PIDE
COVID-19 BULLETIN
No. 18

Explaining the R_0 in the COVID-19 Projections

PIDE COVID-19 Bulletin No.10, "What Do Confirmed Numbers Tell Us? Using an Adapted SEIR Model for Estimation of COVID-19 in Pakistan", had an all-important term R_0. Despite seeing this term in all projections, not many know what it means. This Bulletin does just that- explains what R_0 means.

A number doing the rounds is R_0. It basically shows the number of people each infected person is going to infect over the course of his infection. Generally, this is used as an important identifier on whether the spread of infection is going to grow or recede over time. If R_0 is more than 1, then it means that the infection is going to spread in the population, while if it is below 1, it should die out.

How does this work? Basically, let us consider the case where 10 people have been infected.

If R_0 is 0.5, that means that by the time those 10 people have recovered (or died), another 5 people will have been infected. Then, by the time those five recover, another 2-3 people will be infected. And over time, this number drops to 1 and then zero. So, an R_0 less than one leads to the infection eventually coming to an end.

On the other hand, if R_0 was 1.5. Then, by the time the initial 10 people have recovered (or died), 15 people will be infected. By the time those 15 people recover, another 22-23 people have been infected and so on. Thus, the infection keeps growing.

Of course, if R_0 is 1, then 10 people are infected by the time the original 10 infected people recover (or die). So over time, there is a constant of 10 people who are infected at any given time.

Most calculations being done are trying to use the value R_0 to estimate the path COVID-19 is going to take. It is worthwhile to look at this figure in a little more detail.

What Impacts the R_0?

There are various factors that can impact the value of R_0. These come from aspects such as social interactions as well as the characteristics of the virus and the course taken by the infection.

To look at the last aspect first, one important factor is how long an affected individual stays in the infectious state or recovers or dies. So, if the disease lasts for a long time before someone recovers or passes away, then that means there is a longer time for that person to cause more infections. On the other hand, if some diseases act rapidly, then the chance of infection spreading is reduced simply because there is a shorter period of time for an infectious person to infect others. Thus, R_0 goes down for diseases that act rapidly but goes up for diseases that last longer.

The R_0 is also affected by transmission rates. This is dependent on factors dependent on the characteristics of the virus and social activity. On the one hand, the more people come into contact with one another, the higher the chance of an infected individual coming into contact with others and

potentially infecting them. On the other, how transmissible the virus is will impact how many of these contacts actually result in the infection being passed on to someone.

As we know, the people who are infected includes those who are asymptomatic carriers i.e. people who show no symptoms of having the infection and are probably unaware that they have it. In fact, given the quarantine procedures being employed in Pakistan, once a person is confirmed as being infected, they are quarantined and are no longer infecting others. In fact, they are soon on their way to developing an immune response and fighting the infection.

Of course, there will be more dimensions that epidemiologists consider for each infection, to get better insight that go into more detail about how each virus and disease operates. But for a novice overview, these are roughly the factors under consideration.

Estimates of R_0 from Various Countries

Princess Diamond	14.8	Rocklöv, J., Sjödin, H., & Wilder-Smith, A. (2020). COVID-19 outbreak on the Diamond Princess cruise ship: estimating the epidemic potential and effectiveness of public health countermeasures. Journal of travel medicine.
Greater Wuhan	2.2	Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., ... & Xing, X. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. New England Journal of Medicine.
Guangdong	4.5	Liu, T., Hu, J., Xiao, J., He, G., Kang, M., Rong, Z., ... & Zeng, W. (2020). Time-varying transmission dynamics of Novel Coronavirus Pneumonia in China. bioRxiv.
South Korea	1.5	Shim, E., Tariq, A., Choi, W., Lee, Y., & Chowell, G. (2020). Transmission potential and severity of COVID-19 in South Korea. International Journal of Infectious Diseases.
Japan	2.6	Kuniya, T. (2020). Prediction of the Epidemic Peak of Coronavirus Disease in Japan, 2020. Journal of Clinical Medicine, 9(3), 789.
UK	2.6	Jarvis, C. I., Van Zandvoort, K., Gimma, A., Prem, K., Klepac, P., Rubin, G. J., ... & CMMID COVID-19 working group. (2020). Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK.

