Cost And Accuracy Analysis Of Group And Individual Testing Strategies: Implications For COVID-19

COST AND ACCURACY ANALYSIS OF GROUP AND INDIVIDUAL TESTING STRATEGIES: IMPLICATIONS FOR COVID-19

By

Ismat Binta Islam (B.Sc. and M.S.)

A THESIS

SUBMITTED TO THE DEPARTMENT OF MATHEMATICS & STATISTICS AND

THE SCHOOL OF GRADUATE STUDIES

MCMASTER UNIVERSITY

PARTIAL FULFILMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

Master of Science (2021) McMaster University

(Mathematics & Statistics) Hamilton, Ontario, Canada

TITLE: Cost and accuracy analysis of group and individual testing strategies: Implications for COVID-19.

AUTHOR: Ismat Binta Islam

B.Sc. & M.S. (University of Chittagong, Bangladesh)

SUPERVISOR: Dr. Stephen Walter

NUMBER OF PAGES: xv, 97

To my family and my ex-husband

ABSTRACT

Background

We compared several group and individual testing strategies in terms of cost and accuracy and then showed which one is more accurate while costing as little as possible for a specified prevalence rate.

Methods

We designed and compared four protocols: 1) Protocol I: Group (first stage) and individual (second stage) testing using an RT-PCR test and a rapid antigen test respectively with and without a dilution effect; 2) Protocol II: Group (first stage) and individual (second stage) testing using only the RT-PCR test with and without a dilution effect; 3) Protocol III(a): Two-stage individual testing using an RT-PCR test and a rapid antigen test respectively; Protocol III(b): Two-stage individual testing using only RT-PCR test; and 4) Protocol IV(a): One stage individual testing with an RT-PCR test; Protocol IV(b): One stage individual testing with a rapid antigen test. We minimized the expected cost of group testing protocols with the optimal group size and estimated the total expected cost of all individual testing protocols for several prevalence rates. Group testing strategies are also compared with the individual testing strategies. We estimated the expected number of false-negative errors for each protocol, either assuming perfect sensitivity or allowing sensitivity to depend on the group size for group testing or assuming imperfect sensitivity for individual testing. Finally, the relationship between the cost and the number of false-negative cases is examined for each protocol based on lower and higher disease prevalence. We illustrate these ideas using testing costs associated with screening for COVID-19,

Results

Our results show that protocol III(b) is the most expensive when $C_1 = 120, 85, 50$ for p = 0.1-5% and $C_1 = 120, 85$ for p = 7-10% and $C_1 = 120$ for p = 15-20%. Protocol I with the dilution effect is the cheapest when p = 0.1-1% for all the cost ratios and when p = 3% and cost ratio is 10. Protocol IV(b) becomes the cheapest for $C_2 = 5$ when p = 5-15% and for $C_2 = 25, 5$ when p = 20%. Relative cost results reveal that protocol I is affordable than protocol III(a) and IV(a) for all scenarios. Protocol II is affordable than protocol III(b) for all scenarios and IV(a) in most cases. However, while comparing with protocol IV(b), protocols I and II become less

affordable in most scenarios except for the lowest prevalences. Protocol II without dilution produces the minimum false-negative test results (26.1-4571.4) per million. Conversely, protocol I with dilution creates the highest false-negative errors (338.8-17870.7) per million. Finally, the trade-off between the cost and accuracy indicates that without dilution protocol II is comparatively more accurate and cheaper for both lower and higher prevalence. While, if dilution exits then protocol IV(b) becomes less costly and produces comparatively fewer errors for any disease prevalence.

Conclusion

Our investigations can assist policymakers in selecting an appropriate protocol in terms of cost or accuracy or both. If minimizing cost is the priority, then protocol I with dilution can be applied for the lowest prevalences. If prevalence becomes higher, then protocol IV(b) can be used. If having greater accuracy is the main concern, then one can use protocol II without dilution for all prevalence rates. If both are the main priority at the same time, then protocol II is the best choice for all prevalences when we do not allow dilution. Whereas, if we allow dilution, it makes protocol IV(b) a relatively better option to use for any prevalences. On the other hand, protocol III(b) or protocol I with dilution are the least recommended strategies in terms of cost or accuracy. However, our methods are general, and so they can potentially be applied to other disease screening situations.

ACKNOWLEDGEMENT

First and foremost, all the praises and gratefulness to the Almighty Allah for giving me strength, patience, and his shower of blessings throughout my research work to complete this dissertation.

I would convey my enormous indebtedness, sincere thanks and bottomless appreciation and regard to my research supervisor, Dr. Stephen Walter, for his continuous inspiration, scholastic guidance, constructive criticism, tireless patience, and necessary support throughout the progress of this thesis. It is only his kind supervision and proper guidance that made it to submit this dissertation.

I am thankful to Dr. Ben Bolker and Dr. Shui Feng for being members of my defence committee and for providing worthwhile comments.

I must express my very profound gratitude to my family and my ex-husband for providing me with tireless support and continuous encouragement throughout my study and the process of researching and writing this thesis. Certainly, this accomplishment would not have been possible without them.

I am also grateful to my well-wishers, my friends and all others who have helped me directly or indirectly in my study and research work.

Finally, a very special acknowledgement to “GOOGLE” without which it would be difficult to do this research work in time.

NOTATION

N = Total population

n = Group Size

$m = \frac{N}{n}$ = Number of groups

$p \equiv$The probability of an individual having a disease of interest or the prevalence rate.

Se = Sensitivity

Sp = Specificity

TP = True Positive

FP = False Positive

TN = True Negative

FN = False Negative

EFN = Expected number of false negative

HIV-1 = Human Immunodeficiency Virus – 1

HCV = Hepatitis C Virus

RT-PCR = Real-Time Reverse Transcription-Polymerase Chain Reaction

COVID = Coronavirus Disease

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Group Testing

Screening all the individuals with a binary trait of interest in a large community can be both costly and time-consuming at the same time. This whole laborious process can be lessened to a great extent by examining a sample or group of the population altogether with the help of a single test. If the whole group is tested negative, then those people in that group can be easily excluded from the community for further screening process which minimizes the cost of the whole process. On the contrary, if the single test turns out to be positive, then examining the members of that group continues at the further stage. This process is known as the group (or pooled) testing process. This process is in need when the objective is to identify and eliminate the defective units or members from a large population when the proportion of being a defective item in a population is relatively low. However, group testing is also used when the prevalence rate is high if the main concern is to save expense and time. Another application of this process is to estimate the proportion of defective items or members in a sample of the population. This procedure was first proposed by Dorfman in 1943 to identify World War II soldiers who were affected by syphilis.

Testing of pooled samples (e.g., blood, swab, urine etc.) through this process commonly used in many different areas. It has been widely used in solving blood bank screening problems for example applying informative retesting procedures for chlamydia and gonorrhea testing for Infertility Prevention Project (Bilder et al., 2010), screening blood donors to detect HIV-1 and HCV virus (Stramer et al., 2004), screening blood and plasma for detecting HCV and HIV-1 (Tabor & Epstein, 2002) and many more.

Group testing has its application in other fields as well for instance discovering new drugs (Salzer et al., 2016), screening West Nile Virus in mosquitoes by using RT-PCR test (Khan et al., 2017), screening DNA chips (Schliep et al., 2003), detecting potato seed virus without (Chiang et al., 2010) and with (S. C. Liu et al., 2011) classification errors, detecting several viruses in pooled bovine milk samples by using ELISA (Græsbøll et al., 2017), demonstrating statistical approaches for detecting individuals who have rare attributes (Venette et al., 2002) and so on.

In general, there are mainly two purposes of the implementation of group testing which are case identification or classification and estimation of proportion. This thesis will mainly focus

on group testing applications in screening the disease. Identification problem was the main goal in Dorfman's group testing algorithm. After that, group testing literature has been covered with countless papers based on identification algorithms varying with numerous complex levels. For example, derivation and comparison of two identification (Hierarchical and Square Array Bases) algorithms (Kim et al., 2007), informative retesting algorithm incorporating the information about all the covariates (Bilder et al., 2010), new informative Dorfman's screening heterogeneous population algorithm (McMahan et al., 2012), an algorithm which is established on the concept of a hypercube in geometry to identify SARS-COV2 infected individuals (Mutesa et al., 2021). There are also several papers related to the estimation of the proportion of the positive individuals in group testing literature. Some of those are: estimating the prevalence of infection in population by testing unequal-sized groups (Walter et al., 1980), estimating the proportion by retesting the groups which were tested positive at the first stage (Hepworth & Walter, 2020) and so forth.

One of the fundamental attributes of group testing is the dilution effect. While forming a group, positive specimens can be diluted with several negative specimens. Because of this dilution,the loss of sensitivity happens which is known as the dilution effect. Dorfman in his procedure did not include the dilution effect. Hwang (1976) included the dilution effect in Dorfman's procedure to estimate the expected cost of group testing after determining the best possible group size. Several papers can be found based on this dilution effect such as screening blood for HIV by considering dilution effect in pools (Wein & Zenios, 1996), a binary regression model including the dilution effect was estimated for both identification and estimation purposes (Mokalled et al., 2021).

In most of the applications of group testing procedure, it is assumed that identifying the status of the specimens are performed without any classification errors. However, in practice, the test can show false results such as identifying positive items as negative or negative items as positive. Consequently, the efficiency of the results of group testing reduces. And the overall cost of group testing reduces or increases to some extent because of false negative or false positive errors respectively. Therefore, many authors started to research group testing policies with classification errors. For example, an extension of Dorfman's procedure considering the classification errors (Graff & Roeloffs, 1972), classification errors depending on the proportion of the faulty samples in a pool (Burns & Mauro, 1987), estimating the prevalence by obtaining the several optimality properties of Dorman's procedure with the presence of test errors (A. Liu et al., 2012).

Most researchers (Malinovsky et al., 2016), (Aprahamian et al., 2019) assumed that the values of test errors do not depend on the group size. In other words, misclassification is defined only for a single or individual test. However, this assumption is not generally reasonablein practice. Haber et al. (2021) showed how this imperfect assumption affects the efficiency of the group testing design and applied a sensitivity function depending on the group size to estimate the expected number of tests required for group testing design.

Group testing becomes worldwide popular because of its cost-effective benefits. The number of tests required for this procedure is usually less compared to the number of tests required for individual testing. As such, it leads to the reduction of test materials to a great extent and thus reduces the cost associated with test resources. An efficient group testing design is one in which minimum expected cost per unit information can be attained with the optimal group size. For example, Turner et al. (2009) formulated such kind of non-linear expected cost model of group testing technique incorporating the pooling and testing costs. They showed how expected costs can be minimized with the optimal group size. Huang et al. (2020) developed an optimal group testing design for estimation purposes allowing different costs and test errors varying on group size. Zhang et al. (2021) proposed a group testing design to estimate the values of biomarkers using prospective-retrospective studies which have higher efficiency in terms of cost than random sampling design.

1.2 SARS-CoV-2

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causes coronavirus (COVID) disease which was first discovered in Wuhan, China in December 2019. This disease can easily be transmitted to people breathing in air contaminated by droplets and infectious particles of this virus. It can also be spread by infected surfaces and liquids. Within a short time, COVID has spread all over the world and has been declared as a pandemic by World Health Organization (WHO) in March 2020 (Wan, 2020). As such, the Covid pandemic adversely has affected the health care and economic system across the globe, especially in developing countries. Since this virus is spreading rapidly, it has led to a shortage of test resources worldwide.

To control the spreading with the limited test supplies, group testing design requiring a reduced number of tests draws the attention of the medical officials for a large-scale screening program. Therefore, a vast amount of research in group testing for screening COVID is going on such as

suggesting a group testing strategy (Bilder et al., 2020) for screening coronavirus in asymptomatic people (Lohse et al., 2020), recommending a rapid near-point-of-careassay for screening program (Becker et al., 2020).

It is also important to check the accuracy of these diagnostic tests. Inaccurately diagnosing COVID makes the whole pandemic situation worse (Woloshin et al., 2020). In most places, the RT-PCR test is used for screening SARS-Cov-2. The accuracy of this test is discussed based on real-life data by Kortela et al. (2021). Nowadays, to get the test results more rapidly at a lower cost (Pilarowski et al., 2021), rapid antigen test has been authorized by Food and Drug Administration (Prince-Guerra et al., 2021)..

1.3 Thesis Structure

In chapter 2, we will consider four protocols in which the first two are group testing protocols with and without dilution effect and the last two are individual testing protocols with different combinations of test kits at both or single stages; then for each group testing protocol, optimal group sizes will be determined according to the several cost parameters; next, the minimized expected cost per person of each group testing protocols will be estimated based on those optimal group sizes; after that total expected cost per person for all types of individuals testing protocols will be estimated; at the end of this chapter relative cost of group testing protocols compared to the individual testing protocols will be discussed. In chapter 3, at first expected number of false-negative test results caused by each protocol will be estimated; then we will describe the number of expected false-negative errors against the expected cost per individual for each protocol according to the different prevalence rates. At the end of this chapter, we will demonstrate which protocol is causing fewer false-negative cases with the low expected cost for higher or lower prevalence rates. At last, in chapter 4, recommendations to a decision-maker based on the overall results for several prevalence rates will be suggested; lastly, this thesis will be concluded with the limitations along with future works of this study.

Chapter 2 Evaluation of the Cost of Group and Individual Testing Protocols

2.1 Identification

Group testing is a technique in which several items combined as a group are tested jointly (Bilder, 2019) to identify the items as positive or negative of interest, based on the testing outcomes (McMahan et al., 2012). This technique is also used to estimate the proportion of positive items in a sample population (Hepworth & Walter, 2020). Hence, there are two main topics of group testing: Case Identification and Proportion Estimation. In this chapter, we will mainly focus on the identification problem with testing carried out in two stages.

Dorfman, 1943 first applied group testing procedure to screen United States soldiers infected by syphilis during World War II for cost-efficiency purposes. Dorfman's program was conducted in two stages. In the first stage, all the non-overlapping groups were tested. If a group was tested as negative at the first stage, then it was assumed that all the soldiers in that group were negative. If a group was tested as positive, then all soldiers in that group were again tested individually at the second stage to detect which soldiers were infected. A group is labelled as positive if at least one member of that group is positive. On the contrary, a group is labelled as negative if none of the members of that group is positive.

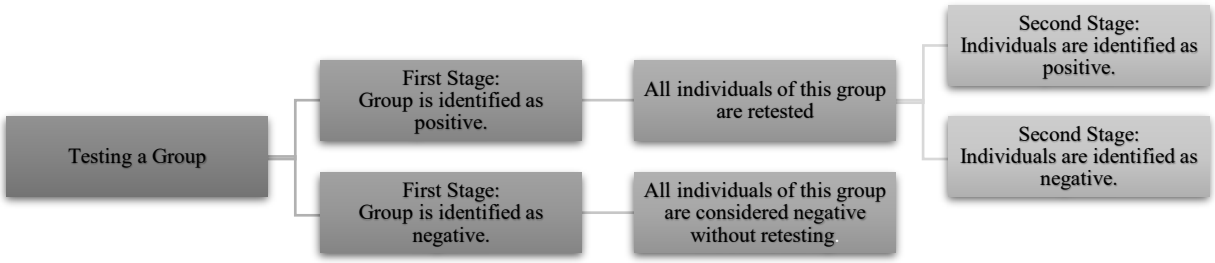


Figure 2. 1: Group testing design for identification with two stages

2.2 Notations and Definitions

Let us define,

N = Total population

n = Group Size

$m = \frac{N}{n}$ = Number of groups

$p \equiv$ The probability of an individual having a disease of interest or the prevalence rate.

$1 - p \equiv$ The probability of an individual not having a disease of interest.

$(1 - p)^n \equiv$The probability of having a negative group in the population.

$p' = (1 - (1 - p)^n) \equiv$ The probability of having a positive group in the population.

AsHaber et al (2021), let us assume that each individual has the same probability of getting infected or the same prevalence rate, p. Also, let us assume, $\tilde{x_i}, i = 1,2, \ldots \ldots \ldots, N;$ be a random variable that represents the presence or absence of disease for ith individual.

In other words, $\tilde{x_i} = \begin{cases} 1; if\ the\ ith\ individual\ is\ positive \\ 0; otherwise \end{cases}$.

x_i is the observed value of $\tilde{x_i} \sim Bernoulli(1, p)$ from an individual test. Here, each individual is independent of each other. A random sample from the population is assembled independently to form a group.

Let $\widetilde{y^{(n)}}$ be a random variable that takes values $= \begin{cases} 1; if\ the\ group\ of\ size\ n\ is\ labelled\ as\ positive \\ 0; otherwise \end{cases}$.

$y^{(n)}$ is the observed value of $\widetilde{y^{(n)}} \sim Bernoulli(1, 1 - (1 - p)^n)$ from a group test.

Graff & Roeloffs (1972) extended Dorfman's procedure of group testing in the presence of classification error to minimize the total cost. This test error occurs when the result of either a group or an individual test is not identical with the true state of that group or individual (Graff & Roeloffs, 1972). There are two types of misclassification errors: False Negatives = (1 – Sensitivity) and False Positives = (1 – Specificity) (S. C. Liu et al., 2011). When these errors depend on the group size, these are called differential misclassifications while if these errors do not depend on the group size, then they are known as non-differential misclassifications (Haber et al., 2021).

Sensitivity for an individual test is defined as

$Se = True\ Positive = \Pr(Positive\ test|\ An\ individual\ is\ diseased)$ (Cowling et al., 1999)

$$= \Pr\ (x_i = 1|\widetilde{x_i} = 1)$$

And specificity for an individual test is defined as

$Sp = True\ Negative = \Pr(Negative\ test|\ An\ idividual\ is\ non - diseased)$ (Cowling et al., 1999)

$$= \Pr\ (x_i = 0|\widetilde{x_i} = 0)$$

Also, for a group test with group size n, sensitivity is defined as

$Se(n) = \Pr(Positive\ Test|\ A\ group\ is\ positive) = \Pr\ (y^{(n)} = 1|\widetilde{y^{(n)}} = 1)$ (Haber et al., 2021)

And specificity is defined as,

$Sp(n) = \Pr(Negative\ test|\ A\ group\ is\ negative) = \Pr\ (y^{(n)} = 0|\widetilde{y^{(n)}} = 0)$ (Haber et al., 2021)

Table 2. 1: In the presence of misclassifications, joint probabilities of an individual test

		Test Result	
		Positive	**Negative**
True Status	**Positive**	$Se.\Pr(\widetilde{x_i} = 1)$ (True Positive)	$(1 - Se)\ \Pr(\widetilde{x_i} = 1)$ (False Negative)
	Negative	$(1 - Sp)\ \Pr(\widetilde{x_i} = 0)$ (False Positive)	$Sp.\Pr(\widetilde{x_i} = 0)$ (True Negative)

Table 2. 2: In the presence of misclassifications, joint probabilities of a group test

		Test Result	
		Positive	**Negative**
True Status	**Positive**	$Se(n)\ \Pr\left(\widetilde{y^{(n)}} = 1\right)$ (True Positive)	$(1 - Se(n))\ \Pr\left(\widetilde{y^{(n)}} = 1\right)$ (False Negative)
	Negative	$(1 - Sp(n))\ \Pr(\widetilde{y^{(n)}} = 0)$ (False Positive)	$Sp(n)\ \Pr\left(\widetilde{y^{(n)}} = 0\right)$ (True Negative)

The probability of an individual tested as a positive is defined by adding two possibilities: a) if an individual is truly positive and also detected as positive b) if an individual is truly negative but incorrectly detected as positive. Therefore, with the help of the definitions and Table 2.1, the probability of a positive individual test is,

$\Pr(x_i = 1) = Se.\Pr(\tilde{x_i} = 1) + (1 - Sp)\ \Pr(\tilde{x_i} = 0)$ [From Table 2.1]

$= Se[p^1(1-p)^{1-1}] + (1 - Sp)[p^0(1-p)^{1-0}]$

$= Se.p\ +\ \left(1 - S_p\right)(1-p)$

…….. (1)

Therefore, $x_i \sim Bernoulli(Se.p\ +\ \left(1 - S_p\right)(1-p))$

Similarly, the probability of a group tested as a positive is defined as

$\Pr\left(y^{(n)} = 1\right) = Se(n)\ \Pr\left(\widetilde{y^{(n)}} = 1\right) +\ (1 - Sp(n))\ \Pr(\widetilde{y^{(n)}} = 0)$ [From Table 2.2] (Haber et al., 2021)

$= Se(n)[\{1 - (1-p)^n\}^1\{1 - 1 + (1-p)^n\}^{1-1}] + \left(1 - Sp(n)\right)[\{1 - (1-p)^n\}^0\{1 - 1 + (1-p)^n\}^{1-0}]$

$= Se(n)(\ 1 - (1-p)^n) +\ \left(1 - Sp(n)\right)(1-p)^n$ (Haber et al., 2021)

…….(2)

Therefore, $y^{(n)} \sim Bernoulli(Se(n)(\ 1 - (1-p)^n) +\ \left(1 - Sp(n)\right)(1-p)^n)$ (Haber et al., 2021)

2.3 Dorfman's Procedure and Several Cost Functions

Most of the group-testing literature focused on Dorfman's procedure over more complicated alternative designs because of the simplicity of the two-stage design. Moreover, the execution of this design is relatively straightforward, and researchers can easily understand the concept. Following Dorfman's procedure, Haber et al. (2021) also showed the expected number of tests (denoted by T) per person, which can be minimized with the choice of group size n written as:

$$E\left(T\middle|p,n,Se(n)\right)=\frac{1}{n}\left[\Pr\left(y^{(n)}=0\right)+\Pr\left(y^{(n)}=1\right)+nPr\left(y^{(n)}=1\right)\right]$$

…….(3)

It means if the group test is negative ($\Pr\left(y^{(n)}=0\right)$), only one test is required. And if the group test is positive ($i.e.,\Pr\left(y^{(n)}=1\right)$) at the first stage, then all the individuals of that group which is 'n' need to be retested at the second stage ($i.e.,nPr\left(y^{(n)}=1\right)$). Hence, $(n+1)$ tests are required for the positive group. In this way, to diagnose individuals we can reduce the number of tests to a great extent.

When the prevalence of having a disease or a defective item is expected to be low or the testing resources are limited, then testing one by one to screen an infection in a community is very ineffective and expensive. In such cases, group testing is a preferable choice to apply because of its cost-effectiveness. Since group testing requires the lowest possible number of tests, it results in the reduction of the overall cost. There are several costs related to group testing such as sampling cost, pooling cost, testing costs, personnel costs etcetera. To begin with, we will consider only costs for testing to diagnose the disease. In this thesis, we have used the numerical values of the testing costs related to diagnosing severe acute respiratory syndrome coronavirus 2 (SARS – COV-2) as an example for illustrative purposes. Besides, we could apply this general methodology to other diseases as well.

Here, our primary focus is to find the optimal group size 'n' in different group testing procedures so that the expected cost per person incorporating testing costs for both stages is minimized. Optimum values of n may vary for different values of p and cost components.

To incorporate the costs, let us consider the following cost parameters:

C_1 = The entire (lab and test) cost of testing a group of samples using a real-time reverse transcription-polymerase chain reaction (RT-PCR) test.

C_2 = The cost of testing (without the lab cost) an individual using a rapid antigen test.

C_3 = The entire (lab and test) cost of testing an individual using an RT-PCR test.

Cost ratio, $r=\frac{C_1}{C_2}$ or $\frac{C_1}{C_3}$.

2.3.1 Protocol I

According to the guideline of the Department of Elementary and Secondary Education (DESE), all k-12 public schools in Massachusetts were following a group testing strategy to diagnose the coronavirus as a cost-saving tactic preparing for the start of the 2020-2021 school year funded by the state (Wu, 2021). In this strategy, samples are mixed in groups or pools at schools and then pooled samples are tested by RT-PCR test at the laboratory (Johnston et al., 2021). If a group is positive then all the members of that group are individually retested with BinaxNOW rapid antigen test at school (Johnston et al., 2021).

Now, the overall expected cost per individual associated with the expected number of tests required in the school program strategy following the equation (3) is given by,

$$E\big(C|p,n,Se(n)\big) = \frac{1}{n}[C_1 \Pr\big(y^{(n)}=0\big) + C_1 \Pr\big(y^{(n)}=1\big) + C_2 nPr\big(y^{(n)}=1\big)]$$

$$= \frac{1}{n}[C_1(\Pr\big(y^{(n)}=0\big) + \Pr\big(y^{(n)}=1\big)) + C_2 nPr\big(y^{(n)}=1\big)$$

$$= \tfrac{1}{n}[C_1 + C_2 nPr\big(y^{(n)}=1\big)] \text{ since } \Pr\big(y^{(n)}=0\big) + \Pr\big(y^{(n)}=1\big) = 1$$

$$= \tfrac{C_1}{n} + C_2 Pr\big(y^{(n)}=1\big)$$

$$= \tfrac{C_1}{n} + C_2[\, Se(n)(\, 1-(1-p)^n) + \big(1-Sp(n)\big)(1-p)^n] \text{ [From equation (2)]}$$

For simplicity, we will assume 100% specificity or $Sp(n)$ = 1 in all situations. However, this assumption may not be true for all situations. Since we are taking into account the differential misclassifications, our sensitivity function depending on the group size can be written as $Se(n) = \frac{p}{1-(1-p)^{n^d}}$ where d is the dilution parameter ranging from 0 to 1 (Hwang, 1976). When d=0, $Se(n)$ becomes 1 (perfect) which means no false negative cases occur at the first stage which is considered as a special case.

Thus, the overall expected cost per person can be written as,

$$E\big(C|p,n,Se(n)\big) = \tfrac{C_1}{n} + C_2[(\tfrac{p}{1-(1-p)^{n^d}})(\, 1-(1-p)^n) + 0] \text{ since } Sp(n) = 1$$

$$= \tfrac{C_1}{n} + C_2[(\tfrac{p}{1-(1-p)^{n^d}})(\, 1-(1-p)^n)]$$

$$= C_2[\frac{C_1}{C_2 n} + (\frac{p}{1-(1-p)^{n^d}})(1-(1-p)^n)]$$

……..(4)

This cost function can be optimized in terms of n for various values of d, p and $r = \frac{C_1}{C_2}$.

For the special case (d=0), the overall expected cost per individual becomes,

$$E(C|p,n,Se(n)=1) = C_2[\frac{C_1}{C_2 n} + (\frac{p}{1-(1-p)^{n^0}})(1-(1-p)^n)]$$

$$= C_2[\frac{C_1}{C_2 n} + (\frac{p}{1-1+p})(1-(1-p)^n)]$$

$$= C_2[\frac{C_1}{C_2 n} + (1-(1-p)^n)]$$

..……(5)

2.3.2 Protocol II

In general, the RT-PCR test is used both at the first stage for group testing and the second stage for individual testing. This is because the RT-PCR test is considered to be the gold or reference standard test because of its excellent sensitivity and specificity (Esbin et al., 2020). Hence, in this situation, the overall expected cost per individual can be written as,

$$E\big(C|p,n,Se(n)\big) = \frac{1}{n}[C_1 \Pr\big(y^{(n)}=0\big) + C_1 \Pr\big(y^{(n)}=1\big) + C_3 n Pr\big(y^{(n)}=1\big)]$$

$$= \frac{1}{n}[C_1 + C_3 n Pr\big(y^{(n)}=1\big)]$$

$$= \frac{C_1}{n} + C_3 Pr\big(y^{(n)}=1\big)$$

$$= \frac{C_1}{n} + C_3[\,Se(n)(1-(1-p)^n) + \big(1-Sp(n)\big)(1-p)^n] \text{ [From equation (2)]}$$

……(6)

$$= \frac{C_1}{n} + C_3[(\frac{p}{1-(1-p)^{n^d}})(1-(1-p)^n) + 0] \quad \text{since } Sp(n) = 1$$

$$= \frac{C_1}{n} + C_3[(\frac{p}{1-(1-p)^{n^d}})(1-(1-p)^n)]$$

$$= C_3[\frac{C_1}{C_3 n} + (\frac{p}{1-(1-p)^{n^d}})(1-(1-p)^n)]$$

which is similar to equation 4 with the different cost parameter C_3 at the second stage.

……..(7)

For the special case (d=0), this expression becomes the same as equation 5 with the cost parameter C_3 instead of C_2 which is

$$E(C|p,n,Se(n)=1) = C_3[\frac{C_1}{C_3 n} + (1-(1-p)^n)]$$

Let us consider, as a special case, that the cost for individual testing using RT-PCR test is the same as the cost for group testing using the same test i.e., $C_1 = C_3$. Therefore, equation (7) with and without dilution effect becomes,

$$E\big(C|p,n,Se(n)\big) = C_1[\frac{1}{n} + (\frac{p}{1-(1-p)^{n^d}})(1-(1-p)^n)] \quad \text{…….(8)}$$

$$E(C|p,n,Se(n)=1) = C_1[\frac{1}{n} + (1-(1-p)^n)] \quad \text{……..(9)}$$

2.3.3 Protocol III

Generally, for screening, individuals can be tested one by one in which the number of tests equals the total number of individuals. When the disease prevalence rate is high, it is reasonable to use individual testing than group testing. Let us consider a situation in which at the first stage all the individuals are tested with the RT-PCR test. Those individuals who are tested positive are again tested individually for reconfirmation at the second stage using a rapid antigen test. We have designed this protocol especially to compare the expected cost of group testing protocols I and II with the cost of this individual testing protocol. Hence the expected number of tests per person is,

$$E(T_i|p,Se) = \frac{1}{N}(N + NPr(x_i = 1))$$

$$= 1 + Pr(x_i = 1)$$

$$= 1 + Se.p + (1-Sp)(1-p) \text{ from equation 1}$$

$$= 1 + Se.p + 0 \text{ when } Sp = 1$$

$$= 1 + Se.p \quad \text{…….(10)}$$

If the sensitivity of individual testing is assumed to be accurate ($Se = 1$), then the expected number of tests per person becomes,

$E(T_i|p, Se) = 1 + p$ ……(11)

Now, the overall expected cost per person will be,

$E(C|p, Se) = C_3 + C_2.Se.p$ [Protocol III(a)]

$= C_1 + C_2.Se.p$ since we assumed earlier $C_1 = C_3$

$= C_2\left(\frac{C_1}{C_2} + Se.p\right)$……(12)

And with perfect sensitivity it becomes

$$E(C|p, Se = 1) = C_2\left(\frac{C_1}{C_2} + p\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (13)$$

Similarly, if we use the RT-PCR test at both stages, then the overall expected cost per person becomes,

$E(C|p, Se) = C_3 + C_3.Se.p = C_3(1 + Se.p) = C_1(1 + Se.p)$ since $C_1 = C_3$ [Protocol III(b)]

……(14)

And with perfect sensitivity it becomes

$E(C|p, Se = 1) = C_1(1 + p) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (15)$

2.3.4 Protocol IV

In practice, individual testing is usually performed only one time. So, this protocol involves a single stage for screening individuals separately. Therefore, the expected number of tests per person becomes,

$$E(T) = \frac{1}{N} \times N = 1$$

The expected cost per person using the RT-PCR test is $E(C) = C_3 = C_1$ [Protocol IV (a)]……(16)

And using a rapid antigen test is

$E(C) = C_2$ [Protocol IV (b)]……(17)

Table 2. 3: Summary of all the protocols

Protocols		Cost Function, E(C)	Test kit and associated cost parameters		How many stages?	Is it group or individual testing?
			First Stage	Second Stage		
I	d = 0	$C_2[\frac{C_1}{C_2 n} + (1-(1-p)^n)]$	RT-PCR; C_1 (Group testing)	Rapid antigen; C_2 (Individual testing)	2	Group
	d ≠ 0	$C_2[\frac{C_1}{C_2 n} + (\frac{p}{1-(1-p)^{n^d}})(1-(1-p)^n)]$				
II	d = 0	$C_1[\frac{1}{n} + (1-(1-p)^n)]$	RT-PCR; C_1 (Group testing)	RT-PCR; C_1=C_3 (Individual testing)	2	Group
	d ≠ 0	$C_1[\frac{1}{n} + (\frac{p}{1-(1-p)^{n^d}})(1-(1-p)^n)]$				
III	a	$C_2\left(\frac{C_1}{C_2} + Se.p\right)$	RT-PCR; C_1 (Individual Testing)	Rapid antigen; C_2 (Individual testing)	2	Individual
	b	$C_1(1+Se.p)$	RT-PCR; C_1 (Individual testing)	RT-PCR; C_1=C_3 (Individual testing)		
IV	a	$C_1 = C_3$	RT-PCR; $C_1 = C_3$ (Individual testing)		1	Individual
	b	C_2	Rapid antigen; C_2 (Individual testing)			

C_1 = The entire (lab and test) cost of testing a group of samples using a real-time reverse transcription-polymerase chain reaction (RT-PCR) test; C_2 = The cost of testing (without the lab cost) an individual using a rapid antigen test; C_3 = The entire (lab and test) cost of testing an individual using an RT-PCR test.

p = Prevalence rate; d = Dilution parameter; n = Group size.

2.4 Estimation of Optimal Group Sizes and Their Respective Minimized Expected Costs per Individual

We have found some real numerical figures for the cost parameters C_1 and C_2 from several online newspapers and government websites. Benjamin Master, a policy researcher and co-author of RAND corporation, said that administering a PCR test in a test center costs $120 (Gewertz, 2021). It means this cost includes the entire cost such as lab cost, test cost and the cost of the test kit. He added that rapid antigen tests which can be processed without a lab can be bought at $5 per test as a bulk rate or $20/$30 per test from a store (Gewertz, 2021). That means this cost includes the cost of the test kit and the test cost but does not include the lab cost. On the Department of Elementary and Secondary Education (2021) of Massachusetts website, it is found that the fee for the PCR test is $50 per group if there are 5-10 samples in a tube or $85 for 11-25 samples. More precisely, these cost figures also include the entire cost

such as lab cost, test cost and the cost of the test kit. Feuer in 2020 provided the information that the cost for using Abbott's rapid antigen test at home is $25 per test which means it only includes the cost of the test kit and at a medical center, the cost is $5 per test which means it might include the cost of the test kit and the cost of the test but might not include the lab cost. Therefore, we can take different values for those cost parameters such that $C_1 = C_3 = (120,85,50)$ and $C_2 = (25,5)$.

Now, for the various values of p (0.001,0.01,0.03,0.05,0.07,0.10,0.15,0.20) and various values of cost ratios $\frac{C_1}{C_2}$ and cost parameters, optimum values of n and reduced expected costs per individual with optimal sizes are presented in the tables for protocols I and II. The graphs of expected cost against group size for locating the optimal n are displayed in the appendix. The estimated expected costs per person are shown in the tables for protocols III and IV.

2.4.1 Protocol I

The sixth column of Table 2.4 depicts that for the lowest cost ratio i.e., $r = \frac{C_1}{C_2} = 2$, optimal n is decreasing until prevalence becomes 15%. However, when the p-value is largest i.e., 20%, the optimal n has become slightly greater than the previous one. We know that if we have a high prevalence rate, the probability of having infected individuals within a community will be high. As such, it is more reasonable to form groups of smaller sizes to screen the disease more accurately. Hence, the decreasing pattern of optimal group sizes when prevalence rates are getting high is obvious. Similarly, for the second smallest cost ratio $r = 3.4$, n is slowly declining when p ranges from 0.1% to 10% (Figure 2.2). For $r = 4.8$, the same is happening until when p is equal to 7%. However, when p is 10%, n becomes larger than the previous one. For $r = 10$, when p is the lowest, i.e., 0.1%, n is 105.39. Then optimal n sharply drops by around 67 while p reaches 1% and gradually decreases to 27.58 when p is 3%.

Table 2. 4: Optimum group size and minimum expected cost per person of protocol I without dilution for different p and cost ratio $\frac{C_1}{C_2}$

	Protocol I $E(C\|p,n,Se(n)) = C_2(\frac{r}{n} + (1-(1-p)^n))$ **when d is 0**					
	$r = \frac{120}{5}$=24	$r = \frac{85}{5}$=17	$r = \frac{50}{5}$=10	$r = \frac{120}{25}$=4.8	$r = \frac{85}{25}$=3.4	$r = \frac{50}{25}$=2
p=0.001 n=(0,200)	n=168.50 E(C)=1.48	n=139.79 E(C)=1.26	n=105.39 E(C)=0.97	n=71.80 E(C)=3.40	n=60.07 E(C)=2.87	n=45.75 E(C)=2.21
p=0.01 n=(0,80)	n=69.18 E(C)=4.24	n=53.93 E(C)=3.67	n=38.22 E(C)=2.90	n=24.75 E(C)=10.35	n=20.38 E(C)=8.80	n=15.23 E(C)=6.83

p=0.03 **n=(0,60)**	n*>60 E(C)*<6.20	n=52.77 E(C)=5.61	n=27.58 E(C)=4.65	n=16.02 E(C)=17.14	n=12.85 E(C)=14.71	n=9.34 E(C)=11.54
p=0.05 **n=(0,40)**	n*>40 E(C)*<7.36	n*>40 E(C)*<6.48	n= 30.63 E(C)=5.59	n=13.77 E(C)=21.38	n=10.72 E(C)=18.50	n= 7.59 E(C)=14.65
p=0.07 **n=(0,30)**	n*>30 E(C)*<8.43	n*>30 E(C)*<7.27	n*>30 E(C)*<6.10	n=13.07 E(C)=24.50	n= 9.75 E(C)=21.40	n= 6.69 E(C)=17.09
p=0.10 **n=(0,25)**	n*>25 E(C)*<9.44	n*>25 E(C)*<8.04	n*>25 E(C)*<6.64	n=14.45 E(C)=27.85	n=9.25 E(C)=24.76	n= 5.97 E(C)=20.05
p=0.15 **n=(0,20)**	n*>20 E(C)*<10.81	n*>20 E(C)*<9.06	n*>20 E(C)*<7.31	n*>20 E(C)*<30.03	n*>20 E(C)*<28.28	n=5.47 E(C)=23.86
p=0.20 **n=(0,15)**	n*>15 E(C)*<12.82	n*>15 E(C)*<10.49	n*>15 E(C)*<8.16	n*>15 E(C)*<32.12	n*>15 E(C)*<29.79	n=5.58 E(C)=26.76

p = The prevalence rate; r = cost ratio = $\frac{C_1}{C_2}$.; n = Optimal group size.

n*=Optimal group sizes that are greater than the upper boundary evaluated and E(C)* is the respective minimized expected cost per person. Thus, red cells indicate that we could not find the exact optimal n and reduced E(C) within that range for those cases.

C_1 = The entire cost of testing a group of samples using a real-time RT-PCR test.

C_2 = The entire cost (without lab cost) of testing an individual using a rapid antigen test.

Protocol I = When RT-PCR test is used at the first stage and rapid antigen test is used at the second stage.

d = Dilution parameter = 0 which means sensitivity is perfect for the group test (Special Case).

But then, n gets bigger (30.63) than the previous one when p is 5%. Therefore, we can say n is not monotonically shrinking. For cost ratio $r = 17$, similarly, optimal n steeply decreases from 105.39 to 53.93 when p=0.01 and slightly decreases to 52.77 when p=0.03. When the cost ratio is the highest i.e., $r = 24$, optimal n reduces from 168.52 to 69.18 when p increases from 0. 1% to 1%. It means, for the lowest prevalence, groups are in huge sizes for all the cost ratios which are fairly not possible to use in a practical situation. However, for the rest of the prevalence rates, we have got reasonable optimal n that is between around 38 and 5 when $r =$ 10, 4.8, 3.4 and 2.

