Predicting the Result of Universal Community Testing

Dah Ming Chiu, September 9, 2020

(revised on September 16, adding a postlogue after the UCTP result came out)

1. Introduction

The Hong Kong government initiated a Universal Community Testing Program (UCTP). Each citizen can voluntarily take a free Covid-19 test within the period from September 1 to 11. Is it possible to predict the number of people turning out to be positive? In particular, can we use simple models for such prediction, with the logic and assumptions clearly spelt out, so that the method can be generally applied?

The time between infection (of Covid-19) to the time the infected person shows symptoms is called the *incubation period*. The initial part of the incubation period when the infected person is not infectious is called the *latent period*. For Covid-19, the latent period is usually shorter than the incubation period, meaning an infected but asymptomatic person can infect others during the second part of the incubation period. This, the existence of numerous asymptomatic transmitters is the main reason it is so hard to control Covid-19, as nicely explained in a recent article [1].

In order to control Covid-19, we must isolate the infected from the non-infected. If we know who the infected are, this is relatively easy: the infected can be sent to hospital or quarantined. But due to the possibility of asymptomatic cases, you don't always know who are infected. Without this knowledge, many governments resorted to draconian measures of physically isolating all people from each other or other strong social distancing policies, which incurred huge economic and social costs. See [2] for a discussion.

Before long, relatively inexpensive and fast testing capability was developed to determine whether a given person is infected or not. Furthermore, group testing techniques can be used to further reduce the per person cost of testing. Since finding out the infected persons benefits the whole community, the government is often willing to foot the bill or heavily subsidize it. How much testing to do, or how much money a community should spend on testing is really a value judgement, and is beyond the scope of our discussion. However, evaluating the result of a testing program, for example in terms of how many infected cases are found, is a useful exercise, since it may help us compare different testing and isolation strategies.

2. Assumptions and methods

Towards developing a simple model, we begin by making some assumptions.

> Assumption 1: Each Covid-19 infected person behaves the same, in terms of incubation and latent period, and ability to transmit to others during the incubation period.

This assumption let us avoid having to model each infected person separately, significantly reducing the complexity of our analysis. In reality, each person is different, and there are so-called super-spreaders who may be infectious for a longer period of time and more socially active. This assumption essentially treats each person as an average person in every respect.

> Assumption 2: For each infected person, the incubation period is m days, and the latent period is zero, so each infected person is equally infectious for each day of the incubation period.

.

This assumption let us treat all the undiscovered asymptomatic infected persons as a group without having to keep track of their days since infection. The reported average incubation period is around 7 days [1] and the most infectious period is the last 2 days of the incubation period. This assumption treats each person in the asymptomatic group as infectious; this over-estimate can be partly compensated by suitably lowering the ability each person infects others.

> Assumption 3: When reaching the end of the incubation period, an infected person always sees a doctor or goes to the hospital, to get tested.

In reality, an infected person with mild symptoms may stay at large for a variety of reasons, and continue to infect others. This situation is similar to having an extra-long incubation period. So to compensate for such cases, we may need to consider setting the average incubation period m to be slightly longer than the reported average.

Based on the above assumptions, we can assume the size of the (asymptomatic) infected group of people to be n(t) on day t. Knowing the value of n(t) on the days of Universal Community Testing Program is statistically equivalent to knowing the number of people testing positive in UCTP, if we assume the tested population is a random[1] sample of the whole population. Before trying to estimate the value of n(t) for a specific day t, let us first look at the factors that affect n(t) from day to day.

In general, the dynamic behavior of n(t) from day to day can be described by an equation, such as:

$$n(t) = fn(n(t-1))$$

The function would include various other parameters than n(t), reflecting social distancing policies, import traveler control policies and other factors reflecting the nature of outbreak. And we know the function must be an increasing function in n(t). For our purposes, preferring a simpler model, we assume a linear model:

$$n(t) = (h(t)-k(t))\ n(t-1) + g(t)$$

where h(t) models the growth of n(t) by the newly infected persons, and k(t) models how n(t) is reduced by discovering existing infected persons via testing. The factor g(t) is simply the number of imported cases undiscovered by border control. Generally speaking, these quantities are all functions of time, reflecting changing

[1] It can be argued that the sampling produced by voluntary participation is not random; those more likely to be infected have a higher incentive to participate. Also, some citizens boycotted the UCTP due to political reasons. So the tested population is not quite random.

environmental factors and adapting government policies. But over a short period of time, the values of h(t), k(t) and g(t) can remain stable. If they are constants, then:

$$n(t) = a\, n(t-1) + b = a^t\, n(0) + b(1-a^t)/(1-a)$$

What can we say about n(t) for such a static model? Basically, the value of n(t) will blow up if $a \geqslant 1$. But if a<1, and b is small, n(t) would hoover or grow slowly in a manageable range. In the real world, governments would not allow n(t) to blow up, hence would adapt social distancing and border control policies to keep a<1 and b very small if not zero, to keep n(t) stable and stay at a minimal level. Hong Kong government calls the policy adaptation (张弛有度). It should be appreciated that this job of policy adaptation is very delicate and difficult, as the effectiveness of policies are unknown, and a small miscalculation can lead to n(t) growing exponentially. With all things considered, the situation in Hong Kong is well controlled.

