

Wide-scale Pooled Sample Testing for COVID-19: Application, Optimisation and Probabilistic Analysis

Xiaolong Su

Wuhan University

Man Zhou (✉ civilcsu@sina.com)

Wuhan University

Research

Keywords: COVID-19, Nucleic acid testing, Individual sample testing, Pooled testing, Mass testing, Probabilistic analysis

Posted Date: September 23rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-871658/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background Currently, coronavirus disease 2019 (COVID-19) has become the most severe infectious disease affecting the world, which has spread around the world to more than 200 countries in 2020. Until COVID-19 vaccine technology completely matures, nucleic acid testing is considered as an effective way to screen virus carriers and control the spread of the virus. Considering that the medical resources and infection rates are different across various countries and regions, if all infected areas adopt the traditional individual nucleic acid testing method, the workload will be heavy and time-consuming. Therefore, this will not lead to the control of the pandemic.

Methods After Wuhan completed a citywide nucleic acid testing in May 2020, China basically controlled the spread of COVID-19 and entered the post-epidemic period. Since then, although some cities in China, such as Qingdao, Xinjiang, Beijing, Dalian, have experienced a local epidemic resurgence, the pandemic was quickly suppressed through wide-scale pooled nucleic acid testing methods. Combined with the successful experience of mass nucleic acid testing in China, this study introduces two main pooled testing methods used in two cities with a population of more than ten million people, Wuhan's "five-in-one" and Qingdao's "ten-in-one" rapid pooled testing methods. This study proposes an improved method for optimising the second round of "ten-in-one" pooled testing, known as "the pentagram mini-pooled testing method", which speeds up the testing process (as a result of reducing the numbers of testing by 40%) and significantly reduces the cost.

Findings Qingdao's optimised "ten-in-one" pooled testing method quickly screens out the infections by running fewer testing. This study also mathematically examines the probabilistic principles and applicability conditions for pooled testing of COVID-19. Herein, the study theoretically determines the optimal number of samples that could successfully be combined into a pool under different infection rates. Then, it quantitatively discusses the applicability and principles for choosing the pooled testing instead of individual testing.

Interpretation Overall, this research offers a reference for other countries with different infection rates to help them in implementing the mass testing for COVID-19 to reduce the spread of coronavirus.

Contributions To The Literature

- The methods and applicable conditions of "multi-in-one" nucleic acid testing is introduced, which is helpful for the detection of COVID-19
- The optimization of "ten - in - one" nucleic acid detection method is proposed, which can reduce detection times
- Conditions for the application of mixed testing are proposed to allow rapid response to COVID-19 testing

Background

Coronavirus disease 2019 (COVID-19) has become the most severe infectious lung disease affecting humanity, which has spread to more than 200 countries and regions. As of January 31, 2021, more than 102.39 million cases have been confirmed globally, and more than 2.21 million people have died because of this disease. Currently, COVID-19 pandemic is considered as one of the most significant health challenges which endangers both human health and economic development. The pathogen causing the new coronavirus pneumonia is severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, hereinafter referred to as COVID-19)¹, which is extremely contagious. Furthermore, it has recently been discovered that this virus has mutated into a new strain called VUI2020/01, with a transmission rate 70% higher than the original coronavirus strain². Many countries, including the United Kingdom, Denmark, the Netherlands, and Japan, have reported infections with such new mutant virus, which has led to further complications to prevention and control methods. In this grim situation, the top priority should be to detect the infections as quickly as possible and exclude noncarriers so that targeted preventive measures can be implemented based on testing results. The testing speed of COVID-19 depends not only on the specific testing protocol but also on the selected testing method. The improvement of testing protocols, such as whole-genome sequencing^{3,4}, RT-PCR⁵⁻⁸, CRISPR⁹⁻¹², and RT-LAMP^{13,14}, can shorten the testing time of individual sample. However, still numerous testing is simultaneously needed since the highly contagious coronavirus has created a large-scale urgent need for testing. Hence, relying only on improved testing protocols will not significantly improve the testing efficiency when the testing equipment and/or testing products are relatively limited. In this circumstance, the choice of testing method becomes crucial. Using appropriate testing methods based on the characteristics of COVID-19 can largely reduce testing times while ensuring testing accuracy. In the nucleic acid screening testing, the conventional test method is used to test samples one by one and it is known as the individual-sample testing. Individual-sample testing ensures timely and effective testing results for relatively small sample sizes, but not for the large sample sizes. Adopting the individual-sample testing for screening requires considerable manpower, material and financial resources to test a large number of samples of highly infectious diseases (e.g., millions of samples). Consequently, it delays the formulation and implementation of related countermeasures because the process is time-consuming. On August 6, 2020, Liverpool, in the northwest of the United Kingdom, launched the first COVID-19 testing pilot project in England to avoid overwhelming local hospitals due to the second wave of the epidemic. The project used conventional individual-sample testing methods, which were expected to complete nucleic acid testing of 500,000 residents of Liverpool within two weeks. However, due to the enormous workload, only 200,000 nucleic acid testing were completed in such two weeks, which squandered an excellent opportunity to prevent the worsening of the epidemic. In this case, the pooled testing method would have been a better option because it significantly reduces testing times and costs and it provides low prevalence of the disease¹⁵⁻¹⁶. Pooled testing refers to mixing multiple samples before testing. If the result is negative, it proves that none of the individuals in the sample are infected; if positive, individual sample method needs to be implemented to identify the infected person¹⁷⁻¹⁸. Majid et al.¹⁹ recognised that pooled testing is not only feasible for low-income countries, but it also provides an efficient use of scarce testing kits. Additionally, Ball et al.²⁰ indicated that pooled testing for SARS-CoV-2 could provide a solution to the UK testing strategy. China was the first country to successfully control the epidemic although Wuhan, with a population of 11 million, was once the most severely affected city in China. In April 2020, Wuhan's nucleic acid testing

used the individual-sample testing method. With an average daily test capacity of approximately 46,000, it was estimated to take nearly eight months to test the entire city's population, which was unfavourable for rapid investigation of infected persons. Accordingly, in May 2020, the “five-in-one” pooled testing method was adopted to improve testing efficiency, and approximately 10 million nucleic acid testing were completed in 19 days. Drawing lessons from Wuhan's successful nucleic acid testing experience, Qingdao's nucleic acid testing adopted the “ten-in-one” pooled testing method in October 2020. The results showed that 10,899,145 nucleic acid testing were completed within five days. Therefore, pooled testing significantly shortens the testing time and improves testing efficiency. Instead of pooling a maximum of five samples as in Wuhan, the “ten-in-one” pooled testing method further improved the testing efficiency. Therefore, choosing a scientific, effective and rapid testing method can quickly and accurately accomplish COVID-19 nucleic acid testing and screening, providing a basis for formulating and implementing subsequent prevention and treatment measures. In this paper, Qingdao's “ten-in-one” pool testing method is optimised, and a more practical and efficient “five-pointed star” pool testing method is proposed for the first time. Moreover, the theoretical basis and applicable conditions of pooled testing are analysed to provide theoretical references for rapid nucleic acid testing in epidemic areas.

Methods

Observational case study.

“Five-in-one” pooled testing for COVID-19 in Wuhan

Wuhan was locked down from January 23, 2020. After the epidemic was controlled, Wuhan lifted travel restrictions on April 8 after an 11-week (77-day) lockdown. However, a small number of local cases emerged, including asymptomatic carriers, within a short period after lifting travel restrictions. The Chinese government was highly concerned about locally spreading cases and worried that this would lead to a second wave of the epidemic. Consequently, Wuhan performed a citywide intensive nucleic acid testing of COVID-19 from May 14 to June 1 and comprehensively checked for asymptomatic infections to control the local epidemic, allowing work and production to resume. To improve the detection efficiency and control the spread of the epidemic, the “five-in-one” pooled testing method was adopted using nucleic acid detection for all the people in Wuhan. Specifically, the method used samples of five people as a set and then sampled the sets separately. As shown in Fig. 1, before collecting the samples, the people's information, including name, identity card number, contact phone number, sample collection site, sample collection date, and sample collection time, was collected and registered. Collection tubes were numbered according to the sets. The following items were collected: throat swabs, nasal swabs, blood samples and stool samples. For large-scale sample collection work, such as in Wuhan, throat swab collection has a significant advantage in operative convenience. The sampling workers used a plastic swab with a polypropylene fibre tip to wipe the subject's bilateral pharyngeal tonsils and posterior pharyngeal wall during the throat swab collection from the population. The swabs were stored in the same sampling tube for nucleic acid testing with the swabs from the other four people.

The principle of “five-in-one” pooled testing is shown in Fig. 2. The samples collected from five people are mixed before testing. If the result of the pooled sample is negative, all five samples in the set are negative, i.e., the five people in the pooled sample are safe. On the opposite, if the pooled sample yielded positive result, the five people are notified as soon as possible. Those people are immediately isolated and tested separately to identify the positive sample. “Five-in-one” pooled testing saves both time and costs and reduces sampling inspections, thereby reducing the workload of the health sector. Especially in large cities such as Beijing, Wuhan and Qingdao, each of which has over 10 million people, pooled testing can be relied on for the general prevention-oriented coronavirus nucleic acid screening. Pooled testing can be promoted when coronavirus activity in the sample and nucleic acid testing sensitivity are not affected by dilution. According to research conducted by the Institute of Virology of Saarland University in Germany, the method of “pooled samples” can merge up to 30 samples, which is sufficient to ensure testing accuracy, although testing sensitivity is slightly reduced. Since the PCR testing initially needs to be amplified, the dilution has almost no effect on the testing results. However, the actual pooled sample ratio should be controlled at 5–10 people, and the maximum should not exceed 20 people to ensure that the testing sensitivity does not decrease²¹.

Wuhan's feat of testing 10 million people in ten days benefited from pooled testing. In addition, it was verified that pooled testing does not affect the virus activity or the nucleic acid testing sensitivity. The testing results were as follows: 9,998,828 people in Wuhan were subjected to nucleic acid testing, of which 9,865,404 had no history of confirmed COVID-19 and 34,424 of previous coronavirus pneumonia patients had recovered. Among subjects without a history of COVID-19, no confirmed cases were found, while 300 asymptomatic infections emerged. The testing rate of asymptomatic infections was 0.303 per 10,000. It is worth noting that large-scale pooled testing is implemented primarily for prevention, meaning that results can only be achieved during the early stages of the epidemic. For high-risk sets, such as symptomatic patients and close contacts in fever clinics, individual collection and individual testing should be implemented. On the opposite, pooled testing is preferred for screening virus carriers in low-risk and high-density populations.