When the prevalence rate is 15% and 20%, except for the last column, exact optimal group sizes could not be found within the chosen range of n. If we look at the other prevalence rates, we did not attain the optimal n (highlighted in red) as well for some cost ratios. This is because, while getting the minimum cost for the different ranges of n in R programming software, the graphs of expected cost against the various group sizes were gradually decreasing in those cases. There is no optimal group size within those reasonable ranges for which the expected cost per person can be minimized. As such, for those cases, the upper level of the chosen interval is always found as an optimal group size which is not the global optimal point. In reality, the exact optimal n will always be greater than those upper boundaries. The respective minimized expected cost will be lesser if the upper boundary of n increases. As it is more

reasonable to have smaller group sizes for higher p values, we have set the boundary level as 20 and 15 for 15% and 20% prevalence respectively. When p is lower (0.001 and 0.01), we have got exact optimal n for all the cost ratios. Besides, each row of Table 2.4 shows that exact optimal group sizes are gradually declining with the decreasing cost ratios. Hence, we can conclude that group size depends on the cost ratio.

In each column, it is also visible that the minimized expected costs per individual are gradually rising with the increasing values of prevalence rates (Figure 2.2). This is because when the prevalence rate is getting larger, the likelihood of having a positive group at the first stage will tend to be higher. As such, at the second stage, more individuals will be retested to diagnose the disease which will incur more cost. Comparing the minimized expected costs when rapid antigen test charges \$5 with \$25, we can see that those costs become greater when the rapid test becomes costly no matter what the value of C_1 and n is. So, the overall expected cost largely depends on the cost of the individual testing at the second stage using a rapid test. In that case, if we want to reduce the overall cost per person we need to use a less expensive rapid test kit. Also, if we have fewer positive groups at the first stage which can happen if the p is lowest, then the number of individual testing will be less. As a result, a fewer number of rapid test kits will be used which will lower the overall expected cost.

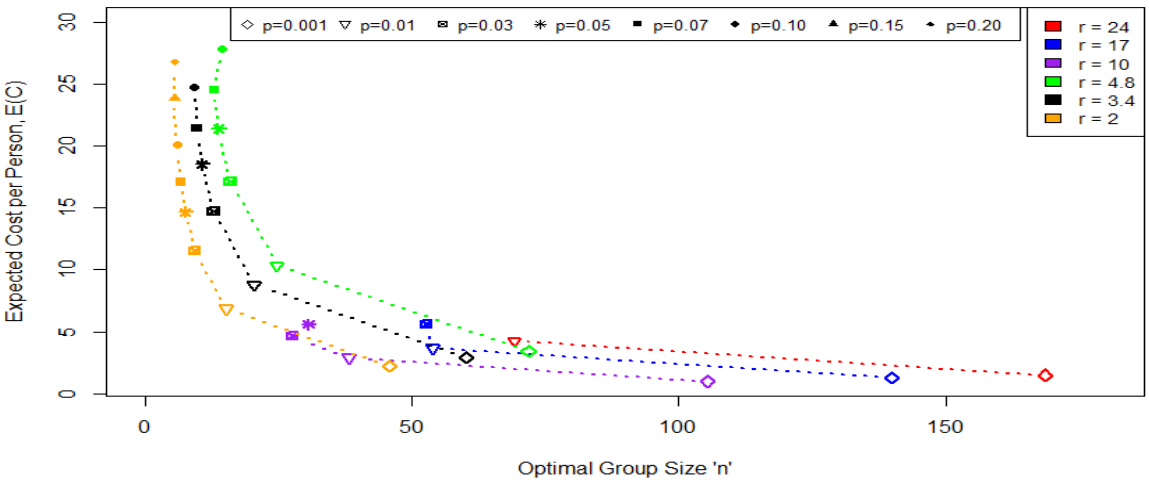


Figure 2. 2: For different prevalence rates, p and cost ratios, $r = \frac{C_1}{C_2}$, minimum expected costs per person (obtained from Table 2.4) against optimal group sizes, n. Here, C_1 is the entire cost of testing a group of samples using a real-time RT-PCR test. C_2 is the entire (without lab cost) cost of testing an individual using a rapid antigen test. In Protocol I, the RT-PCR test is used at the first stage without dilution and the rapid antigen test is used at the second stage.

Keeping C_2 constant, if we look at the values of C_1 separately, both optimal n and expected cost are decreasing with the decreasing C_1. For example, when $C_2 = 5\ and\ 25$, r = 10 and 2 respectively provide the lowest expected costs with the smallest group sizes. Hence, it is better to use the RT-PCR test which costs the lowest (i.e., \$50) for any C_2.

Moreover, changing the values of $C_1 and\ C_2$ keeping the cost ratio constant does not affect the optimal group size but affects the expected cost per individual. For example, if $r = \frac{25}{12.5} = 2$, for 10% prevalence, the expected cost per person becomes half which is 10.02 and n remains the same (5.97). So, if the cost components become half of the previous values, E(C) also becomes half for all prevalences. On the other hand, if $r = \frac{100}{10} = 10$, for 3% prevalence, only E(C) becomes double i.e., 9.31. So, E(C) becomes double if we double the cost parameters. And, in none of the cases, optimal n will change. Therefore, we can say optimal n is just a function of cost ratio and prevalence rate.

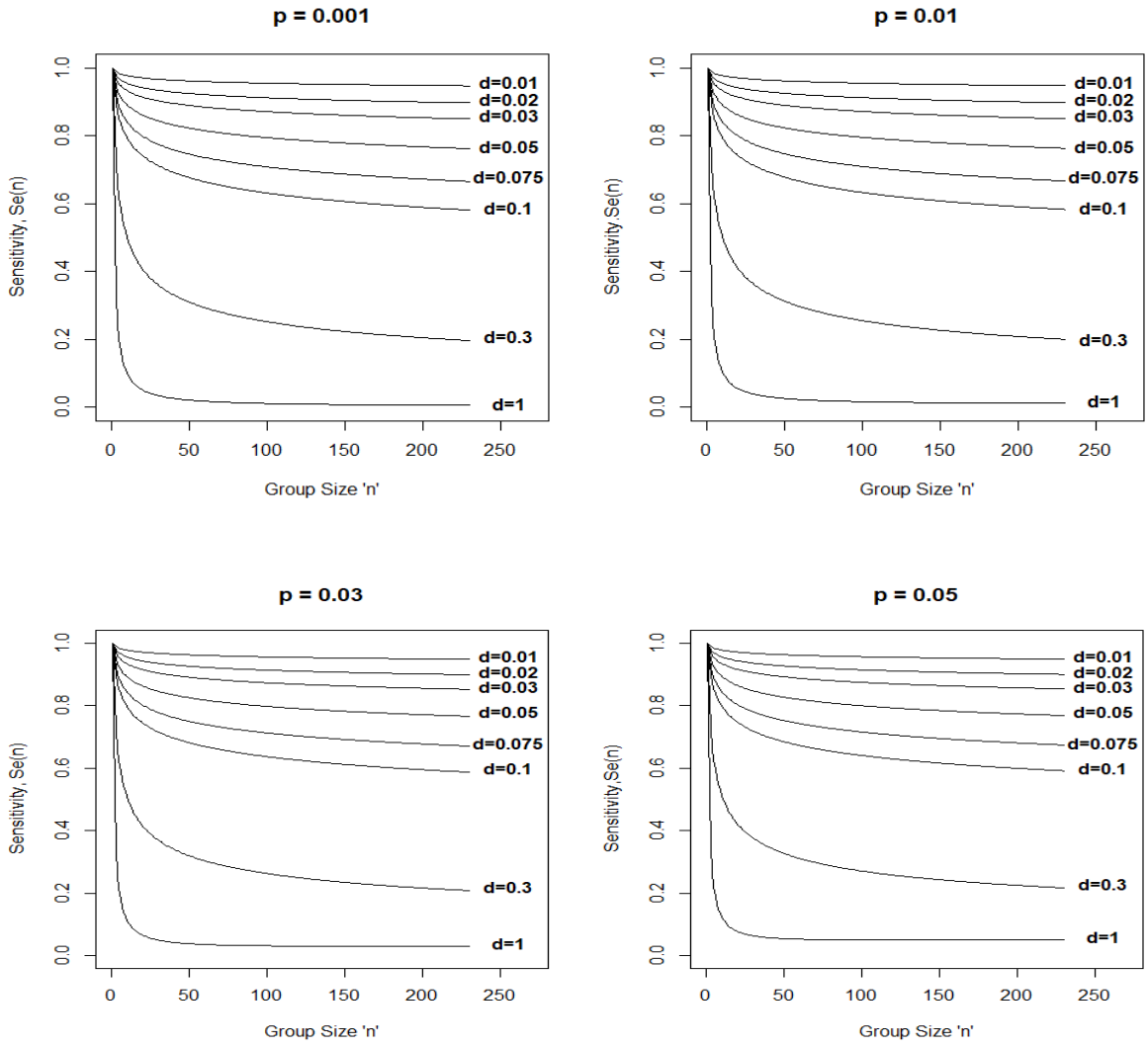
2.4.2 Dilution Effect

Generally, a group is formed by mixing a small segment of every individual's sample in a tube. Since a small portion of a pathogen is used in a group, a positive sample can be easily diluted with several negative individual samples. So, dilution occurs when positive samples are mixed with negative samples. In this situation, the biomarker concentration level of the group could be below the detection level of the test. As such, while diagnosing the group, it may lead to misclassification. Consequently, the reduction in the sensitivity of the group testing happens. This is known as the dilution effect (denoted by d) in group testing.

While focusing on group testing procedures, we should keep in mind that there might be some dilution effect in a group. The chances of misclassifying a positive individual in a group become higher if the group size becomes larger. Our sensitivity function, $Se(n) = \frac{p}{1-(1-p)^{n^d}}$ (Hwang, 1976) when $d \neq 0$ is a decreasing function of a group size. If the group size increases, then sensitivity decreases.

Let us plot this sensitivity function by assuming various values of d (0.01,0.02,0.03,0.05,0.075,0.1,0.3,1) against several values of group size (from 0 to 250) for different values of p (0.001,0.01,0.03,0.05,0.07,0.10,0.15,0.20) to find out how the sensitivity changes with the group size.

From Figure 2.3, we can see that the sensitivity curve is decaying with the increasing group sizes for all prevalence rates. Thus, these graphs indicate if we have a smaller group size then sensitivity will be nearly perfect. Moreover, we can see that as the prevalence rate gets higher, curves are shifting upwards slightly. That means sensitivity gets marginally better for higher prevalence rates at fixed n. For the highest values of d (d=0.1, 0.3 & 1), sensitivity lies between 70% and 10% approximately. Conversely, for the lowest values of d (0.01,0.02, 0.03, & 0.05) sensitivity reaches nearly 98%. Likewise, Haber et al. did in 2021, here we will choose d=0.075.



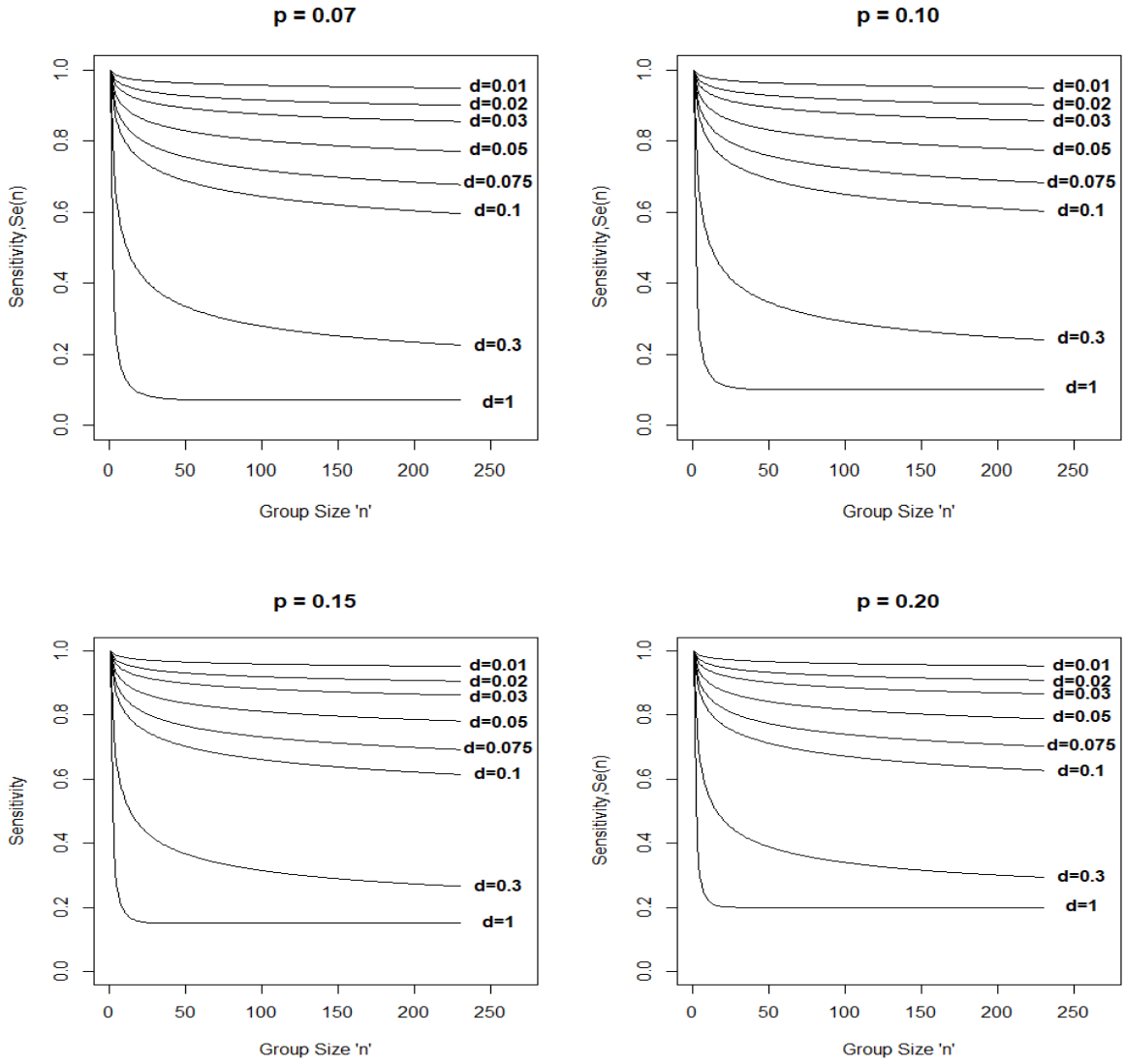


Figure 2. 3: Line graphs of $Se(n) = \frac{p}{1-(1-p)^{nd}}$ versus various group sizes (n) for several prevalence rates (p) and several values of d (Dilution parameter).

2.4.3 Protocol I with Dilution Effect

Now, with the dilution effect (when d = 0.075), we will check what happens to the optimal group sizes and the respective minimized overall expected costs in protocol I for various prevalence rates (0.001,0.01,0.03,0.05,0.07,0.10,0.15,0.20) and cost ratios. The results are presented in Table 2.5.

In Table 2.5, we can see the optimal group sizes with dilution are significantly largerthan those without dilution in groups. For example, without dilution optimal n ranges from 45.74 (p=0.1%) to 5.47 (p=15%) while with dilution it ranges from 55.63 to 7.27 when $\frac{C_1}{C_2} = \frac{50}{25}$. In the special case, sensitivity was perfect i.e., 100%. On the other hand, allowing dilution

prompts the sensitivity to reduce to around 70%. Consequently, less sensitivity makes the optimal n larger since our sensitivity function decreases with the increasing n. Implicitly it means when we have dilution, we will have more false-negative results from group tests which will reduce the overall expected cost andas a result, we can afford to form larger groups. So, if we want to have better sensitivity, we should form smaller groups. Moreover, like the special case, we can see a similar pattern in optimal group sizes and expected costs per individual for different cost ratios and prevalence rates (Figure 2.4). Here also with the dilution for $\frac{C_1}{C_2} = \frac{50}{25}$, we have found precise optimal group sizes which are decreasing until p becomes 10%. Nevertheless, when p is 15%, optimal n becomes slightly larger and when p is 20% optimal n is not exact. With the dilution, we can see 20% prevalence rate does not provide exact optimal n for any cost ratios within the chosen range of n. On the contrary, 0.1% and 1% prevalence rates have provided the precise optimal group sizes for all the cost ratios. For cost ratio 4.8 , we have found the exact optimal sizes until p becomes 5%. For $\frac{C_1}{C_2} = \frac{85}{25}$, optimal n is found until p is 7%.

Table 2. 5: Optimum group sizes and minimum expected costs per person of protocol I with dilution for different p and cost ratio $\frac{C_1}{C_2}$

	Protocol I $E(C\|p,n,Se(n)) = C_2(\frac{r}{n} + \frac{p}{1-(1-p)^{n^d}}(1-(1-p)^n))$ **when d=0.075**					
	$r = \frac{120}{5}$=24	$r = \frac{85}{5}$=17	$r = \frac{50}{5}$=10	$r = \frac{120}{25}$=4.8	$r = \frac{85}{25}$=3.4	$r = \frac{50}{25}$=2
p=0.001 n=(0,300)	n=221.30 E(C)=1.20	n= 180.99 E(C)=1.03	n= 133.92 E(C)=0.81	n= 89.28 E(C)=2.87	n=74.02 E(C)=2.44	n= 55.63 E(C)=1.9
p=0.01 n=(0,200)	n=106.98 E(C)=3.45	n=74.46 E(C)=3.05	n=49.10 E(C)=2.48	n=30.23 E(C)=9.05	n=24.49 E(C)=7.77	n=17.94 E(C)=6.11
p=0.03 n=(0,100)	n*>100 E(C)*<4.59	n*>100 E(C)*<4.24	n*>100 E(C)*<3.89	n=20.08 E(C)=15.14	n=15.57 E(C)=13.17	n=10.95 E(C)=10.15
p=0.05 n=(0,60)	n*>60 E(C)*<5.54	n*>60 E(C)*<4.96	n*>60 E(C)*<4.37	n=18.38 E(C)=18.87	n=13.29 E(C)=16.63	n=8.92 E(C)=13.43
p=0.07 n=(0,40)	n*>40 E(C)*<6.62	n*>40 E(C)*<5.75	n*>40 E(C)*<4.87	n*>40 E(C)*<21.12	n=12.62 E(C)=19.23	n=7.94 E(C)=15.73
p=0.10 n=(0,30)	n*>30 E(C)*<7.77	n*>30 E(C)*<6.60	n*>30 E(C)*<5.43	n*>30 E(C)*<23.83	n*>30 E(C)*<21.66	n=7.23 E(C)=18.50
p=0.15 n=(0,20)	n*>20 E(C)*<9.92	n*>20 E(C)*<8.17	n*>20 E(C)*<6.42	n*>20 E(C)*<25.58	n*>20 E(C)*<23.83	n=7.27 E(C)=22.00
p=0.20 n=(0,10)	n*>10 E(C)*<15.83	n*>10 E(C)*<12.33	n*>10 E(C)*<8.83	n*>10 E(C)*<31.16	n*>10 E(C)*<27.66	n*>10 E(C)*<24.16

p = The prevalence rate; ; r = cost ratio = $\frac{C_1}{C_2}$.; n = Optimal group size; d = Dilution parameter = 0.075.

n*=Optimal group sizes that are greater than the upper boundary evaluated and E(C)* is the respective minimized expected cost per person. Thus, red cells indicate that we could not find the exact optimal n and reduced E(C) within that range for those cases.

C_1 = The entire cost of testing a group of samples using a real-time RT-PCR test.

C_2 = The entire cost (without lab cost) of testing an individual using a rapid antigen test.
Protocol I = When RT-PCR test is used at the first stage with dilution and rapid antigen test is used at the second stage.

If we look at the cost values, we notice that the total expected costs per person have been reduced slightly than the special case (d=0). For example, without dilution E(C) ranges from 2.21 (p=0.1%) to 23.86 (p=15%) while with dilution it ranges from 1.9 to 18.50 when $\frac{C_1}{C_2} = \frac{50}{25}$. This is also because of the dilution in a group. More precisely, with the presence of dilution, there is a higher chance of missing detection of the positive groups at the first stage. As such, at the second stage, the number of individual testing will be less compared to the special case. That is why the overall cost becomes lesser when d=0.075. Overall, this table is suggesting that with dilution expected costs per person of protocol I have got reduced compared to the special case which results in having larger groups but causing more false-negative errors. In the next chapter, we will explore more about the relationship between these costs and the number of FN errors.

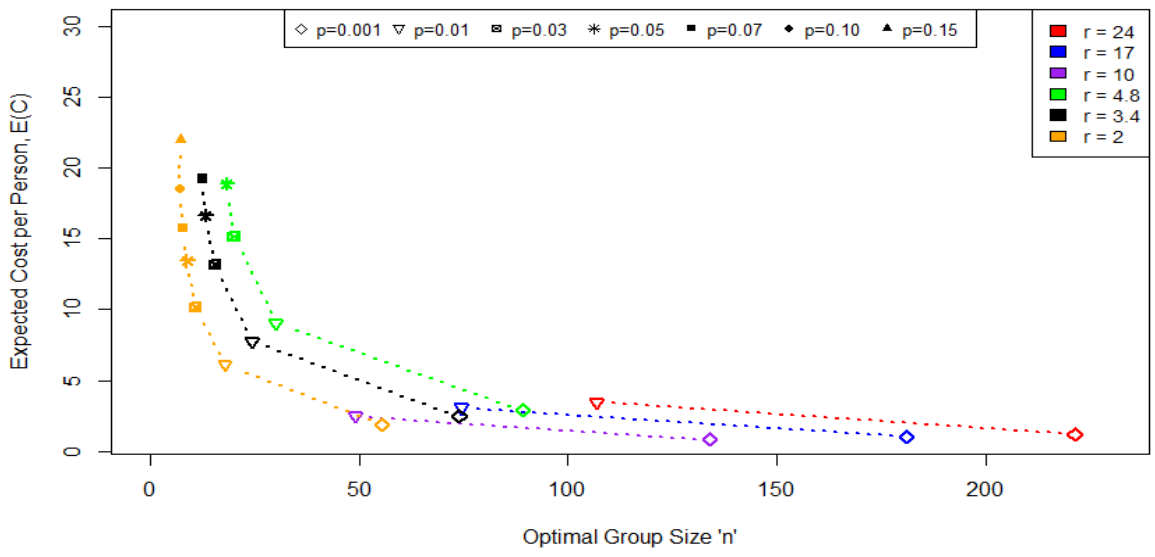


Figure 2. 4: For several values of prevalence rate, p and cost ratios $\frac{C_1}{C_2}$, minimum expected costs (obtained from Table 2.5) against several optimal values of n (Group size). Here, C_1 is the entire cost of testing a group of samples using a real-time RT-PCR test. C_2 is the entire (without lab cost) cost of testing an individual using a rapid antigen test. In Protocol I, the RT-PCR test is used at the first stage with dilution and the rapid antigen test is used at the second stage.

Again, we have examined the values of n and E(C) for different cost parameters keeping the cost ratio constant. Similar to the case without dilution, it does not affect the optimal group size but affects the expected cost per individual. For example, if $r = \frac{25}{12.5} = 2$, for 10% prevalence, the expected cost per person becomes half which is 9.25 and n remains the same (7.23). So, if the cost components become half of the previous values, E(C) also becomes half for all prevalences. On the other hand, if $r = \frac{100}{10} = 10$, for 1% prevalence, only E(C) becomes double i.e., 4.95. So, E(C) becomes double if we double the cost parameters. In none of the cases, optimal n has changed. Therefore, we can say optimal n is just a function of cost ratio and prevalence rate.

Since we are not interested in the units of the cost parameters and C_2 is the common factor of the cost function, let us assume C_2 be the unit cost at the second stage to mainly focus on the cost ratio. Our cost ratios $\frac{C_1}{C_2} = r = (24,17,10,4.8,3.4,2)$ are the same as previous cost ratios. Here the only difference is our C_2 is 1. For example, r = $\frac{C_1}{C_2} = \frac{24}{1} = 24$ means the cost of testing a group using RT-PCR test (C_1) are 24 times the unit cost of a rapid antigen test for individuals testing (C_2) at the second stage.

Since optimal n depends only on the cost ratio, the unit cost for individual testing will not affect the optimal group size. From Tables 2.6 & 2.7, we can see that minimized expected cost per person reduces than before because we have scaled the cost function by assuming C_2 = 1. We have not found the exact optimal n and minimized expected cost per person for the same values of p and r within those ranges of n.

When the cost of group testing is 10 times the unit cost of individual testing for p = 5%, the reduced expected cost for a person for this protocol is 1.12 times the cost of overall individual testing (Table 2.6). Thus, we can say, by adopting this group testing protocol we are not saving any money. Therefore, we will prefer to test all the populations separately in this case. Similarly, the individual testing strategy would be a better plan when p is 3%, 10% and 20% and r =17, 4.8 and 2 respectively. Also, we can see as p goes up, E(C) also goes up. So, we can assume some of the red highlighted E(C) especially when r is higher (in which we did not get exact optimal n) can be greater than 1. Group testing will not be the preferred one in those cases. On the other hand, for the rest of the cases, protocol, I when d=0 is found to be the appropriate one.

Table 2. 6: Optimum group size and minimum expected cost per person of protocol I without dilution for different p and r when $C_2 = 1$

	Protocol I $E(C\|p,n,Se(n)) = (\frac{r}{n} + (1-(1-p)^n))$ when d = 0					
	$r = 24$	$r = 17$	$r = 10$	$r = 4.8$	$r = 3.4$	$r = 2$
p=0.001	E(C)=0.30	E(C)=0.25	E(C)=0.19	E(C)=0.14	E(C)=0.11	E(C)=0.09
p=0.01	E(C)=0.85	E(C)=0.73	E(C)=0.58	E(C)=0.41	E(C)=0.35	E(C)=0.27
p=0.03	E(C)*<1.24	E(C)=1.12	E(C)=0.93	E(C)=0.69	E(C)=0.59	E(C)=0.46
p=0.05	E(C)*<1.47	E(C)*<1.30	E(C)=1.12	E(C)=0.86	E(C)=0.74	E(C)=0.59
p=0.07	E(C)*<1.69	E(C)*<1.45	E(C)*<1.22	E(C)=0.98	E(C)=0.86	E(C)=0.68
p=0.10	E(C)*<1.89	E(C)*<1.61	E(C)*<1.33	E(C)=1.11	E(C)=0.99	E(C)=0.80
p=0.15	E(C)*<2.16	E(C)*<1.81	E(C)*<1.46	E(C)*<1.20	E(C)*<1.13	E(C)=0.95
p=0.20	E(C)*<2.56	E(C)*<2.10	E(C)*<1.63	E(C)*<1.28	E(C)*<1.19	E(C)=1.07

p = The prevalence rate; r = cost ratio = $\frac{C_1}{C_2}$; d = Dilution parameter.

E(C)* is the minimized expected cost per person in which exact optimal n was not found within the boundary of n which are also coloured as red.

C_1 = The entire cost of testing a group of samples using a real-time RT-PCR test.

C_2 = The entire cost (without lab cost) of testing an individual using a rapid antigen test.

Protocol I = When RT-PCR test is used at the first stage without dilution and rapid antigen test is used at the second stage.

With the dilution effect, protocol I seems to be the appropriate one to apply over individual testing for all the prevalences in which we have found exact optimal n (Table 2.7). Similar to protocol I without dilution, here for higher prevalence and higher cost ratios we might get E(C) greater than 1 as well since it is increasing with increasing values of p. In those cases, group testing will cost more money than individual testing. As such, in those situations, individual testing will be better than the group testing procedure.

Table 2. 7: Optimum group size and minimum expected cost per person of protocol I with dilution for different p and r when $C_2 = 1$

	Protocol I $E(C\|p,n,Se(n)) = (\frac{r}{n} + \frac{p}{1-(1-p)^{nd}}(1-(1-p)^n))$ when d=0.075					
	$r = 24$	$r = 17$	$r = 10$	$r = 4.8$	$r = 3.4$	$r = 2$
p=0.001	E(C)=0.24	E(C)=0.21	E(C)=0.16	E(C)=0.11	E(C)=0.10	E(C)=0.08
p=0.01	E(C)=0.69	E(C)=0.61	E(C)=0.50	E(C)=0.36	E(C)=0.31	E(C)=0.24
p=0.03	E(C)*<0.92	E(C)*<0.85	E(C)*<0.78	E(C)=0.61	E(C)=0.53	E(C)=0.42
p=0.05	E(C)*<1.11	E(C)*<0.99	E(C)*<0.87	E(C)=0.75	E(C)=0.67	E(C)=0.58
p=0.07	E(C)*<1.32	E(C)*<1.15	E(C)*<0.97	E(C)*<0.84	E(C)=0.77	E(C)=0.63
p=0.10	E(C)*<1.55	E(C)*<1.32	E(C)*<1.09	E(C)*<0.91	E(C)*<0.87	E(C)=0.74
p=0.15	E(C)*<1.98	E(C)*<1.633	E(C)*<1.28	E(C)*<1.02	E(C)*<0.95	E(C)=0.88
p=0.20	E(C)*<3.17	E(C)*<2.47	E(C)*<1.77	E(C)*<1.25	E(C)*<1.11	E(C)*<0.97

p = The prevalence rate; r = cost ratio = $\frac{C_1}{C_2}$; d = Dilution parameter=0.075.

E(C)* is the minimized expected cost per person in which exact optimal n was not found within the boundary of n which are also colored as red.

C_1 = The entire cost of testing a group of samples using a real-time RT-PCR test.
C_2 = The entire cost (without lab cost) of testing an individual using a rapid antigen test.
Protocol I = When RT-PCR test is used at the first stage with dilution and rapid antigen test is used at the second stage.

2.4.4 Protocol II

Here, the optimal group sizes of protocol II do not depend on the price of the RT-PCR test (Table 2.8). This is because this protocol involves only cost parameter C_1 which is a common factor in the cost function. Optimal group sizes vary according to the prevalence rates only. As p goes up, optimal n is going down gradually (Figure 2.5). For the highest prevalence rate (p = 20%), optimal group sizes are the lowest (2.94 & 3.39). Conversely, for the lowest prevalence rate (p = 0.1%) values of optimal n are the highest (32.13 & 38.45). These results are understandable since if the probability of having a disease is high then it is easier to screen the disease when the group size is small. On the other hand, if the prevalence is too low then to save the test resources and money, it is more sensible to form a larger group for screening. As in protocol I, group sizes with dilution are larger than without dilution. However, the difference between the optimal n with and without dilution is minimal (Figure 2.5). For example, when p is 0.1%, the difference is only nearly 6. And when prevalence gets higher, the difference gets lower. Such as, when p is 20%, the difference becomes less than 1. It means if group size depends only on the prevalence rate, then having dilution does not affect the group size to a large extent. Overall, results are indicating that if the same test kit (RT-PCR) is used at both stages, all the prevalence rates will provide exact optimal group sizes which are smaller than those of protocol I.

Table 2. 8: Optimum value of n and minimum expected cost per person of protocol II with and without dilution for different values of p and C_1

	Protocol II					
	$E(C_{PCR}\|p,n,Se(n)) = C_1\left(\frac{1}{n}+(1-(1-p)^n)\right)$ when d=0			$E(C_{PCR}\|p,n,Se(n)) = C_1\left(\frac{1}{n}+(\frac{p}{1-(1-p)^{n^d}})(1-(1-p)^n)\right)$ when d=0.075		
	$C_1 = 120$	$C_1 = 85$	$C_1 = 50$	$C_1 = 120$	$C_1 = 85$	$C_1 = 50$
p=0.001 n=(0,50)	n=32.13 E(C)=7.53	n= 32.13 E(C)=5.33	n= 32.13 E(C)=3.14	n= 38.45 E(C)=6.57	n= 38.45 E(C)=4.65	n= 38.45 E(C)=2.74
p=0.01 n=(0,20)	n=10.52 E(C)=23.45	n=10.52 E(C)=16.61	n=10.52 E(C)=9.77	n= 12.13 E(C)=21.33	n= 12.13 E(C)=15.11	n= 12.13 E(C)=8.89
p=0.03 n=(0,10)	n=6.31 E(C)=40.00	n=6.31 E(C)=28.33	n=6.31 E(C)=16.67	n= 7.18 E(C)=37.09	n= 7.18 E(C)=26.28	n= 7.18 E(C)=15.46

p=0.05 n=(0,10)	n=5.02 E(C)=51.15	n=5.02 E(C)=36.23	n=5.02 E(C)=21.31	n= 5.69 E(C)=47.85	n= 5.69 E(C)=33.89	n= 5.69 E(C)=19.94
p=0.07 n=(0,10)	n=4.35 E(C)=60.07	n=4.35 E(C)=42.55	n=4.35 E(C)=25.03	n= 4.92 E(C)=56.51	n= 4.92 E(C)=40.03	n= 4.92 E(C)=23.55
p=0.10 n=(0,10)	n=3.75 E(C)=71.17	n=3.75 E(C)=50.41	n=3.75 E(C)=29.65	n= 4.25 E(C)=67.33	n= 4.25 E(C)=47.69	n= 4.25 E(C)=28.05
p=0.15 n=(0,10)	n=3.22 E(C)=86.16	n=3.22 E(C)=61.03	n=3.22 E(C)=35.90	n= 3.67 E(C)=81.99	n= 3.67 E(C)=58.08	n= 3.67 E(C)=34.16
p=0.20 n=(0,10)	n=2.94 E(C)=98.55	n=2.94 E(C)=69.81	n=2.94 E(C)=41.06	n= 3.39 E(C)=94.11	n= 3.39 E(C)=66.66	n= 3.39 E(C)=39.21

p = The prevalence rate; n= Optimal group size; d = Dilution parameter.
C_1 = The entire cost of testing a group of samples using a real-time RT-PCR test.
C_3 = The entire cost of testing an individual using a real-time RT-PCR test.
Protocol II = When RT-PCR test is used both for group testing (1st stage) with and without dilution and individual testing (2nd stage) without and with dilution effect. In protocol II, we assumed $C_1 = C_3$ for simplicity.

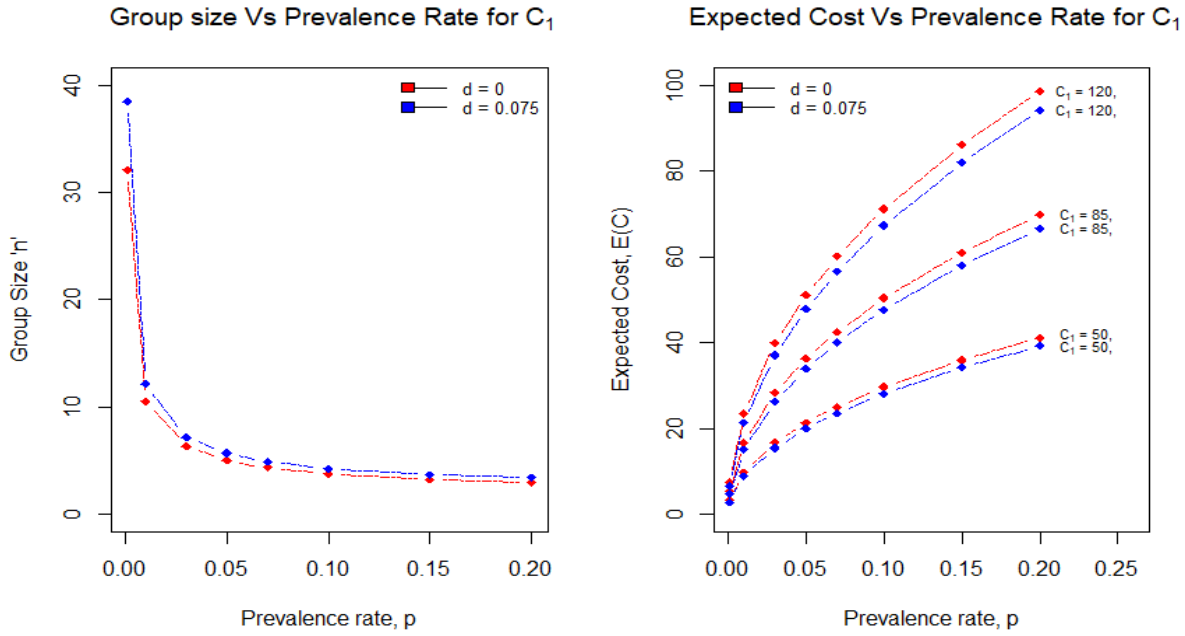


Figure 2. 5: In protocol II with and without dilution, the line graph of optimal group sizes against several prevalence rates are shown on the left side when group sizes are equal for $C_1 = 120, 85, \& 50$. And the line graph of expected costs per person with and without dilution against several prevalence rates for three values of C_1 are shown on the right side.

Now, if we look at the minimum overall expected cost per person, we notice that this protocol is more expensive than protocol I. This is because we are only using the RT-PCR test for both stages and this test is costly than the rapid antigen test. As the prevalence rate gets higher, the cost of this procedure gets higher (Figure 2.5). This is happening because, with the higher prevalence rate, more individual testing takes place at the second stage after detecting more positive groups at the first stage. With the dilution effect, the total expected cost per person

also decreases a little (Figure 2.5). For example, when p is 0.1%, the difference between E(C) with and without dilution is only around 1. When prevalence gets higher, the difference also gets higher. Such as, when p is 20%, the difference becomes around 4. Since there is a possibility of missing true positive cases, the number of tests will be less which will save some money. Since the optimal group size does not get affected by the dilution that much, we can concentrate only on how much cost per person on average is required. Moreover, we can easily say that this protocol is more expensive than Protocol I. Additionally, we can also explore how many true cases are missed under this protocol and how it will affect the overall expective cost which will be shown in the next chapter.

2.4.5 Protocol III

Since we are only performing individual testing at both stages, we do not need to form groups in Protocol III. We are only concerned about what the overall expected cost per person will be using an individual testing protocol which is shown in Tables 2.9, 2.10 and 2.11. Since there is no group, so dilution effect does not exist in this protocol.

Table 2. 9: Total expected cost per person of protocol III(a) for different p and $\frac{C_1}{C_2}$ when Se=100%

	Protocol III $E(C\|p,Se) = C_2\left(\frac{C_1}{C_2}+p\right)$					
	$r=\frac{120}{5}$=24	$r=\frac{85}{5}$=17	$r=\frac{50}{5}$=10	$r=\frac{120}{25}$=4.8	$r=\frac{85}{25}$=3.4	$r=\frac{50}{25}$=2
p=0.001	E(C)=120.01	E(C)=85.01	E(C)=50.01	E(C)=120.03	E(C)=85.03	E(C)=50.03
p=0.01	E(C)=120.05	E(C)=85.05	E(C)=50.05	E(C)=120.25	E(C)=85.25	E(C)=50.25
p=0.03	E(C)=120.15	E(C)=85.15	E(C)=50.15	E(C)=120.75	E(C)=85.75	E(C)=50.75
p=0.05	E(C)=120.25	E(C)=85.25	E(C)=50.25	E(C)=121.25	E(C)=86.25	E(C)=51.25
p=0.07	E(C)=120.35	E(C)=85.35	E(C)=50.35	E(C)=121.75	E(C)=86.75	E(C)=51.75
p=0.10	E(C)=120.50	E(C)=85.50	E(C)=50.50	E(C)=122.50	E(C)=87.50	E(C)=52.50
p=0.15	E(C)=120.75	E(C)=85.75	E(C)=50.75	E(C)=123.75	E(C)=88.75	E(C)=53.75
p=0.20	E(C)=121.00	E(C)=86.00	E(C)=51.00	E(C)=125.00	E(C)=90.00	E(C)=55.00

p = The prevalence rate; Se = Sensitivity
C_1 = The entire cost of testing a group of samples using a real-time RT-PCR test.
C_2 = The entire cost (without lab cost) of testing an individual using a rapid antigen test.
C_3 = The entire cost of testing an individual using a real-time RT-PCR test.
Protocol III = When RT-PCR test is used for individual testing (1st stage) and rapid antigen test at 2nd stage is used again for individual testing to reconfirm. In Protocol III, we assumed $C_1 = C_3$ for simplicity.

