PIDE

COVID-19 BULLETIN

No. 10

WHAT DO CONFIRMED NUMBERS TELL US? Using an Adapted SEIR Model for Estimation of COVID-19 in Pakistan

Bulletin Series Editors' Note: The current Bulletin presents estimates for those affected by the coronavirus pandemic in Pakistan. The estimates are based on a well-established model (SEIR), but given the small number of available data points, these numbers should be treated with caution for now. The model would be updated/revised as more data become available with time.

Since the model involves a lot of technical work, for the sake of transparency, the complete paper is also available on the PIDE website in the PIDE COVID-19 Archives section.

This Bulletin, and the paper, should also help understand the figures presented in the PIDE COVID-19 Dashboard on our website.

While the government publishes daily numbers of confirmed COVID-19 cases, it's understood that these obviously do not reflect the exact total number of people infected by the virus to date. This is primarily because tests are reserved for people who are 'suspected' to be infected by the virus. This means that many people who have been exposed to the virus may not be tested because they have yet to show symptoms.

We know that many carriers of the SARS-CoV-2 virus have a period of time where they are asymptomatic. In the initial stages of this period, they are not yet infectious but later they become asymptomatic carriers. These are the people likely to missed in the count if testing is targeted only towards symptomatic cases.

We are working on developing a model to try and estimate the actual numbers using the official numbers published every day.

Many countries that have to ration test kits will do so by targeting only people with symptoms (and possibly a travel history). This raises an interesting problem in terms of estimating the actual number of infected people at any time, given that we only know the confirmed cases. We need actual numbers if we are to project the course of this infection and plan accordingly.

Furthermore, individuals who test positive are likely to be quarantined. This will have some impact on the dynamics of the contagion in the population and makes our estimation problem more intriguing.

What is SEIR?

The basic model used for understanding the course of an infection in a population is the SEIR model. The standard *SEIR* assumes all people in the population are in one of four states – *(S)usceptible, (E)xposed, (I)nfected or (R)emoved*.

S= Susceptible- Population that is in the susceptible state*,*
E= Exposed- Those who have come into contact with infected. they are carrying the virus, but not yet infecting anyone else.
I=Infected- After a period of exposure they get infected and can infect others. They are now actively infecting other people.
R= Removed, either cured or dead.

So, a small number of Infected people come into contact with *Susceptible* people who are then moved to *Exposed* and the infection begins to grow. The numbers of *infected* people begin to accelerate initially as more and more infected people are around to infect other susceptible people around them. It slows down later when most infected people are just interacting with other *Infected* people and rarely come into contact with *Susceptible* ones. How quickly *Susceptible* people are transferred to the Exposed stage depends, among other things, on the *transmission rate,* which is the product of the contact rate and the probability of transmission given contact.

From there on, a series of equations tells us what proportion of the population is in each of the four stages, based on how long each individual stays in each stage. So, for example, if the period of time spent Exposed is very short and the duration of infection is long, then more and more people will start to accumulate in the Infected state. If the exposed state was also longer then people would enter the Infectious state more gradually.

Some proportion of the Removed cases will have recovered and presumed to have developed an immunity (at least temporarily) from the infection, while some portion of them will die. Mortality rates can be fed into these equations as some proportion of the Removed cases. This is were it is important to know how fatalities are being calculated in an SEIR models because they will be affected by various factors – for one, the entire population may not end up being infected. Or the model may not take into account the degradation in medical services with high number of infections.

A good explanation of such models can be found at:

https://www.tandfonline.com/doi/full/10.1080/23737867.2018.1509026.

It is on this base that further models are developed, with additions for particular cases. We have attempted to do the same in Pakistan to answer the question about the relationship between confirmed cases and actual cases.

SEIR in Pakistan

For a naive estimate of how far along the infection we are and how it is going to progress, we could simply assume that the confirmed cases are the number of infected people. Feeding this into the model should tell us how many people to expect the next day and so on.

