ISSN 2616-5775 Vol. 5, Issue 9: 22-27, DOI: 10.25236/AJCIS.2022.050904

Optimized Solution for Nucleic Acid Detection Based on Group Testing Model

Bo Shu[1,#], Ruochong Xiong[1,#], Hongjin Xiao[2,#]

[1]College of Science, Minzu University of China, Beijing, 100081 China
[2]College of Information Engineering, Minzu University of China, Beijing, 100081, China
[#]These authors contributed equally.

Abstract: *Under the critical situation of recurrent new crown outbreaks, large-scale full-scale nucleic acid testing has become the norm across the country. When the total number of people in a region is large and the infection rate is small, mixed testing can greatly improve testing efficiency and save costs compared with one-person testing. For the optimization of the mixed testing model, the Group Testing strategy is used to analyze the mixed testing method with the minimum number of tests, build a model and calculate the results with the help of Matlab, and finally obtain the optimal ratio of mixed testing in each region.*

Keywords: *Nucleic acid mixing test; Group Testing; Overlap test; Optimization model*

1. Introduction

The New Coronavirus outbreak was the fastest spreading, most widespread infection, and most difficult to prevent and control major public health emergency that has occurred since the founding of New China.

Nucleic acid testing is used to determine if a patient is infected with NCCV by looking for the presence of nucleic acids from foreign invasive viruses in the patient's respiratory specimen. If the test is "positive" for nucleic acid, it proves the presence of the virus in the patient's body. Nucleic acid testing allows for timely detection of confirmed, suspected, and asymptomatic cases of Neoplastic pneumonia, and rapidly cuts off the pathway of virus transmission [1].

There are two types of mixed tests: sample mixing and swab mixing. Typically, mixed tests are "5-mix" and "10-mix", where samples from 5 or 10 people are mixed. In the "10-mix" collection technique, for example, swabs from 10 individuals are mixed in a single collection tube for nucleic acid testing, and if a positive test is found in the mixed collection tube, the department is immediately notified to temporarily isolate the 10 subjects in the mixed collection tube and re-collect a single swab for review to determine if there is a case among the 10 subjects. subjects to determine if there is a case among them. A negative test result from a mixed collection tube means that all ten samples are negative and that the mix is safe for the individual. This type of testing maximizes the efficiency of testing, and the results are accurate, with no missed or false detections, and isolates patients in advance to reduce transmission and save social costs.

Assuming different incidence rates ($r_1, \cdots, r_6$ per 10,000) in six districts of Beijing city, the number of people in the region $M_1, \cdots, M_6$ per 10,000 people under different conditions, the optimal mixing method $N_1, \cdots, N_6$ is given, assuming that the testing capacity of six districts of Beijing city is $W_1, \cdots, W_6$ per10,000 tests every 5 hours, and the total time required for one round of testing and three rounds of testing in this way is discussed.

For this problem, we review the literature and find that the Group Testing strategy proposed by Dorfman in the 20th century can be a good reference for solving this problem. We modify and build a mathematical model based on the Group Testing strategy, we base on the grouping idea of Group Testing to group the samples in one dimension, and the grouped samples After the testing, the positive samples were tested individually to determine the best way to test the samples in Beijing. We then improved the samples on this basis, considering the overlapping samples, we considered grouping the samples by columns and rows, testing the samples of columns and rows separately, and then testing the samples of crossed rows and columns individually, which may optimize the overall number of tests[2]. After obtaining the optimization we then give the total time for one and three rounds of the relevant tests in the two testing orders, respectively, to build the model.

2. Assumptions and notations

2.1. Assumptions

We use the following assumptions [3].

(1) All the reagents give correct results for the assay.

(2) The specific data used are kept constant after being given.

(3) The specified test time is guaranteed to be stable and other relevant factors do not affect the test time.

(4) Nucleic acid testing is performed only in the region and identified positives are not tested again.

(5) The infection rate is the incidence rate.

2.2. Notations

The primary notations used in this paper are listed as Table 1.