As we can see in Tables 2.9, 2.10 and 2.11, protocol III is more expensive than protocols I and II. Since we are not forming any groups here, this protocol is costing more than the other two protocols. The expected cost per individual has increased by only one dollar when the

prevalence rate ranges from the lowest (0.1%) to the highest (20%) and C_2 is \$5 (Table 2.9). Therefore, for a \$5 rapid antigen test, the overall cost of this protocol almost remains the same for lower and higher prevalence rates. However, when C_2 is \$25 the difference of the expected cost per person between the lowest and highest p becomes around \$5. It might be because rapid test charges 5 times more than before. When 20 people are affected among 100 individuals and $\frac{C_1}{C_2} = \frac{120}{25} = 4.8$, on average, testing a person individually twice by an RT-PCR test first and a rapid test to reconfirm costs \$125 which is the highest (Figure 2.6). On the other hand, when only 1 person is affected among 1000 individuals and $\frac{C_1}{C_2} = \frac{50}{5} = 10$, the cost is nearly \$50. It means expected cost depends largely on the cost of the RT-PCR test (C_1). So, if we want to minimize this cost, we have to use a less expensive RT-PCR test kit for this protocol.

If the sensitivity of the rapid antigen test is 93.3% (Pilarowski et al., 2021), the only difference is that the expected costs per person for all p and $\frac{C_1}{C_2}$ become slightly less (less than 1 unit) (Table 2.10) than the expected costs when assuming the sensitivity of the rapid test is perfect.

Table 2. 10: Total expected cost per person of protocol III(a) for different p and $\frac{C_1}{C_2}$ when Se=93.3%

	Protocol III(a) $E(C\|p,Se) = C_2\left(\frac{C_1}{C_2} + Se.p\right)$					
	$\frac{C_1}{C_2} = \frac{120}{5}$	$\frac{C_1}{C_2} = \frac{85}{5}$	$\frac{C_1}{C_2} = \frac{50}{5}$	$\frac{C_1}{C_2} = \frac{120}{25}$	$\frac{C_1}{C_2} = \frac{85}{25}$	$\frac{C_1}{C_2} = \frac{50}{25}$
p=0.001	E(C)=120.00	E(C)=85.00	E(C)=50.00	E(C)=120.02	E(C)=85.02	E(C)=50.02
p=0.01	E(C)=120.05	E(C)=85.05	E(C)=50.05	E(C)=120.23	E(C)=85.23	E(C)=50.23
p=0.03	E(C)=120.14	E(C)=85.14	E(C)=50.14	E(C)=120.70	E(C)=85.70	E(C)=50.70
p=0.05	E(C)=120.23	E(C)=85.23	E(C)=50.23	E(C)=121.17	E(C)=86.17	E(C)=51.17
p=0.07	E(C)=120.33	E(C)=85.33	E(C)=50.33	E(C)=121.63	E(C)=86.63	E(C)=51.63
p=0.10	E(C)=120.47	E(C)=85.47	E(C)=50.47	E(C)=122.33	E(C)=87.33	E(C)=52.33
p=0.15	E(C)=120.70	E(C)=85.70	E(C)=50.70	E(C)=123.50	E(C)=88.50	E(C)=53.50
p=0.20	E(C)=120.93	E(C)=85.93	E(C)=50.93	E(C)=124.67	E(C)=89.67	E(C)=54.67

p = The prevalence rate; Se = Sensitivity;
C_1 = The entire cost of testing a group of samples using a real-time RT-PCR test.
C_2 = The entire cost (without lab cost) of testing an individual using a rapid antigen test.
C_3 = The entire cost of testing an individual using a real-time RT-PCR test.
Protocol III(a) = When RT-PCR test is used for individual testing (1st stage) and rapid antigen test at 2nd stage is used again for individual testing to reconfirm. In Protocol III(a), we assumed $C_1 = C_3$ for simplicity.

In Table 2.11, we can see the expected cost per person gets higher since we are using the RT-PCR test at both stages. When the prevalence rate is highest, this strategy becomes the most

expensive. When the sensitivity becomes 97.2%, the cost slightly changes (Figure 2.7). More precisely costs decrease by less than 1 unit. Therefore, with or without perfect sensitivity this strategy of individual testing costs almost the same. Hence, to lower the expected cost per person, we should choose the RT-PCR test charging $50.

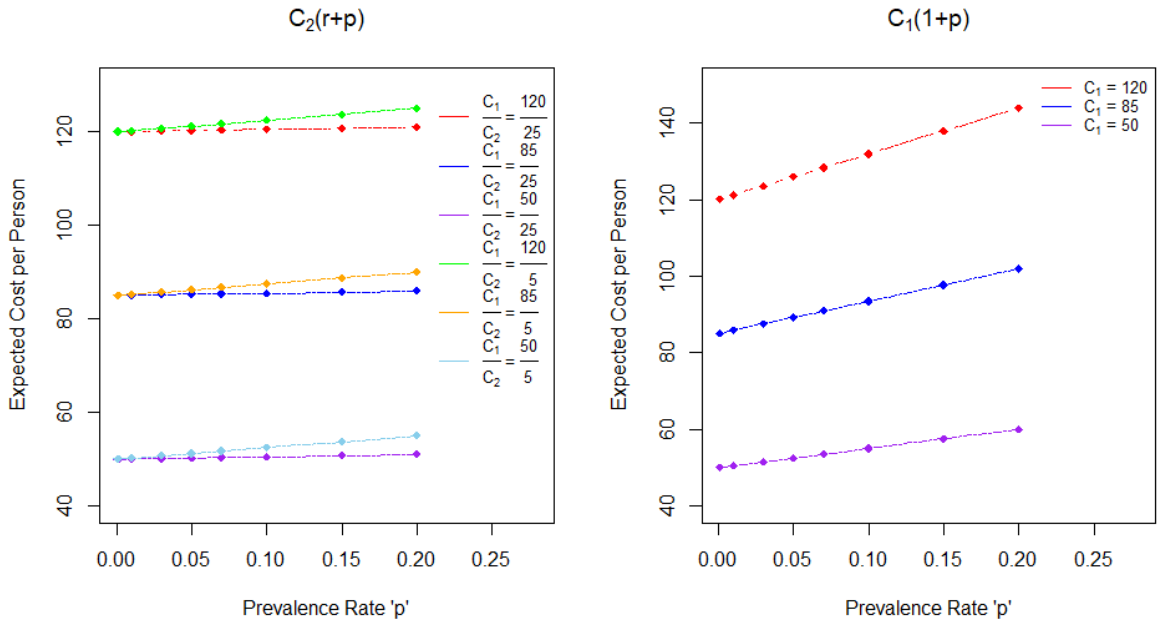


Figure 2. 6: Total expected cost per person against various prevalence rates for all the cost ratios $r = \frac{C_1}{C_2}$ (Left side) and all cost parameter C_1(Right side) when sensitivity is 100% in Protocol III(a) and III(b).

Table 2. 11: Expected cost per person of protocol III(b) for different p and C_1 when Se =100% and 97.2%

	Protocol III(b) $E(C\|p,Se) = C_1(1+p)$			Protocol III(b) $E(C\|p,Se) = C_1(1+Se.p)$		
	$C_1 = 120$	$C_1 = 85$	$C_1 = 50$	$C_1 = 120$	$C_1 = 85$	$C_1 = 50$
p=0.001	E(C)=120.12	E(C)=85.09	E(C)=50.05	E(C)=120.12	E(C)=85.08	E(C)=50.05
p=0.01	E(C)=121.20	E(C)=85.85	E(C)=50.50	E(C)=121.17	E(C)=85.83	E(C)=50.49
p=0.03	E(C)=123.60	E(C)=87.55	E(C)=51.51	E(C)=123.50	E(C)=87.48	E(C)=51.46
p=0.05	E(C)=126.00	E(C)=89.25	E(C)=52.50	E(C)=125.83	E(C)=89.13	E(C)=52.43
p=0.07	E(C)=128.40	E(C)=90.95	E(C)=53.50	E(C)=128.16	E(C)=90.78	E(C)=53.40
p=0.10	E(C)=132.00	E(C)=93.50	E(C)=55.00	E(C)=131.66	E(C)=93.26	E(C)=54.86
p=0.15	E(C)=138.00	E(C)=97.75	E(C)=57.50	E(C)=137.50	E(C)=97.39	E(C)=57.29
p=0.20	E(C)=144.00	E(C)=102.00	E(C)=60.00	E(C)=143.33	E(C)=101.52	E(C)=59.72

p = The prevalence rate; Se = Sensitivity
C_1 = The entire cost of testing a group of samples using a real-time RT-PCR test.
Protocol III(b) = When RT-PCR test is used for both stages. In Protocol III(b), we assumed $C_1 = C_3$ for simplicity.

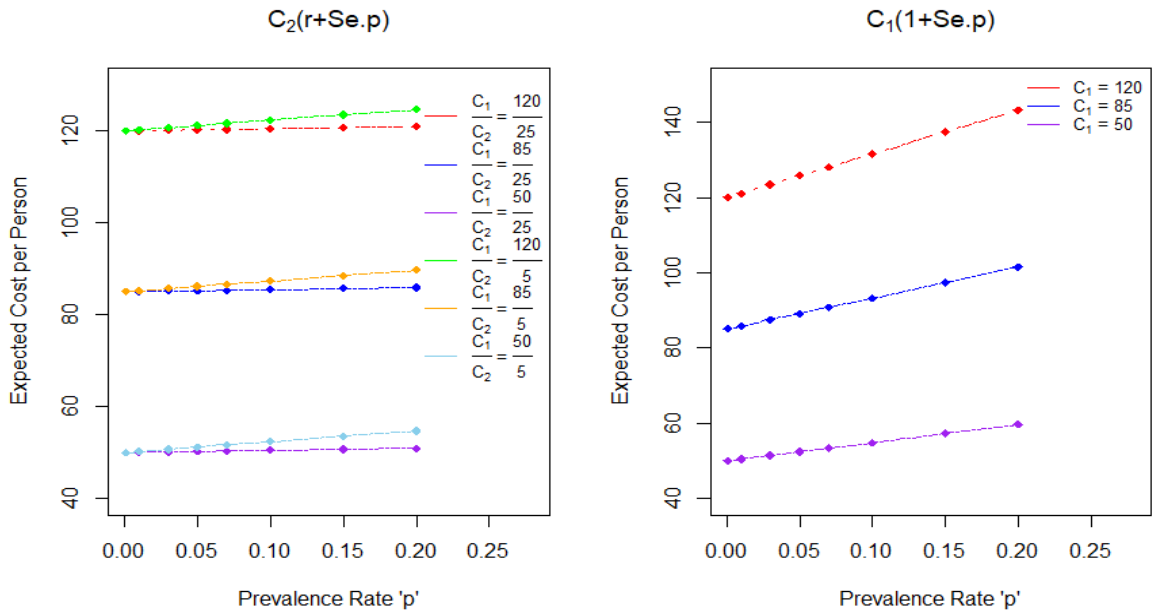


Figure 2. 7: Total expected cost per person against various prevalence rates for all the cost ratios $r = \frac{C_1}{C_2}$ (Left side) and all cost parameter C_1(Right side) when sensitivity is 93.3% (Left side) and 97.2% (Right side) in Protocol III(a) and III(b).

2.4.6 Protocol IV

Now, if the individuals are testing separately only one time (single-stage) to screen the disease, then the expected cost per person will be equal to the testing cost only. If an RT-PCR test is used which is named protocol IV(a), the E(C) will be \$120 or \$85 or \$50 (Equation 16). Otherwise, if a rapid antigen test is used which is named protocol IV(b), then E(C) will be \$5 or \$25 (Equation 17). Among all the individual strategies, we can see that single-stage individual testing using a rapid antigen test is the cheapest one.

2.5 Relative Cost

2.5.1 Protocol I and Protocol III(a)

The cost of group testing protocol I (without dilution) relative to the cost of the individual testing protocol III(a) is defined as $\frac{C_2(\frac{r}{n}+(1-(1-p)^n))}{C_2(r+Se.p)} = \frac{\frac{r}{n}+(1-(1-p)^n)}{r+Se.p}$.

Table 2.12 shows that protocol I is saving money in all cases while compared to protocol III(a). Protocol I is the best when the prevalence of a disease is 1 in 1000 people and the cost ratio is 24 or 17**.** In these two cases, group testing is costing only 1% of the cost of individual testing. Group testing becomes slightly worse when 20 among 1000 people are affected and r is 2. Even though protocol I is worse in this case, but still, it saves 51% of the cost. Generally, for higher prevalence rates, individual testing is more desirable to use for screening more accurately. But, in this case, we can see group testing is saving almost half of the money. So in terms of cost, here group testing is preferred over individual testing. However, in terms of sensitivity, we know individual testing usually produces fewer test errors than group testing. In this situation, it arises a question of which protocol will be best to use in terms of having fewer errors with low cost. This matter will be discussed in the next chapter.

Table 2. 12: The cost of protocol I (d=0) compared to the cost of protocol III(a) for different values of p and $\frac{C_1}{C_2}$

Relative Cost=$\frac{C_2(\frac{r}{n}+(1-(1-p)^n))}{C_2(r+Se.p)}$ when d=0							
		$Cost\ Ratio, r = \frac{C_1}{C_2}$					
p	**Se**	$\frac{120}{5} = 24$	$\frac{85}{5} = 17$	$\frac{50}{5} = 10$	$\frac{120}{25} = 4.8$	$\frac{85}{25} = 3.4$	$\frac{50}{25} = 2$
0.001	**100%**	0.012	0.015	0.019	0.028	0.034	0.044
	93.3%	0.012	0.015	0.019	0.028	0.034	0.044
0.01	**100%**	0.035	0.043	0.058	0.086	0.103	0.136
	93.3%	0.035	0.043	0.058	0.086	0.103	0.136
0.03	**100%**	--	0.066	0.093	0.142	0.172	0.227
	93.3%	--	0.066	0.093	0.142	0.172	0.228
0.05	**100%**	--	--	0.111	0.176	0.215	0.286
	93.3%	--	--	0.111	0.176	0.215	0.286
0.07	**100%**	--	--	--	0.201	0.247	0.330
	93.3%	--	--	--	0.201	0.247	0.331
0.10	**100%**	--	--	--	0.227	0.283	0.382
	93.3%	--	--	--	0.228	0.283	0.383
0.15	**100%**	--	--	--	--	--	0.444
	93.3%	--	--	--	--	--	0.446
0.20	**100%**	--	--	--	--	--	0.487
	93.3%	--	--	--	--	--	0.490

p = The prevalence rate; Se = Sensitivity; d = Dilution parameter.
C_1 = The entire cost of testing a group of samples using a real-time RT-PCR test.
C_2 = The entire cost (without lab cost) of testing an individual using a rapid antigen test.
C_3 = The entire cost of testing an individual using a real-time RT-PCR test.
Protocol I = When RT-PCR test is used at the first stage without dilution and rapid antigen test is used at the second stage.
Protocol III = When RT-PCR test is used for individual testing (1st stage) and rapid antigen test at 2nd stage is used again for individual testing to reconfirm. In Protocol III, we assumed $C_1 = C_3$ for simplicity.

The cost of group testing protocol I (with dilution) in terms of the cost of the individual testing protocol III (a) is defined as $\frac{C_2(\frac{r}{n}+\frac{p}{1-(1-p)^{n^d}}(1-(1-p)^n))}{C_2(r+Se.p)} = \frac{\frac{r}{n}+\frac{p}{1-(1-p)^{n^d}}(1-(1-p)^n)}{r+Se.p}$.

With dilution, group testing protocol I is also saving money in all cases (Table 2.13). Here, with 15 affected people among 1000 people, protocol I is saving 59% of the total cost of individual testing when the cost ratio is 2. While prevalence is the lowest and r is 24 and 17, protocol I is saving 90% money. Likewise without dilution effect, protocol I with dilution is saving money for all prevalences. Now it is a matter of discussion whether group testing protocol is producing fewer errors than individual testing with these costs. We will explore this matter in the next chapter.

Table 2. 13: The cost of protocol I (d≠0) compared to the cost of protocol III(a) for different values of p and $\frac{C_1}{C_2}$

		Relative Cost = $\frac{\frac{r}{n}+\frac{p}{1-(1-p)^{n^d}}(1-(1-p)^n)}{r+Se.p}$ when d=0.075					
		$Cost\ Ratio, r = \frac{C_1}{C_2}$					
p	**Se**	$\frac{120}{5} = 24$	$\frac{85}{5} = 17$	$\frac{50}{5} = 10$	$\frac{120}{25} = 4.8$	$\frac{85}{25} = 3.4$	$\frac{50}{25} = 2$
0.001	**100%**	0.010	0.012	0.016	0.024	0.029	0.038
	93.3%	0.010	0.012	0.016	0.024	0.029	0.038
0.01	**100%**	0.029	0.036	0.049	0.075	0.091	0.122
	93.3%	0.029	0.036	0.049	0.075	0.091	0.122
0.03	**100%**	--	--	--	0.125	0.154	0.207
	93.3%	--	--	--	0.125	0.154	0.207
0.05	**100%**	--	--	--	0.156	0.193	0.262
	93.3%	--	--	--	0.156	0.193	0.262
0.07	**100%**	--	--	--	--	0.222	0.304
	93.3%	--	--	--	--	0.222	0.305
0.10	**100%**	--	--	--	--	--	0.352
	93.3%	--	--	--	--	--	0.354
0.15	**100%**	--	--	--	--	--	0.409
	93.3%	--	--	--	--	--	0.411
0.20	**100%**	--	--	--	--	--	--
	93.3%	--	--	--	--	--	--

p = The prevalence rate; Se = Sensitivity; d = Dilution parameter.
C_1 = The entire cost of testing a group of samples using a real-time RT-PCR test.
C_2 = The entire cost (without lab cost) of testing an individual using a rapid antigen test.
C_3 = The entire cost of testing an individual using a real-time RT-PCR test.
Protocol I = When RT-PCR test is used at the first stage with dilution and rapid antigen test is used at the second stage.

Protocol III(a) = When RT-PCR test is used for individual testing (1st stage) and rapid antigen test at 2nd stage is used again for individual testing to reconfirm. In Protocol III, we assumed $C_1 = C_3$ for simplicity.

2.5.2 Protocol II and Protocol III(b)

The cost of group testing protocol II (without dilution) in terms of the cost of the individual testing protocol III (b) is defined as $\frac{C_1[\frac{1}{n}+(1-(1-p)^n)]}{C_1(1+Se.p)} = \frac{\frac{1}{n}+(1-(1-p)^n)}{1+Se.p}$.

The cost of group testing protocol II (with dilution) in terms of the cost of the individual testing protocol III (b) is defined as $\frac{C_1[\frac{1}{n}+(\frac{p}{1-(1-p)^{nd}})(1-(1-p)^n)]}{C_1(1+Se.p)} = \frac{\frac{1}{n}+(\frac{p}{1-(1-p)^{nd}})(1-(1-p)^n)}{1+Se.p}$.

Table 2.14 also implies that if we use protocol II with and without dilution, then in all cases it will save money while compared to the cost of the two-stage individual testing protocol using only PCR test.

Table 2. 14: The cost of protocol II (d=0 and d≠0) compared to the cost of protocol III(b) for different values of p and C_1

		Relative Cost= $\frac{\frac{1}{n}+(1-(1-p)^n)}{1+Se.p}$ when d=0			Relative Cost= $\frac{\frac{1}{n}+(\frac{p}{1-(1-p)^{nd}})(1-(1-p)^n)}{1+Se.p}$ when d=0.075		
p	Se	$C_1 = 120$	$C_1 = 85$	$C_1 = 50$	$C_1 = 120$	$C_1 = 85$	$C_1 = 50$
0.001	100%	0.06	0.06	0.06	0.05	0.05	0.05
	97.2%	0.06	0.06	0.06	0.05	0.05	0.05
0.01	100%	0.19	0.19	0.19	0.18	0.18	0.18
	97.2%	0.19	0.19	0.19	0.18	0.18	0.18
0.03	100%	0.32	0.32	0.32	0.30	0.30	0.30
	97.2%	0.32	0.32	0.32	0.30	0.30	0.30
0.05	100%	0.40	0.40	0.40	0.38	0.38	0.38
	97.2%	0.40	0.40	0.40	0.38	0.38	0.38
0.07	100%	0.47	0.47	0.47	0.44	0.44	0.44
	97.2%	0.47	0.47	0.47	0.44	0.44	0.44
0.10	100%	0.54	0.54	0.54	0.51	0.51	0.51
	97.2%	0.54	0.54	0.54	0.51	0.51	0.51
0.15	100%	0.62	0.62	0.62	0.59	0.59	0.59
	97.2%	0.63	0.63	0.63	0.60	0.60	0.60
0.20	100%	0.68	0.68	0.68	0.65	0.65	0.65
	97.2%	0.69	0.69	0.69	0.66	0.66	0.66

p = The prevalence rate; Se = Sensitivity; d = Dilution parameter.
C_1 = The entire cost of testing a group of samples using a real-time RT-PCR test.
C_3 = The entire cost of testing an individual using a real-time RT-PCR test.
Protocol II = When RT-PCR test is used both for group testing (1st stage) and individual testing (2nd stage) without and with dilution effect. In protocol II, we assumed $C_1 = C_3$.

Protocol III(b) = When RT-PCR test is used for both stages of individual testing. Here, we assumed $C_1 = C_3$ for simplicity.

2.5.3 Protocol I without Dilution and Protocol IV (a)

The cost of group testing protocol I (without dilution) relative to the cost of the individual testing protocol IV(a) is defined as $\frac{C_2(\frac{r}{n}+(1-(1-p)^n))}{C_1}$.

From Table 2.15, we can see, group testing protocol I without dilution is saving money for all the lower and higher prevalence rates. As p goes up, the amount of money saved by protocol I compared to the one-stage individual testing using PCR test is going down. For example when p is 0.1% and r is 2 and $C_1 = 120$, protocol I is saving 98% of the total cost per person of protocol IV (a). Whereas when p becomes 20%, for the same cost ratio, this amount comes down to 78%.

Table 2. 15: The cost of protocol I (d=0) compared to the cost of protocol IV(a) for different values of p and $\frac{C_1}{C_2}$

Relative Cost=$\frac{C_2(\frac{r}{n}+(1-(1-p)^n))}{C_1}$ when d=0							
		$Cost\ Ratio, r = \frac{C_1}{C_2}$					
p	C_1	$\frac{120}{5} = 24$	$\frac{85}{5} = 17$	$\frac{50}{5} = 10$	$\frac{120}{25} = 4.8$	$\frac{85}{25} = 3.4$	$\frac{50}{25} = 2$
0.001	**120**	0.01	0.01	0.01	0.03	0.02	0.02
	85	0.02	0.01	0.01	0.04	0.03	0.03
	50	0.03	0.03	0.02	0.07	0.06	0.04
0.01	**120**	0.04	0.03	0.02	0.09	0.07	0.06
	85	0.05	0.04	0.03	0.12	0.10	0.08
	50	0.08	0.07	0.06	0.21	0.18	0.14
0.03	**120**	--	0.05	0.04	0.14	0.12	0.10
	85	--	0.07	0.05	0.20	0.17	0.14
	50	--	0.11	0.09	0.34	0.29	0.23
0.05	**120**	--	--	0.05	0.18	0.15	0.12
	85	--	--	0.07	0.25	0.22	0.17
	50	--	--	0.11	0.43	0.37	0.29
0.07	**120**	--	--	--	0.20	0.18	0.14
	85	--	--	--	0.29	0.25	0.20
	50	--	--	--	0.49	0.43	0.34
0.10	**120**	--	--	--	0.23	0.21	0.17
	85	--	--	--	0.33	0.29	0.24
	50	--	--	--	0.56	0.50	0.40
0.15	**120**	--	--	--	--	--	0.20
	85	--	--	--	--	--	0.28
	50	--	--	--	--	--	0.48
0.20	**120**	--	--	--	--	--	0.22

	85	--	--	--	--	--	0.31
	50	--	--	--	--	--	0.54

p = The prevalence rate; d = Dilution parameter.
C_1 = The entire cost of testing a group of samples using a real-time RT-PCR test.
C_2 = The entire cost (without lab cost) of testing an individual using a rapid antigen test.
C_3 = The entire cost of testing an individual using a real-time RT-PCR test.
Protocol I = When RT-PCR test is used at the first stage without dilution and rapid antigen test is used at the second stage.
Protocol IV(a) = When RT-PCR test is used for a single stage individual testing. In Protocol IV(a), we assumed $C_1 = C_3$.

2.5.4 Protocol I with Dilution and Protocol IV (a)

The cost of group testing protocol I (with dilution) in terms of the cost of the individual testing protocol IV (a) (one-stage) is defined as $\frac{C_2(\frac{r}{n}+\frac{p}{1-(1-p)^{nd}}(1-(1-p)^n))}{C_1}$.

Likewise, protocol I without dilution, here from Table 2.16, we can see, group testing protocol I with dilution is saving money for all the lower and higher prevalence rates. As p goes up, the amount of money saved by protocol I compared to the one-stage individual testing using PCR test is going down. For example when p is 0.1% and r is 2 and $C_1 = 120$, protocol I is saving 98% of the total cost per person of protocol IV (a). Whereas when p becomes 15%, for the same cost ratio, this amount comes down to 82%.

Table 2. 16: The cost of protocol I (d≠0) compared to the cost of protocol IV(a) for different values of p and $\frac{C_1}{C_2}$

Relative Cost=$\frac{C_2(\frac{r}{n}+\frac{p}{1-(1-p)^{nd}}(1-(1-p)^n))}{C_1}$ when d=0.075							
		$Cost\ Ratio, r = \frac{C_1}{C_2}$					
p	C_1	$\frac{120}{5} = 24$	$\frac{85}{5} = 17$	$\frac{50}{5} = 10$	$\frac{120}{25} = 4.8$	$\frac{85}{25} = 3.4$	$\frac{50}{25} = 2$
0.001	**120**	0.01	0.01	0.01	0.02	0.02	0.02
	85	0.01	0.01	0.01	0.03	0.03	0.02
	50	0.02	0.02	0.02	0.06	0.05	0.04
0.01	**120**	0.03	0.03	0.02	0.08	0.06	0.05
	85	0.04	0.04	0.03	0.11	0.09	0.07
	50	0.07	0.06	0.05	0.18	0.16	0.12
0.03	**120**	--	--	--	0.13	0.11	0.08
	85	--	--	--	0.18	0.15	0.12
	50	--	--	--	0.30	0.26	0.20
0.05	**120**	--	--	--	0.16	0.14	0.11
	85	--	--	--	0.22	0.20	0.16
	50	--	--	--	0.38	0.33	0.27
0.07	**120**	--	--	--	--	0.16	0.13
	85	--	--	--	--	0.23	0.19
	50	--	--	--	--	0.38	0.31

0.10	**120**	--	--	--	--	--	0.15
	85	--	--	--	--	--	0.22
	50	--	--	--	--	--	0.37
0.15	**120**	--	--	--	--	--	0.18
	85	--	--	--	--	--	0.26
	50	--	--	--	--	--	0.44
0.20	**120**	--	--	--	--	--	--
	85	--	--	--	--	--	--
	50	--	--	--	--	--	--

p = The prevalence rate; d = Dilution parameter.
C_1 = The entire cost of testing a group of samples using a real-time RT-PCR test.
C_2 = The entire cost (without lab cost) of testing an individual using a rapid antigen test.
C_3 = The entire cost of testing an individual using a real-time RT-PCR test.
Protocol I = When RT-PCR test is used at the first stage with dilution and rapid antigen test is used at the second stage.
Protocol IV(a) = When RT-PCR test is used for a single stage individual testing. In Protocol IV(a), we assumed $C_1 = C_3$.

2.5.5 Protocol I without Dilution and Protocol IV (b)

The cost of group testing protocol I (without dilution) relative to the cost of the individual testing protocol IV(b) is defined as $\frac{C_2(\frac{r}{n}+(1-(1-p)^n))}{C_2} = \frac{r}{n} + (1-(1-p)^n)$.

Here from Table 2.17, we can see in some cases individual testing protocol IV with one stage using the rapid test is much cheaper than protocol I without dilution. For example, for lower and higher prevalences if C_2 *is* 5 then protocol IV (b) is much cheaper than protocol I when r ranges from 17 to 2. As prevalence is getting higher, for example, 20%, protocol IV (b) is cheaper than protocol I even if C_2 *is* 25. Hence, we can say, in some cases, the protocol I without dilution can be more expensive than protocol IV (b)

Table 2. 17: The cost of protocol I (d=0) compared to the cost of protocol IV(b) for different values of p and $\frac{C_1}{C_2}$

Relative Cost=$\frac{r}{n} + (1-(1-p)^n)$ when d=0							
		***Cost Ratio*, $r = \frac{C_1}{C_2}$**					
p	C_2	$\frac{120}{5} = 24$	$\frac{85}{5} = 17$	$\frac{50}{5} = 10$	$\frac{120}{25} = 4.8$	$\frac{85}{25} = 3.4$	$\frac{50}{25} = 2$
0.001	**5**	0.30	0.25	0.19	0.68	0.57	0.44
	25	0.06	0.05	0.04	0.14	0.11	0.09
0.01	**5**	0.85	0.73	0.58	2.07	1.76	1.37
	25	0.17	0.15	0.12	0.41	0.35	0.27
0.03	**5**	--	1.12	0.93	3.43	2.94	2.31
	25	--	0.22	0.19	0.69	0.59	0.46
0.05	**5**	--	--	1.12	4.28	3.70	2.93
	25	--	--	0.22	0.86	0.74	0.59

0.07	**5**	--	--	--	4.90	4.28	3.42
	25	--	--	--	0.98	0.86	0.68
0.10	**5**	--	--	--	5.57	4.95	4.01
	25	--	--	--	1.11	0.99	0.80
0.15	**5**	--	--	--	--	--	4.77
	25	--	--	--	--	--	0.95
0.20	**5**	--	--	--	--	--	5.35
	25	--	--	--	--	--	1.07

p = The prevalence rate; d = Dilution parameter.
C_1 = The entire cost of testing a group of samples using a real-time RT-PCR test.
C_2 = The entire cost (without lab cost) of testing an individual using a rapid antigen test.
Protocol I = When RT-PCR test is used at the first stage without dilution and rapid antigen test is used at the second stage; Protocol IV(b) = When a rapid antigen test is used for a single stage individual testing.
Red cells indicate when protocol I without dilution becomes expensive than protocol IV(b).

2.5.6 Protocol I with Dilution and Protocol IV (b)

The cost of group testing protocol I (with dilution) in terms of the cost of the individual testing protocol IV (b) is defined as $\frac{C_2(\frac{r}{n}+\frac{p}{1-(1-p)^{nd}}(1-(1-p)^n))}{C_2} = \frac{r}{n}+\frac{p}{1-(1-p)^{nd}}(1-(1-p)^n)$.

Here from Table 2.18, for the lower cost ratios, as the p goes up, and if we use \$5 rapid antigen for protocol IV(b), we can see protocol I with dilution is becoming costly than protocol IV(b). For example, for the lower cost ratios (4.8,3.4,2), when p becomes greater than 0.1%, individual testing using a rapid test costing only \$5 for one stage always spends less money than Protocol I with dilution. On the contrary, for the highest r such as 24,17 and 10 and lowest prevalences such as 0.1% and 1%, protocol I with dilution still spends less money than protocol IV (b). For example, when r is 24, p is 0.1% and C_2 *is* 25, the protocol I with dilution saves 95% of the total cost per person of protocol IV(b). It saves 90% when r is 10, p is 1% and C_2 *is* 25.

Table 2. 18: The cost of protocol I (d≠0) compared to the cost of protocol IV(b) for different values of p and $\frac{C_1}{C_2}$

Relative Cost = $\frac{r}{n}+\frac{p}{1-(1-p)^{nd}}(1-(1-p)^n)$ when d=0.075							
		***Cost Ratio*, $r = \frac{C_1}{C_2}$**					
p	C_2	$\frac{120}{5}=24$	$\frac{85}{5}=17$	$\frac{50}{5}=10$	$\frac{120}{25}=4.8$	$\frac{85}{25}=3.4$	$\frac{50}{25}=2$
0.001	**5**	0.24	0.21	0.16	0.57	0.49	0.38
	25	0.05	0.04	0.03	0.11	0.10	0.08
0.01	**5**	0.69	0.61	0.50	1.81	1.55	1.22
	25	0.14	0.12	0.10	0.36	0.31	0.24
0.03	**5**	--	--	--	3.03	2.63	2.03
	25	--	--	--	0.61	0.53	0.41

0.05	**5**	--	--	--	3.77	3.33	2.69
	25	--	--	--	0.75	0.67	0.54
0.07	**5**	--	--	--	--	3.85	3.15
	25	--	--	--	--	0.77	0.63
0.10	**5**	--	--	--	--	--	3.70
	25	--	--	--	--	--	0.74
0.15	**5**	--	--	--	--	--	4.40
	25	--	--	--	--	--	0.88
0.20	**5**	--	--	--	--	--	--
	25	--	--	--	--	--	--

p = The prevalence rate; d = Dilution parameter.
C_1 = The entire cost of testing a group of samples using a real-time RT-PCR test.
C_2 = The entire cost (without lab cost) of testing an individual using a rapid antigen test.
Protocol I = When RT-PCR test is used at the first stage with dilution and rapid antigen test is used at the second stage; Protocol IV(b) = When a rapid antigen test is used for a single stage individual testing.
Red cells indicate when protocol I with dilution becomes expensive than protocol IV(b).

2.5.7 Protocol II and Protocol IV(a)

The cost of group testing protocol II (without dilution) in terms of the cost of the individual testing protocol IV (a) is defined as $\frac{C_1[\frac{1}{n}+\,(1-(1-p)^n)]}{C_1} = \frac{1}{n} + \,(1-(1-p)^n)$.

The cost of group testing protocol II (with dilution) in terms of the cost of the individual testing protocol IV (b) is defined as $\frac{C_1[\frac{1}{n}+(\frac{p}{1-(1-p)^{nd}})\,(1-(1-p)^n)]}{C_1} = \frac{1}{n} + (\frac{p}{1-(1-p)^{nd}})\,(1-(1-p)^n)$.

If we look at Table 2.19, we can see if $C_1\ is\ 50$ in protocol II without dilution for all the prevalence rates, protocol II is less costly than protocol IV(a). As p goes up, the amount of money that protocol II without dilution saves is decreasing. For example, when p is 0.1% and $C_1\ is\ 50$ for both protocols, protocol II is saving 94% of the cost of protocol IV(a). When p becomes 20%, it saves only 18%. So, in those cases, protocol II without dilution is the recommended one in terms of cost. On the other hand, if $C_1\ is\ 120\ or\ 85$ in protocol II without dilution, when p ranges from 5% to 20%, protocol IV(a) costing \$85 or \$50 per person is cheaper than protocol II without dilution. For instance, when p is 15% or 20%, and when $C_1\ is\ 120$ in protocol II and $C_1\ is\ 85$ in protocol IV(a), protocol II without dilution costs 1.02 times or 1.16 times of the overall cost of protocol IV(a). So, in those cases, protocol IV(a) is the recommended one to use in terms of cost.

If we look at the costs of protocol II with dilution compared to the cost of protocol IV(a), similar to without dilution, for all the prevalence rates, protocol II is less costly than protocol IV(a) when $C_1\ is\ 50$ in protocol II. On the other hand, if $C_1\ is\ 120\ or\ 85$ in protocol II, when

p ranges from 7% to 20%, protocol IV(a) costing $50 per person is cheaper than protocol II. For instance, when p is 15% or 20%, and when C_1 *is* 85 in protocol II and C_1 *is* 50 in protocol IV(a), protocol II with dilution costs 1.16 times or 1.33 times of the overall cost of protocol IV(a). So, in those cases, protocol IV(a) is the recommended one to use in terms of cost.

Table 2. 19: The cost of protocol II (d=0 and d≠0) compared to the cost of protocol IV(a) for different values of p and C_1

		Relative Cost= $\frac{1}{n}+(1-(1-p)^n)$ when d=0			Relative Cost= $(\frac{p}{1-(1-p)^{n^d}})(1-(1-p)^n)$ when d=0.075		
p	**C_1**	**$C_1=120$**	**$C_1=85$**	**$C_1=50$**	**$C_1=120$**	**$C_1=85$**	**$C_1=50$**
0.001	**120**	0.06	0.04	0.03	0.05	0.04	0.02
	85	0.09	0.06	0.04	0.08	0.05	0.03
	50	0.15	0.11	0.06	0.13	0.09	0.05
0.01	**120**	0.20	0.14	0.08	0.18	0.13	0.07
	85	0.28	0.20	0.11	0.25	0.18	0.10
	50	0.47	0.33	0.20	0.43	0.30	0.18
0.03	**120**	0.33	0.24	0.14	0.31	0.22	0.13
	85	0.47	0.33	0.20	0.44	0.31	0.18
	50	0.80	0.57	0.33	0.74	0.53	0.31
0.05	**120**	0.43	0.30	0.18	0.40	0.28	0.17
	85	0.60	0.43	0.25	0.56	0.40	0.23
	50	1.02	0.72	0.43	0.96	0.68	0.40
0.07	**120**	0.50	0.35	0.21	0.47	0.33	0.20
	85	0.71	0.50	0.29	0.66	0.47	0.28
	50	1.20	0.85	0.50	1.13	0.80	0.47
0.10	**120**	0.59	0.42	0.25	0.56	0.40	0.23
	85	0.84	0.59	0.35	0.79	0.56	0.33
	50	1.42	1.01	0.59	1.35	0.95	0.56
0.15	**120**	0.72	0.51	0.30	0.68	0.48	0.28
	85	1.01	0.72	0.42	0.96	0.68	0.40
	50	1.72	1.22	0.72	1.64	1.16	0.68
0.20	**120**	0.82	0.58	0.34	0.78	0.56	0.33
	85	1.16	0.82	0.48	1.11	0.78	0.46
	50	1.97	1.40	0.82	1.88	1.33	0.78

p = The prevalence rate; d = Dilution parameter.
C_1 = The entire cost of testing a group of samples using a real-time RT-PCR test.
C_2 = The entire cost (without lab cost) of testing an individual using a rapid antigen test.
C_3 = The entire cost of testing an individual using a real-time RT-PCR test.
Protocol II = When RT-PCR test is used both for group testing (1st stage) and individual testing (2nd stage) without and with dilution effect. In protocol II, we assumed $C_1 = C_3$.
Protocol IV(a) = When RT-PCR test is used for a single stage individual testing. In Protocol IV(a), we assumed $C_1 = C_3$.
Red cells indicate when protocol II without and with dilution becomes expensive than protocol IV(a).

2.5.8 Protocol II and Protocol IV(b)

The cost of group testing protocol II (without dilution) in terms of the cost of the individual testing protocol IV (a) is defined as $\frac{C_1[\frac{1}{n}+(1-(1-p)^n)]}{C_2}$.

The cost of group testing protocol II (with dilution) in terms of the cost of the individual testing protocol III (b) is defined as $\frac{C_1[\frac{1}{n}+(\frac{p}{1-(1-p)^n d})(1-(1-p)^n)]}{C_2}$.

If we compare the cost of protocol II with and without dilution effect to the cost of the individual testing protocol with one stage using a rapid test, we can see for most of the prevalence rates and most of the combinations of C_1 and C_2, protocol IV(b) is cheaper than the cost of protocol II. However, if we use \$25 rapid antigen test for protocol IV(b) and 0.1% and 1% are prevalence rates, then group testing protocol II with and without dilution will be the cheaper one for all values of C_1. Also, for some other cases which are few in numbers, protocol II is cheaper than protocol IV(b). Such as, without dilution when C_1 is 50 and C_2 is 25, protocol II is recommended to use when p is 0.01 and 0.05. With dilution, protocol II is cheap when C_1 is 50 and C_2 is 25 for 3%,5% and 7% prevalences.