“Ten-in-one” pooled testing for COVID-19 in Qingdao

On October 11, 2020, after three asymptomatic cases of COVID-19 were found in Qingdao, Shandong Province, the Qingdao government immediately organised large-scale investigations and classified testing to prevent the spread of the epidemic. At 23:00 on October 11, six confirmed cases and six asymptomatic infection cases were identified in Qingdao. This was caused by the sharing of CT (computed tomography) scan rooms with patients in general wards during infected hospitalisation. For that reason, the Qingdao government formulated and launched a full-staff nucleic acid testing programme. It was necessary, but formidable, to complete an assessment of six million people in the primary urban area within three days and to cover the entire city within five days. Therefore, based on Wuhan's “five-in-one” pooled testing method, the Qingdao government adopted the improved “ten-in-one” pooled testing for nucleic acid testing. The Chinese government has currently issued the new coronavirus nucleic acid “ten-in-one” standardized technical specifications for pooled testing. Compared to Wuhan's “five-in-one” pooled testing technology, the “ten-in-one” pooled testing method significantly expands testing capacity, which increases its efficiency. In addition,

the “ten-in-one” pooled testing method can detect infections in advance and isolate them, which is conducive in reducing the virus spread. The principle of the “ten-in-one” pooled testing method, which increases pooled samples, is similar to that of the “five-in-one” pooled testing method. The “ten-in-one” pooled testing method of Qingdao strictly follows the requirements of “*Technical Specifications for the Detection of COVID-19 with Ten in One Pooled Testing*” promulgated by China for nucleic acid sampling and testing. As shown in Fig. 3, pooled samples collected from ten people are mixed before testing. If the pooled sample is negative, all ten samples are negative, and the ten people in the pooled testing are safe. However, if the result of the pooled sample is positive, the ten people are notified immediately, isolated and individually-checked to identify those who are positive. Compared to the “five-in-one” pooled testing method in Wuhan, the “ten-in-one” pooled testing method increases the number of pooled samples and the efficiency of the process. At 18:00 on October 16, 10,899,145 nucleic acid testing were completed in Qingdao, from which nearly 10 million people had been tested within five days. The testing speed was twice as fast as that in Wuhan. All nucleic acid testing results of the citizens in Qingdao were negative, which virtually eliminated the risk of community transmission of the epidemic.

Optimisation of the “ten-in-one” testing method: the pentagram mini-pooled testing method

As noticed above, the “five-in-one” pooled testing method used in Wuhan and the “ten-in-one” pooled testing method utilised in Qingdao provide new strategies for large-scale nucleic acid testing of the COVID-19 epidemic for other countries. These two methods, using different numbers of pooled samples, have significantly increased the speed of nucleic acid testing in the first round of testing. However, both methods still use “individual-sample testing” in the second round of testing. Especially in Qingdao's “ten-in-one” pooled testing method, when a sample testing becomes positive in the first round of nucleic acid testing, all 10 individual specimens require individual-sample testing in the second round. This individual-sample testing method not only increases the cost of testing by significantly increasing the number of testing kits and the workload on the health sector, but also reduces the efficiency of the nucleic acid testing and controlling the spread of the epidemic. Accordingly, some scholars have explored the second-round testing method of pooled testing to improve the testing efficiency. As early as the 1940s, Dorfman²², an economist at Harvard University, proposed pooled testing for screening the syphilis carriers among soldiers in World War II. Dorfman suggested that after the individual blood sera were drawn, they were pooled in groups of N and that the groups rather than the individual sera were subjected to chemical analysis. For positive samples in the first round of testing, the individuals constituting the pool must be retested to determine which of the members are infected. Subsequently, some scholars have improved the pooled testing method. Many people in South Africa are infected by human immunodeficiency virus (HIV) every year, and the prevalence is almost 17% among adults 15–49 years of age. Therefore, many scholars in South Africa have investigated HIV testing methods. van Zyl et al.²³ adopted matrix strategies to reduce the costs of virologic monitoring. Specifically, using nine specimens as an example, the specimens are labelled to form a 3×3 matrix. Each row and column of the matrix is detected by “three-in-one” pooled testing. If the testing results of two pooled samples are positive, bearing in minds that the rows and columns intersect in the matrix platform, then the intersection identifies the positive sample. This method only requires that six testing be performed on nine specimens to determine an individual or more confirmed patients. Recently, scientists in

Rwanda proposed a hypercube algorithm testing method suitable for areas with low infection rates to suppress infections of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)²⁴. This method quickly identifies and isolates individuals infected with the virus. Taking 27 ($N = 3^3 = 27$) testing samples as an example, the hypercube ($3 \times 3 \times 3$) of 27 samples are sliced into 3 slices in each of the 3 principal directions. Each of the slices contains nine samples, and “nine-in-one” pooled testing is performed on each slice to identify the coordinates of the positive sample. Thus, in this example, only nine testing can be used to uniquely identify an individual infected person among 27 people. For the optimised testing methods proposed by the above scholars, the number of pooled samples in the first round is n^2 or n^3 (n is a natural number), usually 3^2 or 3^3 . These values can be divided equally in multiple directions, and in each direction, they can be divided in a multidimensional matrix platform. Then, positive patients can be identified through the intersection of different positive pooled samples. However, in Qingdao’s “ten-in-one” pooled testing, the number of pooled samples cannot be sorted to form $n \times n$ matrix pools. To reduce the number of testing for the second round of “ten-in-one” pooled testing without increasing the time and cost, this paper innovatively proposes the pentagram mini-pooled testing to identify an individual positive patient, as shown in Fig. 4.

Since pooled testing is suitable for areas with low infection rates, a high number of individuals who are positive in the random “ten-in-one” testing is a small probability event from a probabilistic point of view. That is, if a “ten-in-one” pooled testing result is positive, the greatest probability is that only an individual person is infected. Therefore, the pentagram mini-pooled testing (see Fig. 4) proposed in this paper focuses on the circumstance in which only one specimen is infected among the ten specimens of the positive pooled sample. As mentioned above, when a “ten-in-one” pooled testing is positive, the ten people should be isolated and retested in a second round. Pentagram mini-pooled testing works as follows in the second round. Before testing, the people are sorted from one to ten, and two throat swab samples are collected from each person. During testing, the twenty swabs are pooled into six samples according to the pentagram mini-pooled testing method (five “three-in-one” pooled samples (S1-S5) and one “five-in-one” pooled sample (S6), as shown in Fig. 5)). Every three specimens are placed in the same collection tube according to the three serial numbers corresponding to the vertices of the five orange triangles of the pentagram. For example, one swab sample is taken from subjects 1, 6, and 7, and these three swab samples are placed into S1 collection tube for the “three-in-one” pooled testing. By analogy, S2-S5 collection tubes are obtained. The “five-in-one” pooled sample tube (i.e., S6 collection tube) refers to the collected swab samples from subjects 1 to 5. Nucleic acid testing is then performed for the six pooled samples (S1-S6). After the testing are returned, the positive patients are identified by comparing the prediction testing results, as shown in Fig. 5.

Theoretical Basis And Applicable Conditions For Pooled Testing

There are certain applicable conditions for large-scale screening of COVID-19 using the pooled testing method. The applicability of this method and the selection of a reasonable number of pooled samples are importantly related to the infection rate of the region^{15,25}. To verify the efficacy and applicable conditions for pooled testing, the quantitative relationship between the value of pooled samples and the virus infection rate

is determined. It is assumed that N is the total population of a certain city; p is the infection rate; x is the number of people in each set of pool testing samples; and Y is the total testing time for the first round of pool testing (Y_1) and the second round of individual-sample testing (Y_2). p is an unknown and dynamically changing value, but it can be estimated based on the number of confirmed cases.

If $x = 1$, each person is tested once. Then, the value of Y is equal to N .

When $x \neq 1$, the testing should be performed in two rounds.

The first round is pooled testing considering that each x people represent a set, so the testing time required for the first round is $Y_1 = N/x$. The first testing result is positive or negative (only if everyone in a given set is uninfected can a negative result be obtained). Therefore, $(1-p)^x$ is the probability of getting a negative result for the set testing of the first round, and $1-(1-p)^x$ is the probability that the result of the set testing is positive for the same round. Hence, the number of sets with positive testing results from the first round can be defined as follows:

$$(N/x) * (1 - (1-p)^x), \quad (2)$$

The second-round testing is for the sets that showed positive testing results in the first round. The total number of people in these sets is as follows:

$$(N/x) * (1 - (1-p)^x) * x = N * (1 - (1-p)^x), \quad (3)$$

Therefore, the time for testing the second round can be determined as:

$$Y_2 = N * (1 - (1-p)^x), \quad (4)$$

After the two-round testing is completed, the total testing time Y can be computed as:

$$Y = Y_1 + Y_2 = N/x + N * (1 - (1-p)^x), \quad (5)$$

Results

Table 1 shows the pentagram mini-pooled testing prediction results. When there is only one positive testing result in S1-S6, it shows an error in the nucleic acid testing collection process (the probability should be zero), requiring swab collection and nucleic acid testing of the 10 subjects again. When two pooled samples of S1 to S6 are positive and the other four are negative, only one confirmed patient can be identified. When more than two positive pooled samples are found in S1-S6, there are at least two positive patients in this group. If more than two samples testing are positive, only the subjects in these positive mini-pools need to be subjected to the third round of individual-sample testing. The relationship between the number of positive testing samples and the number of confirmed patients is expressed by formula (1). In summary, the pentagram mini-pooled testing method can accurately determine the only positive patient among 10 people by testing six mixed samples. Compared with the individual-sample testing utilised in the second round of

Qingdao's "ten-in-one" pooled testing method, the proposed method significantly improves the efficiency of the second round of nucleic acid testing, and the accuracy of the testing results is guaranteed.

Table 1
Pentagram mini-pooled testing results prediction table

Confirmed sample	Sample testing results					
	S1	S2	S3	S4	S5	S6
1	+	-	-	-	-	+
2	-	+	-	-	-	+
3	-	-	+	-	-	+
4	-	-	-	+	-	+
5	-	-	-	-	+	+
6	+	-	-	-	+	-
7	+	+	-	-	-	-
8	-	+	+	-	-	-
9	-	-	+	+	-	-
10	-	-	-	+	+	-
(-: the sample is negative; +: the sample is positive)						

1

This method is suitable for rapidly screening large-scale low-risk populations and local epidemic rebound areas. It ensures that infected people can be quickly identified in the second round of pooled testing, hence it quickly interrupts the possible spread of the epidemic. However, for detecting high-risk populations, such as symptomatic patients in fever clinics and those who were in close contacts with confirmed patients, individual testing should be adopted.

This work focused on large cities, which require 10 million testing to quantitatively analyse the relationship among the total population of city N , the infection rate p and the number of samples x . Specifically, the infection rate p is set as 0.0001, 0.001, 0.01, and 0.1. Each of the values is round then studied with parameterisation. In the following, Figs. 6–9 show the testing time relationships (considering the first-round, second-round and total times) of different infection rates.