From the factors discussed above, we can see that social distancing covers essentially one dimension of the factors influencing R_0. Essentially, it is trying to reduce contact rate in the population to push R_0 below 1 and contain the infection from exploding. This raises difficult questions about what happens when social distancing ends. It is possible that R_0 will jump back up because of the many unidentified cases of infection (possibly asymptomatic), and we could see a second wave of infection hitting the population. This 'second wave' problem is what the UK had in mind when they initially announced they would be opting for herd-immunity.

To add to the complexity of the problem, we are also unclear about the type of immunity developed to COVID-19. This complicates matters further as we cannot simply rule out people who have already been infected as being susceptible to future infection.

Why Calculating R_0 is Difficult?

From the factors above, we can see some rough measures that can be theoretically used to calculate R_0.

The most basic one, for a simple SIR model, is $R_0 = \frac{transmission\ rate\ (\beta)}{recovery\ rate\ (\gamma)}$.

Here the transmission rate is the effective rate of transmission of disease per interaction between individuals. Essentially, this equation is saying that if infected people recover more quickly than new people are infected, then R_0 will be less than 1 and the infection will not grow rapidly. On the other hand, if new infections are happening more rapidly than older ones are recovering then we will see "exponential growth".

This formula gets more complicated with more detailed SEIR models and we will do one application to an adapted SEIR model later. Essentially, it must account for the population at different stages of infection.

Most models such as SIR and SEIR, make certain simplifying assumptions that mean that their results need to be taken with a grain of salt. They operate under the assumptions that all individuals in the population have some probability of coming into contact. Real world populations do not operate in such a way and are actually more likely to come in contact with certain people and not others. This will be due to geographic constraints and social norms.

Furthermore, many of the models we are seeing are non-stochastic. This means they assume fixed numbers for characteristics of the disease that are variable from person to person, e.g. the incubation period of the disease, the time before recovery etc. could vary. It is not immediately clear that such a process "averages out" and that we can use fixed values as substitutes in such a process.

Thus, such models are more useful to illustrate broader phenomenon than for accurate estimation/prediction.

From an empirical perspective, estimating R_0 in the middle of a pandemic is a very difficult task. The most exact calculation would, of course, trace each single case to see how many more cases it led to. At the moment these numbers do not appear to be public knowledge. In fact, it is unclear how the government may actually be going about doing this, and whether the infection is currently at a stage where the numbers are still manageable.

Falling short of that, if one knew details on social activity of individuals, one could theoretically couple it with information about the characteristics of the virus and disease to calculate R_0 - but such information is also difficult/impossible to come by.

Tracking Networks

Smart tracking is possible even in Pakistan with considerable accuracy. Mobile phone companies can geo track and geolocate via cell phones. Given cell phone permeation this could be a fairly good indicator of physical networks of disease.

The disease progresses in networks of physical contact as the virus jumps from host to host. With geo tracking of mobile phones of those who test positive and fairly generous testing of these network, it is possible to get a better handle on the transmission rates as well as geographical concentration of the infection.

This smart tracking could be of great use in the corona war.

Given official numbers of confirmed cases, and its progression, one might be tempted to take a stab at guessing transmission rates and couple it with data on recovery rates to guess what R_0 is. So, for example, if one assumed that the disease started with just one individual, and that numbers are growing at some fixed rate, then one might try to calculate R_0 simply treating the total confirmed cases as a sort of geometric series.