Table 2. 20: The cost of protocol II (d=0 and d≠0) compared to the cost of protocol IV(b) for different values of p and C_1

		Relative Cost= $\frac{C_1[\frac{1}{n}+(1-(1-p)^n)]}{C_2}$ when d=0			Relative Cost= $\frac{C_1[\frac{1}{n}+(\frac{p}{1-(1-p)^n d})(1-(1-p)^n)]}{C_2}$ when d=0.075		
p	C_2	$C_1 = 120$	$C_1 = 85$	$C_1 = 50$	$C_1 = 120$	$C_1 = 85$	$C_1 = 50$
0.001	**5**	1.51	1.07	0.63	1.31	0.93	0.55
	25	0.30	0.21	0.13	0.26	0.19	0.11
0.01	**5**	4.69	3.32	1.95	4.27	3.02	1.78
	25	0.94	0.66	0.39	0.85	0.60	0.36
0.03	**5**	8.00	5.67	3.33	7.42	5.26	3.09
	25	1.60	1.13	0.67	1.48	1.05	0.62
0.05	**5**	10.23	7.25	4.26	9.57	6.78	3.99
	25	2.05	1.45	0.85	1.91	1.36	0.80
0.07	**5**	12.01	8.51	5.01	11.30	8.01	4.71
	25	2.40	1.70	1.00	2.26	1.60	0.94
0.10	**5**	14.23	10.08	5.93	13.47	9.54	5.61
	25	2.85	2.02	1.19	2.69	1.91	1.12
0.15	**5**	17.23	12.21	7.18	16.40	11.62	6.83
	25	3.45	2.44	1.44	3.28	2.32	1.37
0.20	**5**	19.71	13.96	8.21	18.82	13.33	7.84
	25	3.94	2.79	1.64	3.76	2.67	1.57

p = The prevalence rate; d = Dilution parameter.
C_1 = The entire cost of testing a group of samples using a real-time RT-PCR test.

C_2 = The entire cost (without lab cost) of testing an individual using a rapid antigen test.
C_3 = The entire cost of testing an individual using a real-time RT-PCR test.
Protocol II = When RT-PCR test is used both for group testing (1st stage) and individual testing (2nd stage) without and with dilution effect. In protocol II, we assumed $C_1 = C_3$.
Protocol IV(b) = When a rapid antigen test is used for a single stage individual testing.
Red cells indicate when protocol II without and with dilution becomes expensive than protocol IV(b).

Overall, we can say, group testing protocols I and II both are cost-effective protocols compared to the two stages of individual testing protocols for both the highest and lowest prevalence rates. So we can say, even for higher prevalence rates group testing strategy (I and II) will be the cheaper one than protocol III (two-stages). Protocol I is also cost-effective compared to single-stage individual testing protocol with PCR only for all types of prevalences. So, for higher and lower prevalence rates, protocol I is recommended over protocol IV(a) in terms of cost. However, protocol IV(a) is recommended over group testing protocol II especially for higher prevalence rates if we want to apply a less expensive protocol. We also have found that single-stage individual testing protocol with the rapid test is cost-effective than both group testing protocol I (especially for the lowest cost ratios) and II (almost for all the cost parameters) for both lower and higher prevalences. So, Protocol IV(b) can be recommended over group testing protocols in some situations.

In this chapter, we have studied how much each protocol I and II costs with the optimal group size and in which situations those are cheaper than individual testing protocols and vice versa. However, we also need to keep in mind that each protocol creates test errors such as false-negative test results in our thesis. Since we have assumed Sp =1, our next step will be to estimate the expected number of false-negative errors for each protocol. Then, we will explore which one is making fewer false-negative errors costing less than the other protocols for a particular prevalence rate. These two problems will be discussed in our next chapter.

Chapter 3 Comparing Protocols With Respect To The Costs and False-Negative Test Results

3.1 Classification Error

A common assumption of the group testing strategy is that testing a group or an individual can be performed without the occurrence of any classification errors. However, in practice, there is always a possibility of having inaccurate test results. There are mainly two forms of classification errors: a) a group or an individual is good but identified as a defective item after the test, known as False Negative (FN); and b) a group or an individual is defective but identified as a good item after the test, known as False Positive(FP). FN occurs when sensitivity is not 100% accurate and FP occurs when specificity is not 100% accurate.

A group testing design that has a higher probability of having FN cases can be unfavourable to societal well-being in terms of screening infectious diseases. The consequence of the FN cases is very harmful since missing true cases will infect the others within the community which will eventually increase the likelihood of spreading the disease. The situation could be worse if the infected person is asymptomatic. On the other hand, because of FP test results, uninfected individuals have to be isolated unnecessarily for a few days which hamper their daily activities, and they need to be in contact with medical officers unnecessarily all the time. So, these facts indicate that in COVID FN test results are alarming to the social level more while FP test results are more alarming to the personal level.Additionally, FP rates are very low in screening for COVID (Prince-Guerra et al., 2021), (Pilarowski et al., 2021). Haber et al. (2021) also assumed perfect specificity for their objective function to obtain the optimal group testing design. For all these reasons and because of simplicity, we are assuming each test result has perfect specificity in our work which means no false-positive test result occurs.

False-negative test result reduces the overall cost of a group testing design carried out with more than one stage since the number of individual testing at the second stage would be less. Besides, getting FN test results means the sensitivity of a test is not accurate. So, the more the FN cases, the less the accuracy of a test. Therefore, to achieve an efficient design, we need to assess the trade-off between the cost of each protocol and the number of FN cases made by each protocol which is the main discussion of this chapter.

3.2 False-Negative Test Result or Error

3.2.1 Expected False Negative Cases in A Group

Let t be the number of true positive cases in a group of size n which follows a binomial distribution with parameters n and p.

Thus, the probability of having t true positive cases in a group is $\Pr(\mathrm{T} = \mathrm{t}) = \binom{\mathrm{n}}{\mathrm{t}} p^t (1-p)^{n-t}; \quad t = 0,1, \dots \dots, n.$

Now, the expected number of true positive cases in a group is $E(T = t) = np$

The expected number of FN cases in a group given that the group has at least one true positive case is (EFN) = $\sum_{t=1}^{n} tPr(T = t)(1 - Se(n))$

As a first approximation, we will assume that there would be at most 1 true positive case in a group of size n for a low prevalence rate. It is because two or more true cases are more unlikely for low prevalence. However, if there are no true cases in a group, then the group is negative and as such there will be no FN case. In other words, for t = 0, $EFN = 0$.

Now, for t = 1, $EFN = 1 \times \Pr(T = 1)\big(1 - Se(n)\big) = 1 \times \binom{\mathrm{n}}{1} p^1 (1-p)^{n-1}\big(1 - Se(n)\big)$

$$= np(1-p)^{n-1}\big(1 - Se(n)\big)$$

$$\cong np\big(1 - Se(n)\big) \text{ since } (1-p) \cong 1$$

Thus, $EFN \cong np\big(1 - Se(n)\big)$

Now, we will find the total number of false-negative cases that occurred in four protocols for various prevalence rates (0.001,0.01,0.03,0.05,0.07,0.10,0.15,0.20).

3.2.2 False-Negative Case Count

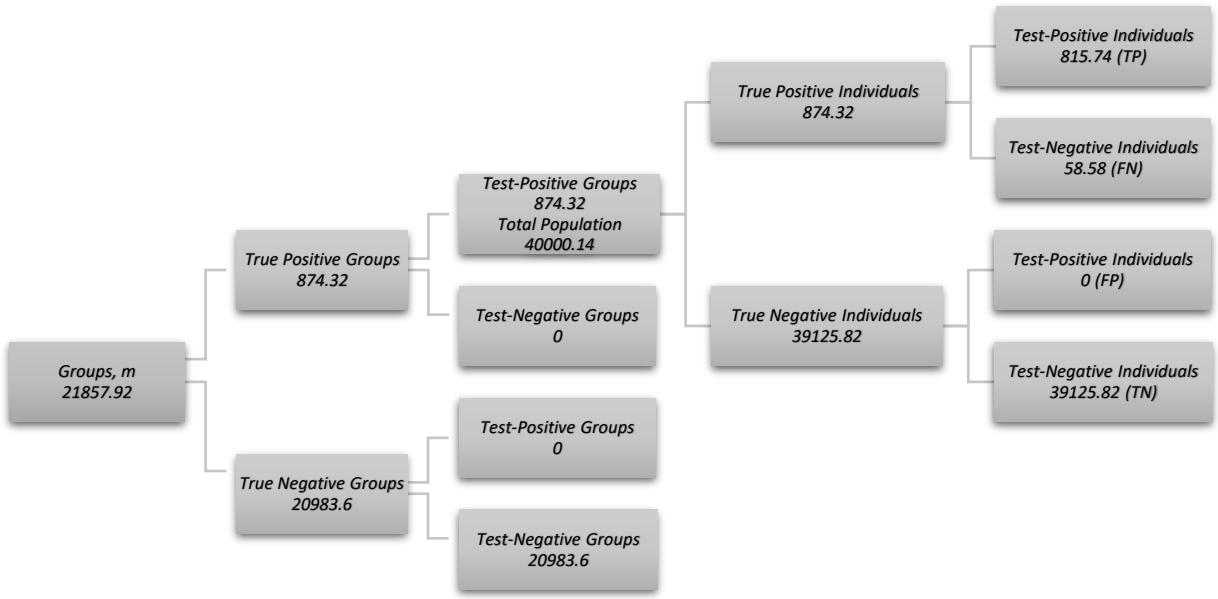
3.2.2.1 Protocol I when d = 0

Let us assume we are testing one million population (N) to screen the disease. Let us consider only 1 individual is infected among 1000 people. For example, let us select the optimal group size 45.75 from Table 2.4 (Chapter 2) when p is 0.001 and $\frac{C_1}{C_2} = \frac{50}{25}$ to illustrate how we have calculated the FN cases. Thus, our total number of groups will be, m = $\frac{1000000}{45.75} = 21857.92$.

Now, the expected number of true positive groups is = mp'= $\mathrm{m}(1 - (1-p)^n)$ =

$21857.92 \times (1 - (1 - 0.001)^{45.75}) = 874.32$ and the expected number of true negative groups is = (21857.92 - 874.32) =20983.60. Since specificity is 1, all true negative groups will be identified correctly as negative groups at the first stage. Se(45.75) will be 1 as d=0..Thus all true positive groups will be tested as positive. The expected number of tested positive (True positive, TP) groups is $874.32 \times 1 = 874.32$ and the test negative (FN) groups are 874.32 - 874.32 = 0.

Now for the second stage, the total population in the test-positive groups are $(874.32 \times 45.75) = 40000.14$. As each group has approximately one true case, the total expected number of positive individuals in those groups is $(874.32 \times 1) = 874.32$. The sensitivity of the rapid antigen test for individual testing is 93.3% (Pilarowski et al., 2021). Thus, the expected number of individuals that test positive (TP) is $(874.32 \times 0.933) = 815.74$ which are shown in Figure (3.1).The rest of them are false-negative individuals (874.32-815.74 = 58.58). As Sp=1, there are no false-positive cases at the second stage. So, the expected number of FP individuals will be 0 at the second stage. Our total FN cases are 58.58 for this example.



Prevalence rate: 0.1% and **Proportion of having a positive group** 0.04. **Sensitivity for a group:** Se(n) =Se(45.75)=1 (RT- PCR Test); **Sensitivity:** Se = 93.3% (Rapid Antigen Test) and **Specificity**: Sp(n) = Sp(45.75) = Sp =1. **Protocol I** = When RT-PCR test is used at the first stage without dilution and rapid antigen test is used at the second stage.

Figure 3. 1: When p=0.001 and n=45.75 in Protocol I when d=0, the expected frequency tree of false-negative cases for both first and second stages in one million people.

Similarly, we have estimated the total number of false-negative individuals of Protocol I without dilution effect for both stages separately for other prevalence rates and cost ratios which are shown in Table 3.1.

Table 3. 1: Summary results of EFN Individuals of Protocol I without dilution for all prevalence rates and cost ratios

p	n	p'	r = $\frac{c_1}{c_2}$	E(C)	Se(n)	FN Individuals		Total FN Individuals
						Stage 1	Stage 2 Se = 93.3%	
	168.5	0.16	24	1.48	1	0	63.6	63.6
	139.8	0.13	17	1.26	1	0	62.3	62.3
	105.4	0.10	10	0.97	1	0	63.6	63.6
	71.8	0.07	4.8	3.40	1	0	65.3	65.3
	60.1	0.06	3.4	2.87	1	0	66.9	66.9
0.001	45.8	0.04	2	2.21	1	0	58.6	58.6
	69.2	0.50	24	4.24	1	0	484.2	484.2
	53.9	0.42	17	3.67	1	0	521.8	521.8
	38.2	0.32	10	2.90	1	0	561.0	561.0
	24.8	0.22	4.8	10.4	1	0	595.6	595.6
	20.4	0.19	3.4	8.80	1	0	624.6	624.6
0.01	15.2	0.14	2	6.83	1	0	615.9	615.9
	52.8	0.80	17	5.61	1	0	1015.7	1015.7
	27.6	0.57	10	4.65	1	0	1384.7	1384.7
	16.0	0.39	4.8	17.1	1	0	1631.1	1631.1
	12.9	0.32	3.4	14.7	1	0	1668.5	1668.5
0.03	9.34	0.25	2	11.5	1	0	1793.4	1793.4
	30.6	0.79	10	5.59	1	0	1728.0	1728.0
	13.8	0.51	4.8	21.4	1	0	2481.5	2481.5
	10.7	0.42	3.4	18.5	1	0	2625.0	2625.0
0.05	7.59	0.32	2	14.7	1	0	2824.8	2824.8
	13.1	0.61	4.8	24.5	1	0	3127.0	3127.0
0.07	9.8	0.51	3.4		1	0	3504.6	3504.6

				21.4				
	6.69	0.38	2	17.1	1	0	3805.7	3805.7
	14.5	0.78	4.8	27.9	1	0	3616.6	3616.6
	9.25	0.62	3.4	24.8	1	0	4490.8	4490.8
0.10	5.97	0.47	2	20.1	1	0	5274.7	5274.7
0.15	5.47	0.59	2	23.9	1	0	7226.7	7226.7
0.20	5.58	0.71	2	26.8	1	0	8525.1	8525.1

p=Prevalence rate; $\boldsymbol{p'}$ = Proportion of having a positive group; **n**= Optimal group size; **r** = cost ratio = $\frac{C_1}{C_2}$; **Se(n)** = Sensitivity depending on the group size = $\frac{p}{1-(1-p)^{nd}} = 1 \; as \; d = 0$.

$\boldsymbol{E(C|p,n,Se(n))} = Minimum\ Expected\ cost\ per\ person = C_2\left(\frac{C_1}{C_2 n} + (1-(1-p)^n)\right)$.

$\boldsymbol{C_1}$ = The entire (lab and test) cost of testing a group of samples using a real-time RT-PCR test kit.
$\boldsymbol{C_2}$ = The cost (without lab) of testing an individual using a rapid antigen test kit.
Protocol I = When RT-PCR test is used at the first stage without dilution and rapid antigen test is used at the second stage

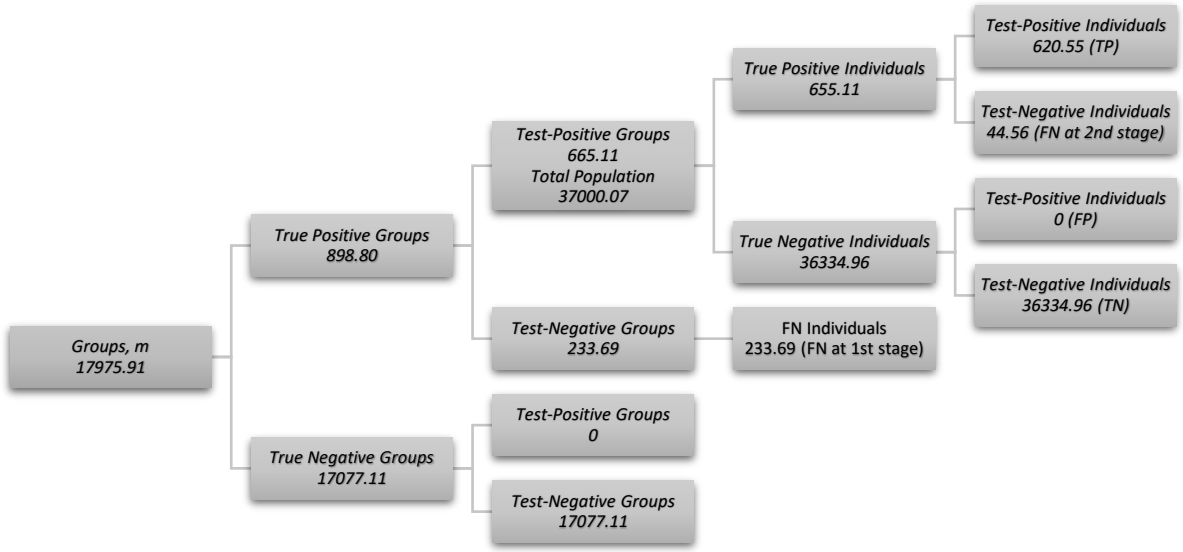
We have assumed the sensitivity of the rapid test at the second stage as 93.3% (Pilarowski et al., 2021). In Table 3.1, we can see with the increasing prevalence, FN cases are also becoming greater. For example, when only 1 person among 1000 is affected and r =2, only around 58 true cases are missed in total. However, if 20 people among 100 are affected, then this number goes beyond 8500. This is because, with the growing prevalences, the proportion of having positive groups is also going up. As such at the second stage, the number of testing individuals becomes higher as well which increases the chance of getting FN errors. Besides,. the overall cost of this testing strategy is also getting higher with the increasing prevalence. It concludes that when prevalence is higher, both these factors get worse simultaneously which makes this protocol less advantageous to use in practice.

3.2.2.2 Protocol I when d ≠ 0

The only difference in Protocol I with dilution effect is that while calculating the total number of FN cases, the accuracy of diagnosing a group at the first stage is not perfect. Thus, we have to calculate the number of missing true cases at this stage as well.

For example, let us select the optimal group size 55.63 from Table 2.5 when p is 0.001 and $\frac{C_1}{C_2} = \frac{50}{25}$ to illustrate how we have calculated the FN cases. Thus, our total number of groups will be, m = $\frac{1000000}{55.63} = 17975.91$. Now, the expected number of true positive groups is = mp'= m$(1-(1-p)^n)$ = $17975.91 \times (1-(1-0.001)^{55.63}) = 898.80$. Sensitivity depending on

n = 55.63 is $Se(n) = Se(55.63) = \frac{p}{1-(1-p)^{nd}} = \frac{0.001}{1-(1-0.001)^{55.63\,0.075}} = 0.74$. At the first stage, the expected number of test-positive groups (TP) is $898.80 \times 0.74 = 665.11$ and the test-negative groups (FN) are 898.8 - 665.11 =233.69. For the FN group, let us assume approximately one true case per group is missed. Thus, the total expected number of FN individuals is 233.69 at the first stage. When t=1; EFN= $np(1 - Se(n))$ = 55.63× 0.001 × $(1 - 0.74) = 0.01$ per group. The rest of the figures are calculated the same way mentioned in the previous section. These results are shown in Figure 3.2.



Prevalence rate: 0.1% and **Proportion of having a positive group:** 0.05. **Sensitivity for a group:** Se(n) =Se(55.63)=0.74 (RT- PCR Test); **Sensitivity:** Se=93.3% (Rapid Antigen Test) and **Specificity**: Sp(n)= Sp(55.63) = Sp =1. **Protocol I** = When RT-PCR test is used at the first stage with dilution and rapid antigen test is used at the second stage.

Figure 3. 2: When p=0.001 and n=55.63 in Protocol I when d=0.075, the expected frequency tree of false-negative cases for both first and second stages in one million people.

Similarly, we have estimated the total number of false-negative individuals of Protocol I with dilution effect for both stages separately for other prevalence rates and cost ratios which are shown in Table 3.2.

Table 3. 2: Summary results of EFN Individuals of Protocol I with dilution for all prevalence rates and cost ratios

p	n	p'	$r = \frac{C_1}{C_2}$	E(C)	Se(n)	FN Individuals		Total FN Individuals
						Stage 1	Stage 2 when Se = 93.3%	
	221.3	0.20	24	1.20	0.67	298.2	40.6	338.8
	181.0	0.17	17	1.03	0.68	300.6	42.8	343.4
	133.9	0.13	10	0.81	0.69	300.9	44.9	345.8
	89.3	0.09	4.8	2.87	0.71	292.3	48.0	340.3
	74.0	0.07	3.4	2.44	0.72	264.8	45.6	310.4
0.001	55.6	0.05	2	1.90	0.74	233.7	44.6	278.3
	107.0	0.66	24	3.45	0.71	1789.1	293.5	2082.6
	74.5	0.53	17	3.05	0.73	1921.8	348.1	2270.0
	49.1	0.39	10	2.48	0.75	1985.7	399.1	2384.9
	30.2	0.26	4.8	9.05	0.78	1892.2	449.5	2341.6
	24.5	0.22	3.4	7.77	0.79	1886.5	475.5	2362.0
0.01	17.9	0.16	2	6.11	0.81	1694.5	484.0	2178.6
	20.1	0.46	4.8	15.1	0.80	4581.7	1227.9	5809.6
	15.6	0.38	3.4	13.2	0.82	4393.1	1340.9	5733.9
0.03	11.0	0.28	2	10.2	0.84	4091.3	1439.1	5530.4
	18.4	0.61	4.8	18.9	0.81	6305.8	1801.1	8106.9
	13.3	0.49	3.4	16.6	0.83	6267.9	2050.3	8318.2
0.05	8.9	0.37	2	13.4	0.85	6222.0	2362.3	8584.3
	12.6	0.60	3.4	19.2	0.83	8082.4	2643.9	10726.3
0.07	7.9	0.44	2	15.7	0.86	7758.2	3193.1	10951.2
0.10	7.2	0.53	2	18.5	0.87	9529.7	4273.0	13802.7
0.15	7.3	0.69	2	22.0	0.87	12338.4	5532.3	17870.7

p=Prevalence rate; $\boldsymbol{p'}$ = Proportion of having a positive group; **n**= Optimal group size; r = cost ratio = $\frac{C_1}{C_2}$; **Se(n)** = Sensitivity depending on the group size = $\frac{p}{1-(1-p)^{n^d}} = 1 \; as \; d = 0$.

$\boldsymbol{E(C|p,n,Se(n))} = Minimum\ Expected\ cost\ per\ person = C_2\left(\frac{C_1}{C_2 n} + (1-(1-p)^n)\right)$.

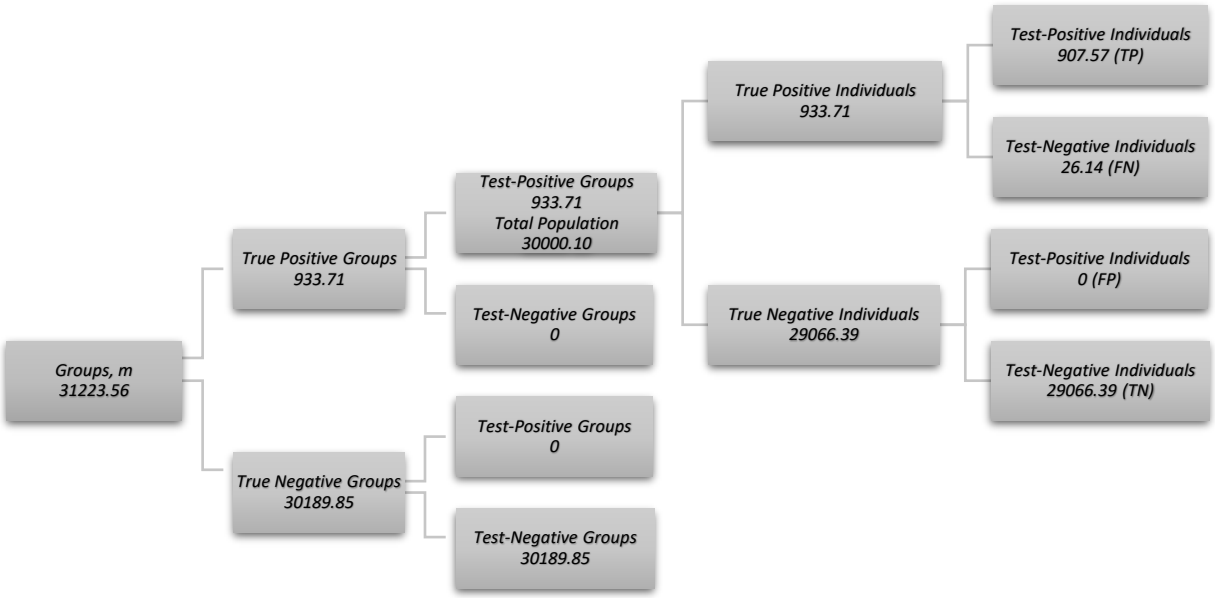
$\boldsymbol{C_1}$ = The entire (lab and test) cost of testing a group of samples using a real-time RT-PCR test kit.
$\boldsymbol{C_2}$ = The cost (without lab) of testing an individual using a rapid antigen test kit.

Protocol I = When RT-PCR test is used at the first stage with dilution and rapid antigen test is used at the second stage

As we can see in Table 3.2, the sensitivity of testing groups ranges between 87% and 67%. Additionally,the sensitivity of a rapid antigen test is 93.3% which is greater than the sensitivity of all the group testing. That is why the number of missing true cases at the 2nd stage is much lower than those at the first stage. Though the sensitivity of the group is getting better with the increasing prevalence, the number of missing true cases is not declining. This is because the likelihood of having positive groups goes up with the growing prevalence rates. Additionally, we notice that the total number of false-negative individuals of protocol I with dilution is greater than those without dilution. For example, with dilution, the number ranges from around 278 to 17870 and without dilution, it ranges from around 58 to 8500. Lastly, the expected cost per person of this protocol is getting higher with the increasing p. Overall, likewise without dilution, when prevalence is higher, with dilution both cost and the number of FN cases also get worse. Though, dilution causes more FN cases with less costs than without dilution effect.

3.2.2.3 Protocol II when d = 0

In Protocol II, only the RT-PCR test kit is used for both stages and the sensitivity of group testing is perfect when d = 0. The total number of false-negative cases of this protocol in 1 million people for various prevalences along with optimal group sizes are estimated (Table 2.8). In Figure 3.3, like in the previous protocol, the total expected FN individuals when p is 0.1% is shown with a hierarchy tree and the rest of them are shown in Table 3.3.



Prevalence rate: 0.1% and **Proportion of having a positive group** 0.03. **Sensitivity for a group:** Se(n) =Se(32.13)=1 (RT- PCR Test); **Sensitivity:** Se = 97.2% (RT-PCR Test) and **Specificity**: Sp(n) = Sp(32.13) = Sp =1. **Protocol II** = When RT-PCR test is used both for group testing (1st stage) without dilution and individual testing (2nd stage).

Figure 3. 3: When p=0.001 and n=32.13 in Protocol II when d=0, the expected frequency tree of false-negative cases for both first and second stages in one million people.

In Table 3.3, we can see that protocol II misses a smaller number of true cases than protocol I for all prevalences. This is because we are using only the RT-PCR test in this protocol which is more sensitive (97.2%) (Böger et al., 2021) than a rapid antigen test for individual testing. However, if the sensitivity of the RT-PCR is 100%, then we will not get any FN. Likewise, in the previous protocol, with the increasing probability of infectious people, the overall expected costs and FN are both becoming larger. For instance, if we form groups of three people approximately and screen a total of 1 million population by this protocol, then it will create the highest FN cases (4571.43) and will also cost the highest for any C_1. Hence, in this case, we need to check whether this protocol is appropriate or not for higher prevalences while compared to individual testing protocols. It will be discussed in the next section.

Table 3. 3: Summary results of EFN Individuals of Protocol II without dilution for Various p and optimal n when $C_1 = 120, 85, 50$

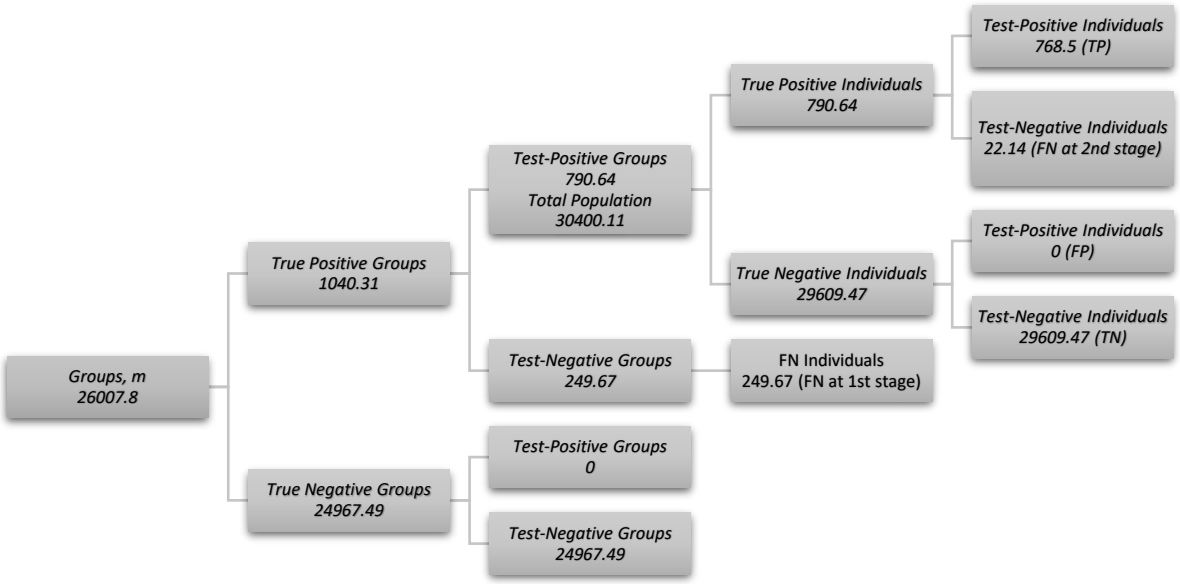
p	p'	n	E(C)= $C_1\left(\frac{1}{n} + (1-(1-p)^n)\right)$			Se(n)	FN Individuals		Total FN Individuals
			C_1=120	C_1=85	C_1=50		Stage 1	Stage 2 when Se= 97.2%	
0.001	0.03	32.1	7.53	5.33	3.14	1	0	26.1	26.1
0.01	0.10	10.5	23.5	16.6	9.77	1	0	266.2	266.2
0.03	0.17	6.31	40.0	28.3	16.7	1	0	754.4	754.4
0.05	0.23	5.02	51.2	36.2	21.3	1	0	1282.9	1282.9
0.07	0.27	4.35	60.1	42.6	25.0	1	0	1737.9	1737.9
0.10	0.33	3.75	70.2	50.4	29.7	1	0	2464.0	2464.0
0.15	0.41	3.22	86.2	61.0	35.9	1	0	3565.2	3565.2
0.20	0.48	2.94	98.6	69.8	41.1	1	0	4571.4	4571.4

p=Prevalence rate; $\boldsymbol{p'}$ = Proportion of having a positive group; **n**= Optimal group size; **E(C)** = Expected cost per person; **Se(n)** = Sensitivity depending on the group size = $\frac{p}{1-(1-p)^{nd}} = 1\ when\ d = 0$.

C_1 = The entire (lab and test) cost of testing a group of samples using a real-time RT-PCR test kit.
Protocol II = When RT-PCR test is used both for group testing (1st stage) without dilution and individual testing (2nd stage).

3.2.2.4 Protocol II when d ≠ 0

When we allow dilution in a group by putting d=0.075, the sensitivity gets changes with the different values of n. Similar to previous protocols, Figure 3.4 is showing how we have got the FN individuals in 1 million people for both stages of protocol II with dilution when p is 0.1% and Table 3.4 is representing the summary for all the p.



Prevalence rate: 0.1% and **Proportion of having a positive group** 0.04. **Sensitivity for a group:** Se(n) =Se(38.45)=1 (RT- PCR Test); **Sensitivity:** Se = 97.2% (RT-PCR Test) and **Specificity**: Sp(n) = Sp(38.45) = Sp =1. **Protocol II** = When RT-PCR test is used both for group testing (1st stage) with dilution and individual testing (2nd stage).

Figure 3. 4: When p=0.001 and n=38.45 in Protocol II when d≠0, the expected frequency tree of false-negative cases for both first and second stages in one million people.

Table 3.4 also shows the same pattern we have seen in Protocol I with dilution. With the increasing probability of infectious people, the overall expected costs and FN are both becoming larger. With the dilution, the number of FN errors is even more than without the dilution effect.

Table 3. 4: Summary results of EFN Individuals of protocol II with dilution for Various p and optimal n when $C_1 = 120, 85, 50$

p	p'	n	E(C)= $C_1\left(\frac{1}{n} + (1-(1-p)^n)\right)$			Se(n)	FN Individuals in		Total FN Individuals
			C_1=120	C_1=85	C_1=50		Stage 1	Stage 2 when Se= 97.2%	
0.001	0.04	38.5	6.57	4.65	2.74	0.76	249.7	22.1	271.8
0.01	0.11	12.1	21.3	15.1	8.89	0.83	1541.6	210.8	1752.4
0.03	0.2	7.18	37.1	26.3	15.5	0.86	3899.7	670.8	4570.5
0.05	0.25	5.69	47.9	33.9	19.9	0.88	5272.4	1082.6	6355.0
0.07	0.3	4.92	56.5	40.0	23.6	0.89	6707.3	1519.5	8226.8
0.10	0.36	4.25	67.3	47.7	28.1	0.90	8470.6	2134.6	10605.2
0.15	0.45	3.67	82.0	58.1	34.2	0.91	11035.4	3124.3	14159.7
0.20	0.53	3.39	94.1	66.7	39.2	0.92	12507.4	4027.4	16534.7

p=Prevalence rate; p' = Proportion of having a positive group; **n**= Optimal group size; **E(C)** = Expected cost per person; **Se(n)** = Sensitivity depending on the group size = $\frac{p}{1-(1-p)^{nd}}$.

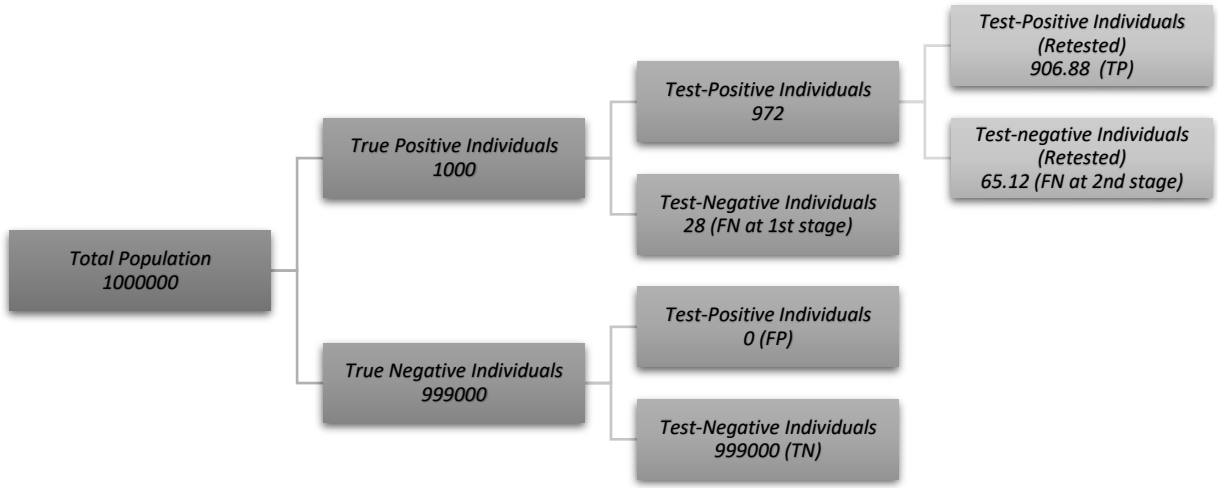
C_1 = The entire (lab and test) cost of testing a group of samples using a real-time RT-PCR test kit.
Protocol II = When RT-PCR test is used both for group testing (1st stage) with dilution and individual testing (2nd stage).

3.2.2.4 Protocol III (Two-Stage)

a) RT-PCR Test (1st Stage) and Rapid antigen Test (2nd Stage)

In this individual testing strategy, one million people are tested separately with the RT-PCR test at the first stage. For example, let us assume only 1 among 1000 people gets affected. So,

the expected number of true positive cases will be $(1000000\times 0.001) = 1000$. At the first stage, with 97.2% sensitivity, the expected number of individuals tested as positive is $(1000\times 0.972) = 972$. (1000-972)=28 persons will be falsely negative at the first stage. Now again, individuals with positive test results are retested with a rapid antigen test at the second stage. In this stage, with 93.3% sensitivity, the expected number of individuals retested as positive is $(972\times 0.933) = 906.88$. (972-906.88)=65.12 persons will be falsely negative at the second stage. The remaining (1000000-1000)=999000 people will be truly non-infected people who will also be tested as negative as the specificity is assumed to be 1 all the time. These results are shown in Figure 3.5. For other prevalences, the total number of expected false-negative results with the expected costs considering Se = 93.3% (Table 2.10) are shown in Table 3.5.



Prevalence rate: 0.1% **Sensitivity:** Sensitivity of a single RT-PCR test, Se=97.2% and sensitivity of a rapid antigen test =93.3% and **Specificity**: Sp =1. **Protocol III** = When RT-PCR test is used for individual testing (1st stage) and a rapid antigen test is used for retesting those individuals (2nd stage).

Figure 3. 5: The expected frequency tree of false-negative cases in one million people for both first and second stages in Protocol III(a) when p=0.001.

In Table 3.5, we can see that as the prevalence is going higher, the number of FN test results is also becoming higher. The same goes for the costs as well. Since the rapid test is less sensitive than the RT-PCR test, we are getting more false-negative cases than in the first stage. Therefore, for higher prevalences, this individual testing is also becoming most expensive and less accurate.

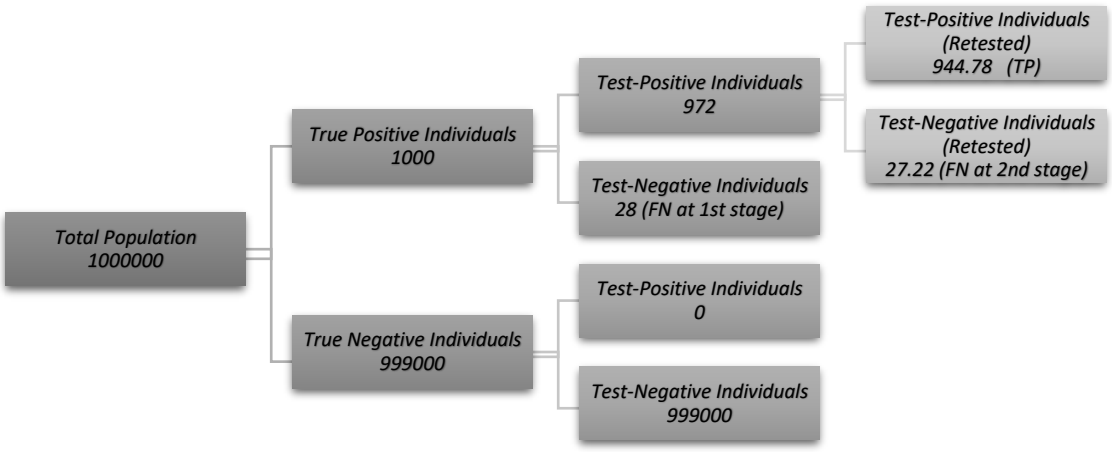
Table 3. 5: Summary results of EFN Individuals of Protocol III(a) for Various p and $\frac{C_1}{C_2}$ when both RT-PCR and Rapid Antigen test is used

p	E(C)= $C_2\left(\frac{C_1}{C_2}+Se.p\right)$						FN Individuals		Total FN Individuals
	Cost Ratio = r = $\frac{C_1}{C_2}$						Stage 1	Stage 2	
	24	17	10	4.8	3.4	2			
0.001	120.0	85.0	50.0	120.0	85.0	50.0	28.0	65.1	93.1
0.01	120.1	85.1	50.1	120.2	85.2	50.2	2800.0	651.2	931.2
0.03	120.1	85.1	50.1	120.7	85.7	50.7	840.0	1953.7	2793.7
0.05	120.2	85.2	50.2	121.2	86.2	51.2	1400.0	3256.2	4656.2
0.07	120.3	85.3	50.3	121.6	86.6	51.6	1960.0	4558.7	6518.7
0.10	120.5	85.5	50.5	122.3	87.3	52.3	2800.0	6512.4	9312.4
0.15	120.7	85.7	50.7	123.5	88.5	53.5	4200.0	9768.6	13968.6
0.20	120.9	85.9	50.9	124.7	89.7	54.7	5600.0	13024.8	18624.8

p=Prevalence rate; **E(C)** = Expected cost per person; **Sensitivity:** Sensitivity of a single RT-PCR test, Se=97.2% and sensitivity of a rapid antigen test =93.3% and **Specificity**: Sp =1.
C_1 = The entire (lab and test) cost of testing a group of samples using a real-time RT-PCR test kit.
C_2 = The cost (without lab) of testing an individual using a rapid antigen test kit.
Protocol III = When RT-PCR test is used for individual testing (1st stage) and a rapid antigen test is used for retesting those individuals (2nd stage).

b) RT-PCR Test for Both Stages

In this strategy, the RT-PCR test is repeated at both stages. Therefore, for both stages, the sensitivity is 97.2%. All the FN cases are calculated the same way mentioned in the previous strategy. Figure 3.6 shows the FN cases when p is 0.1%. For other prevalences, the total number of expected false-negative results with the expected costs considering Se = 97.2% (Table 2.11) are shown in Table 3.6.