In Figs. 6–9, the testing time Y_1 in the first round is an inverse proportional function, meaning that as the number of samples x increases, the testing time Y_1 in the first round gradually decreases. It is worth noting that the testing time reduction does not correlate with the infection rate p . On the opposite, the testing time Y_2 in the second round is an exponential function, and it is observed that it is highly correlated with the infection rate p . When the infection rate p is very small ($p = 0.0001$), although it is not clear, the testing time Y_2 in the second round increases with increasing x sample numbers. As the infection rate p gradually increases, the testing time Y_2 in the second round increases sharply. Consequently, if pooled testing is adopted, when the number of infected people in a certain area increases under the constant number of samples, the time required for the second round of testing will increase with the increasing number of infected people, and it will show an obvious dramatic increase. Accordingly, this increases the number of nucleic acid testing and results in more resources and time being consumed. If the virus cannot be controlled in a short time, it will form a vicious circle, whereby an increase in the number of infected people will lead to a sharp increase in the testing time. Such increased testing time puts tremendous pressure on the testing personnel and medical system. Therefore, the best opportunity to screen infected people and control the epidemic is to perform nucleic acid testing when the infection rate is still low.

From the above, it can be concluded that as the infection rate p increases, the testing time Y_2 in the second round increases dramatically, which significantly affects the total testing time Y . The value Y is plotted when the infection rate p is 0.0001, 0.001, 0.01, 0.1, and 0.2 to analyse the relationship between the total testing time Y and the number of samples x under different infection rates p (see Fig. 10). When the infection rate p is low ($p \leq 0.001$), as the number of samples x increases, the total testing time Y drops sharply. However, as the number of samples x continues to increase, the total testing time Y stabilises. Figure 10 shows that further increase in the number of samples x when $p \leq 0.001$ barely reduces the testing time. When the infection rate p increases ($p \geq 0.01$), it can be noticed that as the number of samples x increases in the initial stage, the total testing time Y decreases significantly. However, as the number of samples x continues to increase, the total testing time Y increases until it approaches or exceeds the total testing time if the traditional individual-sample testing is used. Consequently, pooled testing is not efficient in these situations.

As can be noticed, the total testing time Y , as shown in Fig. 10, initially decreases rapidly. This is because when the number of sets is small, the first round of set testing time N/x dramatically reduces the total testing time. Since the number of samples x is small, the testing time for positive sets in the next round remains relatively small. Therefore, the second-round testing time Y_2 in the total testing time Y is relatively small. On the contrary, when the number of samples x is large, the value N/x significantly reduces the testing time. However, the testing time of positive sets in the second round also increases significantly as it becomes exponentially, which adds to the total testing time. Therefore, there is a reasonable value range for the number of samples x under different infection rates p . From Fig. 5, it could be observed that as the infection rate p increases, the most reasonable number of samples x decreases.

According to Eq. 5, this study calculates the times of testing needed in pooled testing for a city with 10 million people under different infection rates, in addition to the most reasonable x , the testing times of each round and the reduction rate relative to traditional individual-sample testing (see Table 2).

Table 2
Testing times under different infection rates

p	x	Y	Y_1	Y_2	Y_2/Y_1	Reduction rate compared to individual-sample testing
0.0001	100	199,507	100,000	99,507	99.51%	98.00%
	101	199,507	99,010	1,004,961	101.50%	98.00%
0.001	32	627,589	312,500	315,089	100.83%	93.72%
0.01	11	1,955,708	909,091	1,046,617	115.13%	80.44%
0.1	4	5,939,000	2,500,000	3,439,000	137.56%	40.61%
0.2	3	8,213,333	3,333,333	4,880,000	146.40%	17.87%
0.3	3	9,903,333	3,333,333	6,570,000	197.10%	0.97%
0.35	3	10,587,083	3,333,333	7,253,750	217.61%	-5.87%

Currently, Fig. 11 provides the optimum number of samples relative to the infection rate. Additionally, the minimum number of testing against the infection rate is illustrated on Fig. 12. The ratio between the second-round testing time to that of the first-round testing is provided in Fig. 13, while Fig. 14 shows the relationship between the reduction rate relative to the individual-sample testing corresponding to the infection rate. From Fig. 11, it is obvious that when the infection rate p is minimal, the value of x (assumed to be the optimal value) corresponding to the minimum value of Y is increasing. As the infection rate p gradually increases, the number of samples x gradually decreases. When the infection rate p reaches 0.01, the “eleven-in-one” pooled testing method becomes the most useful testing method. When the infection rate $p = 0.1$, only the “four-in-one” pooled testing method can be used to reduce testing times. As the infection rate p increases, the corresponding minimum testing times also increases (see Fig. 12). As the infection rate p increases, the value of Y_2 increases sharply. The proportion of testing times in the second round also increases distinctly compared to that of the first round (see Fig. 13). In addition, from Fig. 14 it can be noticed that as the infection rate p increases, the advantages of pooled testing compared with those of individual-sample testing gradually decrease. When the infection rate p is low (e.g., when the infection rate is 0.001), the pooled testing method should be used, which reduces the total testing time by more than 90% compared with individual-sample testing. When the infection rate p is 0.01, the “eleven-in-one” pooled testing method is adopted, which reduces the total testing time Y by 80.44%. When the infection rate p reaches 0.3, the pooled testing times are close to that of individual-sample testing, and the reduction rate compared to individual-sample testing is minimal. As the infection rate p further increases, the total time required for pooled testing surpasses that of individual-sample testing, which increases the sampling and testing costs. Therefore, pooled testing has significant advantages over individual-sample testing when the infection rate p is low; however, with increasing infection rate, testing efficiency gradually decreases, and its advantages further decrease. This also theoretically explains why the “ten-in-one” pooled testing method was used in Qingdao, while the “five-in-one” pooled testing method was used in Wuhan. Specifically, the infection rate p in Qingdao was low, and the use of “ten-in-one” pooled testing significantly reduced the testing times. Therefore, it took only five days to conduct nucleic acid testing on more than 10 million people. On the

opposite, due to the higher infection rate in Wuhan, the “five-in-one” pooled testing method was adopted with fewer combinations. From the above discussion, it could be concluded that when the infection rate p is low, pooled testing has undeniable advantages. However, as the infection rate p increases, its testing advantages gradually decrease. When the infection rate p reaches a certain value ($p > 0.3$), pooled testing becomes no longer applicable. However, COVID-19 continues to spread around the world. Before the vaccine is developed and widely used, nucleic acid testing and screening for infected persons is still one of the most effective control measures. Therefore, in regions with few infected people, earlier pooled testing should be adopted to reduce the numbers and time required for nucleic acid testing, which is beneficial for controlling COVID-19. On the opposite, higher infection rate increases testing times, and the dramatic increasing trend increases the difficulty and cost of testing.

Discussion

To date, COVID-19 is still spreading worldwide, and nucleic acid testing is considered as one of the most effective ways to screen virus carriers. Based on the successful experience of China's fight against COVID-19, this study introduced a promising mass-scale coronavirus testing strategy for infection control and surveillance; the “five-in-one” pooled testing method in Wuhan and the “ten-in-one” pooled testing method in Qingdao. Pooled testing speeds up the testing process and expands screening capacity because it is far less time-consuming than obtaining a sample from an individual and then putting it through all the necessary lab processes.

Conclusion

The practical application showed that the pooled testing method is suitable for mass nucleic acid testing in regions with low infection rates. In addition, this study optimised the individual-sample testing program used in the second round of the “ten-in-one” pooled testing method to slash costs and preserve supplies and testing chemicals. Therefore, a more efficient “ten-in-one” second-round testing, named the pentagram mini-pooled testing method, was proposed. Specifically, before the second round of “ten-in-one” pooled testing, ten persons are ranked 1–10, and two throat swab samples are collected from each person. The twenty throat samples are divided into six sets combining different samples (as discussed in Sect. 3), and a pooled testing is performed on each set. In this way, only six times testing in the second round are performed to identify the positive infection. Compared with the current individual-sample testing in the second round of Qingdao's “ten-in-one” method, the optimised method increases testing efficiency by 40%.

This study rigorously and intensively discussed the theoretical basis and applicable conditions for pooled testing of COVID-19 based on probability theory and statistics. The results showed that when the infection rate is low ($p \leq 0.001$), pooling of samples can drastically amplify testing capacity and quickly screen out infected people, and the reasonable number of samples that can be combined in a given pool can be in the range of 10 to 20. In addition, it was found that the times of testing needed in the second round of pooled testing is much lower than the times needed in the first round when the infection rate is lower than 0.001, and the total times of testing using the pooled testing method is consequently far fewer than the number of testing using the traditional individual-sample testing method. However, as the infection rate increases, the

advantage of using pooled testing gradually decreases. The times of testing needed in the second round increases exponentially as the infection rate increases, which could lead to a dramatic increase in the times of total testing. When the infection rate exceeds a certain critical value ($p \geq 0.3$), the times of testing needed in pooled testing are close to the times using individual-sample testing. As the positivity increases, more time and materials are used to analyse and identify the positive samples. Hence, pooled testing becomes meaningless. Generally, pooled testing is an effective approach for mass nucleic acid detection in regions with low rates of COVID-19 (usually $p \leq 0.01$). In this range, it can boost testing capacity and dramatically reduce the times of testing required in specific populations compared with traditional individual-sample testing.

Declarations

Acknowledgements

Thank you to all the health care workers who are fighting COVID-19

Authors' contributions

XLS conceived the study, analysed the data and wrote the manuscript. MZ drafted the manuscript, then circulated among all authors for comments and revision. All authors read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (Grant 51808559). The authors also express their gratitude for the funding provided by the Natural Science Foundation of Hunan Province (Grant 2019JJ50770).

Competing interest declaration

We declare there is no interest statement.

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Ethical approval for this study was obtained from the Wuhan University in 2020.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹School of Health Science, Wuhan University, Wuhan, China

²School of Mathematics and Statics, Wuhan University, Wuhan, China.