However, such an approach is unlikely to give accurate/meaningful results. Here are some obvious problems with this approach:

a. Given the limited amount of testing, there is likely to be significant uncertainty on the actual prevalence and spread of disease. As it stands, we just cannot use numbers of confirmed cases as a reflection of the total number of infected people in Pakistan. The situation is compounded by the fact that we have asymptomatic carriers in the form of people who are yet to develop symptoms, and those who never develop symptoms.
b. Social behavior would vary vastly depending on various factors, including density of population and social norms. The transmission rate, (described above) will vary drastically in different neighborhoods, based on a multitude of factors. Furthermore, this averaging of behavior ignores incidents such as cases of "super-spreaders". Hence, it is very difficult to use aggregated data to try and estimate transmission rates.
c. There is also uncertainty on the sources of each case at the moment. To consider an extreme example, if Pakistan has ~4500 cases and all of them came from abroad, then that means that R_0 in the country is essentially zero as it indicates that none of them spread the infection to anyone else. On the other hand, if there was one case that led to these 4500 cases over time, then that means R_0 is much higher.
d. The process is dynamic, and infections show up with delays. Furthermore, the characteristics of the disease are not fixed, and can vary person to person. This means that the duration of being contagious is not fixed for each individual. So, if we knew that one person was infectious on Day-1 and 4 more people were infected on Day-10, it would be unclear whether that one person led to three more infections directly, or whether they led to one or two more, which were then further transmitted. Thus, even if we knew the exact numbers in the population (which we do not), if would be difficult to find R_0 unless we had detailed information on geographic spacing and subjected it to rigorous mathematical techniques, estimates on R_0 might not be in the same ballpark as the actual number.
e. Without a detailed mathematical framework and detailed data that takes multiple factors into consideration, any estimates for R_0 are difficult to trust. At the moment, the approach in play is to assume it is similar to other countries and other situations, but it is unclear at the moment whether this approach is justified.
f. Another thing to keep in mind, along with our limited testing, is the fact that anyone who tests positive is then hospitalised/quarantined. This means that the moment the one confirmed case is established, it ceases to be a contributor to the further spread of infection. This also distorts calculations somewhat.

Keeping all of this in mind, we need more detailed information along multiple dimensions from the government to be made available to get a clearer picture on where we stand with R_0. ***Minor errors in calculation can compound and completely change the picture. After all, there is a difference of 0.2 between $R_0 = 0.99$ and $R_0 = 1.1$. In one case, the infection dies out, in the other it grows exponentially***.

Take-away Lesson on R_0

While we might not be able to do exact calculations to ascertain the value of R_0 or the impact of lockdown/social distancing policies on R_0, we can operate on a few basic principles.

Firstly, reducing contact rate by any factor should reduce R_0 by that factor. So, if everyone came into contact with 1 person every day, and now we meet 1 person every 2 days, then R_0 should reduce by half. Simply put, suppose an infected person was infecting one person every day and would be responsible for 20 new infections over the course of 20 days. Now, he'd be responsible for 10.

If this is enough to push R_0 below 1, then we can expect the infection to gradually die out over time (this would turn up with a lag in the official confirmed cases numbers). If it does not push R_0 below 1, it would still be significant in slowing down the spread of the infection and "flattening the curve", thus reducing the strain on medical facilities, and most importantly saving lives.

Some R_0 Calculations

We used an adjusted SEIR model to do some calculations. As stated above, it suffers from certain simplifying assumptions, and is therefore to be used for illustrative purposes.

We have based our model on explanations presented by Blackwood and Childs (2018)[1]. As can be seen from Section 3.1.1 of their paper, '*A Cautionary Tale of Computing* R_0*'*, our model allows for importation, which makes it different from what is stated in the section above.

