing to Estimate a Proportion and to Test the Binomial Model. *Biometrics* v. 46, p. 1035-1046.
- Chen, C., Swallow, W., (1995). Sensitivity Analysis of Variable-size Group Testing and its Related Continuous Models. *Biometrical Journal*, v. 37, p. 173-181.
- Chick, S., (1996). Bayesian Models for Limiting Dilution Assay and Group Test Data. *Biometrics*, v. 52, p. 1055-1062.
- Christopher R. Bilder, Joshua M. Tebbss, (2005). Empirical Bayesian Estimation of the Disease Transmission Probability in Multiple-Vector-Transfer Designs. *Biometrical Journal*, v. 47, p. 502-516.
- Dorfman, R., (1943). The Detection of Defective Members of Large Populations. *The Annals of Mathematical Statistics*, v.14, no. 4, p. 436-440.
- Fang, Xiang, Stroup, Walter W. and Zhang, Shunpu, (2007). Improved Empirical Bayes Estimation in Group Testing Procedure for Small Proportions. *Communications in Statistics: Theory and Methods*, 36, 2937-2944
- Hepworth, G., (2005). Confidence Intervals for Proportions Estimated by Group Testing with Groups of Unequal Size. *Journal of Agricultural, Biological, and Environmental Statistics*, v.10, no. 4, p. 478-497.
- Huang, M., Swallow, W. H., (1999). Robustness of Group Testing in the Estimation of Proportions. *Biometrics*, v. 55, p. 231-237.
- Hughes-Oliver, J., Swallow, W., (1994). A Two-stage Adaptive Procedure for Estimating Small Proportions. *Journal of the American Statistical Association*, v. 89, p. 982-993.

- Joshua M. Tebbs, Christopher R. Bilder, Barry K. Moser (2003). An Empirical Bayes Group-Testing Approach to Estimating Small Proportions. *Communications in Statistics - Theory and Methods*, 32:5, 983-995
- Kline, R., Brothers, T., Brookmeyer, R., Zeger, S., Quinn, T. (1989). Evaluation of HIV seroprevalence in population surveys using pooled sera. *Journal of Clinical Microbiology* 27:1449-1452.
- Mohd Hanafiah K, Groeger J, Flaxman AD, Wiersma ST., (2013) Global Epidemiology of Hepatitis C Virus Infection: New Estimates of Age-specific Antibody to HCV Seroprevalence. *Hepatology*; 57:1333-1342.
- Monzon, O., Paladin, F., Dimaandal, E., Balis, E. (1992). Relevance of antibody content and test format in HIV testing of pooled sera. *AIDS* 6:43-48.
- Murrall, D., Nault, L., Hoy, C., Madden, L., and Miller, S., (1996). Effects of temperature and vector age on transmission of two Ohio strains of aster yellow phytoplasma by the aster leafhopper. *Journal of Economic Entomology* 89, 1223-1232.
- Ornaghi, J., March, G., Boito, G., Marinelli, A., Beviacqua, J., Giuggia, J., and Leonardon, S., (1999). Infectivity in natural populations of *Delphacodes kuscheli* vector of Mal Rio Cuarto virus. *Maydica* 44, 219-223.
- Neill, A., Conradie, J. (1992). Low cost anti-HCV screening of blood donors. *Lancet* 340:1096.
- Neill, A., Conradie, J. (1994). Pool tests for hepatitis C virus. *Lancet* 343:738.
- Romanow, L., Moyer, J., and Kennedy, G., (1986). Alteration of efficiencies of acquisition and inoculation of watermelon mosaic virus 2 by plant resistance to the virus and to an aphid vector. *Phytopathology* 76, 1276-1281.
- Swallow, W., (1985). Group Testing for Estimating Infection Rates and Probabilities of Disease Transmission. *Phytopathology*, v. 75, p. 882-889.
- Swallow, W., 1987: Relative mean squared error and cost considerations in choosing group size for group testing to estimate infection rates and probabilities of disease transmission. *Phytopathology* 77, 1376-1381.
- Tebbs, J. M., Bilder, C. R., Moser, B.K., (2003). An Empirical Bayes Group Testing Approach to Estimating Small Proportions. *Communications in Statistics - Theory and Methods*, v. 32, no. 5, p. 983-995.

Xiang Fang, Walter W. Stroup, Shunpu Zhang, (2007). Improved Empirical Bayes Estimation in Group Testing Procedure for Small Proportions. *Communications in Statistics - Theory and Methods*, v. 36, p. 2937-2944.