References

1. Oberfeld B, et al. Snapshot: COVID-19. *Cell*. 2020;181:954–4.
2. Patrick K, et al. *What is the Kent Covid strain and why is it more contagious?* <https://www.thesun.co.uk/news/13532260/where-new-strain-covid-why-more-contagious/> (2020).
3. Dorp LV, et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection Genetics Evolution*. 2020;83:104351.
4. Zhang T, et al. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Current Biology*. 2020;30:1346–51.
5. Chung HY, et al. Novel Dual Multiplex Real-time RT-PCR Assays for the Rapid Detection of SARS-CoV-2, Influenza A/B, and Respiratory Syncytial Virus using the BD Max Open System. *Emerging microbes infections*. 2020;10:161–6.
6. Albastaki A, et al. First confirmed detection of SARS-COV-2 in untreated municipal and aircraft wastewater in Dubai, UAE: The use of wastewater based epidemiology as an early warning tool to monitor the prevalence of COVID-19-ScienceDirect. *Science of The Total Environment*. 2021;760:143350.
7. Luo GC, et al. High-quality RT-PCR with chemically modified RNA controls-ScienceDirect. *Talanta*. 2021;224:121850.
8. Deka S, et al. Effectiveness of Sample Pooling Strategies for SARS-CoV-2 Mass Screening by RT-PCR: A Scoping Review. *Journal of Laboratory Physicians*. 2020;12:212–8.
9. Guo L, et al. SARS-CoV-2 detection with CRISPR diagnostics. *Cell Discovery*. 2020;6:1.
10. Wang R, et al opvCRISPR. One-pot visual RT-LAMP-CRISPR platform for SARS-cov-2 detection. *Biosensors Bioelectronic*. 2021;172:112766.
11. Palaz F, et al CRISPR-based tools: Alternative methods for the diagnosis of COVID-19. *Clinical Biochemistry*. (2021).

12. Safari F, et al. CRISPR systems: Novel approaches for detection and combating COVID-19 - ScienceDirect. *Virus Research*. 2021;294:198282.
13. Alekseenko A, et al. Direct detection of SARS-CoV-2 using non-commercial RT-LAMP reagents on heat-inactivated samples. *Scientific Reports*. 2021;11:1820.
14. de Oliveira KG, et al Rapid molecular diagnostics of COVID-19 by RT-LAMP in a centrifugal polystyrene-toner based microdevice with end-point visual detection. *The Analyst*. (2021).
15. Regen F, et al. A simple approach to optimum pool size for pooled SARS-CoV-2 testing. *International Journal of Infectious Diseases*. 2020;100:324–6.
16. Lim K, L, et al. A novel strategy for community screening of SARS-CoV-2 (COVID-19): Sample pooling method. *PLOS ONE*. 2020;15:e0238417.
17. Gopalkrishnan M, et al. Pooling samples to increase SARS-CoV-2 testing. *J Indian Inst Sci*. 2020;. **100**:787–92.
18. Aragon-Caqueo D, et al. Optimization of group size in pool testing strategy for SARS-CoV-2: A simple mathematical model. *Journal of Medical Virology*. 2020;92:1988–94.
19. Majid F, et al. Optimising SARS-CoV-2 pooled testing for low-resource settings. *The Lancet Microbe*. 2020;1:101–2.
20. Ball J, et al. Pooled testing for SARS-CoV-2 could provide the solution to UK's testing strategy. *BMJ-British Medical Journal*. 2021;371:1–2.
21. Lohse S, et al. Pooling of samples for testing for SARS-CoV-2 in asymptomatic people. *Lancet. Infect Dis*. 2020;20:1231–2.
22. Dorfman R. The detection of defective members of large populations. *Ann Math Stat*. 1943;14:436–40.
23. van Zyl GU, et al. Pooling strategies to reduce the cost of HIV-1 RNA load monitoring in a resource-limited setting. *Clin Infect Dis*. 2011;52:264–70.
24. Mutesa L, et al A pooled testing strategy for identifying SARS-CoV-2 at low prevalence. *Nature*. 1–5. (2020).
25. Ben-Amotz D. Optimally pooled viral testing. *Epidemics*. 2020;33:100413.

Figures

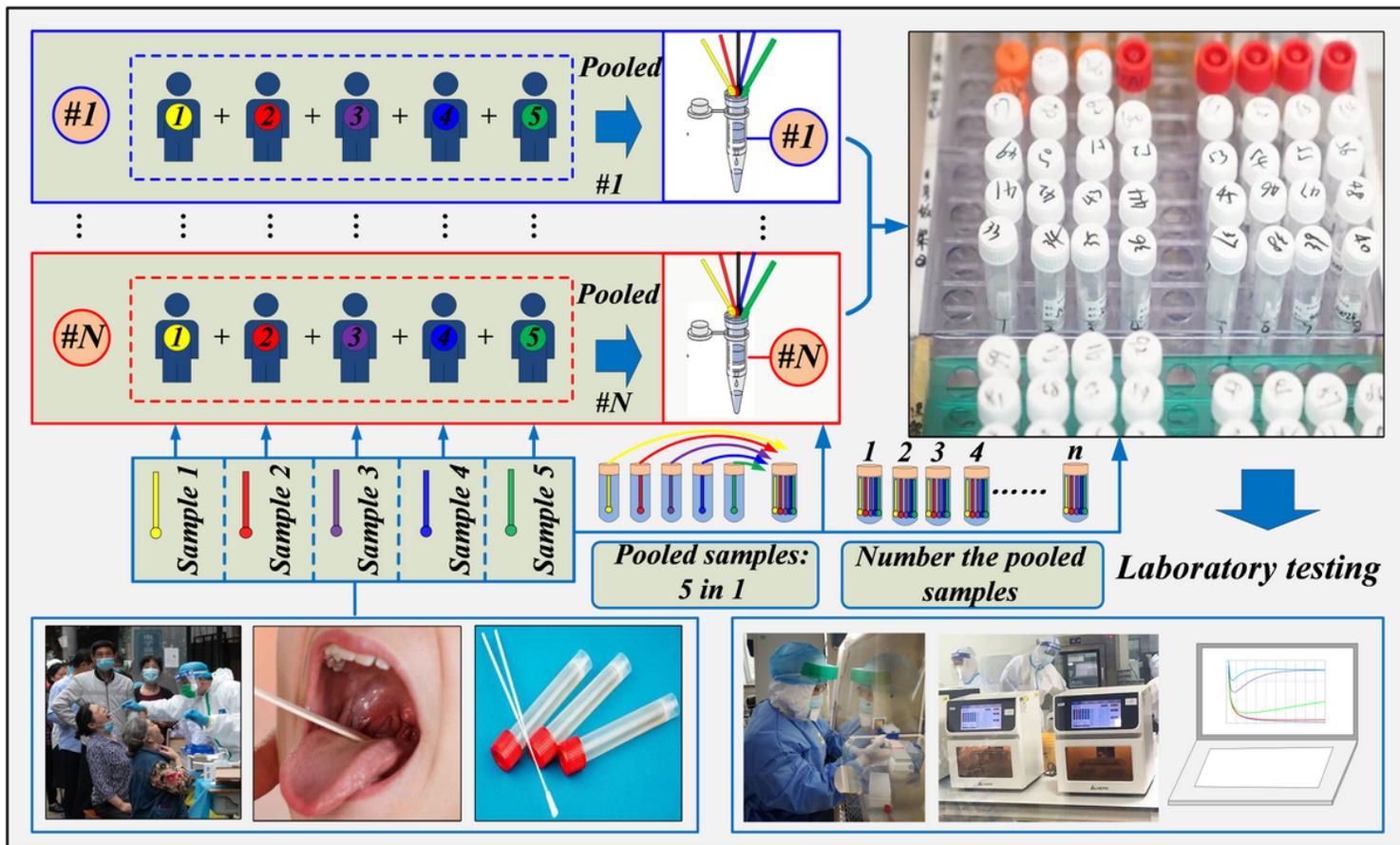
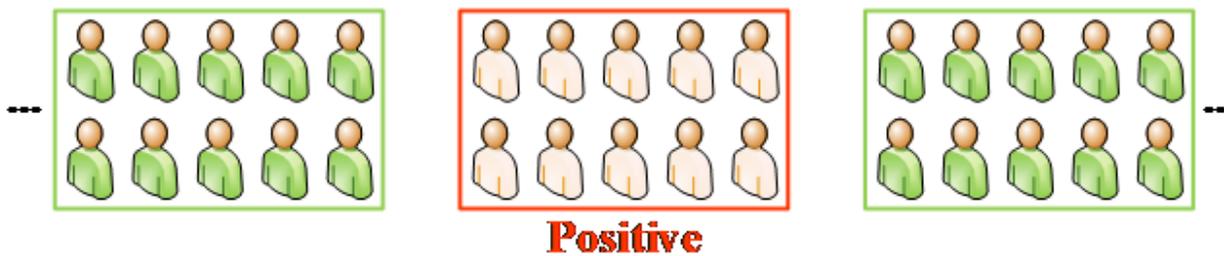


Figure 1

Primary rounds of sampling and testing

Round 1: 10-in-1 pooled sample testing



Round 2: 10 tests

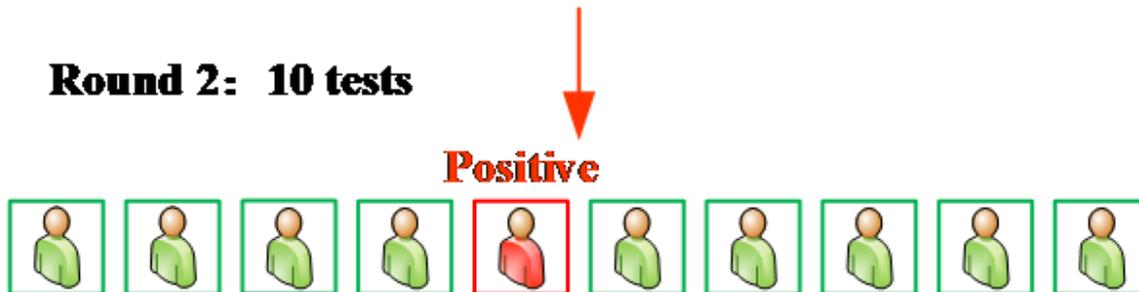
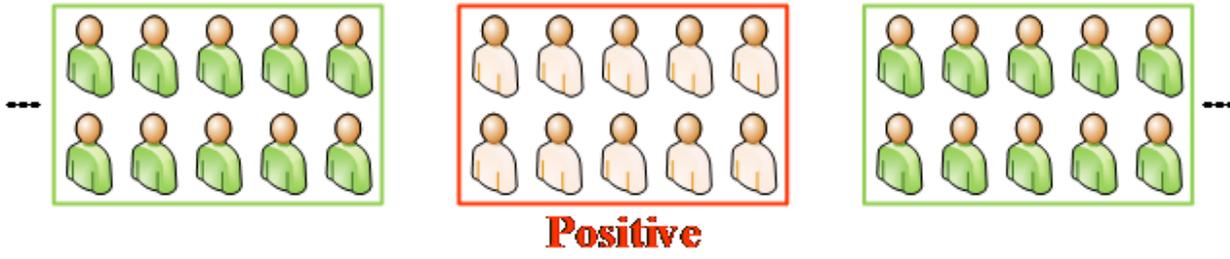