However, this misses some important aspects. Firstly, it misses the fact that people are asymptomatic carriers for a considerable period of time. Thus, official numbers are going to be missing infected people and furthermore, are going to miscalculate how many more infected people are going to appear the next day. In fact, not even all people who show symptoms are going to be tested. Thus, there is going to be some probability that an infected person does actually get tested.

Secondly, we seem to have a system where anybody who tests positive is put into quarantine. Thus, this person is infected, but is not causing any more infections until he moves to Removed.

Finally, for several days at the start of the infection, we weren't dealing with a fixed population of people. Individuals at each stage S, E, I and R were entering the country until international flights were banned.

Thus, we made the following modifications to the model. We now assume that people in the Infected stage do not show symptoms for several days. So the Infected individuals who have been identified as infected are now moved to the quarantine. During this time, they are not infecting others. We also allow for E and I individuals entering the country for some days before flights were suspended.

We then try to use this in conjunction with actual numbers to try and estimate the rate of testing in the country, and the rate of transmission. This should tell us how far along we are in the progress of the infection and what we should expect next.

How do Testing and Transmission Rates Affect Confirmed Cases?

As explained earlier, transmission rates basically affects how quickly one infected person infects those around him. A high transmission rate would thus automatically imply a rapid increase of infection in the population.

On the other hand, suppose we see the number of confirmed cases rising quickly, what does that mean?

How we interpret a rapid increase could depend on what proportion of infected people we think are being tested. If a very large proportion of infected people was being tested then a rapid increase may not be such as bad - it would mean that infected people are being rapidly identified and quarantined, potentially slowing down the infection.

On the other hand, if we think that testing is low, then high rates of increase mean that the actual numbers are much larger than being identified and quarantined and we can expect the infection to progress rapidly.

However, we also have one other measure at our disposal that may help identify which of the two scenarios it is – the number of fatalities. After all, given a certain mortality rate, which we know to be around 1% for Covid-19, if we see a small number of fatalities, we may be more inclined to believe that it is the first case and not the second one.

Initial Estimation

Using the adapted SEIR model described above and parameters for the Covid-19 virus that are being shared from different studies, we can estimate what the numbers of people in each of the five states – S,E,I,Q,R – should be at this stage in time, along with how many fatalities we should expect.

Further, from the status of official numbers of reports and fatalities on 27th March (the current number of actual positive cases is 1235, and number of deaths is 9), we vary the values of the testing and transmission to see which ones are most likely to produce the real world situation. Using this technique, we can get some rough idea of what the actual situation might be.

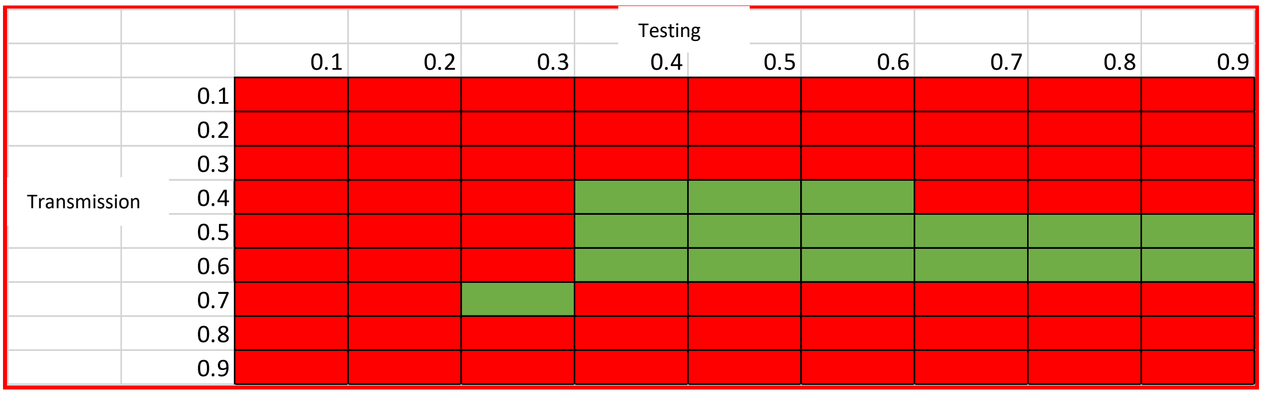
Mortality Numbers

There is an important question on the mortality rates being used by the government in terms of how well hospitals are equipped to identify cases of COVID-19 as cause of death. Thus, we can take two approaches to how we interpret the numbers of people dead:

A. We can consider them to be the actual total number of people who have died
B. We can assume that they are the number of people who have died after they were identified and quarantined.