Prevalence rate: 0.1% **Sensitivity:** Sensitivity of a single RT-PCR test, Se=97.2% and **Specificity**: Sp =1.
Protocol III = When RT-PCR test is used for individual testing at both stages.

Figure 3. 6: The expected frequency tree of false-negative cases in one million people for both first and second stages in Protocol III(b) when p=0.001.

Since the RT-PCR test is more sensitive than the rapid test and is used at both stages in this protocol, this individual testing strategy makes less number of test errors than protocol III(a) for all the prevalences. If the sensitivity of the RT-PCR test is perfect, then there would be no FN test results. Additionally, as the entire test cost of RT- PCR is more expensive than any rapid test, the expected costs for all p are also more costly than protocol III(a). Besides, with the increasing prevalence, this protocol is also becoming worse in terms of both cost and accuracy.

Table 3. 6: Summary results of EFN Individuals of protocol III(b) for Various p and $C_1 = 120, 85, 50$ when RT-PCR is used for both stages

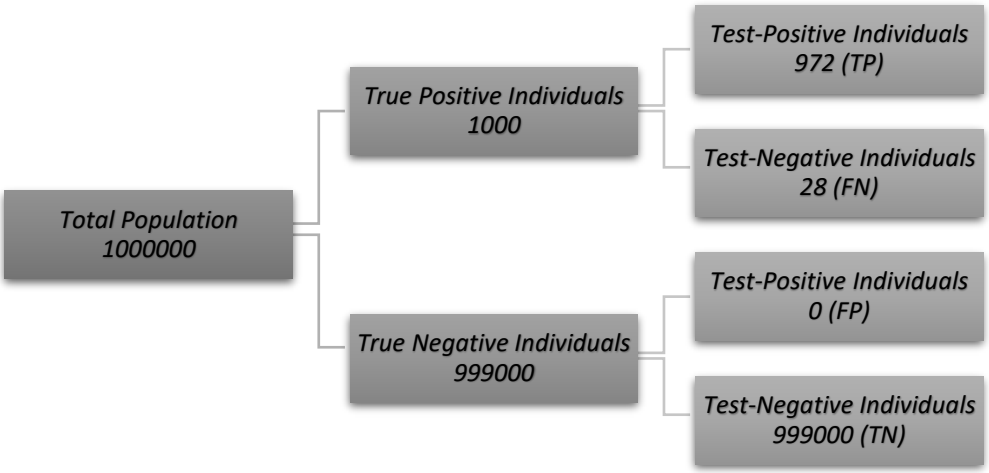
p	E(C)= $C_1(1 + Se.p)$			Sensitivity	FN Individuals		Total FN Individuals
	C_1=120	C_1=85	C_1=50		Stage 1	Stage 2	
0.001	120.1	85.1	50.1	0.972	28.0	27.2	55.2
0.01	121.2	85.8	50.5	0.972	280.0	272.2	552.2
0.03	123.5	87.5	51.5	0.972	840.0	816.5	1656.5
0.05	125.8	89.1	52.4	0.972	1400.0	1360.8	2760.8
0.07	128.2	90.8	53.4	0.972	1960.0	1905.1	3865.1
0.10	131.7	93.3	54.9	0.972	2800.0	2721.6	5521.6
0.15	137.5	97.4	57.3	0.972	4200.0	4082.4	8282.4
0.20	143.3	101.5	59.7	0.972	5600.0	5443.2	11043.2

p=Prevalence rate; **E(C)** = Expected cost per person; **Sensitivity:** Sensitivity of a single RT-PCR test, Se=97.2% and **Specificity**: Sp =1. C_1 = The entire (lab and test) cost of testing a group of samples using a real-time RT-PCR test kit. **Protocol III** = When RT-PCR test is used for individual testing at both stages. Assuming, $C_1 = C_3$.

3.2.2.4 Protocol IV (Single Stage)

a) RT-PCR Test Only

In this individual testing protocol, individuals are examined separately only once by RT-PCR test. For example, if the total population is one million and 1 among 1000 people is infected, then a total of 1000 persons will be truly infected. Among them after testing $(1000\times 0.972) = 972$ will be identified as positive and the rest of them will be incorrectly identified as negative (Figure 3.7). For other prevalences, results are presented in Table 3.7.



Prevalence rate: 0.1% **Sensitivity:** Sensitivity of a single RT-PCR test, Se=97.2% and **Specificity**: Sp =1. **Protocol IV** = When RT-PCR test is used only once for individual testing.

Figure 3. 7: The expected frequency tree of false-negative cases in one million people for a single stage in Protocol IV(a) when p=0.001.

Table 3. 7: Summary results of EFN individuals of protocol IV(a) for various p and $C_1 = 120, 85, 50$ when RT-PCR is used for a single-stage

p	C_1	$E(C) = C_1$	Sensitivity, Se	EFN Individuals
0.001	120	120.0	0.972	28.0
	85	85.0		
	50	50.0		
0.01	120	120.0	0.972	280.0
	85	85.0		
	50	50.0		
0.03	120	120.0	0.972	840.0
	85	85.0		
	50	50.0		
0.05	120	120.0	0.972	1400.0
	85	85.0		
	50	50.0		
0.07	120	120.0	0.972	1960.0
	85	85.0		
	50	50.0		
0.10	120	120.0	0.972	2800.0
	85	85.0		
	50	50.0		
0.15	120	120.0	0.972	4200.0
	85	85.0		
	50	50.0		
0.20	120	120.0	0.972	5600.0
	85	85.0		
	50	50.0		

p=Prevalence rate; **E(C)** = Expected cost per person; **Sensitivity:** Sensitivity of a single RT-PCR test, Se=97.2% and **Specificity**: Sp =1. $C_1 =$ The entire (lab and test) cost of testing a group of samples using a real-time RT-PCR test kit.**Protocol IV** = When RT-PCR test is used for individual testing (single-stage).

In Table 3.7, we can see with the increasing prevalence, the number of FN test results is increasing. However, the expected cost remains the same for all prevalence rates.

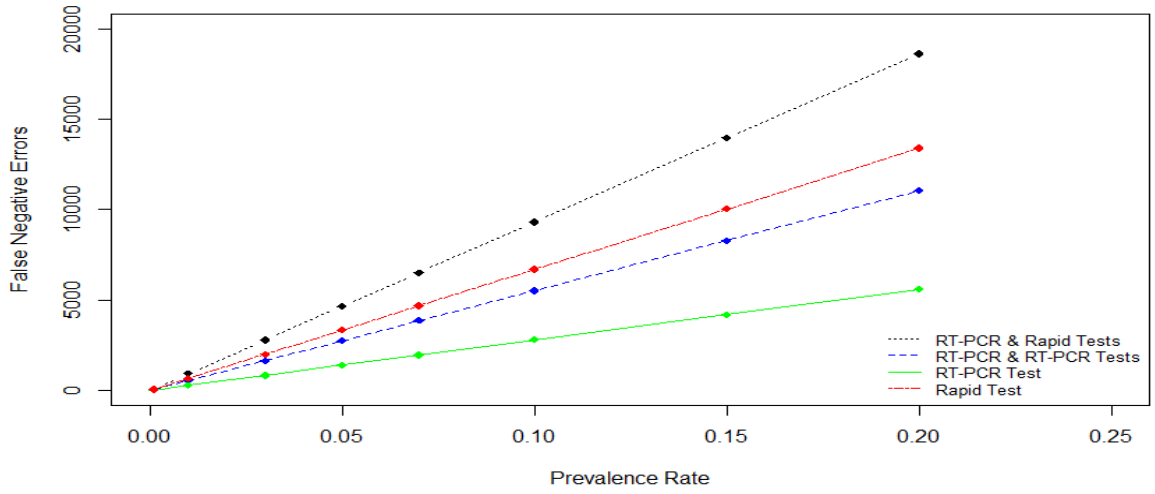
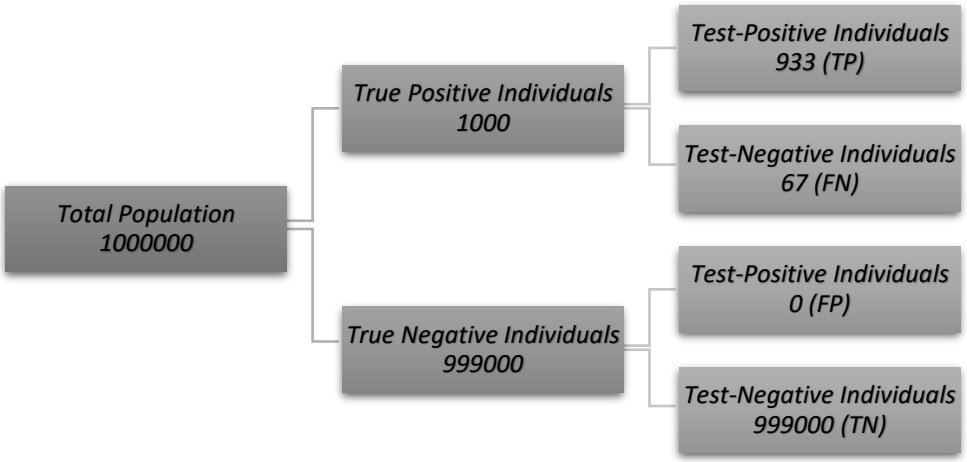


Figure 3. 8: False-negative errors against all the prevalence rates of all types of individual testing strategies (Protocol III and Protocol IV).

b) Rapid Antigen Test Only

In this individual testing protocol, the rapid test is used only once for screening people separately. For example, let us assume our total population is 1 million and prevalence is 0.1%. As the sensitivity is 93.3%, then $(1000000\times .001) \times 0.933 = 933$ people in total will be identified as positive and the rest of them will be detected as negative which is considered as an FN test error (Figure 3.9). For other prevalences, the results are shown in Table 3.8.



Prevalence rate: 0.1% **Sensitivity:** Sensitivity of a single rapid antigen test, Se=93.3% and **Specificity**: Sp =1.
Protocol IV = When rapid antigen test is used only once for individual testing.

Figure 3. 9: The expected frequency tree of false-negative cases in one million people for a single stage in Protocol IV(b) when p=0.001.

In Table 3.8, we can see if we use a rapid test only for a single stage, the number of false-negative errors is increasing with the increasing prevalence rates. Since the cost does not depend on the prevalence, the overall cost remains the same for all p.

Table 3. 8: Summary results of EFN individuals of protocol IV(b) for various p and $C_2 = 5, 25$ when rapid antigen is used for a single-stage

p	C_2	$E(C) = C_2$	Sensitivity, Se	EFN Individuals
0.001	5	5.0		
	25	25.0	0.933	67.0
0.01	5	5.0		
	25	25.0	0.933	670.0
0.03	5	5.0		
	25	25.0	0.933	2010.0
0.05	5	5.0		
	25	25.0	0.933	3350.0
0.07	5	5.0		
	25	25.0	0.933	4690.0
0.10	5	5.0		
	25	25.0	0.933	6700.0
0.15	5	5.0		
	25	25.0	0.933	10050.0
0.20	5	5.0		
	25	25.0.	0.933	13400.0

p=Prevalence rate; **E(C)** = Expected cost per person; **Sensitivity:** Sensitivity of a single rapid antigen test, Se=93.3%% and **Specificity**: Sp =1. C_2 = The cost (without lab) of testing an individual using a rapid antigen test kit. **Protocol IV** = When rapid antigen test is used for individual testing (single-stage).

Overall, from Figure 3.8 it is evident that among all individual testing strategies, protocols III(a) and IV(a) produce the highest and the lowest number of errors respectively. Since we are retesting the positive individuals at the second stage and also the sensitivity (93.3% for the rapid test) at the second stage is lower than 97.2% (RT-PCR test) in protocol III(a), we are getting the higher number of FN with the two-stage protocol using both tests. In protocol III(b), we are retesting the positive cases separately at the second stage with 97.2% sensitivity of the RT-PCR test at both stages. With the same sensitivity we are testing all the individuals only once in protocol IV(a). That is why protocol III(b) produces higher FN errors than protocol IV(a). On the other hand, we can see, if we test all of them with the 93.3% sensitive rapid test only once (IV(b)), then it produces higher FN than protocol III(b) and IV(a). It means the rapid

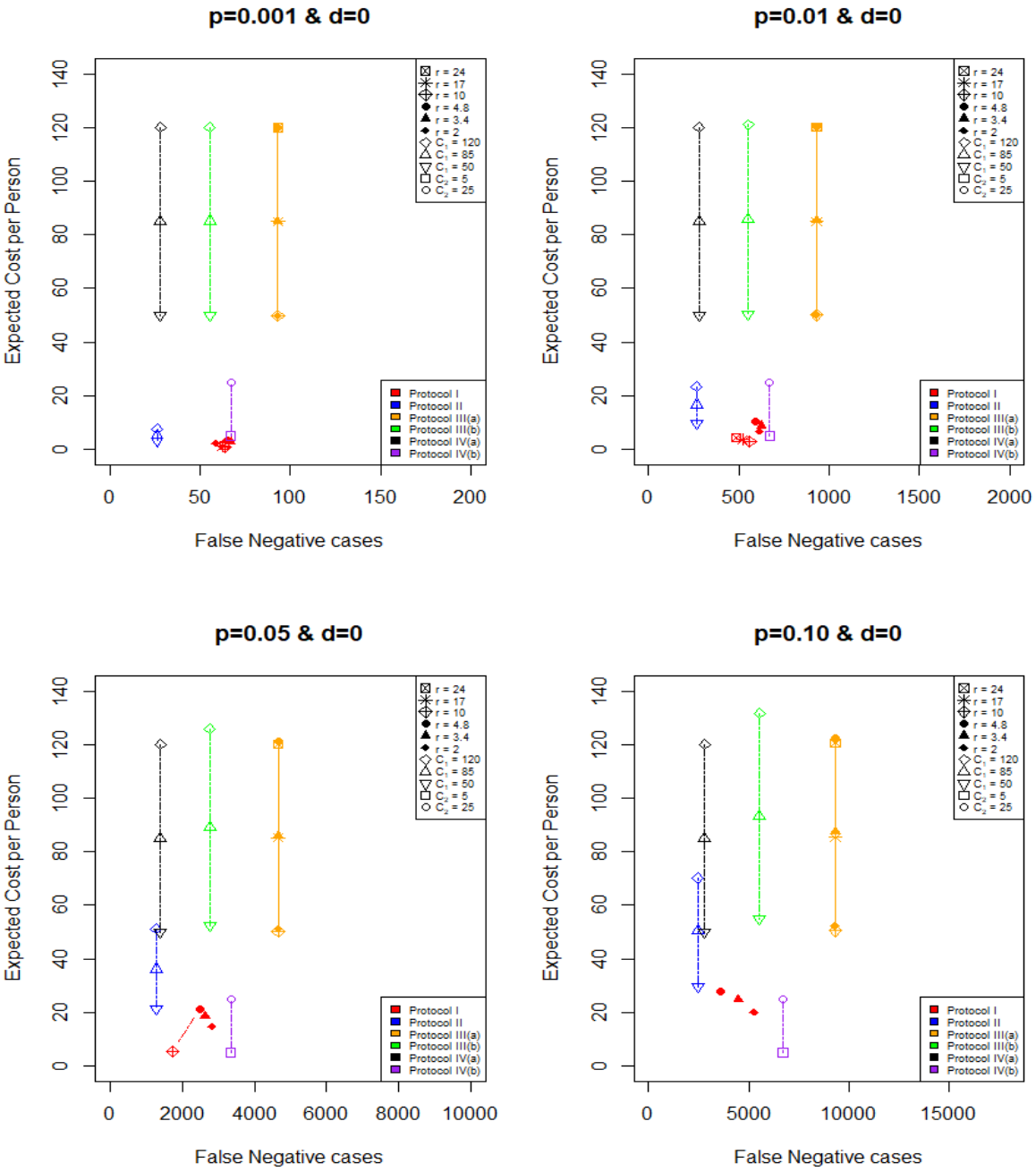
test with the lower sensitivity will be less accurate than the RT-PCR test whether being used either once or twice.

3.3 Relationship between Expected Cost per Person and False Negative Error

Now we will explore the relationship between the false-negative error and the expected cost per person of all the protocols for the prevalence rates (p = 0.001,0.01,0.05,0.10). We will find out which protocol is giving the best result for the chosen prevalences in terms of test errors and costs. Here for the lowest prevalence rates, we have selected 0.1% and 1% and 5% is for the middle and for the highest prevalence rate 10% is selected. 15% or 20% prevalence rate is not chosen because we have not found exact optimal group sizes in Protocol I for these rates.

3.3.1 Without Dilution Effect

When prevalence is the lowest that is only 1 among 1000 people is affected, protocol II misses the fewest number of true cases (26.14) among all protocols. If we compare the cost of protocol II with others, we can see protocol II is the cheapest for any value of C_1 than protocols III (a and b) and protocol IV(a). So, clearly two-stage group testing protocol with RT-PCR test repeated for any values of C_1 is the most appropriate to apply in practice in terms of cost and accuracy while compared to these three individual testing strategies. Whereas, compared to the cost of protocols I and IV(b), when C_1 is 50 protocol II costs almost the same as protocol I for all the cost ratios and protocol IV(b) when a rapid test costs only \$5. However, if the cost of protocol IV(b) is \$25, then protocol II for all values of C_1 becomes cheaper than protocol IV(b). Thus, we can say again protocol II is the most suitable one than these two strategies, especially in the above-mentioned situations in terms of both cost and accuracy.



Protocol I: When RT-PCR test is used at the first stage without dilution and rapid antigen test is used at the second stage.
Protocol II = When RT-PCR test is used both for group testing (1st stage) without dilution and individual testing (2nd stage).
Protocol III(a) = When RT-PCR test is used for individual testing (1st stage) and a rapid antigen test is used for retesting those individuals (2nd stage).
Protocol III(b) = When RT-PCR test is used for individual testing at both stages.
Protocol IV(a) = When RT-PCR test is used for individual testing (single-stage).
Protocol IV(b) = When rapid antigen test is used for individual testing (single-stage).
r = cost ratio.

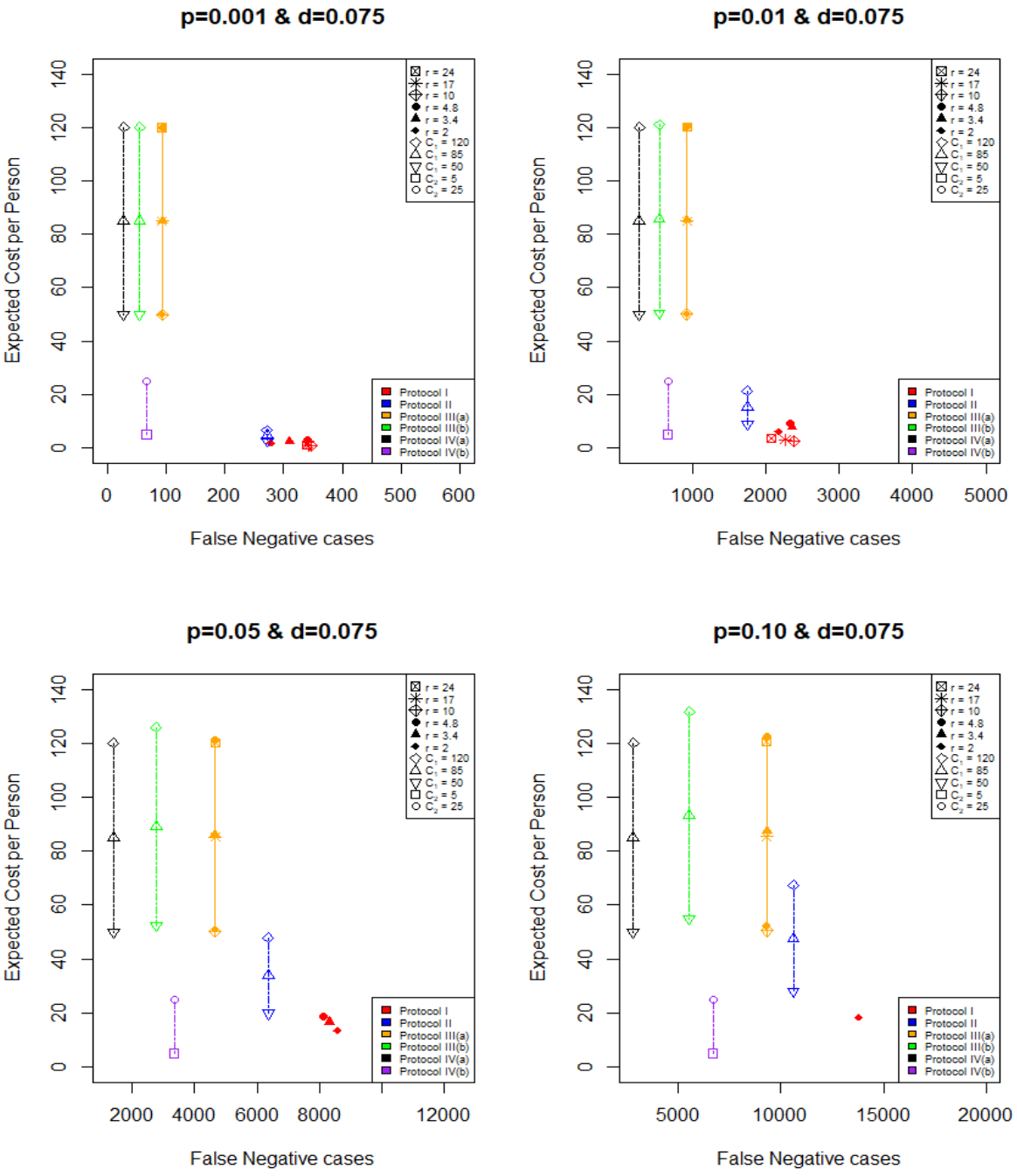
Figure 3. 10: Line graphs of false-negative error against expected cost per person for all the protocols without dilution effect when prevalence rate is 0.1%,1%,5% and 10%.

Now, when prevalence becomes the largest such as when 10 among 100 people are affected, still protocol II misses the lowest number of true cases (2464) than other protocols. However, for the highest prevalence, protocol II is less costly than protocols III (a and b) and IV(a) only when $C_1\ is\ 50$. Thus, protocol II when the entire cost of the RT-PCR test is \$50 is the appropriate one to use over these three strategies in terms of cost and accuracy. Similarly, when $C_1\ is\ 50$, protocol II is relatively the most suitable one to use in comparison to protocols I and IV(b). This is because the costs of both protocols I when r is 4.8 and IV(b) when $C_1 = 25$ are almost the same as the cost of protocol II when $C_1\ is\ 50$. But other two protocols in this case produce more FN errors than protocol II. Therefore, overall, we can say when there is no dilution effect in a group, protocol II is the best choice for all prevalence rates from both cost and accuracy points of view.

On the other hand, we can see that protocol III(a) causes the highest number of FN test results than all other protocols for all prevalence rates. Also, it is costly than protocols I and IV(b). When p is lower, protocol III(a) is costly than protocol II. However, as p becomes higher i.e., 5% or 10%, the cost of protocol II when $C_1\ is\ 120$ or 85 becomes the same or higher than the cost of protocol III(a) when r is 10 or 2. And the cost of protocol III(a) is almost the same as protocols III(b) and IV(a) for all p. So, protocol III(a) is not cheaper than most of the protocols. Thus, overall, Figure 3.10 suggests that protocol III(a) is the least preferable to use in terms of cost and accuracy for any p.

3.3.2 With Dilution Effect

From Figure 3.11 we can see that if we allow dilution in groups, the number of FN test results caused by protocol IV(a) is the minimum among all the protocols for all prevalences. In terms of the cost, when prevalence is lower such as 0.1% or 1%, it is more costly than protocols I, II and IV(b). The cost is almost the same as the cost of protocols III(a and b). Thus, if our main aim is to avoid FN errors as much as possible, then we need to pay more money for protocol IV(a) in comparison to group testing protocols and individual testing protocol IV(b). For example, when p is 0.1%, if we pay around \$40 or \$45 extra money to perform protocol IV(a) for C_1=50 compared to the cost of protocols I and II then we can reduce 262 or 322 number FN errors. We can have 39 fewer errors if we use protocol IV(a) costing \$50 instead



Protocol I: When RT-PCR test is used at the first stage with dilution and rapid antigen test is used at the second stage.
Protocol II = When RT-PCR test is used both for group testing (1st stage) with dilution and individual testing (2nd stage).
Protocol III(a) = When RT-PCR test is used for individual testing (1st stage) and a rapid antigen test is used for retesting those individuals (2nd stage).
Protocol III(b) = When RT-PCR test is used for individual testing at both stages.
Protocol IV(a) = When RT-PCR test is used for individual testing (single-stage).
Protocol IV(b) = When rapid antigen test is used for individual testing (single-stage).

Figure 3. 11: Line graphs of false-negative error against expected cost per person for all the protocols with dilution effect when prevalence rate is 0.1%,1%,5% and 10%.

of using protocol IV(b). But to achieve that we have to pay \$25 or \$45 extra if protocol IV(b) costs \$25 or \$5. Whereas compared to protocol III (a and b), we can have around 65 or 27 fewer errors spending almost the same amount of money for protocol IV(a). Therefore, protocol IV(a) can be a suitable one to apply in terms of accuracy for the lowest prevalence if there is no money constraint.

Similarly, we can conclude the same for the highest prevalence (10%). For example, in this case, we need to pay nearly \$30 or \$25 extra money to have around 11000 or 7800 less errors if we use protocol IV(a) when C_1=50 instead of protocol I when r =2 or protocol II when C_1=50 respectively. Also, we can reduce 3900 FN errors if we pay \$25 or \$45 more money for protocol IV(a) when C_1=50 compared to protocol IV(b). For other protocols, with the same cost, we can reduce around 6500 or 2721 FN test results. Therefore, we can say by paying a few dollars extra or the same for protocol IV(a), we can reduce a large amount of FN errors which is a great advantage.

Now, if there is a money restriction, then we can choose protocol IV(b). This is because it is less costly than protocols III (a and b) and IV(b) for all the prevalence rates. For lower prevalence (0.1%), the cost of protocol IV(b) when C_2 is 5 is the same as protocol I and II. And for higher prevalence, it becomes cheaper. In terms of accuracy, for all p, protocol IV(b) makes less FN errors than protocols I, II and III(a) while more FN errors than protocols IV(a) and III(b). Therefore, if there is a money constraint, then protocol IV(b) can be reasonably a better choice to use in practice.

Conversely, for all prevalence rates, we can see from Figure 3.11 that both group testing protocols make first and second highest FN errors. Thus in terms of accuracy, protocols I and II with dilution are the least preferable to use.

Overall, if we look at both Figures 3.10 and 3.11, it concludes that group testing protocol II is appropriate to use in terms of both cost and accuracy when there is no dilution. On the other hand, when there is dilution, the performance of group testing protocols gets worse in terms of accuracy. Consequently, in terms of accuracy protocol IV(a) or in terms of both protocol IV(b) become suitable to use.

Chapter 4 Discussion

4.1 Extension of Dorfman's Procedure

Robert Dorfman first proposed the group testing procedure for screening US soldiers for World War II as an alternative to individual testing to reduce the cost and to save time. His main goal was to identify the soldiers who had syphilitic antigens (Dorfman, 1943). The procedure was carried out in two stages. At the first stage, groups were formed independently by collecting blood samples of all the soldiers and then those groups were tested to minimize the number of tests. For example, a group of size n is randomly chosen and among the N population, total $m = \frac{N}{n}$ groups were selected and tested instead of testing the N population individually. If a group is tested positive, then all soldiers in that group were tested individually at the second stage which means n number of tests were performed to finally determine separately who were affected by syphilis. If a group was tested negative, then all the soldiers of that group were assumed to be infection-free. Dorfman did not assume any classification errors in his procedure. The dilution effect is common in group testing which leads to misclassifications. Graff & Roeloffs in 1972 extended the Dorfman procedure by including the misclassifications classification errors. Haber et al. in 2021 also implemented the Dorfman procedure including the test errors to find out what would happen to the minimum number of tests required for this procedure after assuming an incorrect sensitivity function. In our study, we have applied the same objective function to estimate the expected number of tests used by Haber et al. As an extension of Haber's paper, we have incorporated several cost components (testing cost, lab cost) into this function considering different group and individual testing protocols. After that, we have minimized the expected cost per person of group testing protocols for different prevalence rates by deriving different optimal group sizes. For individual testing protocols we have estimated the total expected cost per person. Then, we have compared the group testing protocols with the individual testing protocols in terms of their expected costs. Next, we have estimated the number of false negative errors encountered by these protocols for smaller and higher prevalence rates. Later, we have studied the relationship between these errors with their respective expected costs. Lastly, we have indicated which protocol is producing comparatively less false negative errors costing less compared to other protocols in terms of several prevalance rates.

4.2 Rationale for Using $Se(n) = \frac{p}{1-(1-p)^{nd}}$

In our group testing protocols I and II, we have assumed that the sensitivity of our groups which are formed at the first stage depends on the group size. For this, we have used a sensitivity function, $Se(n) = \frac{p}{1-(1-p)^{nd}}$ in which d is the dilution parameter, p is the prevalence rate and n is the group size. Here, $Se(n)$ is the probability of detecting a group correctly as a positive or a defective group. This function was first introduced by Hwang in 1976. Misclassification arises when there is a dilution effect in a group. Dilution effect takes place when positive samples are diluted with negative samples in a group. As a result, a loss in sensitivity happens. This function is a decreasing function with the increasing group sizes. If a group of size is 1, , there is no chance of diluting a positive sample with a negative sample. In other words, there will be no dilution effect i.e., d=0. So, the sensitivity will be perfect or Se(1) = 1. However, if the group size increases, the chances of the dilution of positive samples with negative samples get higher and the sensitivity of that group becomes lower. In other words, as n gets high or d ≠ 0, $Se(n)$ becomes less than 1. Also, if the group size gets larger, $Se(n)$ will be close to the prevalence rate (Hwang, 1976). The sensitivity function satisfied the above two conditions (Hwang, 1976). Haber et al. (2021) mentioned this sensitivity function in their paper because for their work they also assumed that the sensitivity will vary according to the group size. They used this function to show how misspecified sensitivity function impact determining the optimal group testing design. However, they have also recommended other sensitivity functions such as $Se(k) = 1 - 0.02(k-1)$ and $Se(k) = 1 - 0.02 \times 2^{\frac{k}{2}}$; where k is the group size. There is an another sensitivity function as well which depends on the proportion of the positive or defective iterms established by Burns & Mauro in 1987. It is defined as $S(y) = a + by^c; a \geq 0, b > 0, 0 \leq c \leq 1 \; and \; 0 \leq y \leq 1$. Here, $a = \alpha_1, b = 1 - \alpha_1 - \alpha_1$ and α_1 is the probability of detecting a negative group as positive and α_2 is the probability of detecting a positive group as negative. Because Hwang (1976) and Haber et al. (2021) both demonstrated their work using this function, $Se(n)$, we have used this function in our work to minimize the expected number of cost with the help of optimal group sizes for different group testing protocols.

4.3 Examples of Our Cost Functions

In our work, to estimate the overall expected cost of all the protocols, we have included the testing and lab cost values for screening SARS-CoV-2 by RT-PCR and rapid antigen test. Because of the pandemic situation of COVID-19 and the need for screening a large population

to stop the spread with the shortage supply of expensive test resources and shortage number of trained personnel, execution of group testing has become an ideal policy in this situation. That is why we have used this example for our cost functions. However, we can also apply these cost functions for screening other diseases too. For example, findings from our functions can also be obtained by screening HIV viral load, screening cancer patients etcetera. In our examples, numerical figures for our cost parameters were \$120 (Gewertz, 2021) or \$85 or \$50 for C_1 based on the information stated on the Department of Elementary and Secondary Education (2021) of Massachusetts website and \$5 or \$25 for C_2 (Feuer,2020). However, the cost figures of testing for novel coronavirus by several types of tests are constantly changing. For instance, the cost of an RT-PCR test ranges from \$194.25 to \$367.49 and the cost of a rapid antigen test ranges from \$110.24 to \$126 for travel outside of or inside Canada (Shepert, 2021). RT-PCR test can cost \$160 tested by company for travellers at Pearson International Airport and this cost can go up to \$1198 charged by Northstream Safety and Rehab in Thunder Bay (Saltzman, 2021). Another source (Seladi-Schulman, 2021) states that the actual cost of molecular tests usually ranges from \$75 to \$100 whereas the actual cost of antigen tests usually ranges from \$5 to \$50. So, we can see that these costs vary depending on the many factors such as the different types of tests (e.g. Molecular test, antigen test, antibody test etc.), the laboratory in which the test is processed, the area of a country, the delivery time of the results etcetera.

4.3.1 Different Types of Tests for COVID-19

There are several types of tests for identifying novel coronaviruses such as molecular tests, antigen tests and antibody tests (Seladi-Schulman, 2021). Molecular and antigen tests can detect the virus in both asymptomatic or symptomatic patients and antibody tests can detect whether a person has ever been affected by this virus or not (Seladi-Schulman, 2021). In our thesis, we have mentioned the RT-PCR test and rapid antigen test. A brief discussion of these two tests is given below.

4.3.1.1 Real-Time Reverse Transcription-Polymerase Chain Reaction (RT-PCR) Test

RT-PCR test for COVID-19 is one kind of molecular test to detect the presence or absence of the viral RNA and is generally administered in the laboratory (Seladi-Schulman, 2021). At first, samples (nasal swab or throat swab or saliva) are collected from asymptomatic or symptomatic patients and then are processed with several chemical solutions in an RT-PCR machine in the lab to detect the virus (Jawerth, 2020). Generally, the result of this test takes 1 to 3 days to get released (Government of Canada, 2021). RT-PCR test is considered to be the gold standard test for diagnosing COVID-19 because of its high sensitivity (Esbin et al., 2020).

4.3.1.2 Rapid Antigen Test

Like the RT-PCR test, rapid tests are also being used to detect the presence or absence of the viral marker COVID-19 antigen of asymptomatic or symptomatic patients. Antigen tests can be administered in different locations such as at home, hospitals, urgent care centers, pharmacies etcetera (Seladi-Schulman, 2021). Usually, samples are stored by using nasal swabs to detect the antigen and the result can be obtained within 15 to 30 minutes (Seladi-Schulman, 2021). It is a reasonably low-priced and simple test with quick results (Khandker et al., 2021) than RT-PCR. Also, it can be implemented by any non-professional (Government of Canada, 2021).

4.4 Summary of Overall Findings

In this thesis, our main objective was to analyze the costs of different protocols for some lower and higher prevalence rates. For this, we have derived different types of cost functions based on two types of group testing protocols and four types of individual testing protocols. Summary descriptions of all protocols are shown in Table 2.3. Another type of group testing protocol can also be assumed by considering rapid antigen tests for both stages. We have not included this because during the time of designing those protocols we did not find any literature stating that a rapid test is also being used for group testing. Most probably because it is not sensible to test a group by a rapid test.

For the group testing protocols, we have minimized the expected costs with the optimal group sizes for various prevalence rates (0.1%,1%,3%,5%,7%,10%,15% and 20%). The optimal group sizes to minimize the cost are the largest for the lowest prevalence and the smallest for the highest prevalence rate. In protocol II, optimal n is monotonically decreasing but in protocol I, it is not. Moreover, in both these protocols, we have noticed optimal n gets bigger when there is dilution in a group. So, overall these results suggest that for lower prevalences, we can form bigger group sizes to minimize the cost. For higher prevalences, we need to form smaller group sizes to do the same. In reality, we know that as prevalence rates get higher, it is better to form smaller groups. Our results are suggesting the same thing. Besides, the dilution effect is making the groups larger by increasing the number of FN errors. It implicitly means we can form larger groups with the dilution but the sensitivity of those groups will become less. Lastly, we have noticed that optimal group sizes in protocol II are smaller than protocol I. Thus, if there is a restriction of forming smaller groups, then one can choose protocol II instead of protocol I in practice.

With the estimated optimal group sizes, the minimized expected costs of both protocols are getting higher as the prevalence gets higher. It means for higher prevalence, we have to spend more money to execute group testing protocols for screening smaller groups. This summary raises the question "Which protocol (individual or group) should we perform to save money?" The answer to this question can be found from the results obtained from the relative cost which will be discussed next. Moreover, in both protocols, when we are allowing dilution, the expected costs become lower than without dilution. Overall, it implies that the dilution effect is allowing us to form bigger group sizes but with less expected costs. Therefore, if we want to form bigger groups and also spend less money on group testing, we can do that by risking the accuracy of the group tests. Now, this raises the question "How many FN errors do we get with these various protocols with or without the dilution ?". This matter was explored in chapter 3. Moreover, we also noticed that protocol I is cheaper than protocol II for all prevalences. So, in terms of costs only, protocol I is a better choice than protocol II.

We also designed four types of individual testing protocols (See Table 2.3). Protocol III(a) and III(b) are designed with two stages so that the cost of protocols I and II can be compared with the costs of protocols III(a) and III(b) in that order. The expected costs of protocol III get higher with the increasing prevalences as well. However, it is not the same for protocol IV since the expected cost does not depend on the p. The costs of group testing protocols have also been compared to protocol IV.