Figure 2

Principle of Wuhan's "five-in-one" testing method

Round 1: 10-in-1 pooled sample testing



Round 2: 10 tests

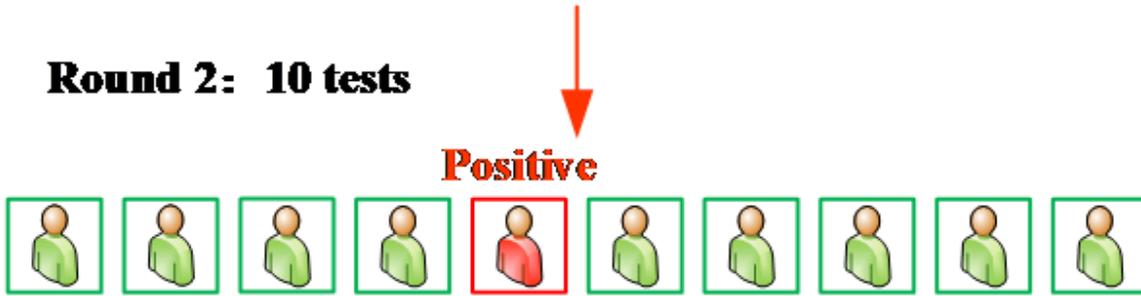
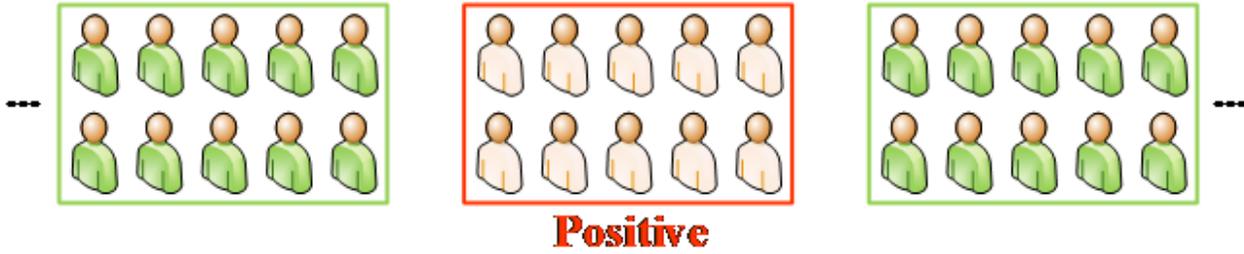