Obviously, this means that, from the options given above, A would be higher than B at all times.

The following diagram shows the estimates for the transmission and testing that are likely/unlikely to have produced the real world numbers.



The values in red show all the parameter values we have eliminated as being too unlikely for given the Pakistani numbers as they are coming out. The green values are the ones that seem likely given what is happening (see attached research document). Thus, it appears that the transmission rate is somewhere between 0.4 and 0.7. Apart from one case, we can see that the testing/quarantine rates are fairly high so far ($testing \geq 0.4$).

This shows that while the transmission rate of Covid-19 is on the high side (as we know), the testing and quarantine rates also seem to be on the high side. This result relies, of course, on the validity of actual government numbers. At high transmission rates, we also see a lower gap between A and B, suggesting that there is reason to believe that official death toll is in accordance with the model.

Obviously, we are dealing with smaller numbers and stochastic processes modelled as deterministic ones, so there is a high chance of error in this model. However, combining confirmed cases to death rates does provide some insight into the underlying picture.

Next, we run simulations using some of these predicted parameters to try and estimate the actual rate of infection and the probable course these figures will take in the coming days. we will be commenting on four of them.

We have estimates for Quarantined/Tested, mortality scenarios A and B, Actual Infected, and Recovered. We have not used the Recovered statistics provided by the government for any estimation, because they only refer to confirmed cases and when the Government actually deems someone "recovered" may be varied. While deaths may also only refer to confirmed cases, they are far less likely to be missed or subjective. Also, for individuals with mild to no symptoms (a vast majority), many will have recovered without anyone knowing.

Forecasts – Two Scenarios

Analysis done at Day 32 from the first confirmed case. Given that the individual had travelled from Iran several days prior, it is difficult to estimate exactly when he got infected or when he switched from being Exposed to Infected.

When looking at forecasts for deaths it is important to understand how mortality in this (and other SEIR models) is calculated. The assumption in an SEIR model is that the entire population is going to eventually be infected. The real questions are on the rate of infection. For the number of deaths, here we simply multiply the number of people who have passed through the infection by a fixed mortality factor. This factor can obviously be tweaked and even endogenized, but for now it is at 1 percent.

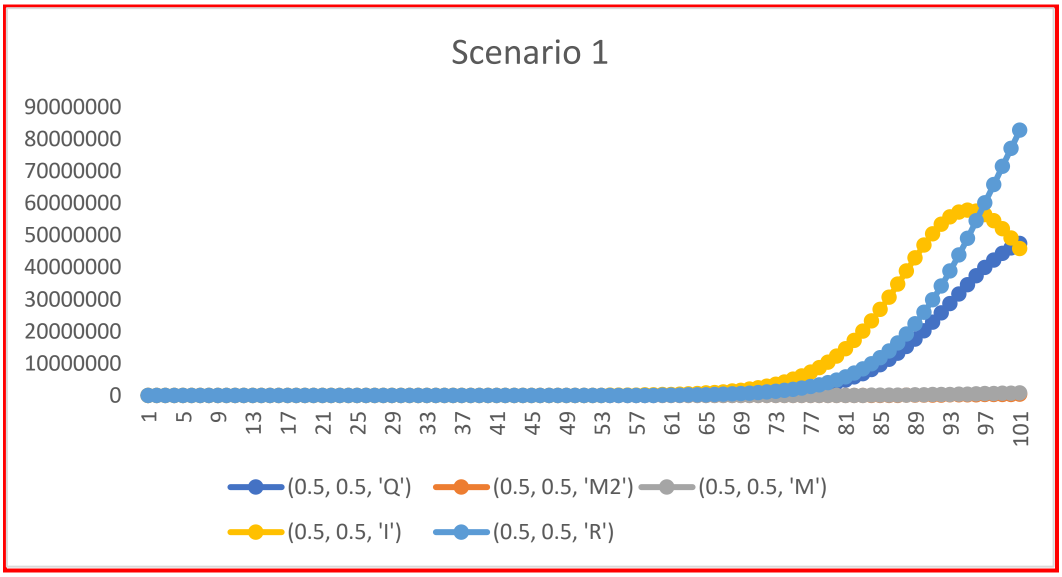
We now look at two of the possible scenarios suggested by our results, by looking at projections of the first 100 days of infection. This is a qualitative exercise to understand what our results may be saying about the future progression of CoV-19 in Pakistan.