Since in general the factors h(t), k(t) are changing, let us consider how to model them. For h(t), we can initially focus on two factors: (1) current social distancing index, S(t), and (2) current average cluster size R(t). So

$$h(t)=S(t)R(t)$$

The social distancing index is some measure between 0 and S*. When the government implements complete lock down or curfew, we can assume S(t) to be almost 0; when there is no social distancing, on the other hand, we can assume the index to be close to S*, which is the average number of persons an asymptomatic carrier will infect in one day. The value of S* would be different in different cities depending on the natural level of social interactions when there is no Covid-19. During a period of time with the same social distancing policy, we would assume the value of S(t) to be roughly a contant.

Once a person is infected, we can assume that a number of people living together with the infected person will also be infected; this is referred to as the cluster size. Note, we assume infection within the cluster is unaffected by social distancing. If the infected person has a normal family, the cluster size R(t) may be 2, for a couple, or 3, 4 or more if the immediate family includes parents and children. Based on this definition R(t) can be assumed to be relatively a constant.

The daily reduction multiplier, k(t), is determined by two important factors as well: (3) U(t), the discovery rate from testing, and (4) V(t), the incubation ending rate:

$$k(t) = U(t) + V(t)$$

That means, we assume U(t)n(t) asymptomatic carriers are discovered by various testing programs implemented at t, i.e. contact tracing, testing of high risk groups, or universal testing programs such as UCTP. If the testing effort stays the same, then U(t) stays relatively constant. The value of V(t) depends on the (infection) age distribution of the carriers, given by $n_1(t)$, $n_2(t)$, …, $n_m(t)$ where the subscript is the age since infection. These values sum to n(t), and $V(t) = n_m(t)/n(t)$. Depending on the social distancing policies and various other factors, the age distribution may vary over time, hence V(t) is usually not a constant. These two factors, U(t) and V(t) do not overlap with each other, and are additive.

Finally, we make one more assumption, that makes g(t)=0:

Assumption 4: All imported cases are detected and quarantined. All local infections are due to asymptomatic carriers undetected locally.

This assumption is, unfortunately, not always true. For a period of about one month when there were no reported cases in Hong Kong; but all of a sudden in early July, new cases started to build up. The new out-break is almost surely due to imported cases. Although Hong Kong has tight border control for regular travelers, there are many exceptions for people working in supply chains and special business and government officials, so the system was not air-tight. Since the new outbreak, the government has tightened policies for the exceptional situations, so it is reasonable to make this assumption now, at the time of UCTP.

3. Data and analysis

The Hong Kong government gives an update on the new cases of Covid-19 each day, giving details such as whether the case is asymptomatic or not, whether it is linked to existing known cases or not (helpful for contact tracing), and various other information about the new cases to help social distancing [3].

t	#local	#symp	#asymp	#unlinked	Date	Est n(t)	Est S(t)
1	16	14	2	8	25/8/20	98	0.42
2	24	14	10	11	26/8/20	98	0.45
3	18	13	5	4	27/8/20	91	0.43
4	10	9	1	3	28/8/20	63	0.42
5	16	12	4	5	29/8/20	84	0.43
6	10	8	2	5	30/8/20	56	0.43
7	7	6	1	2	31/8/20	42	0.42

The above table lists some relevant statistics in the daily update for the week before the start of the UCTP. We have not listed the number of imported cases, which we assume all get quarantined and do not contribute to the hidden carrier group. Column 2 is the total number local cases, out of which some are symptomatic and some are asymptomatic as shown in the 3^{rd} and 4^{th} columns. The 5^{th} column shows the number of cases not linked to earlier cases (source of infection is unknown).