One thing we can do is estimate R_0 assuming no importation. In this case, assuming that the quarantine state is not part of the infectious state as people in quarantine are no longer infecting others, we can adjust the formula for calculation as:

$$\frac{\beta\sigma}{\left(\gamma+\frac{q}{k}+\mu\right)(\sigma+\mu)}$$

Where:
β = transmission rate
σ = rate at which people move from being exposed to the infection
γ = rate at which infected people move to recovered/deaths
μ = rate at which people are leaving the total population
q = probability that an infected person is testing and quarantined
k = number of days before an infected person exhibits symptoms

This poses some problems because we are estimating for values of β and q, our estimates can be skewed by how many of our cases are accounted for by importation as opposed to infectious spread.

As you can see, this model ignores cases of infectious people who never exhibit symptoms.

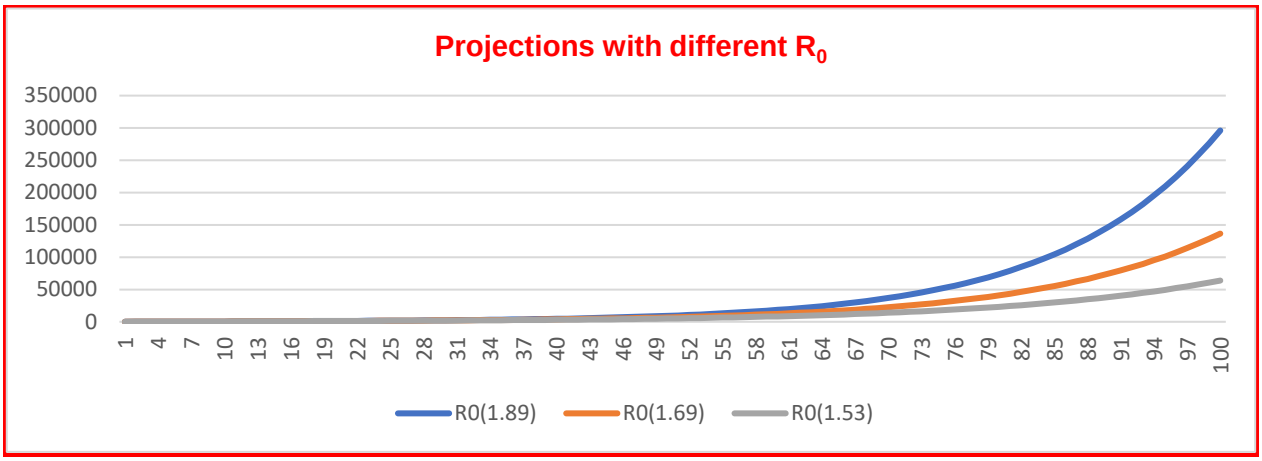
A high number of imported cases would imply a low β as it means most cases are being accounted for through importation. We have across this volunteer site (https://COVID19.pk/) that claims giving numbers on cases for Foreign which we presume are imported. Using this, we run estimates on the values of β and q. There is a large range of values here that they consider to be "unknown".

We allow for the number of incoming cases to be around 800 and run the simulation, assuming Day 43 has 2708 cases with 41 deaths, we get an estimate of $\beta = 0.2$ and testing probability between 0.4 and 0.6. This also implies that actual infection is somewhere around the 4000 to 6,500 range. ***This implies*** $\boldsymbol{R_0 = 1.35 \sim 2.03}$***.*** Earlier estimates would have put this at a higher value, but this might be showing some convergence to a new value after the lockdown precautions.

We must be careful when discussing R_0 It is extremely sensitive to the underlying parameters. Slight change in the transmission, rate, the infection rate or any other parameter can lead to large results

[1] Blackwood, J. C. and L. M. Childs (2018), "An introduction to compartmental modeling for the budding infectious disease modeler", *Journal Letters in Biomathematics*. Volume 5(1), pp: 195-221. https://www.tandfonline.com/doi/full/10.1080/23737867.2018.1509026.

