Modelling & Analysis of 2019-nCoV in India

A study on Attack & Recovery Rate, Basic & Effective Reproductive Number, Herd Immunity Threshold & Case Fatality Ratio of Coronavirus Covid-19 in India.

Tanishque P. Malik¹

¹Institute & Faculty of Actuaries (IFoA)

Email : maliktanishque@gmail.com

This paper focuses on building a predictive mathematical model in order to understand (and predict) Coronavirus Covid-19 in India.

Abstract

Mathematical modelling of any epidemic plays a crucial role in quantifying the impact of such pathogens. This paper focuses on building a Stochastic SIR Model with non-linear parameters (to account for the effect of lockdowns) to gain a broader cognition of the 2019 novel Coronavirus pathogen (2019-nCoV), widely known as Covid-19, in India. Such models help in gauging the virulence and fecundity of pathogens. Based on early transmission dynamics the basic reproductive number (R_0) is computed to be 1.605. Whereas, effective reproductive number (R_t) is computed to be 4.880 as on 19 March, 2.756 as on 19 April, and 1.995 as on 19 May. Furthermore, the proportion of population that needs to be immunized (through inoculation, recovery, or death) to halt the infection spread is estimated to be 37.69%, ergo, the Herd Immunity Threshold is estimated to be 51.36 crores recoveries, if the R_t remains below 2. R_t is expected to fall below 2, and the Case Fatality Ratio (CFR) to fall to 4.5%, circa early-September (assuming minimal or no medical breakthroughs). The formulated model also provides inferential evidence manifesting the extent to which lockdowns contained the spread of the virus.

Keywords : Covid-19, severe acute respiratory syndrome coronavirus 2, Reproductive Number, Attack & Recovery Rate, Herd Immunity Threshold, Case Fatality Ratio.

Covid-19 in India

The 2019 novel Coronavirus pathogen (2019-nCoV), widely known as Covid-19, is a cause of concern across the world due to its uncontrollable infection rate. This disease is caused by the virus Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). The virus first manifested in Wuhan, China, has travelled across continents. The spike in the number of deaths in USA, Italy, UK and other countries has alarmed both the developing and developed countries.

As of 15 May 2020, more than 40 lakh cases have been reported across 216 countries and territories, resulting in over 3 lakh deaths and over 15 lakh recoveries [\[01\]](#).

Common symptoms include fever, cough, fatigue, shortness of breath, and loss of smell and taste [\[02\]](#). While the majority of cases result in mild symptoms, some progress to an unusual form of acute respiratory distress syndrome (ARDS) likely precipitated by cytokine storm, multi-organ failure, septic shock, and blood clots.

According to the World Health Organization (WHO), currently, there are no available vaccines or specific antiviral treatments for the disease [\[03\]](#). Prevention involves frequent hand-washing, maintaining physical distance from others, coughing into the bend of the elbow, staying indoors whenever sick and wearing a cloth face covering if unable to practice social distancing [\[04\]](#).

WHO declared the Covid-19 outbreak a Public Health Emergency of International Concern (PHEIC) on 30 January 2020, and a pandemic [\[05\]](#) on 11 March 2020 and countries around the world are struggling to contain the surge of confirmed cases. Meanwhile, scientists are exploring potential treatments and are undertaking clinical trials to test new vaccines and therapies for the same. And hospitals are taking necessary actions to ramp up their capabilities to care for increasing numbers of infected patients.

The first case of the Covid-19 in India was reported on 30 January 2020, there are over 1 lakh [\[06\]](#) confirmed cases as on 20 May 2020.

The outbreak was declared an epidemic in numerous states & union territories, where provisions of the Epidemic Diseases Act, 1897 [\[07\]](#) have been invoked. Educational institutions, gyms movie halls, malls and places of social gathering have also been shut down. India has suspended all tourist visas, as a majority of the (initial) confirmed cases were linked to other countries.

On 22 March 2020, India observed a 14-hour voluntary public curfew at the plea of the prime minister Mr. Narendra Modi. The prime minister, on 24 March, ordered a nationwide lockdown for 21 days. Further, On 14 April, the lockdown was extended till 3 May (Lockdown 2.0). On 1 May, lockdown across the country was further extended by two more weeks till 17 May 2020 (Lockdown 3.0).

Currently, there is no vaccine for Covid-19. But the Indian Council of Medical Research has approved a Lopinavir-Ritonavir fixed dose combination (FDC), that was originally prescribed for the treatment of HIV/AIDS, for Covid-19 patients in India [\[08\]](#).