Figure 3

Qingdao's "ten-in-one" testing method

Round 1: 10-in-1 pooled sample testing



Round 2: 6 tests

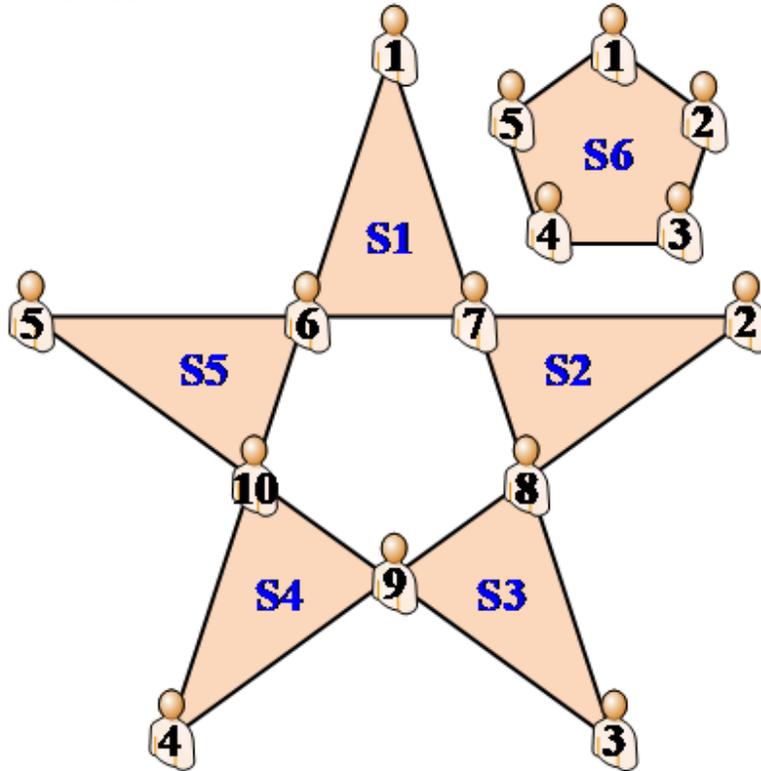


Figure 4

Pentagram mini-pooled testing method

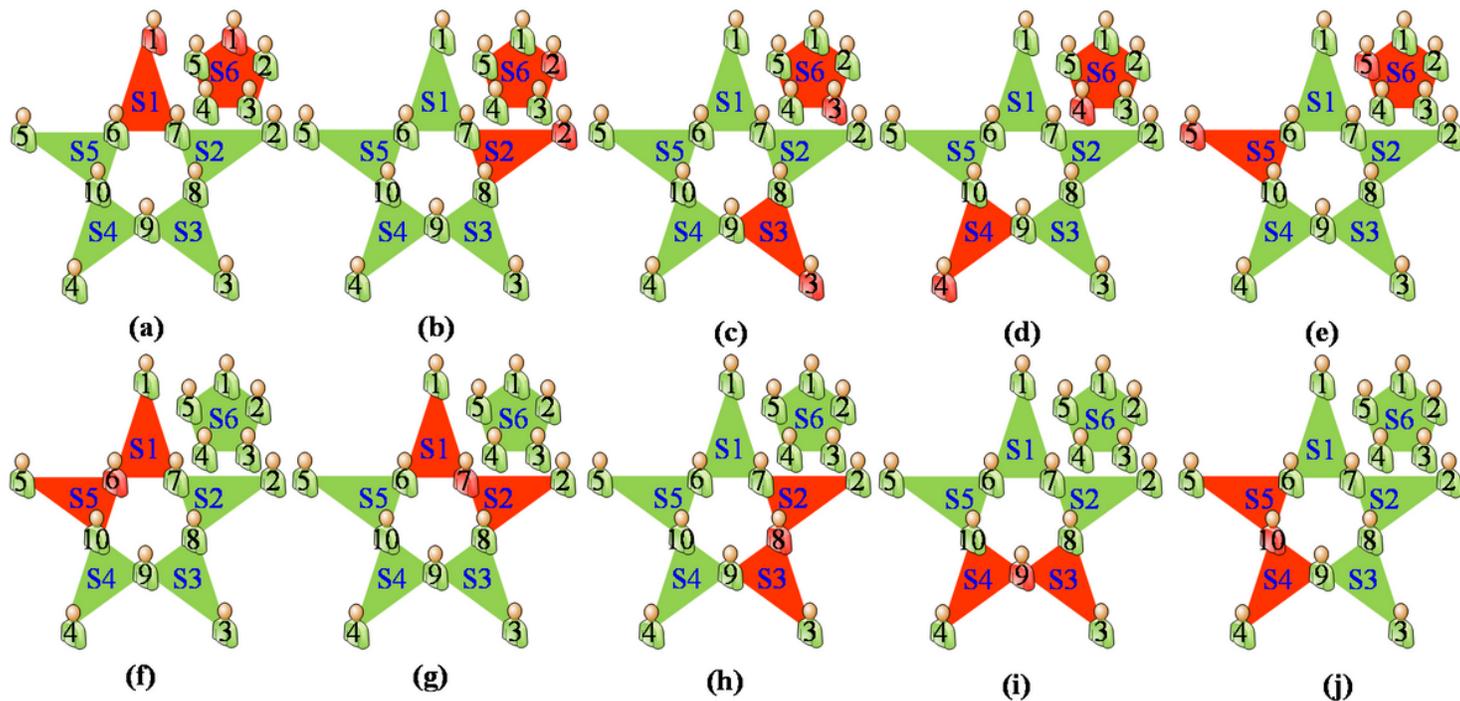


Figure 5

Pentagram mini-pooled testing results prediction chart

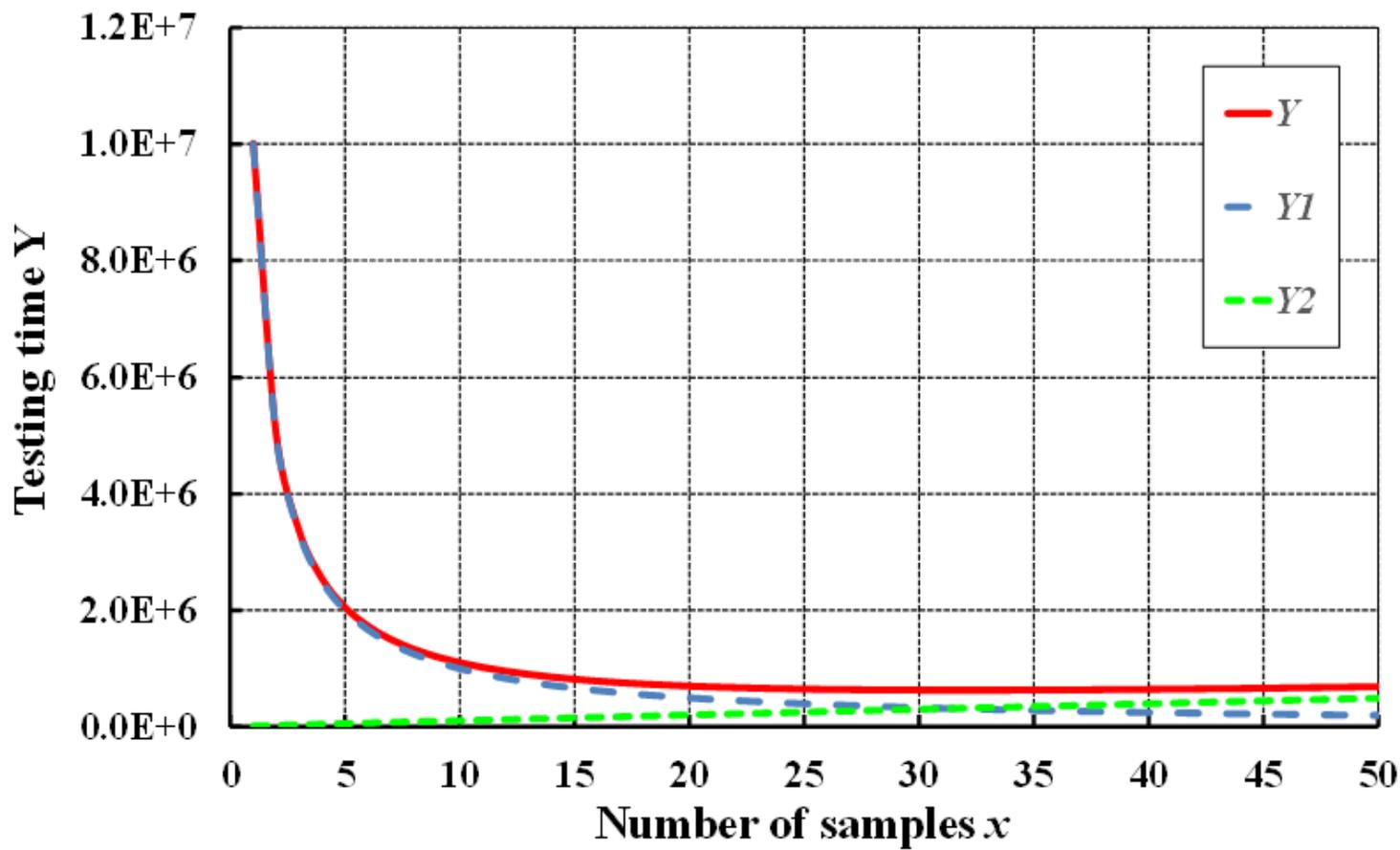


Figure 6

Testing time when $p=0.0001$