Table 1: Notations

Symbols	Meaning	Unit
M_i	Number of people in the region	10,000 people
r_i *or* p_i	Morbidity (infection rate)	1
N_i	Optimal mixed test method	Person/inspection
W_i	Testing capacity every five hours	10,000 reagents/five hours
$E(X)$	Mathematical expectations of X	1
T_i	The time required for the ith round of detection	Hours
Δt	Time required for a single nucleic acid reagent test	Hours/Dose

3. Model building and solving

3.1. Model building

The number of people in the six regions of Beijing is $M_1,\cdots,M_6$ million and the incidence rate is $r_1,\cdots,r_6$, per 10,000, thus, let the number of people in each region of Beijing be $M_i{}'$' and the infection rate in each region be $p_i(i=1,\cdots,6)$ its.

$$M_i{}'=10000M_i,\ p_i=\frac{r_i}{10000} \tag{1}$$

In the mixed testing mode, the average number of testing reagents *E(X)* used per individual is independent of the total number of people in the area $M_i{}'$. Therefore, we mainly consider the effect of infection rate on the optimal mixing method $N_1,\cdots,N_6$. We reviewed the literature and found that this grouping method fits better with the Group testing algorithm proposed by Dorfman, and we modified it based on this algorithm to make it fit our model better [4].

Our strategy is to divide M_i individuals in each region into $N_i$groups and mix each group of samples to perform mixed testing, we test all group samples one by one, if there is no virus in the group, we exclude everyone in this group, and if there is, we rank all individuals in the sample again. The relevant demonstration is shown in Figure 1 below.

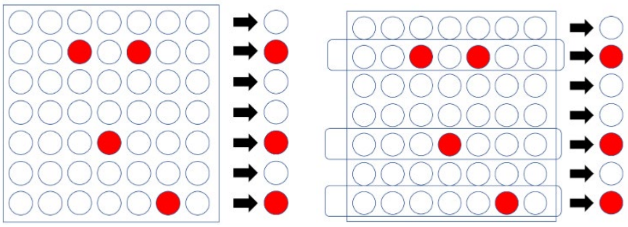

Figure 1: Related Demos

The specific model is developed as follows.

If the mixed collection tube in which the sample is located has a negative test result, the average number of test reagents used by each individual in the group is $\frac{1}{N_i}$, and the random event can be considered as the result of $N_i$independent repetitions of the Bernoulli test with probability $(1-p_i)^{N_i}$.

If the mixed collection tube in which the sample is located tests positive, the average number of test reagents used per person in the group is $\frac{1}{N_i}+1$ with probability $1-(1-p_i)^{N_i}$.

Then the probability distribution of X is.

$$X\sim\begin{pmatrix} \frac{1}{N_i} & \frac{1}{N_i}+1 \\ (1-p_i)^{N_i} & 1-(1-p_i)^{N_i} \end{pmatrix} \tag{2}$$

Thus the mathematical expectation of X is/

$$(X)=\frac{1}{N_i}*(1-p_i)^{N_i}+\left(\frac{1}{N_i}+1\right)*[1-(1-p_i)^{N_i}]=1-(1-p_i)^{N_i}+\frac{1}{N_i} \tag{3}$$

By calculating the relationship between *E(X)* and N_i, we can obtain the optimal solution for N_i at different p_i The optimal solution is obtained by the above equation under the condition of the incidence rate of $r_1,\cdots,r_6$ per 10,000 in each area of Beijing urban area given by the question.

We also assigned specific values to p_i through Matlab and gave feedback on the results to obtain the following results, which can be observed that the higher the infection rate, the less this strategy saves, while Table 2 can provide a reference for the optimal mixed detection ratio N_i in the case of changing incidence rates in various regions of Beijing.

Table 2: The best number of mixed inspections under non-repetitive mixed inspections

Infection rate p_i	Number of tests per capita$E(X)$	Optimal group size N_i
0.84%~1%	0.2	12
0.7%~0.83%	0.17	13
0.6%~0.69%	0.15	14
0.52%~0.59%	0.14	15
0.45%~0.51%	0.13	16
0.4%~0.44%	0.12	17
0.35%~0.39%	0.116	18
0.31%~0.34%	0.109	19
0.28%~0.3%	0.105	20