Scenario 1 $(mid - level\ transmission\ and\ testing)$

The last case we consider is the one where our model leads the real-world data slightly. It could imply that the infection had started a few days before the actual confirmed case. So, on Day 36 and Day 37 of the infection, we see that between 1194 and 1445 positive tests, the undiagnosed infection numbers are approximately in the 3800 to 4600 range.

Day	Quarantine (Tested)	Deaths (B)	Deaths (A)	Actual Infected(I)	Recovered (R)
36	1194.902	7.086651	14.1733	3849.26	1417.33
37	1445.385	8.570228	17.14046	4657.458	1714.046

As seen below, with this transmission rate the cycle will not be completed within 100 days. The fact that transmission is low and there is a reasonable rate of testing, we can see that the spread of infection is delayed. Another important thing to note, is that the official numbers would continue to increase even when the infection had peaked and was on the decline. This is because of the lag in the appearance of symptoms.



Scenario 2 (high transmission & low testing)

We now consider a scenario where the transmission rate is very high and the testing rate is relatively low. This case seems less likely as it implies that real world data is leading the model i.e. the model is predicting the current numbers at a faster rate than they are actually occurring.

Also, it points to a much higher mortality rate than currently exists (if deaths are not going unrecorded).

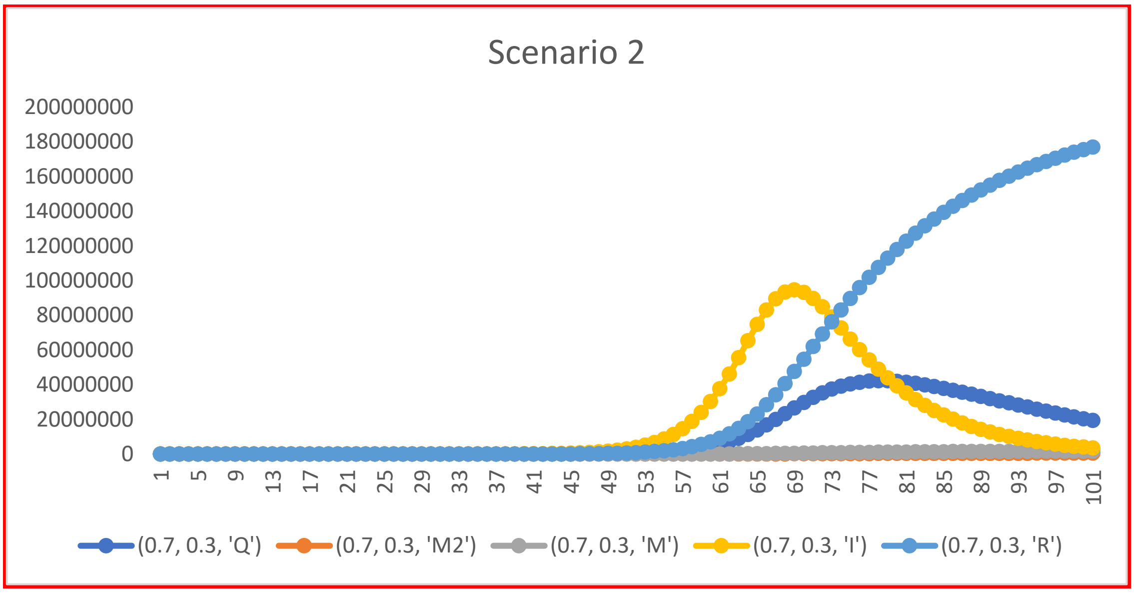
It is still worthwhile to consider this scenario as well, though keeping in mind that it is an exception. It estimates that when confirmed cases stand at 919 -1597, the actual undiagnosed cases are somewhere between 6923 to 12039.