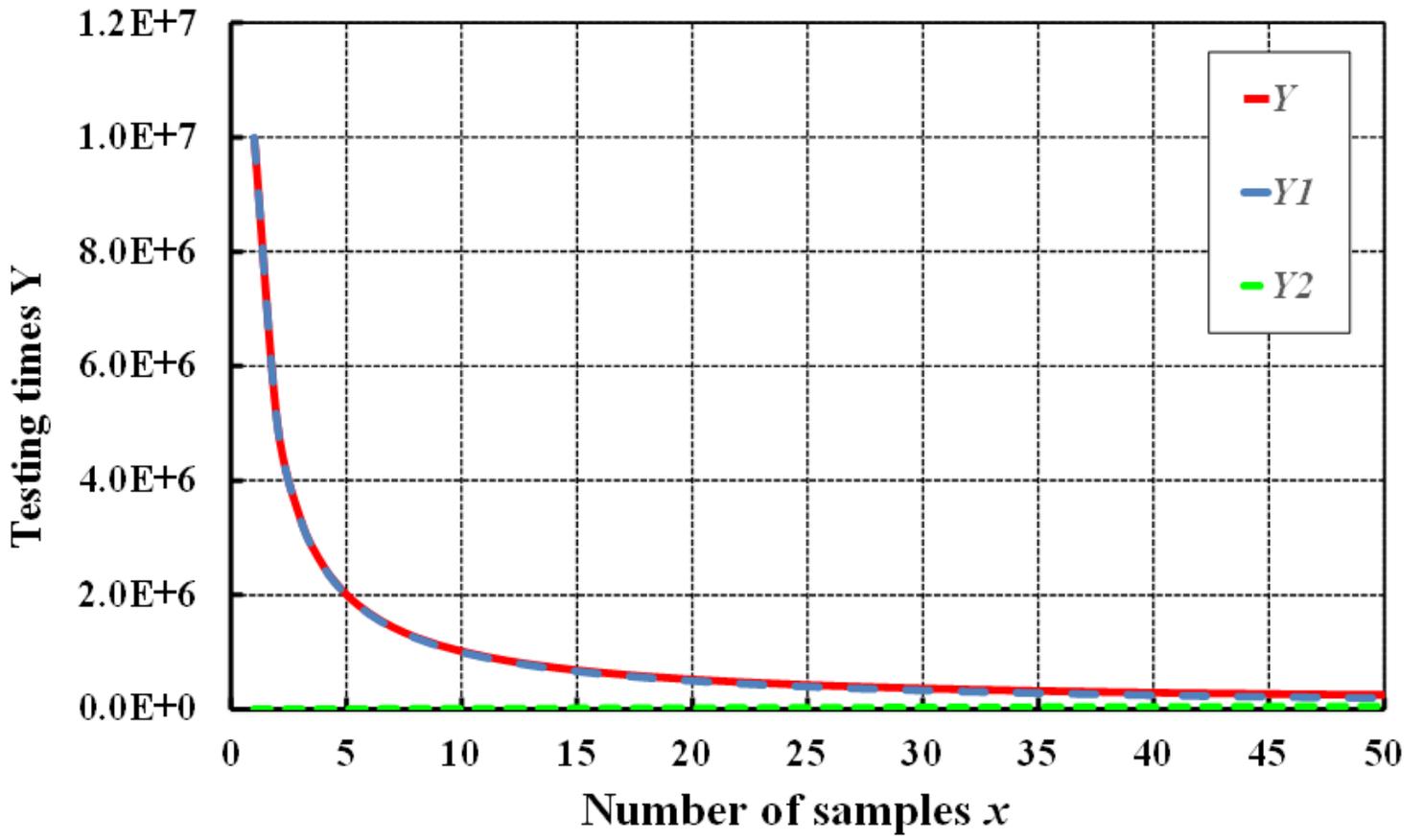


Figure 7

Testing time when $p=0.001$

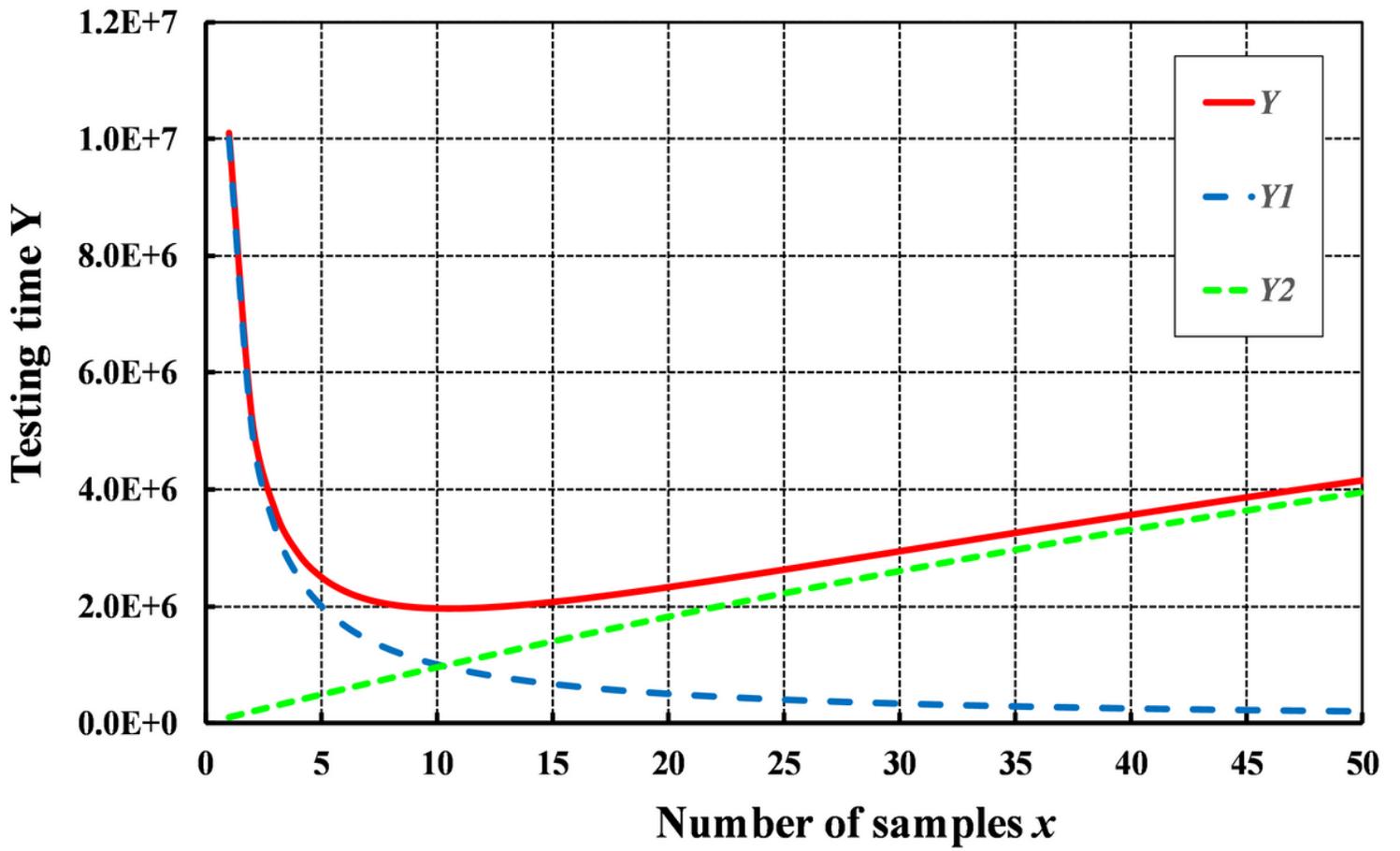


Figure 8

Testing time when p=0.01

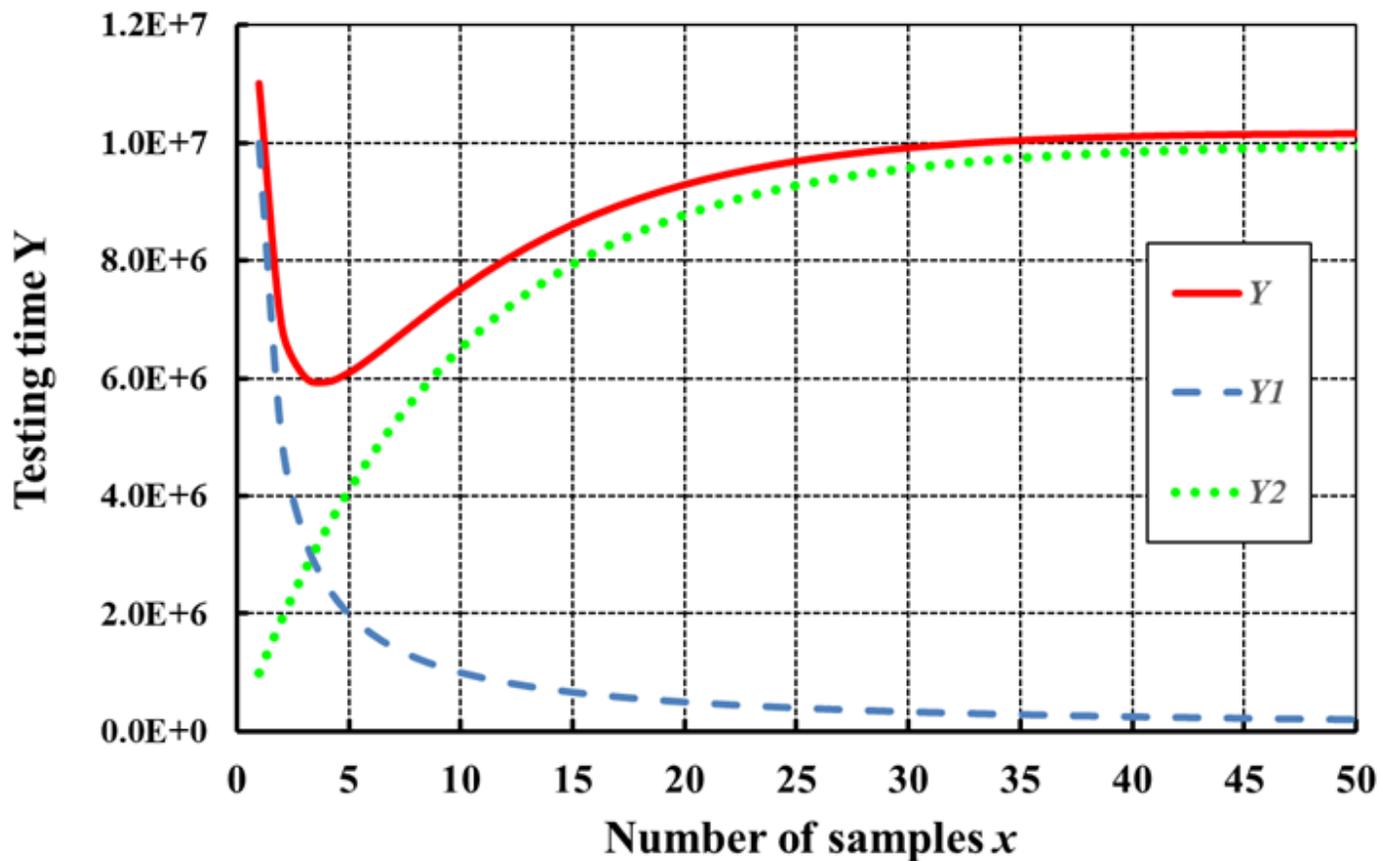


Figure 9

Testing time when p=0.1

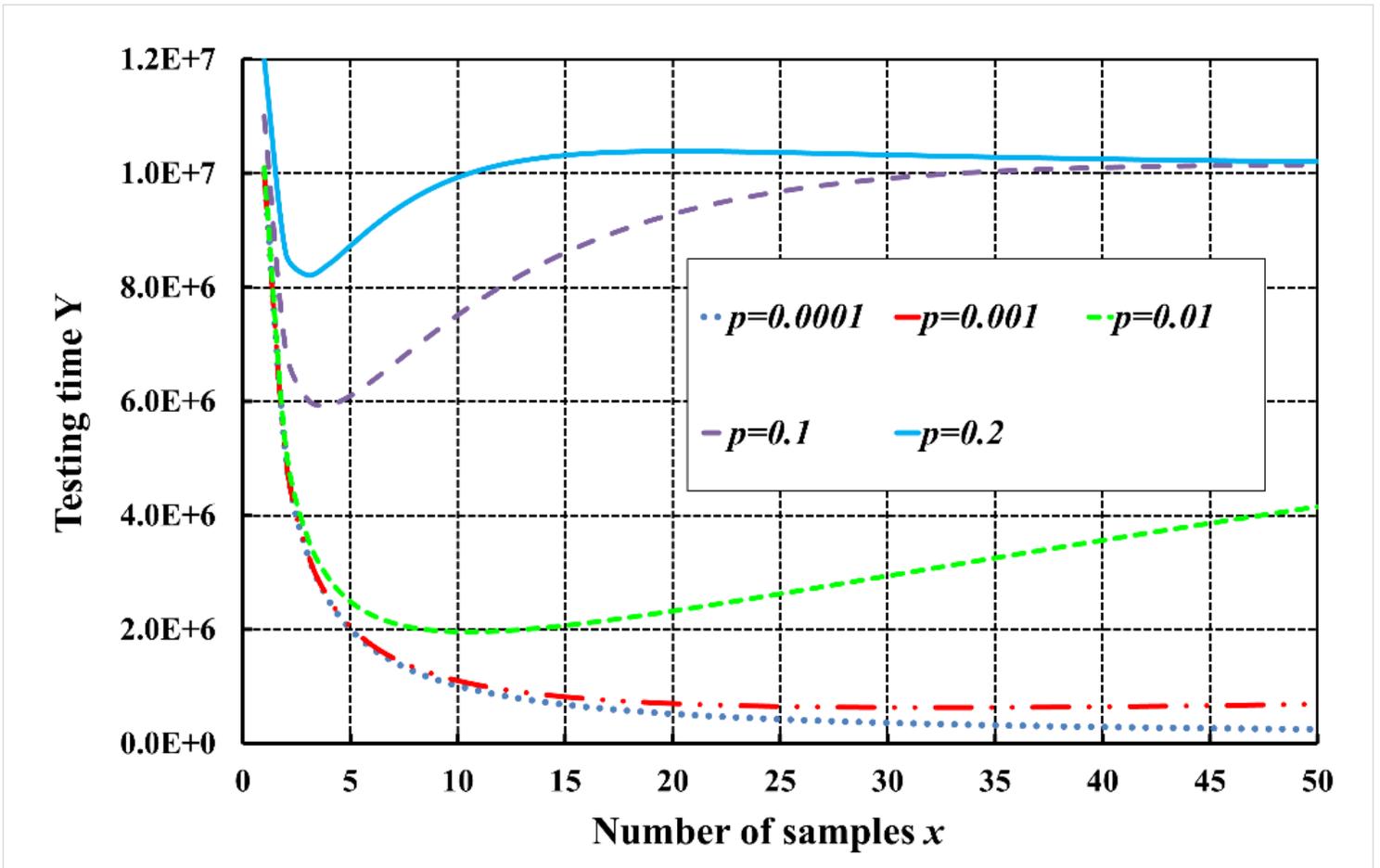


Figure 10

Testing times at different infection rates p

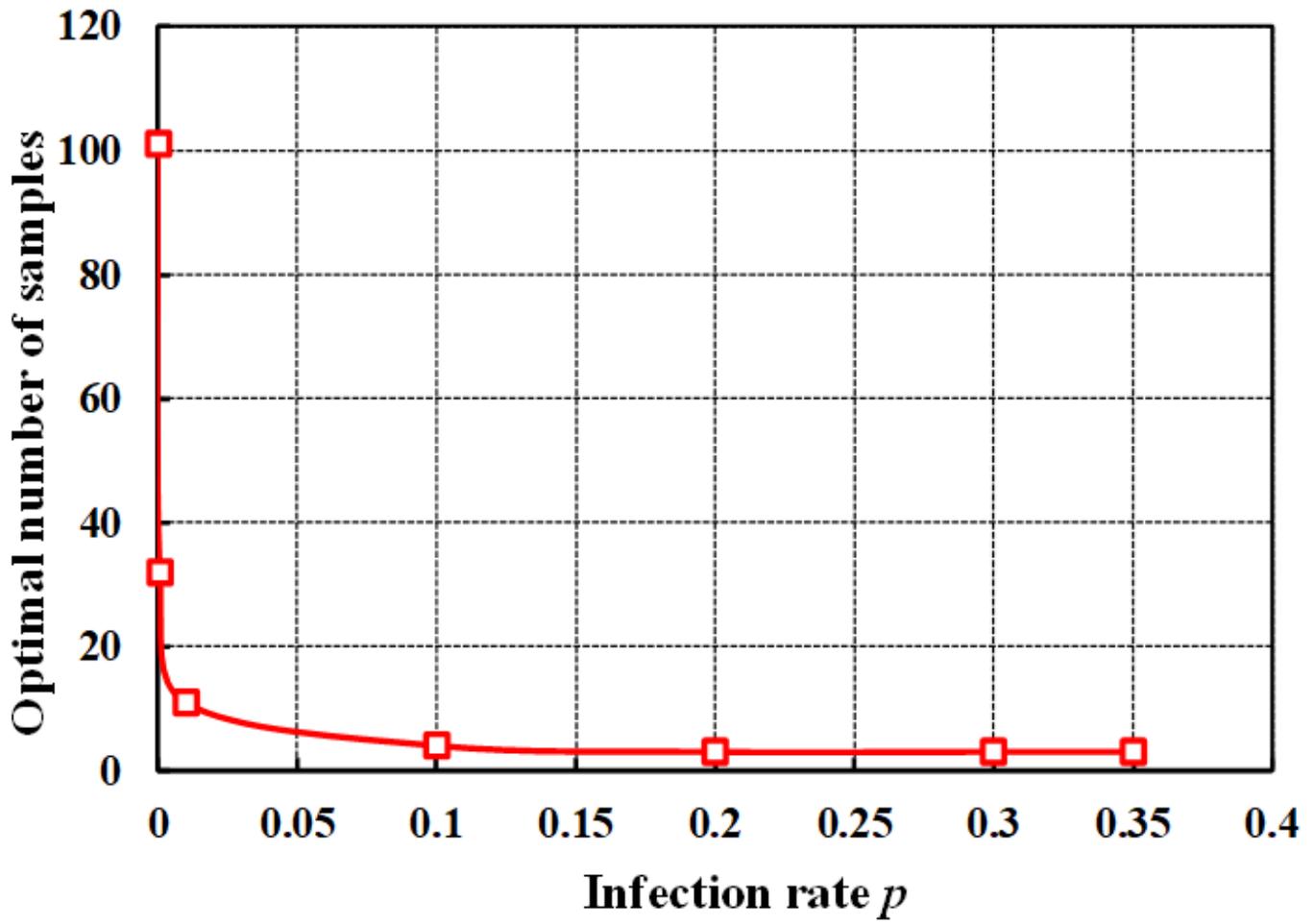


Figure 11