From the table, and our simple model, we first observe that both the total number of local cases, and the number of cases with symptom (let us denote that by $W(t)$), follow a slowly declining trend. Based on this pattern, we make a conjecture - the age distribution of the asymptomatic carrier group is relatively flat. If the trend for local cases is not declining, but flat, then the age distribution is most likely declining, since some of the cases are discovered before they reach the end of incubation period. But the declining trend of local cases should compensate towards making the age distribution flat. Given this conjecture, we can estimate $n(t)$ as

$$n^*(t) \sim m\, W(t)$$

where $m=7$ is the average length of incubation period. We list this estimated $n^*(t)$ in column 7.

Another observation is that due to the rather smooth declining trend, we can fit an average declining rate of

a = 0.89

In our model

a = (RS – U – V)

The values of R and V have been assumed to be roughly constant during the short-term:

V = 1/m = 1/7

R = 2.5

which means on average 2.5 immediate family members are assumed to be infected together. The value of U(t) is given in the daily update, column 4 divided by column 7:

U(t) = #asymp / n*(t)

This allows us to estimate S(t) as

S*(t) = (0.89+U(t)+1/7)/2.5

The result is listed in column 8. S(t) means the number of persons an asymptomatic carrier would infect under the social distancing rule on day t, on average. Since the social distancing rule did not change during that week, we would expect S(t) to stay relatively constant, and indeed we see a rather steady infection rate. Note, this estimated value of S(t) would be lower, if the family cluster size (currently set to 2.5) is high, and vice versa.

Based on the above table, we would predict that the UCTP will discover fewer than 42 positive cases. If only a fraction x of the population will go for the UCTP test, and if x is a random sample of the population, then the number of positive would be 42x.

4. Postlogue

After the UCTP is over, on September 15, the Hong Kong government gave the result. A total of 32 positive cases[2] were discovered by testing a total of about 1.78M people. So the result falls within our prediction of (0,42). If the people tested is a random sample, then the 32 positive would indicate the asymptomatic carrier group size to be:

n(1-Sept) = 32*7000000/1780000 = 126

Since our model is a very rough one, intended to find a ball-park estimate of the asymptomatic carrier group size, the discrepancy is within our expectations. There are a couple of reasons that our prediction is even better than it looks. First, the group that went for the test is most probably not a random sample, with the higher risk people more likely to have volunteered to get tested. Secondly, since the test went on for two weeks, additional people could have been infected, increasing the pool of people to be found positive.

The government also reported the cost for Hong Kong: 530M Hong Kong dollars,

[2] Besides these 32 cases, there were another 10 other positive cases discovered by UCTP. Five of the 10 were old cases who have already recovered from Covid-19; the other five either had symptoms or were linked cases that were first discovered by testing in hospitals.

covering mostly the pay for medical staff to collect the specimens for the test, transportation and publicity. The test analysis was carried out by a team from mainland China, and did not incur cost to Hong Kong.

5. Discussion and conclusions

In this note, we showed some simple model/methods to analyze daily updates of Covid-19 cases, to estimate the number of asymptomatic cases found by a Universal Community Testing Program, like the one recently implemented in Hong Kong. We elaborated on some assumptions used in our analysis, to clarify the soundness of the very rough analysis.

There is certainly room for further analysis. Some of the other statistics published in the daily updates seem quite relevant, but we have not exploited in this analysis. For example, the number of local cases not linked to any earlier cases; this should be correlated to the asymptomatic group size, as well as the effectiveness of current social distancing policies. The government report also gives the larger clusters they found, for example at some nursing homes, or special work places where social distancing rules are relaxed. The existence of these larger groups should also affect our analysis since we assume each case is independent and take averages in various parameters.

The purpose of our analysis is to help better evaluation and review of testing policies. There has been a lot of criticism and counter arguments with respect to the UCTP. The most common criticism is that the UCTP is not cost effective. If we look at a single policy, such as the UCTP, cost-effectiveness becomes a value judgement, of how much you are willing to pay to discover each asymptomatic carrier out there, at a given time. Instead, we can look more closely at alternative testing policies. For example, instead of a UCTP, if we spread our money and efforts over a longer period of time, instead of testing 1.5M people all at one time, we put the money to test 60K high risk people each week for 25 weeks, would it be more effective? After all, no one would expect Covid-19 to completely go away after the UCTP, considering that we will implement travel bubble and open up schools in the coming months. Comparing alternative testing strategies is no longer just value judgement, and can lead to better policies.

Reference

[1] Nicholas A Christakis, "Nicholas Christakis on fighting Covid-19 by truly understanding the virus", The Economist, Aug 10, 2020

[2] Dah Ming Chiu, "Isolation control and its challenges", https://staff.ie.cuhk.edu.hk/~dmchiu/isolation2.pdf, Mar 8, 2020

[3] Latest Local Situation of Covid-19, https://www.chp.gov.hk/files/pdf/local_situation_covid19_en.pdf