Since the rate of testing and quarantine is low, the number of undiagnosed cases is significant.

Day	Quarantine (Tested)	Deaths (B)	Deaths (A)	Actual Infected(I)	Recovered (R)
28	919.859923	4.360815	14.53605	6923.39409	1453.604986
29	1211.92022	5.744919	19.14973	9129.694057	1914.972869
30	1597.115498	7.569909	25.23303	12039.03617	2523.303121

Furthermore, if we look at how the disease progresses when testing and quarantine is low, we can see the number of infected people peaks much sooner and at a much higher value. This means that the sudden burden on medical resources would be much higher. Again, we see that official number and quarantining lags the actual peak of infection number.

Without delving into numbers, and simply analysing the mechanism of how disease spreads, we can see a policy of simply testing and quarantining symptomatic cases will inevitable lead to an explosion of the infection. In order to curb spread, the government needs to get ahead of the curve and start testing asymptomatic cases (which it may already be doing through a focus on contact tracing).



Conclusions

This is a first pass at a possible technique to use official numbers to try and estimate actual ones. With more data coming in, and better tuning of the model, we should start getting better results. We have tried to do a very rough technique of parameter estimation with simple rules of the thumb. Given the time sensitivity of the matter, we think this provides a good base to see where things are heading.

We can draw a couple of conclusions about our results.

- Rate of testing and quarantine does not seem to be horrendously bad. This seems to be coming from (and is borne out of) the fact that the death rate is currently not very high for the number of confirmed cases. To counter any underreporting of deaths, we have put in a cushion of them being valid for at least up to a factor of 2 i.e. even if the number of deaths are double than the reported value, these results should hold.

- The other thing it points to is a fairly high rate of transmission. This again, should not be surprising as the defining feature of SAR-CoV-2 is its high transmissibility, with near or more to half the population getting affected till the outbreak hits its peak.

- Depending on which of the two given scenarios might be valid, the model estimates about 2000 to 9000 undetected cases. This might influence how the government decides to deal with the problem.

- If we persist with only testing symptomatic cases, there is no way the actual number of infections will not explode over time. That's exactly why, the *PIDE COVID-19 Bulletin No. 9* advocates the use of mass-testing.

It should be borne in mind that the SEIR model is an aggregated approximation of how infection behaves in one population. At the moment, we have treated Pakistan as one large population. When we break it down to a more granular level, where we inspect data from smaller geographical areas (i.e. cities, villages etc.), we should be able to get better estimates, especially if we can include information of movement between these cities.

Furthermore, it cannot (presently) account for fluidity of the counter-measures against the infection. For now, the focus of this bulletin should be taken as the expounding of a technique which can estimate the actual numbers of infection using official data. Actual numbers are important because they can be used to present more accurate predictions of the course the infection is going to take than confirmed numbers. This technique is constantly being improved and can be used to increasingly better estimates in the future.

If left unchecked the SEIR model shows large and alarming numbers of infections and quarantines. These are important for policy to keep in its purview to develop strategies for checking the pandemic. Such models are the starting point of understanding the pandemic and we will continue to improve them and inform the policymaker and you the reader.

By:
Amin Hussain
Teaching Fellow, LUMS
and
Visiting Fellow, PIDE.

PIDE COVID-19 Bulletin is an initiative by the Institute in response to the current pandemic, which is bound to have serious consequences for the country, specifically for its economy. The Bulletin would carry research that would aid in an informed policymaking to tackle the issue.

Pakistan Institute of Development Economics
Web: www. pide.org.pk, Twitter: @PIDEpk, Facebook: PIDEIslamabad