According to these findings, which protocol is the most expensive and which one is the cheapest for various prevalence rates are summarised in Table 4.1. Among all protocols, the amount of money needed to execute protocol III(b) is always the highest for all values of C_1 when prevalence rate ranges from the lower to medium higher (0.1-5%) and for $C_1 = 85$ and 120 when the prevalence rate becomes larger than 5%. So, clearly, we can say protocol III(b) is the least preferable in terms of the cost perspective to apply for any prevalence rates. On the other hand, for the lowest prevalence rates (0.1% and 1%), protocol I with dilution

Table 4. 1: Summary of findings from the expected cost per person of all protocols

p	**0.001**	**0.01**	**0.03**	**0.05**	**0.07**	**0.10**	**0.15**	**0.20**
Which protocol is the most expensive?	III(b) for all C_1	III(b) for all C_1	III(b) for all C_1	III(b) for all C_1	III(b) when C_1 =85,120	III(b) when C_1 =85,120	III(b) when C_1 =120	III(b) when C_1 =120

Which protocol is cheapest?	I when d=0.075 for all r	I when d=0.075 for all r	I when d=0 and r = 10	IV(b) when C_2 *is* 5	IV(b) when C_2 *is* 5	IV(b) when C_2 *is* 5	IV(b) when C_2 *is* 5	IV(b) when C_2 *is* 5 *or* 25

Protocol I = When RT-PCR test is used at the first stage with and without dilution and rapid antigen test is used at the second stage.
Protocol II = When RT-PCR test is used both for group testing (1st stage) with and without dilution and individual testing (2nd stage).
Protocol III(a) = When RT-PCR test is used for individual testing (1st stage) and rapid antigen test at 2nd stage is used again for individual testing to reconfirm.
Protocol III(b) = When RT-PCR test is used for both stages.
Protocol IV(a) = When RT-PCR test is used for a single stage individual testing.
Protocol IV(b) = When a rapid antigen test is used for a single stage individual testing.

for any cost ratios compared to other protocols is the most suitable one to use because it is the cheapest one. For 3% prevalence, group testing protocol I without dilution is the cheapest one when the cost ratio is 10. As the prevalence rate is getting higher than 3%, the cost of group testing protocols I and II are gradually getting higher than the costs of individual testing protocols. As such, one stage individual testing protocol using a rapid antigen test costing only \$5 becomes the cheapest one among all other protocols When p becomes 20%, protocol IV(b) costing \$5 or \$25 both are the cheapest. So, for higher prevalences, if a single-stage individual testing protocol with a rapid test is applied, then the maximum amount of money can be saved.

Relative costs in chapter 2 showed us in which cases protocols I and II are cheaper than protocols III and IV. Table 4.2 depicts that protocol I with and without dilution is less costly than protocol III(a) and IV(a) for all prevalences. Certainly, protocol I is the preferred one to protocols III(a) and IV(a) for all p in terms of cost. Besides, protocol I spends less money than protocol IV(b) as long as prevalence is the lowest (0.1%). Protocol I is also less costly when cost ratios are 24, 17, and 10 with dilution and 24 without dilution when p is 1%. When p is getting higher, and the cost ratio is getting lower, we can see protocol IV(b) saves money when C_2 is \$5. So, for higher prevalences, this group testing strategy is not preferred to use in terms of cost compared to protocol IV(b). Instead, it is better to use protocol IV(b) for higher prevalences despite knowing that this strategy will take more time to screen but will definitely save money than protocol I.

Table 4. 2: Preferred protocol based on the relative costs while comparing protocol I with protocol III(a), IV(a and b)

	$r = 24$	$r = 17$	$r = 10$	$r = 4.8$	$r = 3.4$	$r = 2$

Protocol I vs Protocol III(a)	**d=0**	I (p=0.1-1%)	I (p=0.1-3%)	I (p=0.1-5%)	I (p=0.1-10%)	I (p=0.1-10%)	I (p=0.1-20%)
	d=0.075	I (p=0.1-1%)	I (p=0.1-1%)	I (p=0.1-1%)	I (p=0.1-5%)	I (p=0.1-7%)	I (p=0.1-15%)
Protocol I vs Protocol IV(a)	**d=0**	I (p=0.1-1%)	I (p=0.1-3%)	I (p=0.1-5%)	I (p=0.1-10%)	I (p=0.1-10%)	I (p=0.1-20%)
	d=0.075	I (p=0.1-1%)	I (p=0.1-1%)	I (p=0.1-1%)	I (p=0.1-5%)	I (p=0.1-7%)	I (p=0.1-15%)
Protocol I vs Protocol IV(b)	**d=0**	I (p=0.1-1%)	I (p=0.1-1%) IV(b) ($C_2 = 5$ and p=3%)	I (p=0.1-3%) IV(b) ($C_2 = 5$ and p=5%)	I (p=0.1%) IV(b) $when\ C_2 = 5(p = 1 - 10\%)$ and 25(p=10%)	I (p=0.1%) IV(b) ($C_2 = 5$ and p=1-10%)	I (p=0.1%) IV(b) $when\ C_2 = 5(p = 1 - 20\%)$ and 25(p=20%)
	d=0.075	I (p=0.1-1%)	I (p=0.1-1%)	I (p=0.1-1%)	I(p=0.1%) IV(b) ($C_2 = 5$ and p=1-5%)	I(p=0.1%) IV(b) ($C_2 = 5$ and p=1-7%)	I(p=0.1%) IV(b) ($C_2 = 5$ and p=1-15%)

Protocol I = When RT-PCR test is used at the first stage with and without dilution and rapid antigen test is used at the second stage.
Protocol III(a) = When RT-PCR test is used for individual testing (1st stage) and rapid antigen test at 2nd stage is used again for individual testing to reconfirm.
Protocol IV(a) = When RT-PCR test is used for a single stage individual testing.
Protocol IV(b) = When a rapid antigen test is used for a single stage individual testing.

We also found from the summary of Table 4.3 that group testing protocol II saves money for all prevalences and all the values of C_1 while compared to the cost of protocol III(b). It means protocol III (b) is expensive for all cases and is not recommended to use if one wants to apply a cheaper protocol. Protocol II is also recommended to use for all p and all values of C_1 while compared to protocol IV(a) costing \$120. However, protocol IV(a) costing \$85 or \$50 is suggested to use over protocol II when C_1 is \$120 or \$85 for the highest prevalence rates only. Protocol IV(b) is preferred to protocol II in terms of cost in almost all cases except for the lowest p especially when IV(b) costs \$25. Therefore, in short, these results conclude that individual testing strategies with two stages are not suitable to use compared to their respective group testing protocols. Particularly, an individual testing strategy with one stage using a rapid antigen test is suitable to use instead of protocols I and II in almost all prevalence rates from the cost point of view.

Table 4. 3: Preferred protocol based on the relative costs while comparing protocol II with protocol III(b), IV(a and b)

			$C_1 = 120$	$C_1 = 85$	$C_1 = 50$
Protocol II vs Protocol III(b)	**d=0**		II (For all p)	II (For all p)	II (For all p)
	d=0.075		II (For all p)	II (For all p)	II (For all p)
Protocol II vs Protocol IV(a)	**d=0**	**$C_1 = 120$ in IV(a)**	II (For all p)	II (For all p)	II (For all p)
		$C_1 = 85$ in IV(a)	II(p=0.1-10%) IV(a) (p=15-20%)	II (For all p)	II (For all p)
		$C_1 = 50$ in IV(a)	II (p=0.1-3%) IV(a) (p=5-20%)	II (p=0.1-7%) IV(a) (p=10-20%)	II (For all p)
	d=0.075	**$C_1 = 120$ in IV(a)**	II (For all p)	II (For all p)	II (For all p)
		$C_1 = 85$ in IV(a)	II (p=0.1-15%) IV(a) (p=20%)	II (For all p)	II (For all p)
		$C_1 = 50$ in IV(a)	II (p=0.1-5%) IV(a) (p=7-20%)	II (p=0.1-10%) IV(a) (p=15-20%)	II (For all p)
Protocol II vs Protocol IV(b)	**d=0**	**$C_2 = 5$ in IV(b)**	IV(b) (For all p)	IV(b) (For all p)	II (p=0.1%) IV(b) (p=1-20%)
		$C_2 = 25$ in IV(b)	II (p=0.1-1%) IV(b) (p=3-20%)	II (p=0.1-1%) IV(b) (p=3-20%)	II (p=0.1-5%) IV(b) (p=15-20%)
	d=0.075	**$C_2 = 5$ in IV(b)**	IV(b) (For all p)	II (p=0.1%) IV(b) (p=1-20%)	II (p=0.1%) IV(b) (p=1-20%)
		$C_2 = 25$ in IV(b)	II (p=0.1-1%) IV(b) (p=3-20%)	II (p=0.1-1%) IV(b) (p=3-20%)	II (p=0.1-7%) IV(b) (p=10-20%)

Protocol II = When RT-PCR test is used both for group testing (1st stage) with and without dilution and individual testing (2nd stage).
Protocol III(b) = When RT-PCR test is used for both stages of individual testing.
Protocol IV(a) = When RT-PCR test is used for a single stage individual testing.
Protocol IV(b) = When a rapid antigen test is used for a single stage individual testing.

Till now, we have explored our protocols in terms of cost only. Nevertheless, it is also important to estimate the number of test errors to check how accurate each protocol is. Since we assumed specificity to be 1, we have estimated the FN errors of each protocol after estimating the costs. If a policymakerwants to apply a design as much accurately as possible without any budget constraints, then according to our results protocol II without dilution is the most accurate one (Table 4.4). Conversely, in general, group testing protocols seem to have a higher probability of making FN errors than individual testing protocols if there is a dilution effect in a group. Agreeing to this fact, protocols I and II with dilution are making the highest

and second-highest number of FN errors and as such these are least desirable to use. However, if a policymaker wants to follow the individual testing strategy and also wants the lowest number of FN errors, then protocol IV(a) is the one to apply since it produces the second-lowest FN errors among all the protocols.

Table 4. 4: Ranking of all protocols from the highest to lowest according to the number of FN errors

p	Protocol I when d=0	Protocol I when d=0.075	Protocol II when d=0	Protocol II when d=0.075	Protocol III(a)	Protocol III(b)	Protocol IV(a)	Protocol IV(b)
0.001	5th	1st	8th	2nd	3rd	6th	7th	4th
0.01	5th	1st	8th	2nd	3rd	6th	7th	4th
0.03	*5th/*6th	1st	8th	2nd	3rd	**5th/**6th	7th	4th
0.05	*5th/*6th	1st	8th	2nd	3rd	**5th/**6th	7th	4th
0.07	6th	1st	8th	2nd	3rd	5th	7th	4th
0.10	6th	1st	8th	2nd	3rd	5th	7th	4th
0.15	6th	1st	8th	2nd	3rd	5th	7th	4th
0.20	5th	--	8th	2nd	1st	4th	6th	3rd

*means exact ranking depends on the cost ratio (r) and ** means exact ranking depends on the entire cost of the RT-PCR test (C_1); p = Prevalence rate.
Protocol I = When RT-PCR test is used at the first stage with and without dilution and rapid antigen test is used at the second stage.
Protocol II = When RT-PCR test is used both for group testing (1st stage) with and without dilution and individual testing (2nd stage).
Protocol III(a) = When RT-PCR test is used for individual testing (1st stage) and rapid antigen test at 2nd stage is used again for individual testing to reconfirm.
Protocol III(b) = When RT-PCR test is used for both stages of individual testing.
Protocol IV(a) = When RT-PCR test is used for a single stage individual testing.
Protocol IV(b) = When a rapid antigen test is used for a single stage individual testing.

At the end of chapter 3, for 0.1%,1%,5% and 10% prevalence rates, we have studied the relationship between the expected costs and the false-negative errors caused by these protocols. And then we showed which protocol is producing fewer false-negative errors with less cost for those selected prevalences. There might be six forms of relationships: suppose a protocol a) might be the cheapest but may produce the highest number of FN errors. b) might be expensive but may produce the smallest number of FN errors. c) might be cheaper and also may produce the smallest number of FN errors (Ideal situation). d) might be most expensive and also may produce the highest FN errors (Worse situation). e) might be less costly and at the same time produce fewer FN errors compared to others (Near to ideal situation). f) might be not cheaper

and at the same time produce a larger number of FN errors compared to others (Near to worse situation).

In this study, our investigation reveals that without dilution effect protocol II can be categorized as the fifth form of relationship for lower prevalence rates (0.1% and 1%). So for lower prevalences, a policymaker can choose protocol II without dilution if the requirement is to select a protocol that will be comparatively cheaper and will produce the minimum FN errors. For higher prevalences, protocol II without dilution is still a better option for a policymaker but this time to reduce the cost it is best to select an RT-PCR test costing $50 or sometimes $85. However, the relationship of protocol III(a) is categorized as 'd' which is the worst case. That is why it is the least preferable for any prevalence.

With the dilution effect, the relationship between cost and FN error of group testing protocols falls into the first category. So, they can no longer be a good choice in terms of both those factors. In addition, protocol III(a) can be categorized as 'f' which is considered as near to the worse case. On the other hand, the relationship of individual testing protocol IV(a) falls into category 'b'. And the relationship of protocol IV(b) falls into category 'e'. More precisely, protocol IV(b) is less costly than all individual testing protocols. And it is less costly or almost the same as group testing protocols when rapid test costs $5. Besides, it produces fewer errors than protocols I, II and III(a) and slightly more errors than III(b) and IV(a). So, overall we can say, a policymaker needs to pay a lot more money to be most accurate for IV(a). Otherwise, he or she needs to choose IV(b) to balance the cost and FN errors. So, the relationship of all these protocols concludes that without dilution protocol II is the best choice in terms of both factors and with dilution protocols IV(a) is the best choice in terms of accuracy or IV(b) is relatively better in terms of both. And overall protocol III(a) is the least recommended protocol.

4.5 Strengths and Limitations

One of the main findings of Haber et al (2021) was to assess the impact of a misspecified sensitivity function or an incorrect assumption of a sensitivity function on the expected number of tests required for their group testing design. They did not incorporate any cost values to investigate the expected cost of that optimal group testing policy. In our study, with the help of their objective function, we derived several cost functions for several group and individual testing strategies. And we focused on assessing the overall expected cost and the optimal group sizes of all strategies and made comparisons of them in terms of these costs.

In other papers, such as in Turner et al., 2009 paper, an expected cost model was developed including pooling and testing costs to generalize the optimal group testing design. This model was applied without including (Chiang et al., 2010) and including (S. C. Liu et al., 2011) the penalty costs of misclassifications in the seed potato certification program to minimize the expected cost for screening the potato viruses by deriving the optimal group sizes. In our research, we not only determined the optimal group sizes to minimize the cost for various group testing protocols with the presence of errors but also estimated the number of FN errors to verify the accuracy of each protocol. Lastly, we focused on finding the balance between those expected cost values and the number of FN errors for lower or higher prevalence rates to find an effective strategy. Therefore, from our findings, a decision-maker can have an idea about which group or individual testing policy for screening the disease is affordable or more accurate or both for lower and higher prevalence rates.

Our findings showed that as the prevalence goes higher, the expected cost and the number of FN test results both get higher whether it is group testing or individual testing strategy. We know that in general group testing is recommended over individual testing strategy for screening large populations because the former is more affordable than the latter ((Dorfman, 1943); (Wein & Zenios, 1996). However, our findings in Chapter 2 suggest that for higher prevalence one-stage individual testing with the rapid test is the most affordable than group testing and all other individual protocols. Therefore, based on our results, for higher prevalence, one can easily reduce some costs by applying protocol IV(b) in practice for screening. Because of the dilution effect, the chance of losing the sensitivity of a group gets higher. In this light, our results showed that protocol I and II when d = 0.075 causes the highest and second-highest number of FN errors. However, if we consider a special case that is no dilution in a group, then group testing protocol with RT-PCR test repeated at both stages becomes the most accurate one. Thus, based on these results, whether the values of d between 0 and 0.075 yield the same conclusion or not would be of interest to examine.

Even though our investigations provide us with some useful findings, these results are only applicable when cost values of C_1 is \$120 or \$85 or \$50 and C_2 is \$25 or \$5 which is the key limitation of this research. We have illustrated our cost functions with the example of test and lab costs for screening Covid-19 by RT-PCR and rapid antigen tests. Nevertheless, there is a wide range of charges for diagnosing this virus depending on various factors. Authorities have been trying constantly to make the overall screening process more affordable to prevent the spread quickly. As such, the results of these specific cost values might not be applicable

elsewhere. So, our preferred protocols may vary according to the different costs values depending on the situation being considered. However, our methodology could be easily adapted to new situations with new cost structures.

Another limitation of our current work is that we explored our protocols only in terms of cost and FN errors. However, to prevent or control the spread of infectious disease it is also vital to have the test results as quickly as possible. A testing strategy will be more effective if it is affordable and at the same time can provide the test results as fast as possible without making too many errors. In our study, we did not examine and compare the time that each protocol will take to provide the test results and how much having a test result immediately can impact the overall cost. For example, in COVID, nowadays, RT-PCR takes generally 48 hours and the rapid test takes usually 15 minutes to provide the result (Du et al., 2021). Considering these times, which protocol will be more feasible to use in terms of time could be another area to research.

One of the assumptions of our testing designs is that there will be no false-positive test results. We have considered this assumption for simplicity. However, false-positive errors also cause several problems such as an increase in the expected number of tests which will eventually increase the overall costs. Besides this false-positive diagnosis can have other harmful effects on patients. For example, the patient will proceed to further treatment unnecessarily which will eventually increase the cost. At the same time, this error will also adversely affect the overall health and psychological condition of the patient. Our first goal in this work was to minimize the expected cost with the optimal group size for all group testing protocols. Thus, in turn, it is worthwhile to derive a cost model including both test errors to obtain an efficient optimal testing design. So, this could be a future topic to research in this area.

In our group testing protocols, we have taken two considerations while estimating the sensitivity function: one is when we assumed there will be no dilution in a group (d=0) which is a special case and another one is there is some dilution in a group (d=0.075). In the sensitivity function, the range of d was between 0 and 1. According to Figure 2.3, we have seen that d = 0.075 shows the moderate decay of sensitivity as the group size increases for all prevalence rates. Based on our findings it would be of interest to study how the relationship between the cost and FN errors of protocols I and II might change for the values between 0 and 0.075 of d. Also, if we allow more dilution or take larger values than 0.075 of d, would this rule out the

group testing completely even more quickly than before? So, another extension of our work would be to explore our findings based on the other values of d.

Another assumption of this research is that we have designed the protocols for homogenous populations. We did not account for the fact that the probability of having a disease can vary for all individuals based on several factors. For instance, prevalence can vary based on sex or gender or age. Also, in this pandemic, the risk of being positive is very high for a whole group if a member of that group is already infected. So, this could be another new approach to explore our cost functions.

4.6 Conclusion

Altogether in this research, we have suggested two group testing and four individual testing strategies and made a comparison in terms of cost and false-negative errors. Our investigations demonstrated that group testing design protocol I with dilution for all the cost ratios can be the most budget-friendly for the lower prevalence while compared to the cost of individual testing protocols. Nevertheless, further research emphasized that allowing for dilution in groups can make the group testing least accurate in terms of FN test results for any prevalence rates. Both these contradictory inquiries influenced us to examine the relationship between the cost and the number of FN errors for all protocols and suggested the appropriate design accordingly. Therefore, our final findings will help a decision-maker to decide which protocol is efficient in terms of both cost and accuracy for lower or higher prevalence rates.

Bibliography

Aprahamian, H., Bish, D. R., & Bish, E. K. (2019). Optimal Risk-Based Group Testing. *Management Science*, *65*(9), 4365–4384.

Becker, M. G., Taylor, T., Kiazyk, S., Cabiles, D. R., Meyers, A. F. A., & Sandstrom, P. A. (2020). Recommendations for sample pooling on the Cepheid GeneXpert® system using the Cepheid Xpert® Xpress SARS-CoV-2 assay. *PLoS ONE*, *15*(11), 1–9. https://doi.org/10.1371/journal.pone.0241959

Bilder, C. R. (2019). Group Testing for Identification. In *Wiley StatsRef: Statistics Reference Online* (pp. 1–11). Wiley. https://doi.org/10.1002/9781118445112.stat08227

Bilder, C. R., Iwen, P. C., Abdalhamid, B., Tebbs, J. M., & McMahan, C. S. (2020). Tests in short supply? Try group testing. *Significance*, *17*(3), 15–16. https://doi.org/10.1111/1740-9713.01399

Bilder, C. R., Tebbs, J. M., & Chen, P. (2010). Informative Retesting. *Journal of the American Statistical Association*, *105*(491), 942–955. https://doi.org/10.1198/jasa.2010.ap09231

Boger, B., Fachi, M. M., Vilhena, R. O., Cobre, A. F., Tonin, F. S., & Pontarolo, R. (2021). Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19. *American Journal of Infection Control*, *49*(1), 21–29. https://doi.org/10.1016/j.ajic.2020.07.011

Burns, K. C., & Mauro, C. A. (1987). Group testing with test error as a function of concentration. *Communications in Statistics - Theory and Methods*, *16*(10), 2821–2837. https://doi.org/10.1080/03610928708829544

Chiang, K. S., Lai, S. H., Chung, W. C., Lin, S. H., & Yang, T. C. (2010). Cost Considerations in Choosing Group Size for Group Testing in the Seed Potato Certification Program. *American Journal of Potato Research*, *87*, 472–478. https://doi.org/10.1007/s12230-010-9149-0

Cowling, D. W., Gardner, I. A., & Johnson, W. O. (1999). Comparison of methods for estimation of individual-level prevalence based on pooled samples. *Preventive Veterinary Medicine*, *39*(3), 211–225. https://doi.org/10.1016/S0167-5877(98)00131-7

Dorfman, R. (1943). The Detection of Defective Members of Large Populations. *Mathematical Statistics*, *14*(4), 436–440.

Du, Z., Pandey, A., Bai, Y., Fitzpatrick, M. C., Chinazzi, M., Pastore y Piontti, A., Lachmann, M., Vespignani, A., Cowling, B. J., Galvani, A. P., & Meyers, L. A. (2021). Comparative cost-effectiveness of SARS-CoV-2 testing strategies in the USA: a modelling study. *The Lancet Public Health*, *6*(3), e184–e191. https://doi.org/10.1016/S2468-2667(21)00002-5

Esbin, M. N., Whitney, O. N., Chong, S., Maurer, A., Darzacq, X., & Tjian, R. (2020). Overcoming the bottleneck to widespread testing: A rapid review of nucleic acid testing approaches for COVID-19 detection. *Rna*, *26*(7), 771–783.

https://doi.org/10.1261/rna.076232.120

Graesboll, K., Andresen, L. O., Halasa, T., & Toft, N. (2017). Opportunities and challenges when pooling milk samples using ELISA. *Preventive Veterinary Medicine*, *139*, 93–98. https://doi.org/10.1016/j.prevetmed.2016.08.001

Graff, L. E., & Roeloffs, R. (1972). Group testing in the presence of test error; an extension of the Dorfman procedure. *Technometrics*, *14*(1), 113–122. https://doi.org/10.1080/00401706.1972.10488888

Haber, G., Malinovsky, Y., & Albert, P. S. (2021). Is group testing ready for prime-time in disease identification? *Statistics in Medicine*, 1–23. https://doi.org/10.1002/sim.9003

Hepworth, G., & Walter, S. D. (2020). Estimation of proportions by group testing with retesting of positive groups. *Communications in Statistics - Theory and Methods*, *49*(22), 5587–5597. https://doi.org/10.1080/03610926.2019.1620280

Huang, S. H., Huang, M. N. Lo, & Shedden, K. (2020). Cost considerations for efficient group testing studies. *Statistica Sinica*, *30*, 285–302. https://doi.org/10.5705/SS.202017.0408

Hwang, F. K. (1976). Group Testing with a Dilution Effect. *Biometrika*, *63*(3), 671–673. https://www.jstor.org/stable/2335750

Khan, S. A., Chowdhury, P., Choudhury, P., & Dutta, P. (2017). Detection of West Nile virus in six mosquito species in synchrony with seroconversion among sentinel chickens in India. *Parasites and Vectors*, *10*. https://doi.org/10.1186/s13071-016-1948-9

Khandker, S. S., Hashim, N. H. H. N., Deris, Z. Z., Shueb, R. H., & Islam, M. A. (2021). Diagnostic accuracy of rapid antigen test kits for detecting SARS-CoV-2: A systematic review and meta-analysis of 17,171 suspected COVID-19 patients. *Journal of Clinical Medicine*, *10*(16). https://doi.org/10.3390/jcm10163493

Kim, H. Y., Hudgens, M. G., Dreyfuss, J. M., Westreich, D. J., & Pilcher, C. D. (2007). Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics*, *63*, 1152–1163. https://doi.org/10.1111/j.1541-0420.2007.00817.x

Kortela, E., Kirjavainen, V., Ahava, M. J., Jokiranta, S. T., But, A., Lindahl, A., Jääskeläinen, A. E., Jääskeläinen, A. J., Järvinen, A., Jokela, P., Kallio-Kokko, H., Loginov, R., Mannonen, L., Ruotsalainen, E., Sironen, T., Vapalahti, O., Lappalainen, M., Kreivi, H.

R., Jarva, H., … Kekalainen, E. (2021). Real-life clinical sensitivity of SARS-CoV-2 RT-PCR test in symptomatic patients. *PLoS ONE*, *16*(5), 1–19. https://doi.org/10.1371/journal.pone.0251661

Liu, A., Liu, C., Zhang, Z., & Albert, P. S. (2012). Optimality of group testing in the presence of misclassification. *Biometrika*, *99*(1), 245–251. https://doi.org/10.1093/biomet/asr064

Liu, S. C., Chiang, K. S., Lin, C. H., Chung, W. C., Lin, S. H., & Yang, T. C. (2011). Cost analysis in choosing group size when group testing for Potato virus y in the presence of classification errors. *Annals of Applied Biology*, *159*, 491–502. https://doi.org/10.1111/j.1744-7348.2011.00513.x

Lohse, S., Pfuhl, T., Berko-Gottel, B., Rissland, J., Geibler, T., Gartner, B., Becker, S. L., Schneitler, S., & Smola, S. (2020). Pooling of samples for testing for SARS-CoV-2 in asymptomatic people. *The Lancet Infectious Diseases*, *20*, 1231–1232. https://doi.org/10.1016/S1473-3099(20)30362-5

Malinovsky, Y., Albert, P. S., & Roy, A. (2016). Reader reaction: A note on the evaluation of group testing algorithms in the presence of misclassification. *Biometrics*, *72*, 299–304. https://doi.org/10.1111/biom.12385

McMahan, C. S., Tebbs, J. M., & Bilder, C. R. (2012). Informative Dorfman Screening. *Biometrics*, *68*, 287–296. https://doi.org/10.1111/j.1541-0420.2011.01644.x

Mokalled, S. C., McMahan, C. S., Tebbs, J. M., Andrew Brown, D., & Bilder, C. R. (2021). Incorporating the dilution effect in group testing regression. *Statistics in Medicine*, *40*, 2540–2555. https://doi.org/10.1002/sim.8916

Mutesa, L., Ndishimye, P., Butera, Y., Souopgui, J., Uwineza, A., Rutayisire, R., Ndoricimpaye, E. L., Musoni, E., Rujeni, N., Nyatanyi, T., Ntagwabira, E., Semakula, M., Musanabaganwa, C., Nyamwasa, D., Ndashimye, M., Ujeneza, E., Mwikarago, I. E., Muvunyi, C. M., Mazarati, J. B., … Ndifon, W. (2021). A pooled testing strategy for identifying SARS-CoV-2 at low prevalence. *Nature*, *589*, 276–280. https://doi.org/10.1038/s41586-020-2885-5

Pilarowski, G., Lebel, P., Sunshine, S., Liu, J., Crawford, E., Marquez, C., Rubio, L., Chamie, G., Martinez, J., Peng, J., Black, D., Wu, W., Pak, J., Laurie, M. T., Jones, D., Miller, S., Jacobo, J., Rojas, S., Rojas, S., … DeRisi, J. (2021). Performance

Characteristics of a Rapid Severe Acute Respiratory Syndrome Coronavirus 2 Antigen Detection Assay at a Public Plaza Testing Site in San Francisco. *The Journal of Infectious Diseases*, *223*, 1139–1144. https://doi.org/10.1093/infdis/jiaa802

Prince-Guerra, J. L., Almendares, O., Nolen, L. D., Gunn, J. K. L., Dale, A. P., Buono, S. A., Deutsch-Feldman, M., Suppiah, S., Hao, L., Zeng, Y., Stevens, V. A., Knipe, K., Pompey, J., Atherstone, C., Bui, D. P., Powell, T., Tamin, A., Harcourt, J. L., Shewmaker, P. L., … Bower, W. A. (2021). Evaluation of Abbott BinaxNOW Rapid Antigen Test for SARS-CoV-2 Infection at Two Community-Based Testing Sites — Pima County, Arizona, November 3–17, 2020. *MMWR. Morbidity and Mortality Weekly Report*, *70*(3), 100–105. https://doi.org/10.15585/mmwr.mm7003e3

Salzer, E., Nixon, E., Drewes, G., Reinhard, F., Bergamini, G., & Rau, C. (2016). Screening Pools of Compounds against Multiple Endogenously Expressed Targets in a Chemoproteomics Binding Assay. *Journal of Laboratory Automation*, *21*(1), 133–142. https://doi.org/10.1177/2211068215595355

Schliep, A., Torney, D. C., & Rahmann, S. (2003). Group testing with DNA chips: Generating designs and decoding experiments. *Computational Systems Bioinformatics. CSB2003. Proceedings of the 2003 IEEE Bioinformatics Conference. CSB2003*, 84–91. https://doi.org/10.1109/CSB.2003.1227307

Stramer, S. L., Glynn, S. A., Kleinman, S. H., Strong, D. M., Caglioti, S., Wright, D. J., Dodd, R. Y., & Busch, M. P. (2004). Detection of HIV-1 and HCV Infections among Antibody-Negative Blood Donors by Nucleic Acid–Amplification Testing. *New England Journal of Medicine*, *351*(8), 760–768. https://doi.org/10.1056/nejmoa040085

Tabor, E., & Epstein, J. S. (2002). NAT screening of blood and plasma donations: evolution of technology and regulatory policy. *Transfusion*, *42*, 1230–1237. https://doi.org/10.1046/j.1537-2995.2002.00183.x

Turner, D. W., Stamey, J. D., & Young, D. M. (2009). Classic group testing with cost for grouping and testing. *Computers and Mathematics with Applications*, *58*, 1930–1935. https://doi.org/10.1016/j.camwa.2009.03.119

Venette, R. C., Moon, R. D., & Hutchison, W. D. (2002). Strategies and Statistics of Sampling for Rare Individuals. *Annual Review of Entomology*, *47*, 143–174.

Walter, S. D., Hildreth, S. W., & Beaty, B. J. (1980). Estimation of infection rates in populations of organisms using pools of variable size. *American Journal of Epidemiology*, *112*(1), 124–128. https://doi.org/10.1093/oxfordjournals.aje.a112961

Wan, W. (2020). WHO declares a pandemic of coronavirus disease covid-19. *The Washington Post*.

Wein, L. M., & Zenios, S. A. (1996). Pooled Testing for HIV Screening : Capturing the Dilution Effect. *Operations Research*, *44*(4), 543–569.

Woloshin, S., Patel, N., & Kesselheim, A. S. (2020). False Negative Tests for SARS-CoV-2 Infection - Challenges and Implications. *The New England Journal of Medicine*, *383*(6).

Zhang, W., Zhang, Z., Krushkal, J., & Liu, A. (2021). Group testing can improve the cost-efficiency of prospective-retrospective biomarker studies. *BMC Medical Research Methodology*, *21*, 1–10. https://doi.org/10.1186/s12874-021-01239-4

Appendix

In chapter 2, we have estimated optimal group sizes to minimize the expected cost of protocols I and II with and without the dilution effects with the help of R programming (Version 4.0.2). While finding the global optima, we first plotted the graph of expected cost against the group size for all prevalence rates. And then from the graph, we noticed the boundary of group size (n) within which the optimal n exists and then we estimated the expected cost based on that n. For example, for cost ratio (r) 4.8 and 0.1% prevalence (p), we can see optimal n lies between 0 and 100 for which the expected cost is minimum. Blue lines indicate that we have got exact optimal n and the reduced cost for that optimal n. But red lines indicate that for those prevalence rates and cost structures, the graph is continuously decreasing with the increasing group sizes. That is why we could not find the exact optimal n within the range of n (0 to 1000) showing on the x-axis. Thus, the expected cost could not be minimized for those cases. These results are shown in Tables 2.4, 2.5 and 2.8.

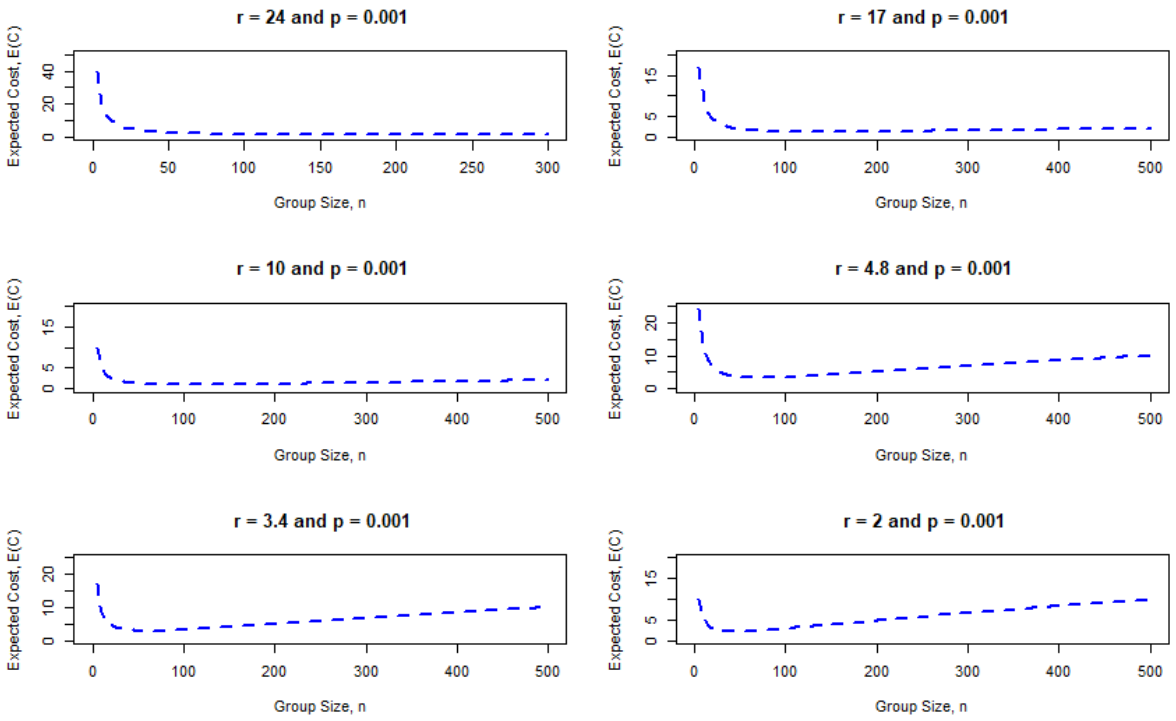


Figure 2.8 Line graphs of the expected cost against group size of protocol I without dilution for all the cost ratios (r) when the prevalence rate (p) is 0.1%.

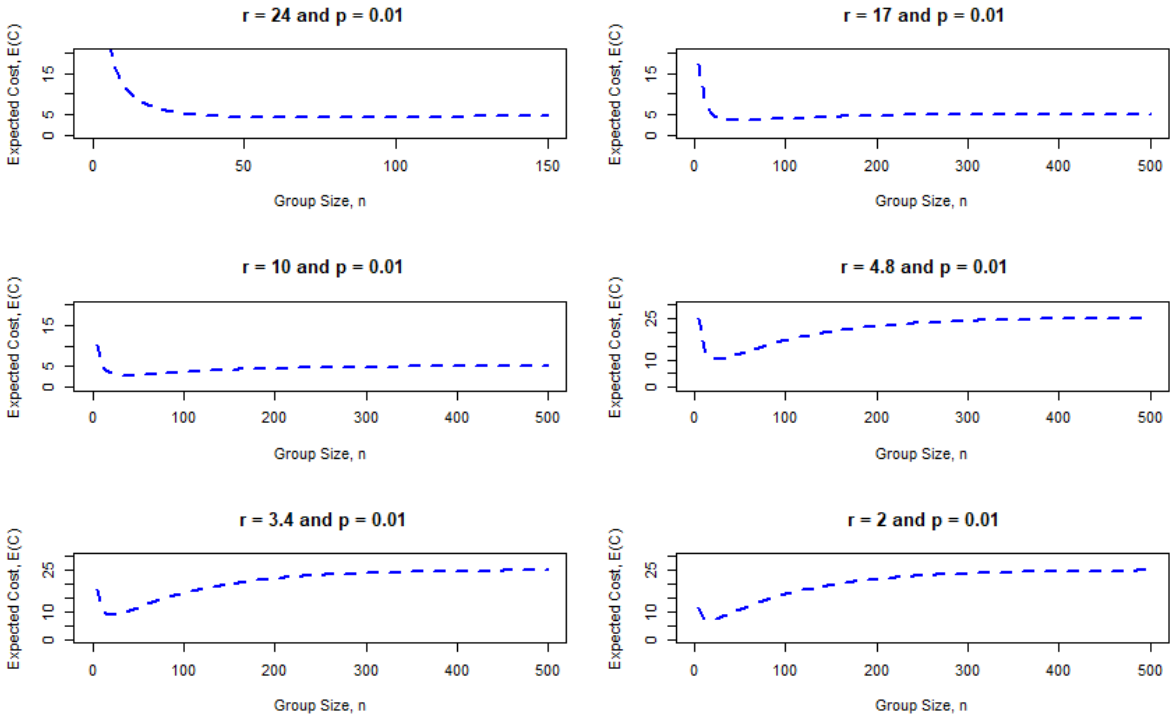


Figure 2.9 Line graphs of the expected cost against group size of protocol I without dilution for all the cost ratios (r) when the prevalence rate (p) is 1%.

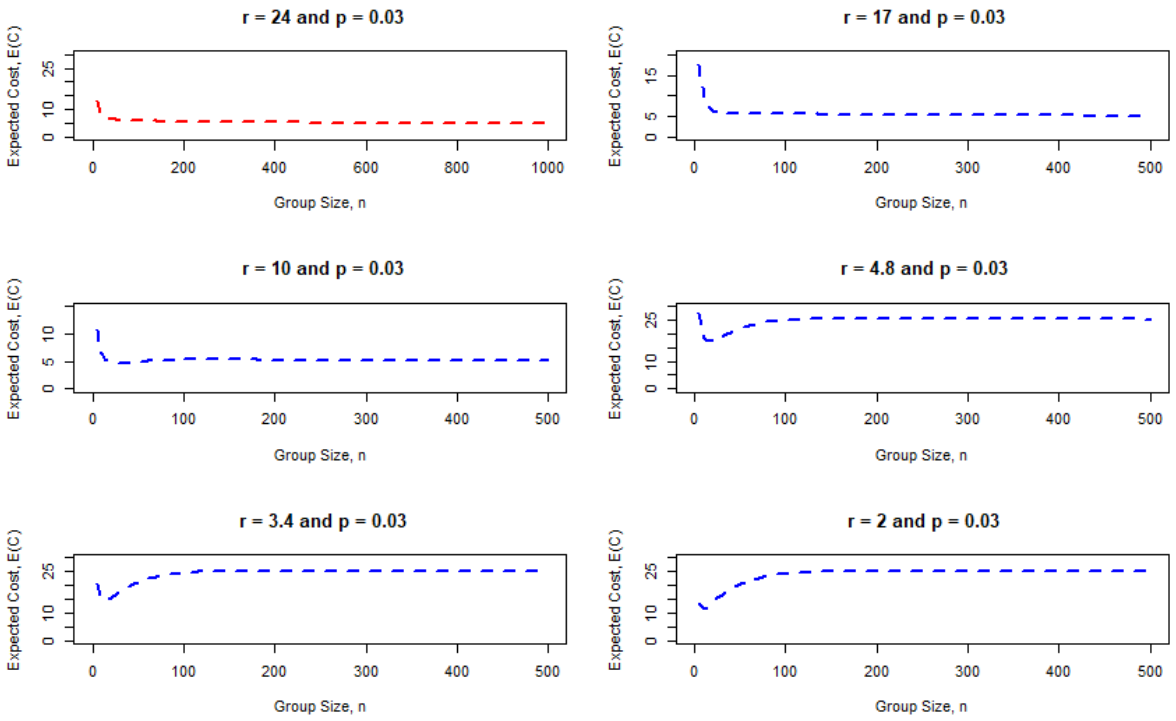


Figure 2.10 Line graphs of the expected cost against group size of protocol I without dilution for all the cost ratios (r) when the prevalence rate (p) is 3%.