3.2. Model optimization

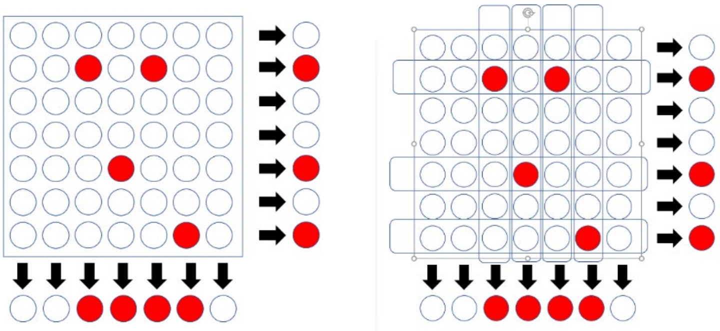

Figure 2: Secondary testing

After obtaining the above results, we consider whether it is possible to reduce the overall number of tests again in the case of mixed tests. If we say that multiple rounds of mixed tests can also reduce the overall number of tests to some extent, but in fact it will add new troubles during the operation and is time-consuming and laborious. In addition, we think of the aforementioned model are not overlapping detection, that is, the mixed inspection combination is not affected by each other. We innovate the mixed-check model, i.e., we consider repeatable mixed-check, as follows.

We take all extracts of nucleic acid tests from all subjects, divide $M_i$individuals in each region into N_i groups by rows and columns, and finally samples missing in columns or rows are filled with negative results, mix each group of samples to perform mixed tests, we test all group samples one by one, if there is no virus in the group, we exclude everyone in this group, if there is, we perform a second test on the relevant crossed rows and columns of All samples are tested twice. This is demonstrated in Figure 2 below.

This method can again reduce the number of reagents required to a certain extent, as in the above idea where we mix the samples by rows and columns to form.

$$Q = 2\left(\frac{M_i' + \alpha}{N_i}\right) \tag{4}$$

Since there are multiple possibilities of containing positive patients in all crossover locations, we consider the total number of tests in the worst case, and the worst case that positive patients are different rows and columns, at this point it is known that we need a total of $(p_i * M_i')^2$ times for nucleic acid testing of all possible positive patient locations, and since we need to test a total of Q samples in the first round of mixed testing, the number of tests per capita.

$$E \leq \frac{Q + (p_i * M_i')^2}{M_i'} = \frac{2N_i + (p_i * N_i^2 - p_i * \alpha)^2}{N_i^2 - \alpha} \tag{5}$$

Since the number of negative samples filled in front of a large sample α has a small effect on the overall, we omit its effect and take the maximum probability of the number of tests per capita E as the value to obtain.

$$E = \frac{2N_i + (p_i * N_i^2)^2}{N_i^2} = \frac{2}{N_i} + p_i^2 * N_i^2 \tag{6}$$

The relationship between the number of detections per capita and N_i is obtained by applying Equation (6) in Matlab, and the relationship between the average number of detections and the infection rate $p_i$can be derived, and the specific results are shown in Table 3.

Table 3: The best number of people under repeated mixed inspections

Infection rate p_i	Number of testing per capita E	Optimal group size N_i
0.94%~1%	0.139	23
0.88%~0.93%	0.133	24
0.83%~0.87%	0.127	25
0.78%~0.82%	0.122	26
0.74%~0.77%	0.117	27
0.70%~0.73%	0.113	28
0.66%~0.69%	0.109	29
0.63%~0.65%	0.102	30
0.6%~0.62%	0.099	31

By comparing this data with the results of 3.1, we found that at an infection rate of less than 6%, the overlapping mixed test is less than the non-overlapping mixed test per capita, but at an infection rate of more than 20%, the overlapping mixed test consumes more reagents than one person per test.

3.3. Model Solving

After obtaining the optimal mixed detection ratio N_i for each district in Beijing, we next calculate the total time required for one or three rounds of nucleic acid census. Before giving the specific calculation process, we first assume that the detection capacity of six districts in Beijing is W_i'every 5 hours, where $W_i' = 10000W_i,\ i = 1,2,\ldots,6$.

Then it can be considered that each nucleic acid reagent needs to be tested.

$$\Delta t = \frac{5}{W_i'} \quad i = 1,2,\ldots,6 \tag{7}$$

On this basis, the time and dose of nucleic acid testing between regions in Beijing are not affected by each other and are not cumulative between regions, i.e., the total time required to perform a round of nucleic acid testing is taken as the longest nucleic acid testing time among the six regions. We also assume that after a positive patient is detected, we take out the positive patient for isolation and do not count it in the next test[5].