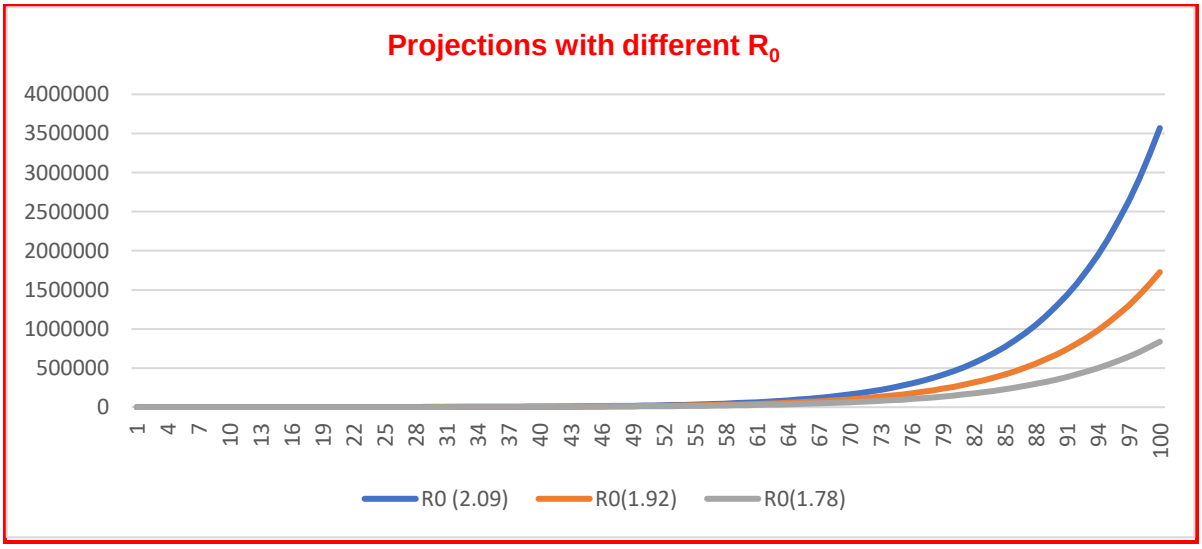
as time goes by. The following chart shows the different courses the infection could be taking over the first 100 days given the different values implied by my results.



As can be seen, the difference is vast at the hundredth day, between around 60,000 total cases and around 300,000 total cases.

On Day 48, we repeat the exercise with 4970 cases and 77 deaths. We get estimates of $\beta = 0.2$ with testing probability ranging between 0.4 and 0.5. Thus, our estimates for R_0 remain in a similar range as before.

If we were to now assume that the number of foreign cases was lower (around 400), then the estimate value of β jumps to 0.3 with a testing/quarantine probability between 0.7 and 0.9. This implies values of R_0 between 1.78 and 2.09. However, since we have now adjusted our assumptions on the sources of the incoming infections, the path will be different.



We see a vast variation within these results between 800,000 and around 3.5 million cases. Basically, since we have reduced the number of imported cases, we are now assuming that more of our cases are indigenous, and given that we have reached 4000 cases, it must be progressing more rapidly than before.

Conclusion

It is important to note that small variations in assumptions can lead to different estimates of R_0. Even more critically, due to the exponential nature of the disease, we can see that small variations in calculating R_0 can lead to very large variations in predictions over the future. Such uncertainty and variations are inevitable in the real world. Something as minor as delays in testing/processing test results could lead to drastically different predictions of the course of the disease.

If we are to seriously estimate R_0, we need to utilise all the data at our disposal, including the number of cases that are local or imported, and the geographic locations of the cases. In terms of testing, we need extensive contact tracing. And apart from that, we need to know what the testing strategy is, so that models and estimates can be adjusted accordingly.

It is possible that specialists in the government are currently taking these factors into account. It would be helpful if they could make this information more accessible.

By:
Amin Hussain
Teaching Fellow, LUMS, & Visiting Fellow, PIDE
with inputs from:
Shaper Mirza
Associate Professor, Department of Biology, LUMS.

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.

PIDE COVID-19 Bulletin Series Editors: Nadeem ul Haque and Durre Nayab

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