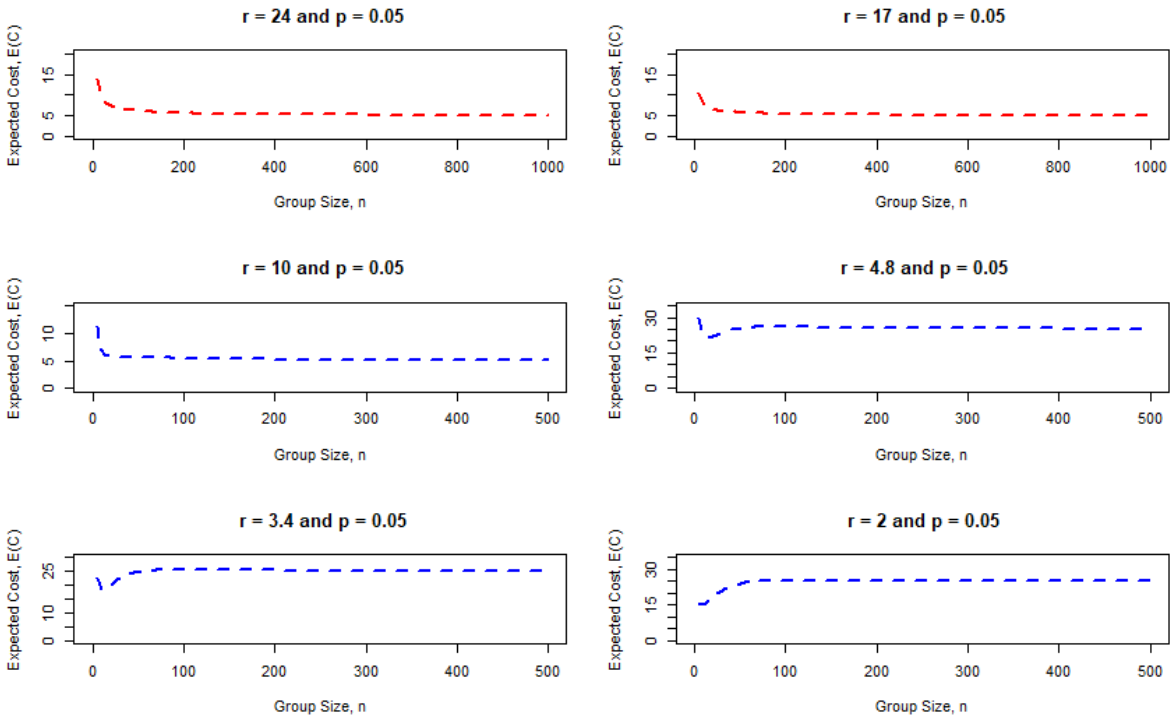


Figure 2.10 Line graphs of the expected cost against group size of protocol I without dilution for all the cost ratios (r) when the prevalence rate (p) is 5%.

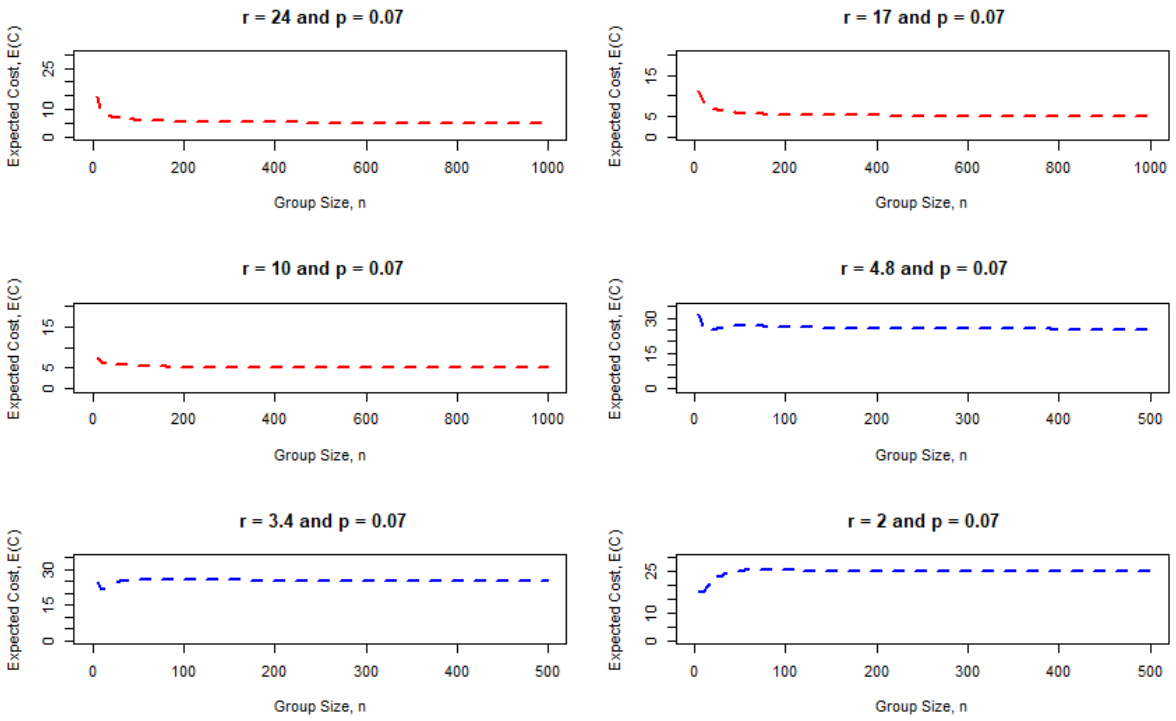


Figure 2.11 Line graphs of the expected cost against group size of protocol I without dilution for all the cost ratios (r) when the prevalence rate (p) is 7%.

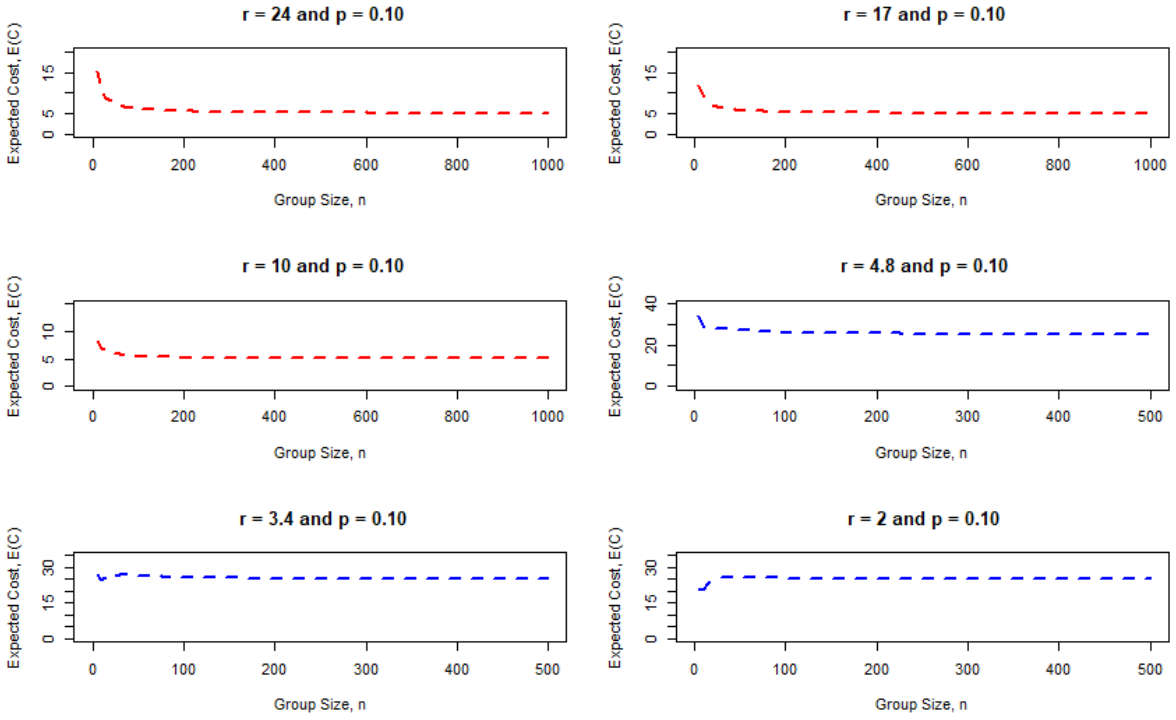


Figure 2.12 Line graphs of the expected cost against group size of protocol I without dilution for all the cost ratios (r) when the prevalence rate (p) is 10%.

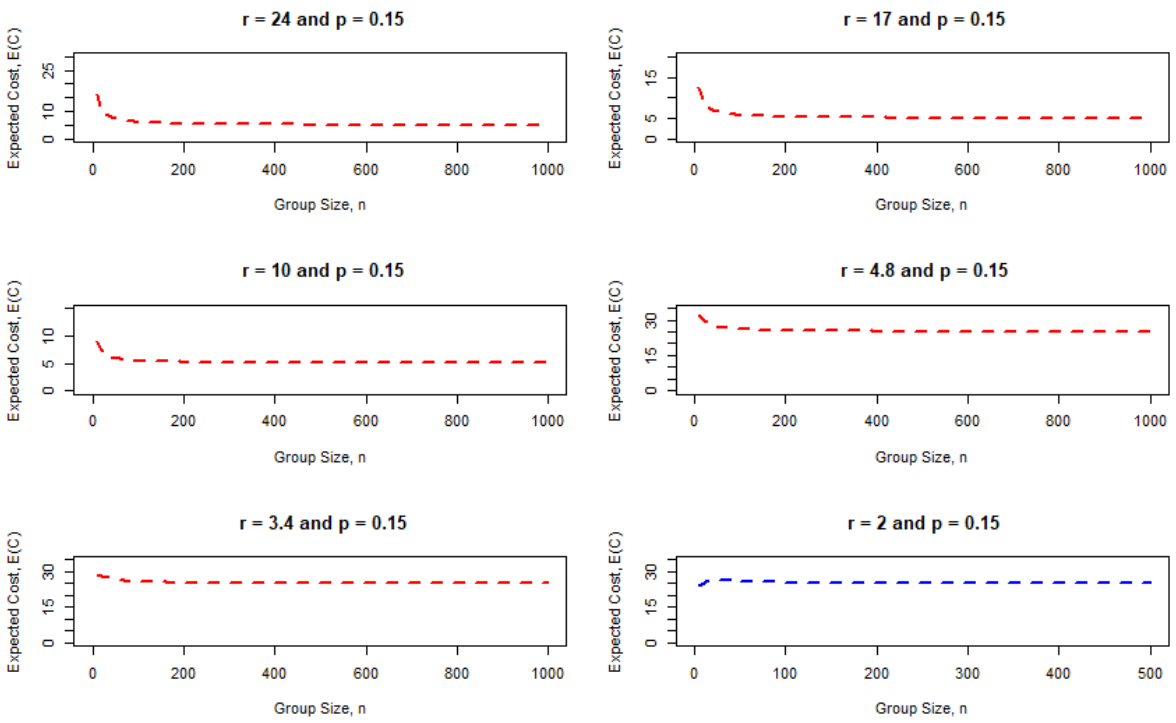


Figure 2.13 Line graphs of the expected cost against group size of protocol I without dilution for all the cost ratios (r) when the prevalence rate (p) is 15%.

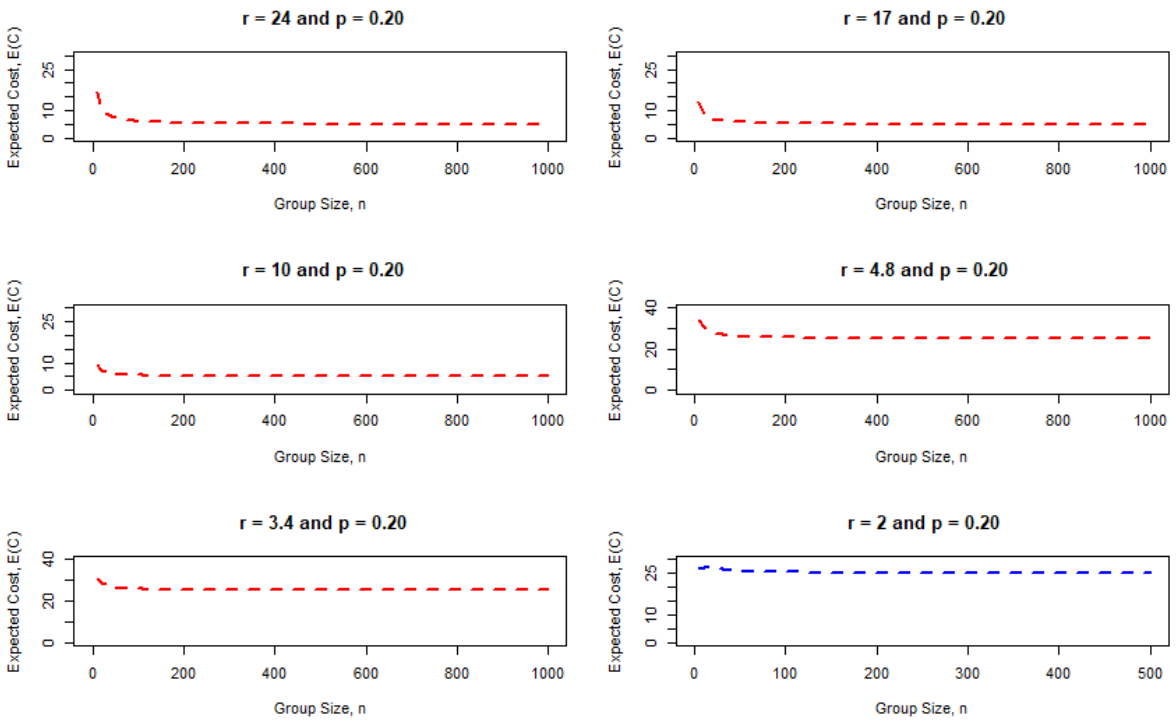


Figure 2.14 Line graphs of the expected cost against group size of protocol I without dilution for all the cost ratios (r) when the prevalence rate (p) is 20%.

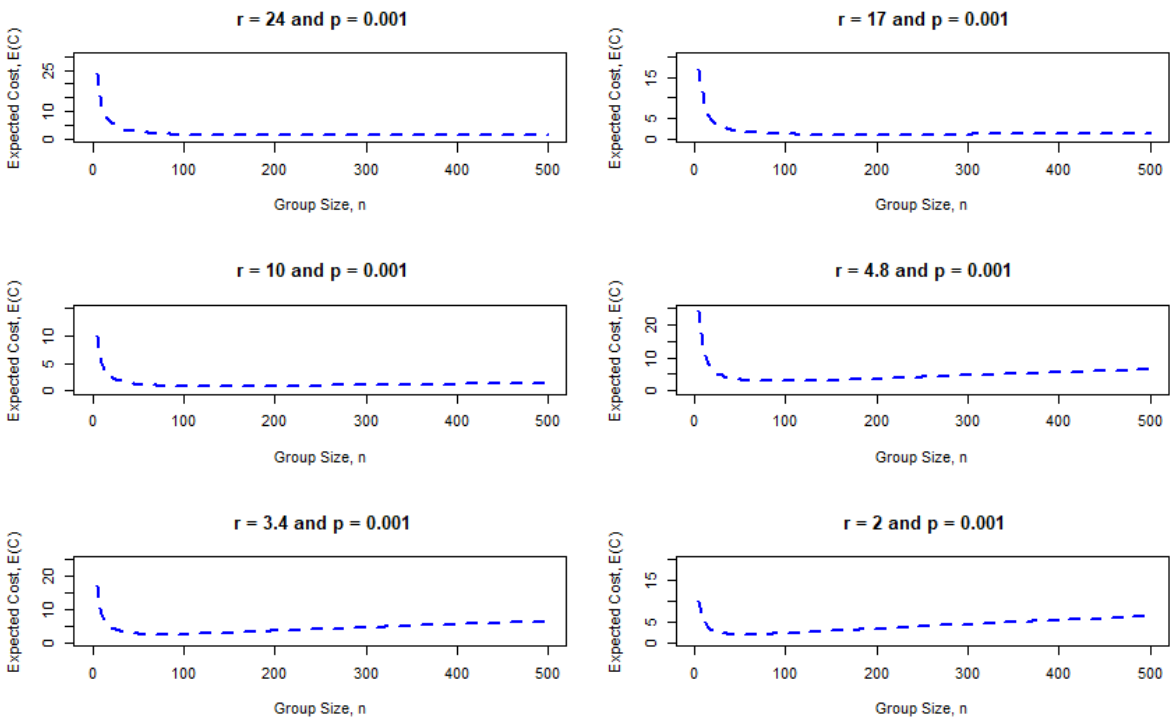


Figure 2.15 Line graphs of the expected cost against group size of protocol I with dilution for all the cost ratios (r) when the prevalence rate (p) is 0.1%.

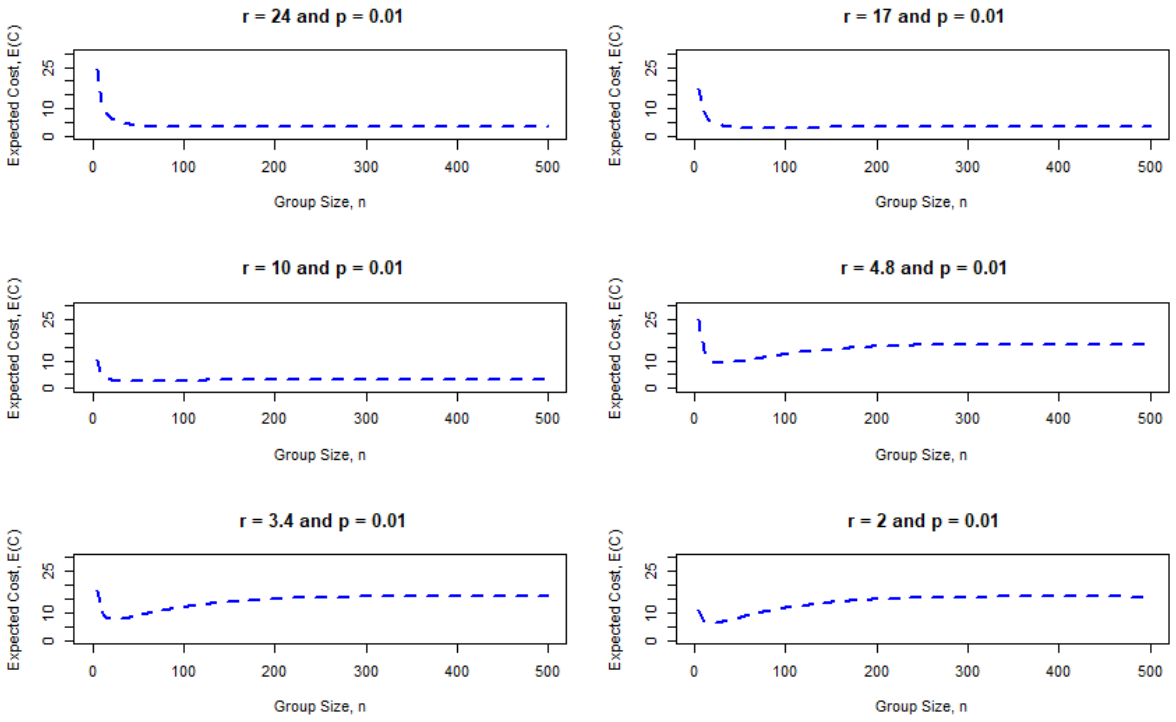


Figure 2.16 Line graphs of the expected cost against group size of protocol I with dilution for all the cost ratios (r) when the prevalence rate (p) is 1%.

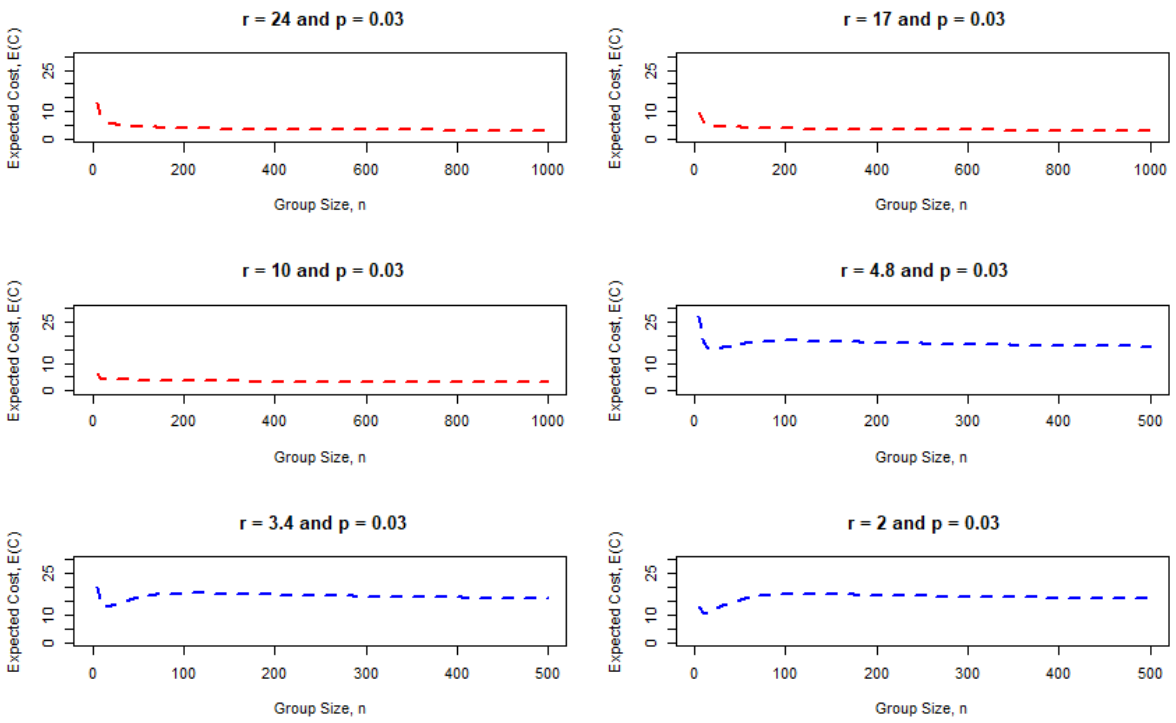


Figure 2.17 Line graphs of the expected cost against group size of protocol I with dilution for all the cost ratios (r) when the prevalence rate (p) is 3%.

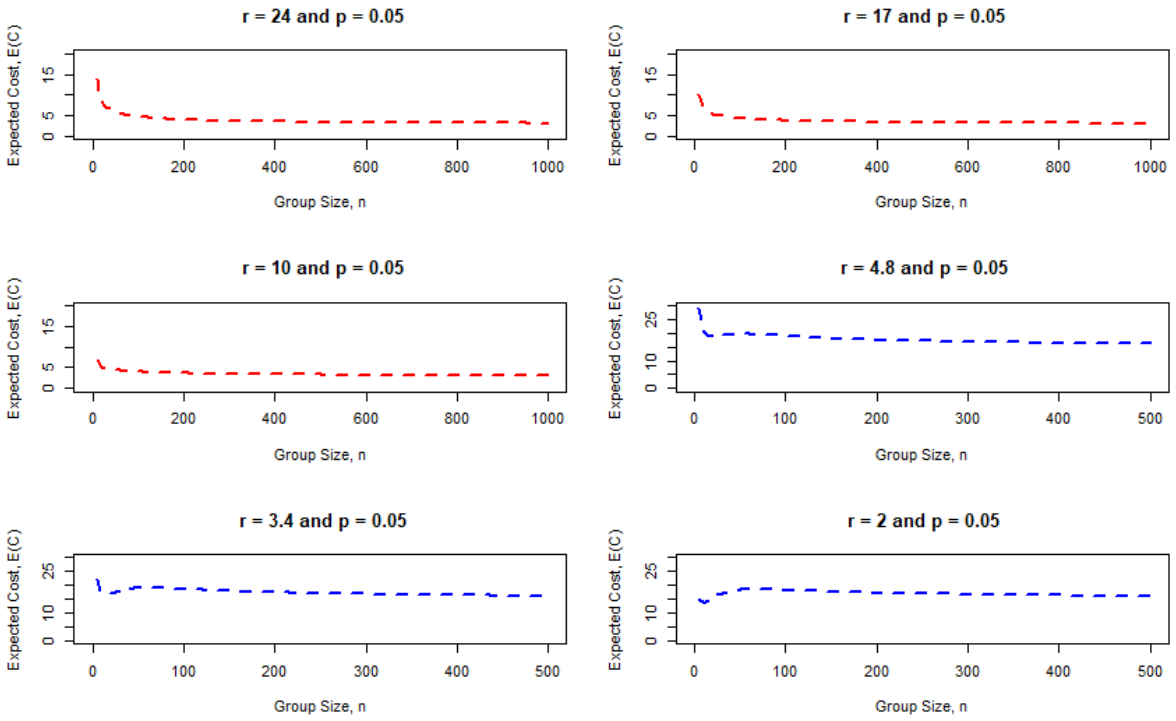


Figure 2.18 Line graphs of the expected cost against group size of protocol I with dilution for all the cost ratios (r) when the prevalence rate (p) is 5%.

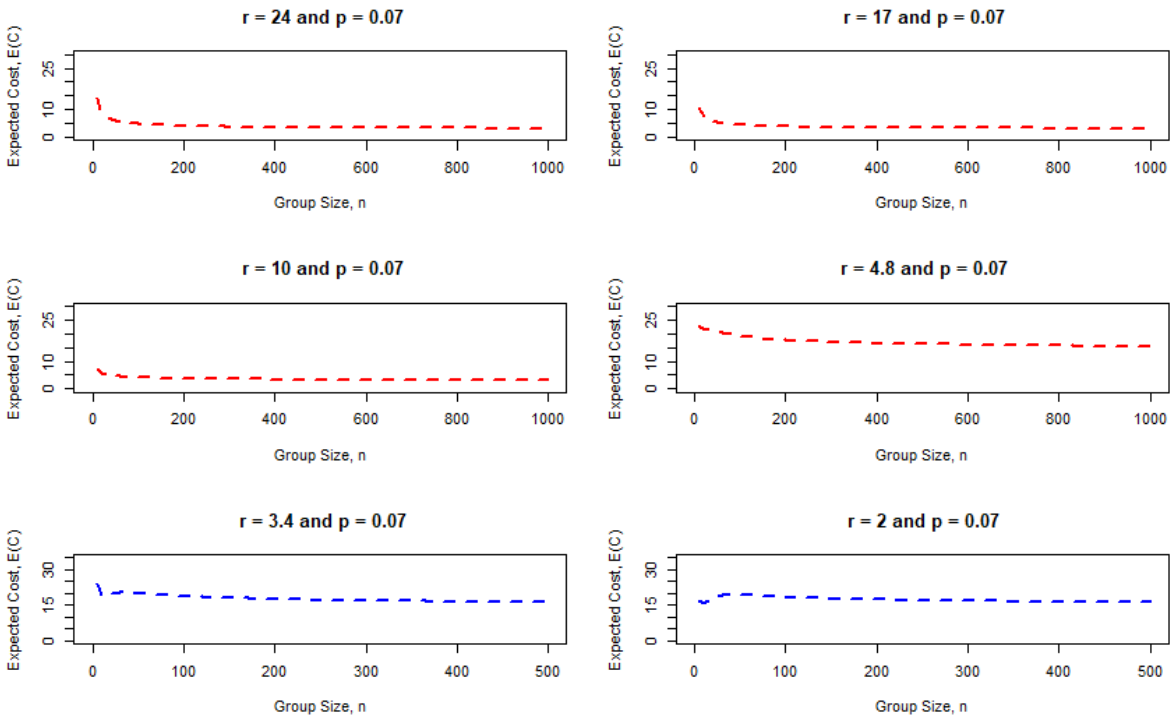


Figure 2.19 Line graphs of the expected cost against group size of protocol I with dilution for all the cost ratios (r) when the prevalence rate (p) is 7%.

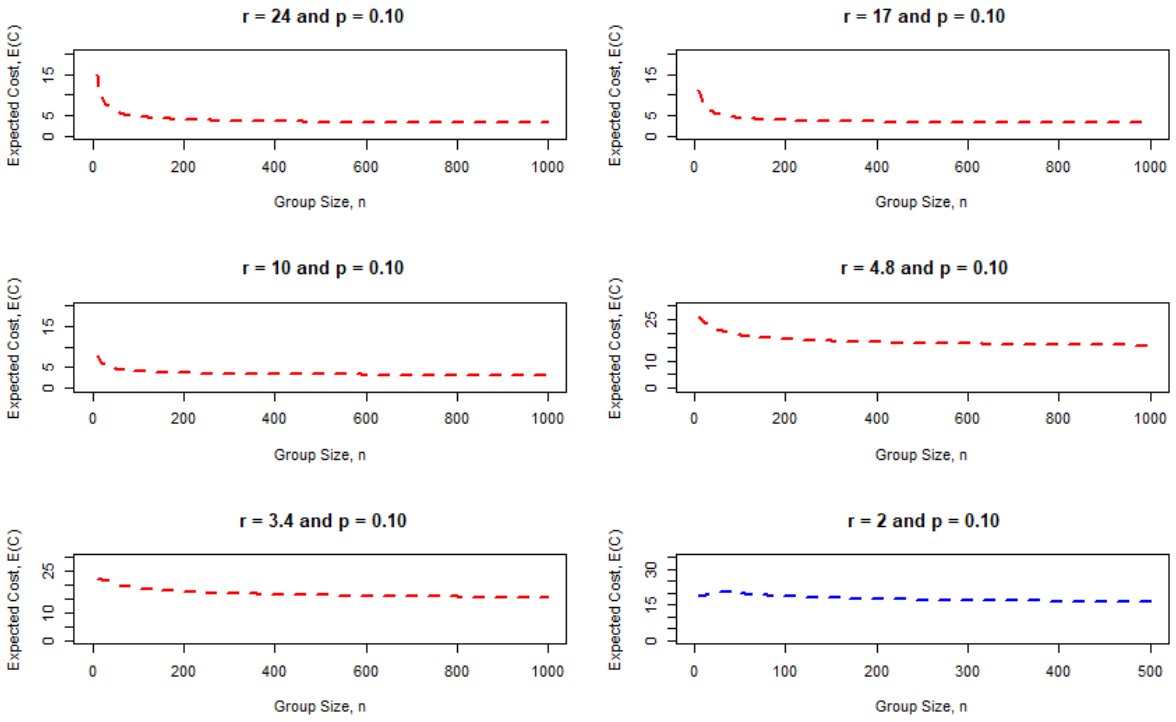


Figure 2.20 Line graphs of the expected cost against group size of protocol I with dilution for all the cost ratios (r) when the prevalence rate (p) is 10%.

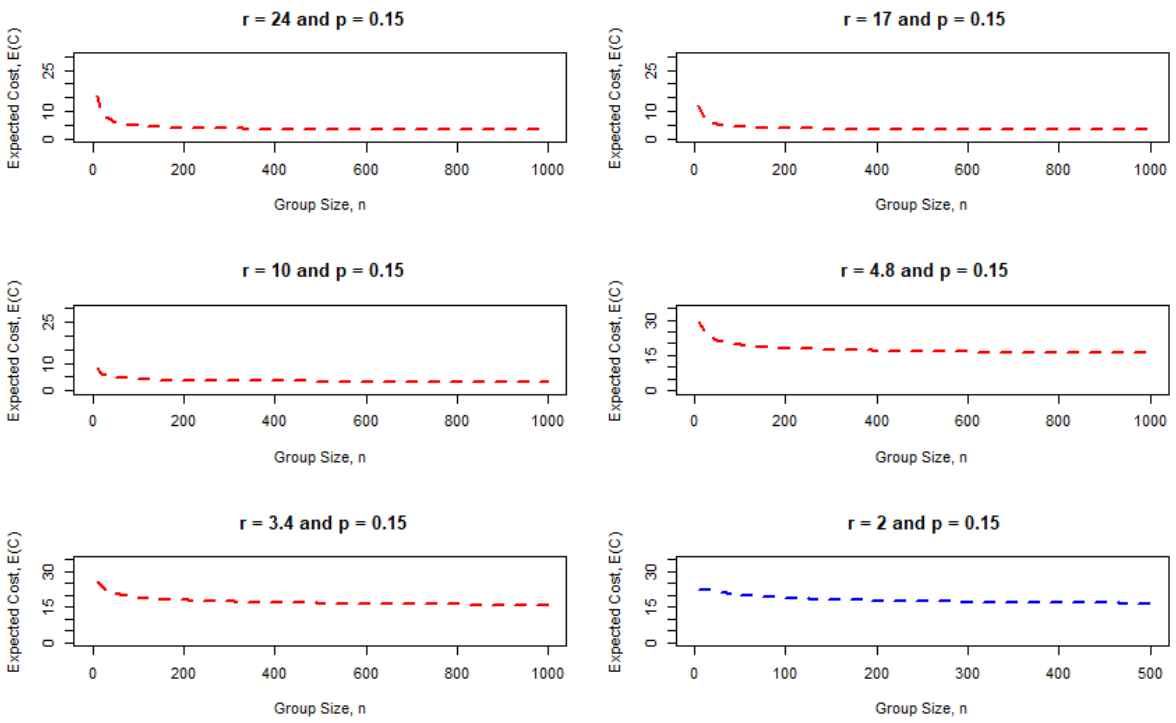


Figure 2.21 Line graphs of the expected cost against group size of protocol I with dilution for all the cost ratios (r) when the prevalence rate (p) is 15%.

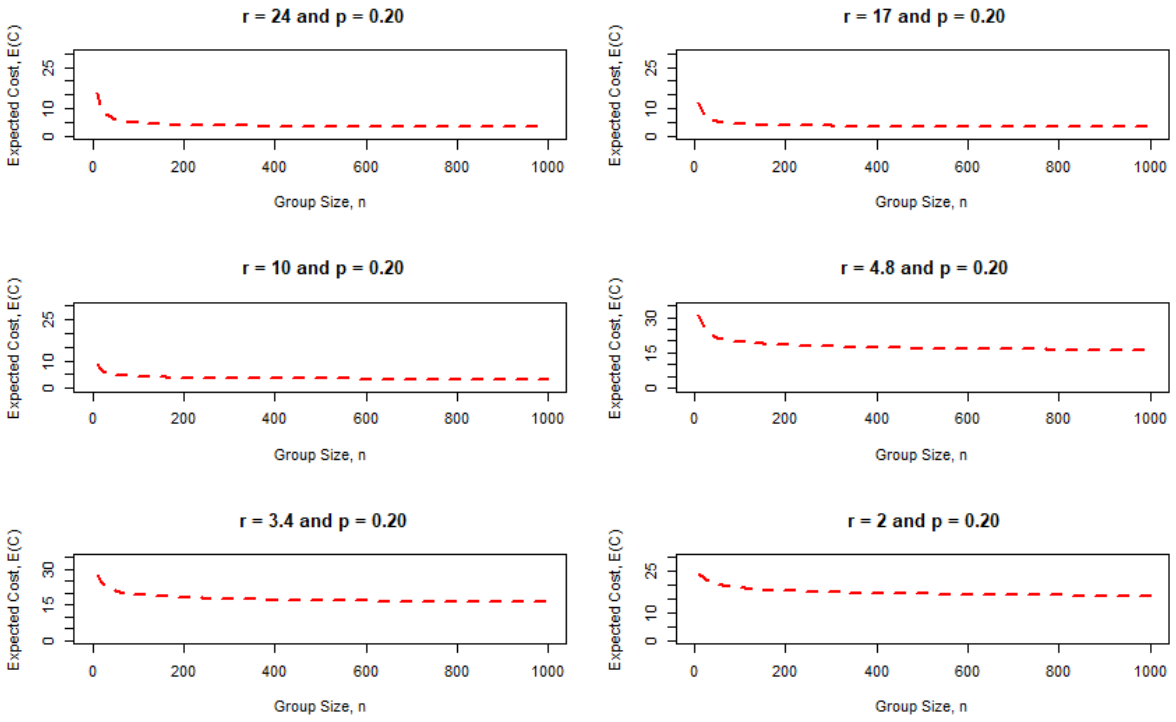


Figure 2.22 Line graphs of the expected cost against group size of protocol I with dilution for all the cost ratios (r) when the prevalence rate (p) is 20%.

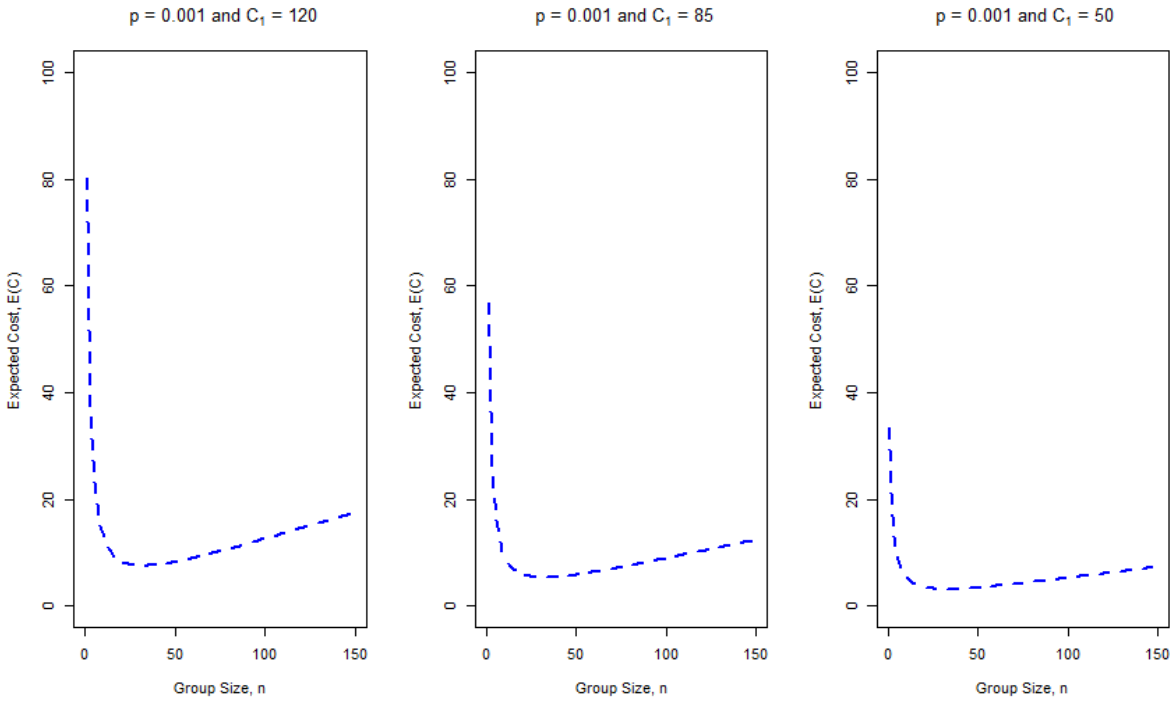


Figure 2.23 Line graphs of the expected cost against group size of protocol II without dilution for $C_1 = 120, 85, 50$ (entire cost of RT-PCR) when the prevalence rate (p) is 0.1%.