Subsequently, we checked the relevant information and learned that one round of nucleic acid testing refers to a mixed test plus a one-person test for problematic samples, so we set $T_i$as the time required for the i-th round of testing. Therefore, the total number of tests required in each round of nucleic acid testing multiplied by the testing time required for each sample is the total time required for each round of nucleic acid testing, which gives us.

$$T_1 = max\left\{\frac{5*M_i'}{W_i'}\left[1-(1-p_i)^{N_i}+\frac{1}{N_i}\right]\right\} \tag{8}$$

$$T_2 = max\left\{\frac{5*M_i'}{W_i'}(1-p_i)\left[1-(1-p_i)^{N_i}+\frac{1}{N_i}\right]\right\} \tag{9}$$

$$T_3 = max\left\{\frac{5*M_i'}{W_i'}(1-p_i)^2\left[1-(1-p_i)^{N_i}+\frac{1}{N_i}\right]\right\} \tag{10}$$

In this case, the value of N_i found in 3.1 is the value that minimizes the total number of detections, or the value of.

$$T_1 = max\left\{\frac{5*M_i'}{W_i'}\left[{p_i}^2 * M_i' + \frac{2}{N_i}\right]\right\} \tag{11}$$

$$T_2 = max\left\{\frac{5*M_i'}{W_i'}(1-p_i)\left[{p_i}^2 * M_i' + \frac{2}{N_i}\right]\right\} \tag{12}$$

$$T_3 = max\left\{\frac{5*M_i'}{W_i'}(1-p_i)^2\left[{p_i}^2 * M_i' + \frac{2}{N_i}\right]\right\} \tag{13}$$

At this point the value of N_i required to satisfy 3.2 is the value that makes the minimum number of total tests, while the time required here is considered as the worst case scenario that all positive patients are in different rows and different columns.

The following is a plot of the total time results with infection rate for our simulation, as shown in Figure 3.

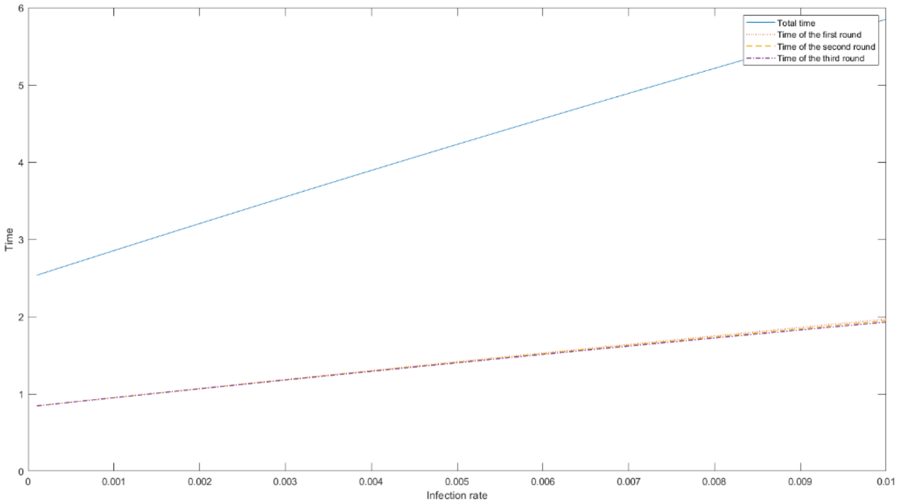

Figure 3: Plot of total time outcome with infection rate

4. Conclusion

The model is clear, easy to understand, practical and readable, and gives the best "N-in-1" test solution for large-scale nucleic acid testing, which greatly reduces the number of samples and provides testing efficiency;

It can be generalized for the selection of testing protocols for similar diseases;

Extend the conventional Group Testing model to two dimensions to reduce testing costs.

For large-scale testing, more parameters are introduced to ensure the accuracy of the model results;

The model formulas are more complex and cumbersome in performing data calculations;

Taking one-dimensional Group Testing, the optimal "N-in-1" is not good enough to reduce the detection cost.

References

[1] Cicalese F. Group Testing [J]. Monographs in Theoretical Computer Science An Eatcs, 2013.

[2] Li, Chou Hsiung (June 1962). "A sequential method for screening experimental variables".Journal of the American Statistical Association. 57(298): 455–477.

[3] Ding-Zhu, Du; Hwang, Frank K. (1993). Combinatorial group testing and its applications. Singapore: World Scientific. ISBN 978-9810212933.

[4] Xie H, Cheng HZ, Niu DX. An algorithm for discretization of continuous attributes of rough sets based on information entropy [J]. Journal of Computer Science, 2005, 28(9): 5.

[5] Xiao Shengxie, Lv Enlin. Mathematical expectations of discrete interval probability random variables and fuzzy probability random variables [J]. Applied Mathematics and Mechanics, 2005, 26(10):8.