The most reasonable pooled-sample numbers relationship

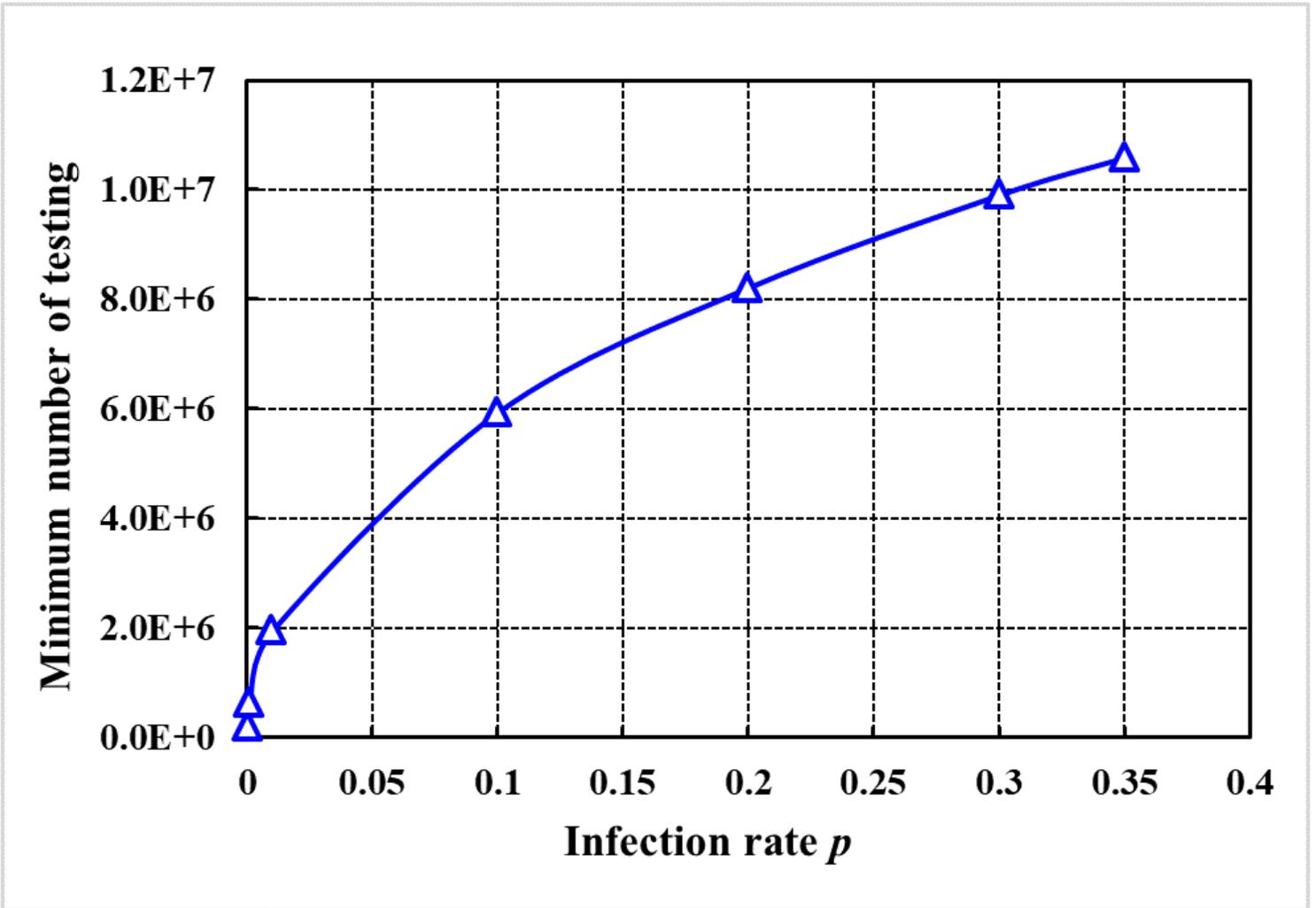


Figure 12

The minimum testing times relationship

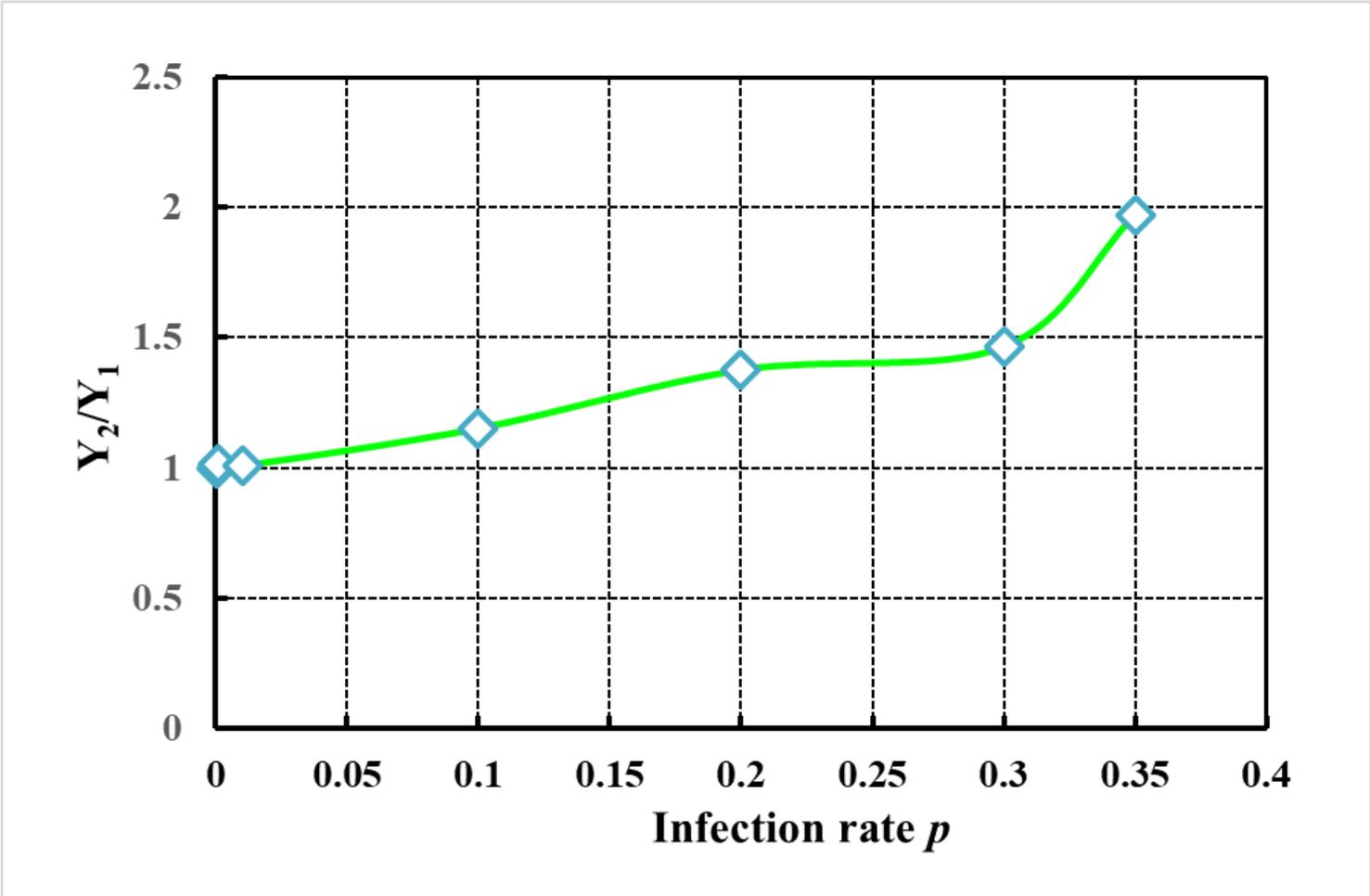


Figure 13

The relationship between the ratio of the second and first rounds of testing times

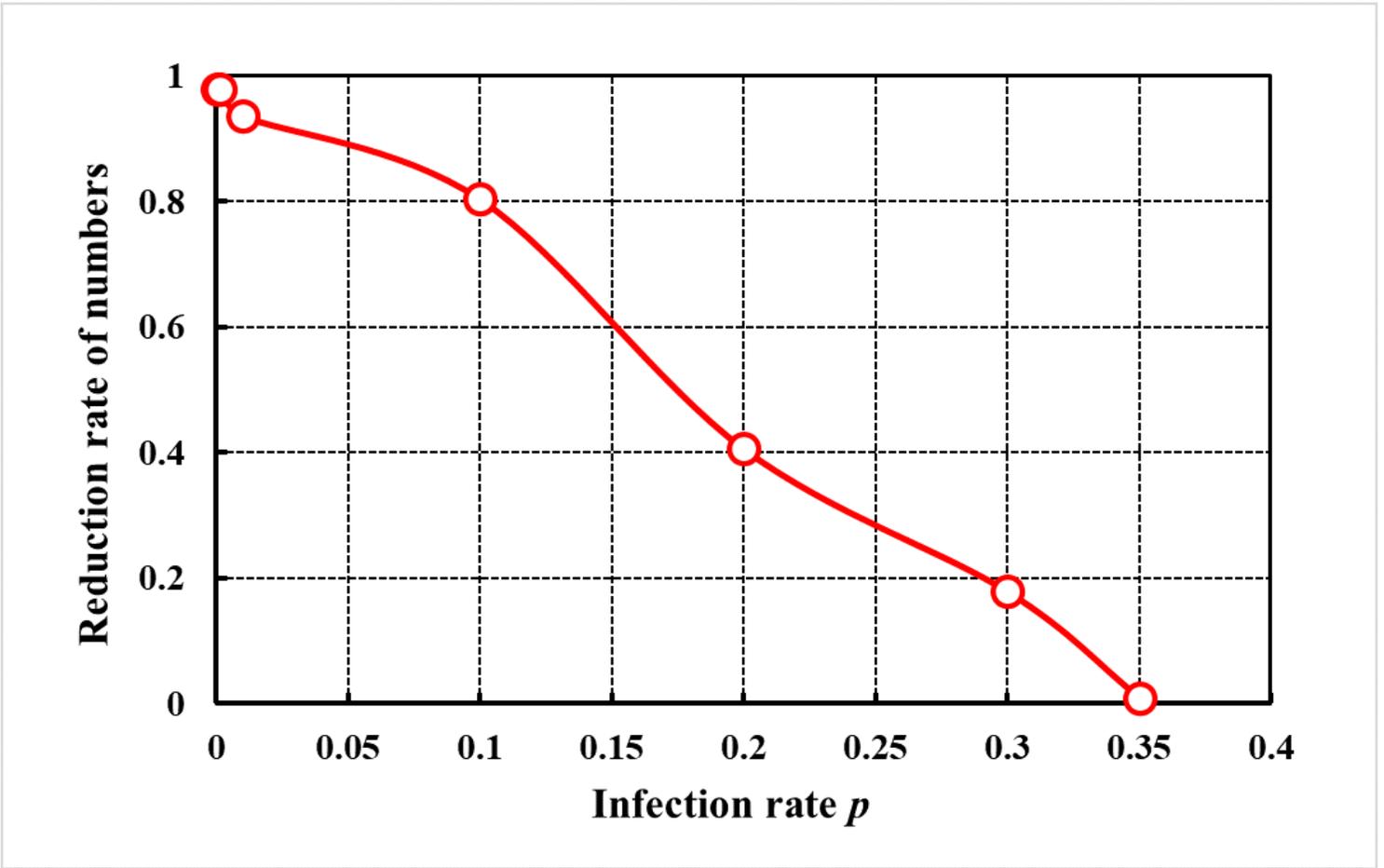


Figure 14

Relative individual-sample reduction-rate relationship