Predicting the Result of Universal Community Testing
Dah Ming Chiu, September 9, 2020

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:

\[
  n(t) = (h(t)-k(t)) \ n(t-1) + i(t)
\]

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 \( i(t) \) is simply number of imported cases undiscovered by border control. These quantities are all functions of time, reflecting changing environmental factors and adapting

---

\(^1\) 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.
government policies. If h(t), k(t) and i(t) are static constants, then we have
\[ n(t) = a n(t-1) + b = a^t n(0) + \sum_{i=0}^{t-1} (a^i b) \]
What can we say about n(t) with such a static model? Basically, the value n(t) will blow up if \( a \geq 1 \). But if \( a<1 \), and b is small, n(t) would hoover 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, so that n(t) is stable and kept at a minimal level. Hong Kong government calls this suitable policy adaptation (张弛有度). It should be observed that this job of policy adaptation is very delicate and difficult, as the effectiveness of different measures are unknown, and a small miscalculation can lead to n(t) starting to grow 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.

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 S(t) may be 2, for a couple, or 3, 4 or more if the immediate family includes parents and children. Based on this definition S(t) can be assumed to be relatively a constant.

The daily reduction multiplier, k(t), is determined by two important factors as well:
(3) discovery rate by testing, U(t), and (4) incubation ending rate V(T)
\[ k(t) = U(t) + V(t) \]
Note, we assume U(t)n(t) asymptomatic carriers are discovered by testing programs aimed at discovering them, 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) can be assumed to be a constant as well. At the end of the incubation period, we assume an asymptomatic carrier starts to show symptoms and immediately gets discovered by testing. During periods when n(t) is relatively stable, we can assume the rate V(t) is also close to 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 i(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 must be imported. 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. After the new outbreak started, the government has tightened policies for the exceptional situations, so it is reasonable to make this assumption now, during the 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].

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

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 3rd and 4th columns. The 5th 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 V(t), follow a slowing 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

\[ n^*(t) \sim m \cdot V(t) \]

where m=7 is the average length of incubation period. We list this estimated number in column 7.

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

\[ a = 0.89 \]
Since our model says
\[ a = (RS - U - V) \]
we can determine S by guessing the values of R, U and V. We have already set
V = 1/m = 1/7. U can be estimated from the estimated n*(t):
\[ U(t) = \text{#asympt} / n*(t) \]
As for R, the number of immediate family members infected together, we can try
setting it to 2.5. This allows us to estimate S(t):
\[ S^*(t) = (0.89 + U(t) + 1/7) / 2.5 \]
as shown 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 see a rather steady infection
rate, as expected. 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.

Now let us come back to our original question: how many positive cases will we find
out of N people tested. The complete result will not be in until after September 14
when the UCTP is over. But based on the results known so far, we already have a
rough answer to our question. For the first 128K people taking the UCTP test, 2 were
found positive. Assuming this first batch is a random sample of the 7M HK population
\[ n(t) = 2*7000000 / 128000 = 109 \]
By September 9, it is reported tentatively that 19 are found positive out of 1.15M
tested, this gives
\[ n(t) = 19*7000000 / 1150000 = 116 \]
which is close to the estimate from the first batch. These results indicate that our
analysis and prediction is a little lower than these results show. The true number is
likely to be in between that indicated by the test results and our prediction. For the
UCTP, since it is voluntary, the ones who went for the test may not be a random
sample, but slightly biased towards the higher risk ones in the population. But all in
all, I think our prediction should definitely in the ball park.

After the UCTP is over, based the subsequent daily update data, we can do a similar
analysis to estimate the value of n(t) post-UCTP, and see if we can link that up with
our pre-UCTP analysis. The difference should be just the number of cases found
during UCTP. We will write an epilogue at that time.

4. 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. The Hong Kong government has not shared the cost information for the UCTP, and alluded that the cost may be largely covered by the mainland government. 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