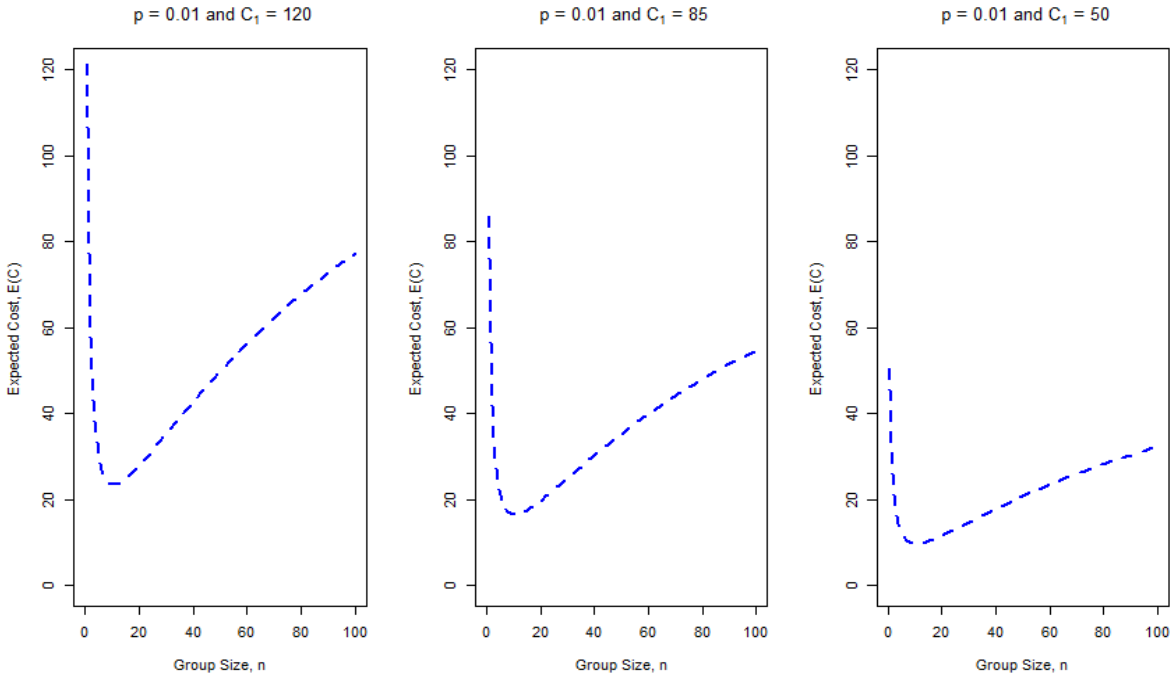


Figure 2.24 Line graphs of the expected cost against group size of protocol II without dilution for $C_1 = 120, 85, 50$ (entire cost of RT − PCR) when the prevalence rate (p) is 1%.

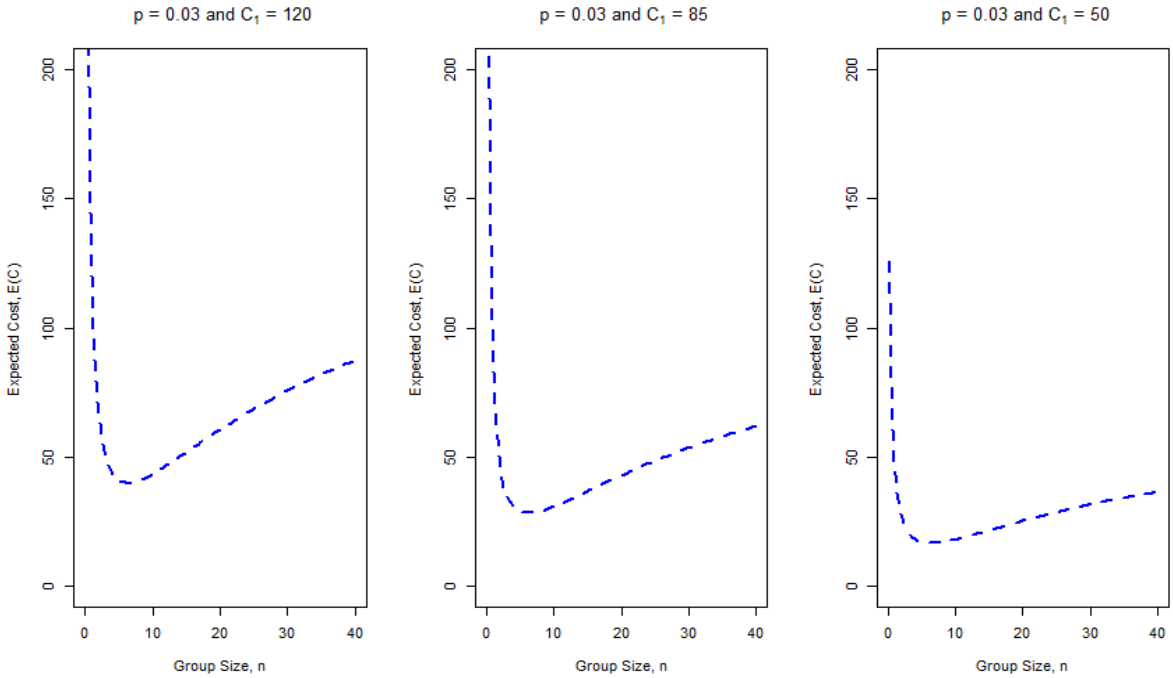


Figure 2.25 Line graphs of the expected cost against group size of protocol II without dilution for $C_1 = 120, 85, 50$ (entire cost of RT-PCR) when the prevalence rate (p) is 3%.

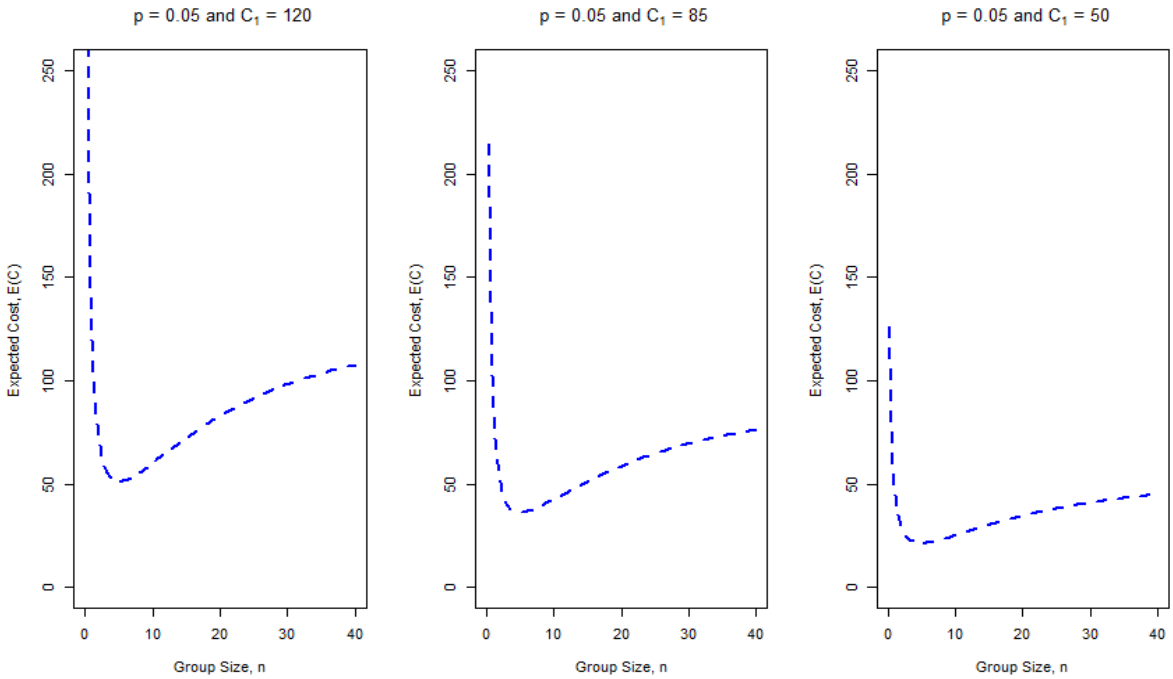


Figure 2.26 Line graphs of the expected cost against group size of protocol II without dilution for $C_1 = 120, 85, 50$ (entire cost of RT-PCR) when the prevalence rate (p) is 5%.

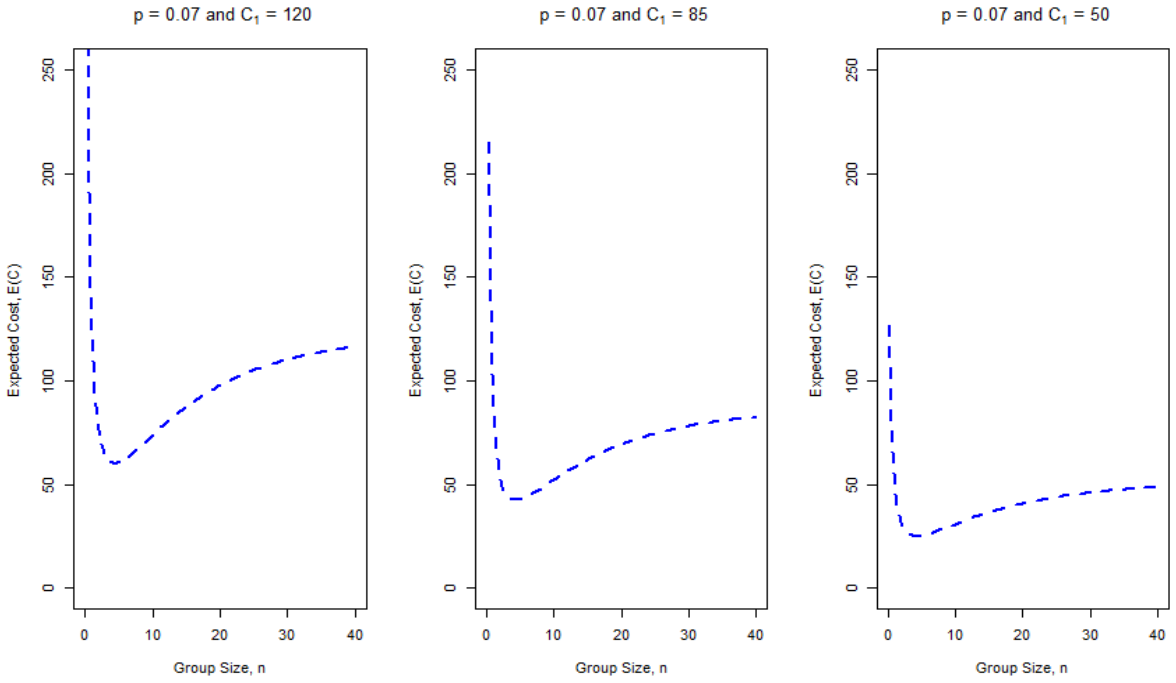


Figure 2.27 Line graphs of the expected cost against group size of protocol II without dilution for $C_1 = 120, 85, 50$ (entire cost of RT-PCR) when the prevalence rate (p) is 7%.

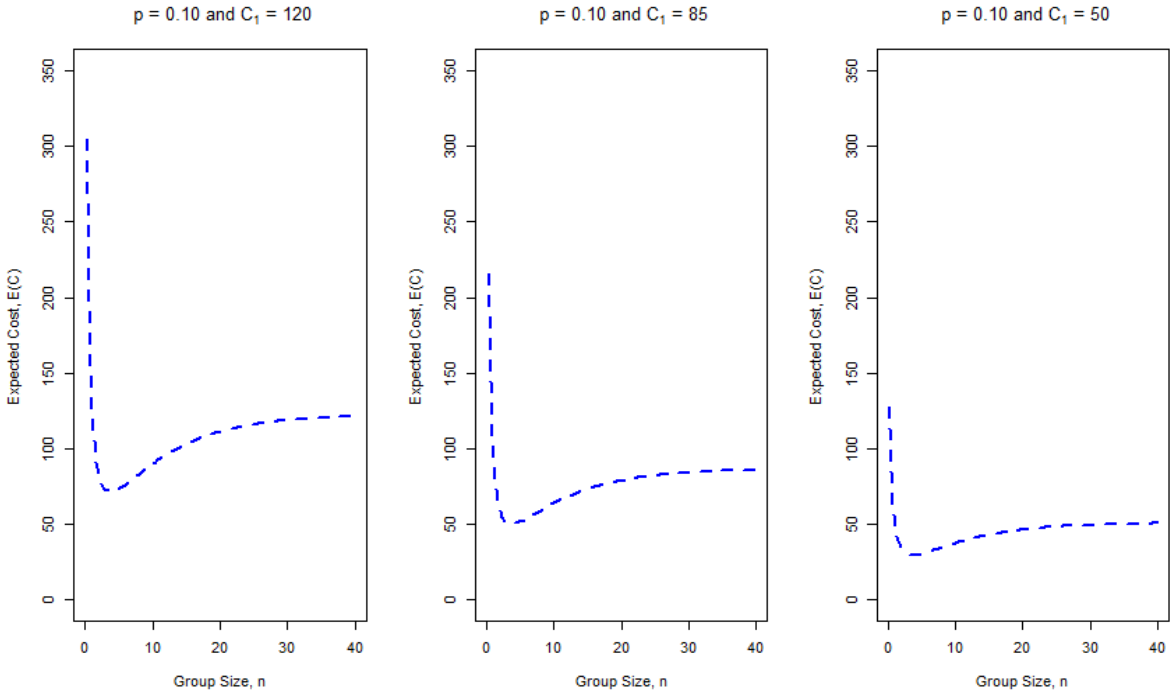


Figure 2.28 Line graphs of the expected cost against group size of protocol II without dilution for $C_1 = 120, 85, 50$ (entire cost of RT-PCR) when the prevalence rate (p) is 10%.

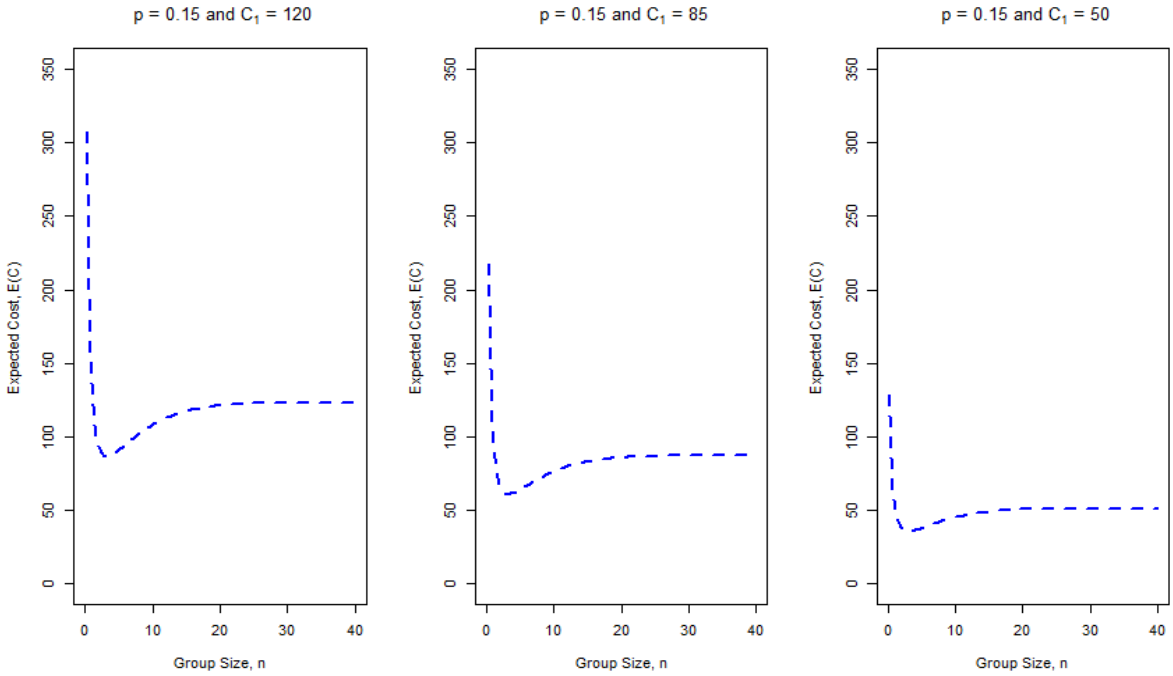


Figure 2.29 Line graphs of the expected cost against group size of protocol II without dilution for $C_1 = 120, 85, 50$ (entire cost of RT-PCR) when the prevalence rate (p) is 15%.

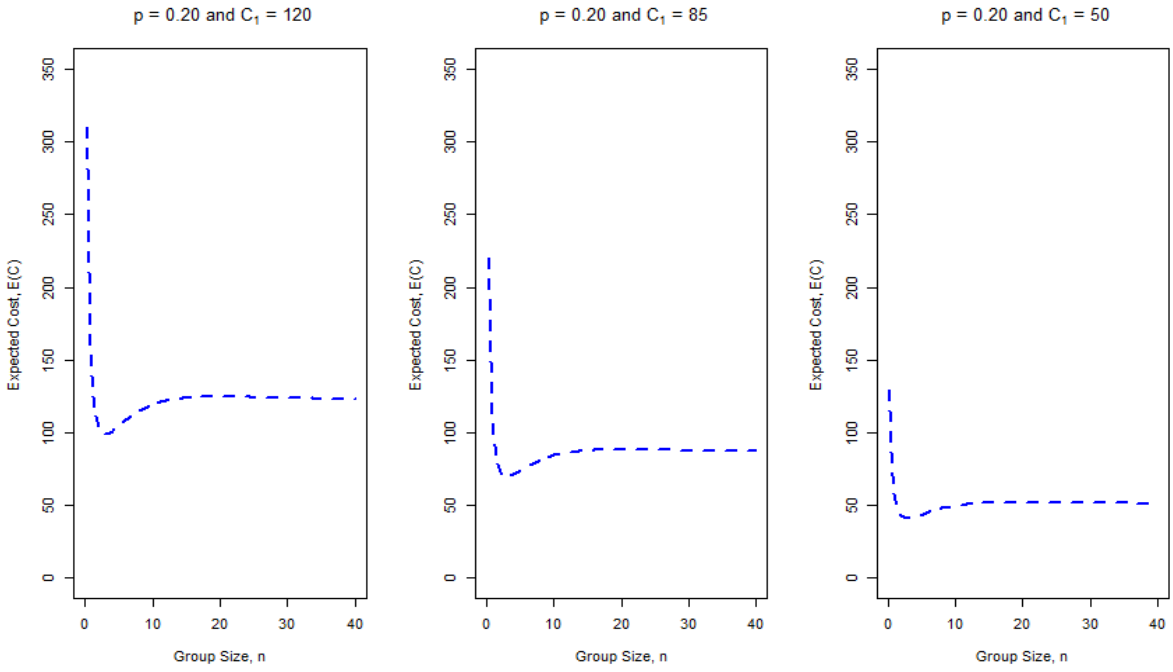


Figure 2.30 Line graphs of the expected cost against group size of protocol II without dilution for $C_1 = 120, 85, 50$ (entire cost of RT-PCR) when the prevalence rate (p) is 20%.

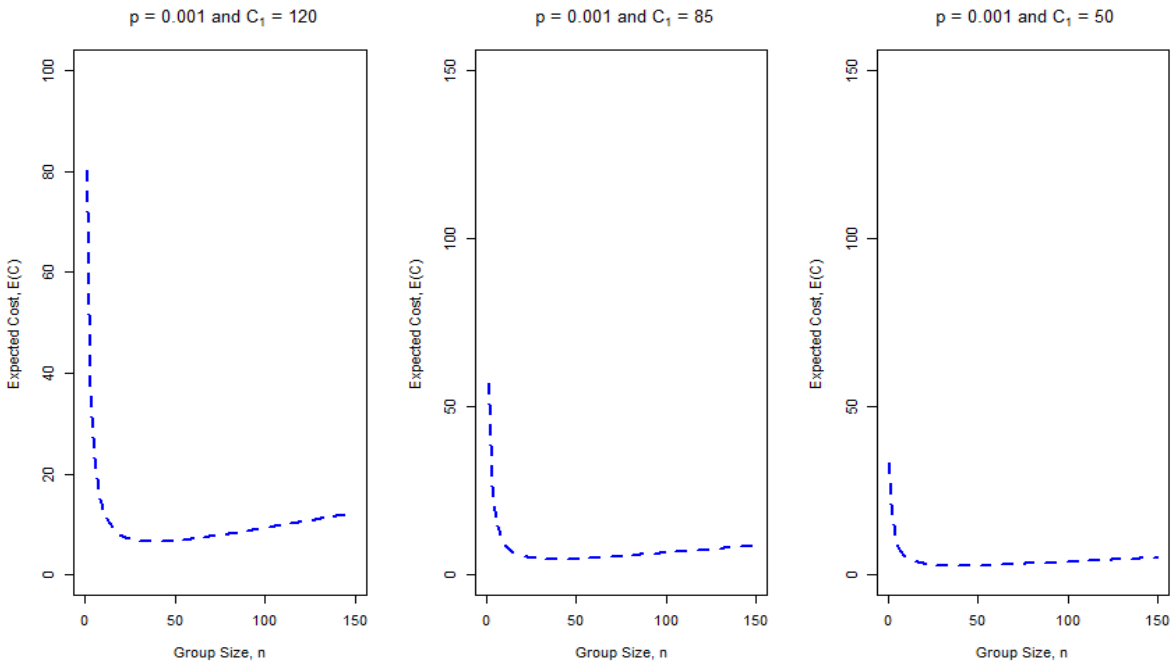


Figure 2.31 Line graphs of the expected cost against group size of protocol II with dilution for $C_1 = 120, 85, 50$ (entire cost of RT-PCR) when the prevalence rate (p) is 0.1%.

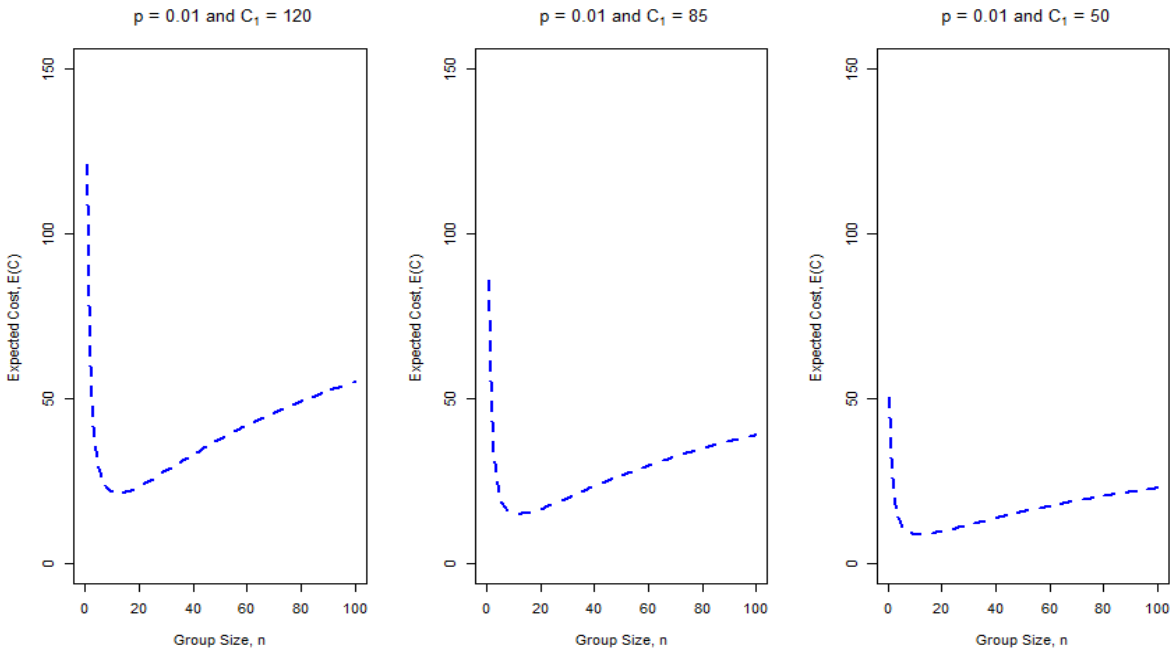


Figure 2.32 Line graphs of the expected cost against group size of protocol II with dilution for $C_1 = 120, 85, 50$ (entire cost of RT-PCR) when the prevalence rate (p) is 1%.

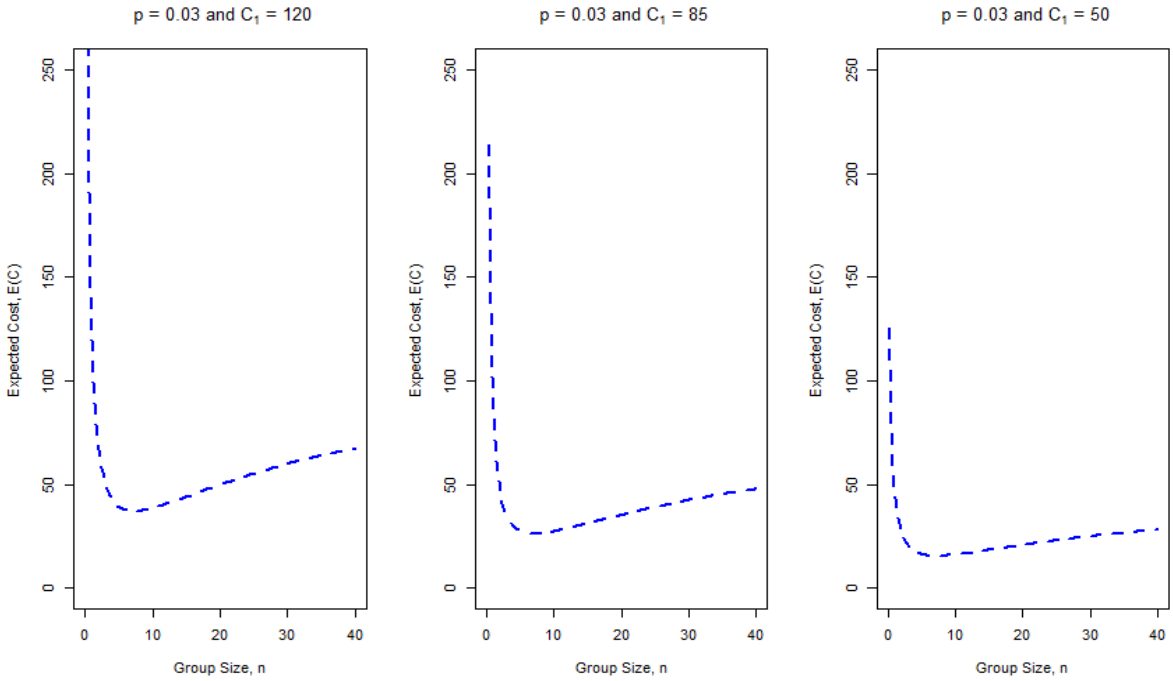


Figure 2.33 Line graphs of the expected cost against group size of protocol II with dilution for $C_1 = 120, 85, 50$ (entire cost of RT-PCR) when the prevalence rate (p) is 3%.

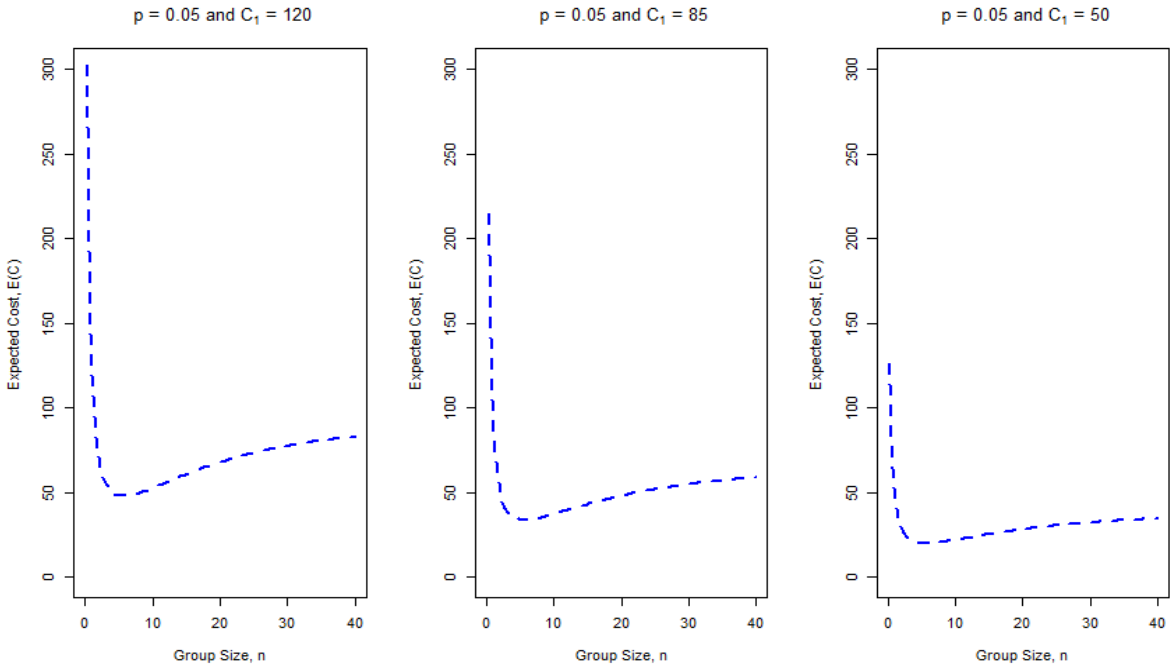


Figure 2.34 Line graphs of the expected cost against group size of protocol II with dilution for $C_1 = 120, 85, 50$ (entire cost of RT-PCR) when the prevalence rate (p) is 5%.

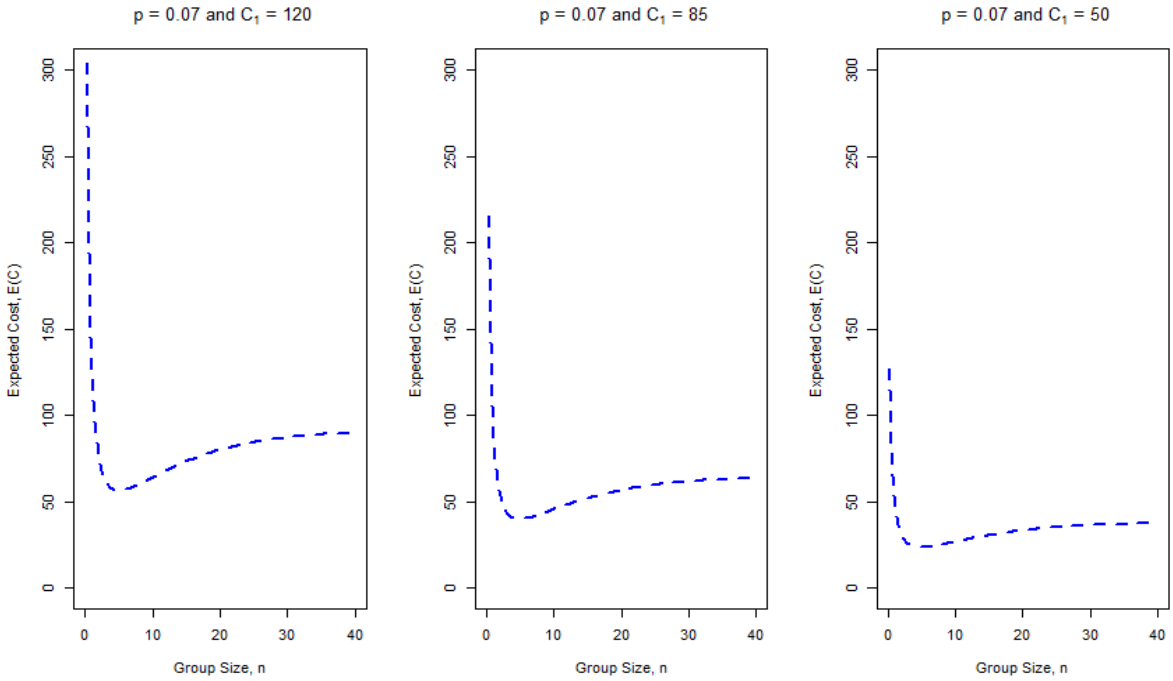


Figure 2.35 Line graphs of the expected cost against group size of protocol II with dilution for $C_1 = 120, 85, 50$ (entire cost of RT-PCR) when the prevalence rate (p) is 7%.

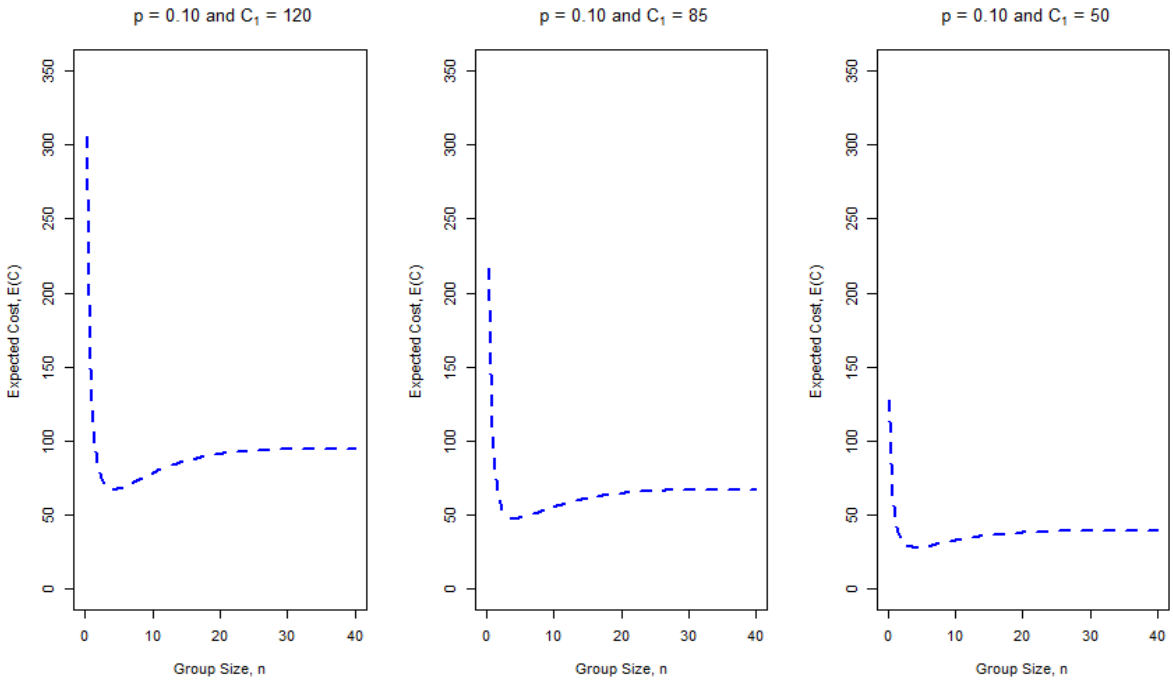


Figure 2.36 Line graphs of the expected cost against group size of protocol II with dilution for $C_1 = 120, 85, 50$ (entire cost of RT-PCR) when the prevalence rate (p) is 10%.

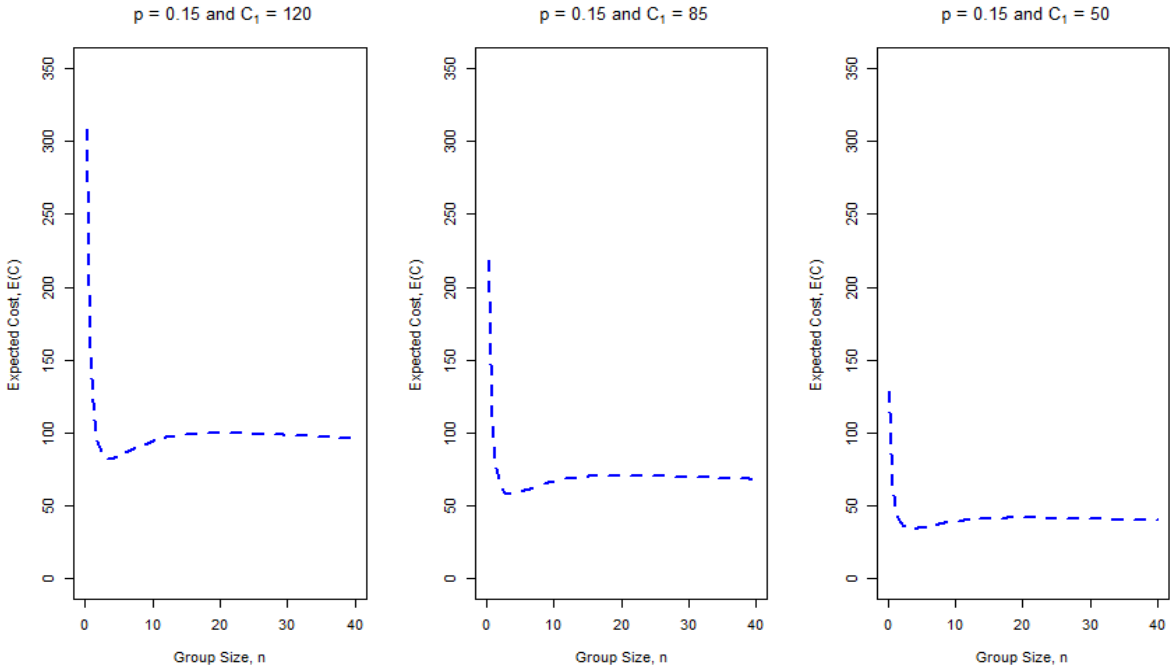


Figure 2.37 Line graphs of the expected cost against group size of protocol II with dilution for $C_1 = 120, 85, 50$ (entire cost of RT-PCR) when the prevalence rate (p) is 15%.

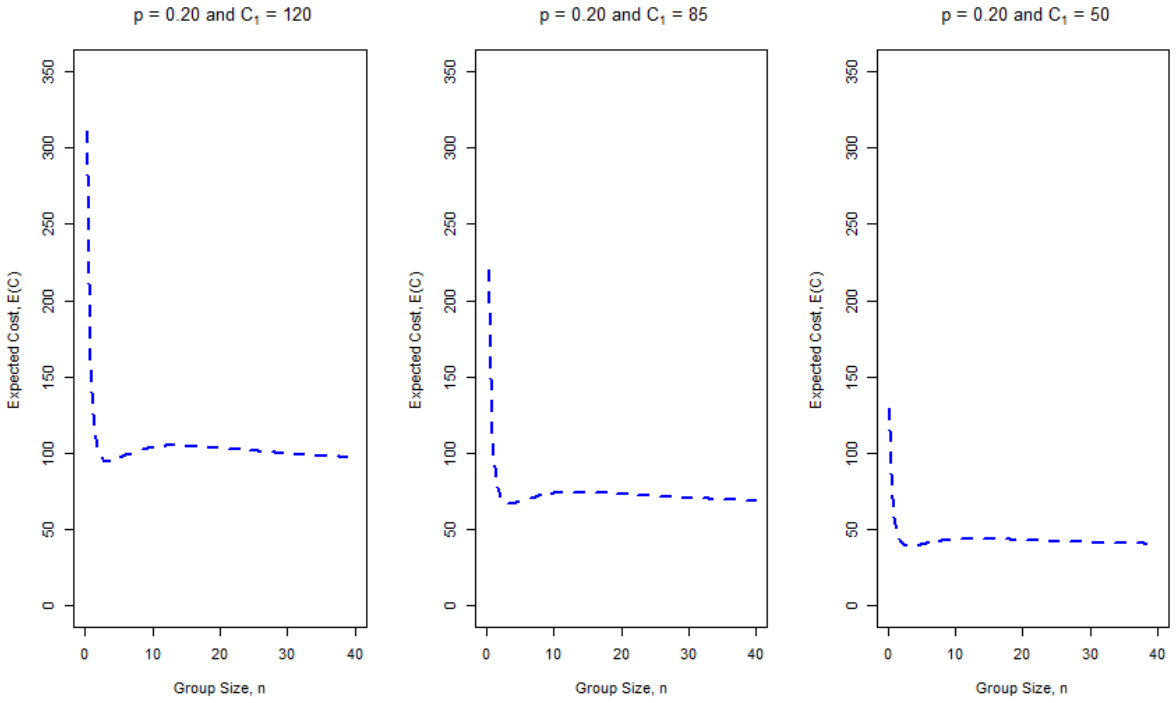


Figure 2.38 Line graphs of the expected cost against group size of protocol II with dilution for $C_1 = 120, 85, 50$ (entire cost of RT-PCR) when the prevalence rate (p) is 20%.