As the number of infected persons, and those recovering is changing differently in different states everyday it is difficult to predict the spread of the virus and its fate in the Indian context. Different states have adopted measures to stop the community spread. Considering the vast size of the country, the population size and other socio-economic conditions of the states, a single uniform policy may not work to contain the disease. In this paper, we discuss a predictive mathematical model that can give us some idea of the fate of the virus, an indicative data and future projections to understand the further course this pandemic can take..

Figure 01 : India 2019-nCoV as on 05 June 2020

Susceptible-Infected-Removed (SIR) Epidemic Model

About the Model

This paper is concerned with the formulation of a Susceptible-Infected-Removed (SIR) Epidemic Model to understand (and predict) the Covid-19 outbreak in India.

The population under consideration is divided into three classes whose constituents change with time t . The *susceptible* class consists of those individuals who can incur the disease but are not yet infective. The *infective* class consists of those who are transmitting the disease to others. The *removed* class consists of those who are removed from the susceptible-infective interaction through recovery or death.

$i(t)$ refers to the newly *infected* persons at any given time t . It is proportionate to number of *susceptible* persons $S(t - 1)$ and the proportion of *newly infected* persons in the population $i(t - 1)$.

The transmission rate comprises of the following components :

- The transmission coefficient (β) is a proportionality constant.
- The overall transmission rate reduces at a reducing rate. A negative exponential term is inscribed to allow for this effect.
- A base transmission rate (β_0) has been inscribed to prevent the transmission rate from going below a certain point.
- The exponential rate at which the infection is spreading is proportionate to the logarithmic scale of $I(t)$. The number of infectives is scaled down to respond to skewness towards its large values. λ represents the proportionality constant.

As mentioned earlier, the net transmission rate decreases at a decreasing rate, ergo, forming a decreasing concave-up curve. This effect is directly analogous to the effect of lockdowns. Simply put, as more persons move to the *infective* class, the more society fears for its safety. This *fear*, of getting infected, discourages such individuals from socializing, thus reducing the effective contact rate.

If $\lambda = 0$, then the effective transmission rate is reduced to $(\beta_0 + \beta)$. This mimics a scenario where no lockdowns have been occurred. The implication of extricating this depreciation would be a steep rise in the number of cases, due to the uncontrolled transmission rate. Thus, λ denotes the extent to which lockdowns contain the spread of the virus.

$r(t)$ refers to the newly *removed* (either through death or recovery) persons, which is proportionate to $I(t - 1)$, at any given time t . $r(t)$ is assumed to follow Poisson distribution.

Here the proportionality constant $(1 - e^{-\gamma})$ refers to the probability of any infected person to recover in a day (where γ is the infections rate). The inverse of this constants returns the duration of infectiousness.

Now, to model the infection process we consider the following equations,

$$S(t) = S(t - 1) - i(t)$$

$$I(t) = I(t - 1) + i(t) - r(t)$$

$$R(t) = R(t - 1) + r(t)$$

$$N = S(t) + I(t) + R(t)$$

Assumptions

$S(t)$, $I(t)$ & $R(t)$ denotes the number of persons in the *susceptible*, *infective* and *removed* class respectively, at any given time t .

In the following assumptions are made :

1. Each class is considered to be a continuous variable due to the large (and assumed constant) population size N . Births & natural deaths have not been accounted for, as India's population growth rate (0.99% annually) would have had negligible effects (for the given time period) on the already massive population.
2. All individuals in the population have the same probability to contract the disease and their age distribution is uniformly distributed between 0 and the life expectancy.
3. The population experiences homogenous interaction. "Interaction" refers to contact that between persons in the population. "Infection" refers to the interaction between a susceptible and infective that results in the transmission of the disease.
4. The pathogen is transmitted at a rate proportional to the number of infectives and susceptible with proportionality constants β & λ . The contact rate is assumed to not vary seasonally.
5. Individuals recover and are removed from the infective class at a rate proportional to the number of infectives with proportionality constant γ , called the daily recovery removal rate. The latent period is zero (it is defined as the period between the time of exposure and the time when infectiousness begins). The recovery rate is assumed to not vary seasonally.

Annotations

- $i(t)$: new infective ($-\Delta S(t)$) persons at any given time t .
- $r(t)$: change in number of removed persons ($\Delta R(t)$) at any given time t .
- N : total population.
- β : transmission coefficient.
- β_0 : base transmission coefficient.
- λ : effects of lockdown on transmission rate.
- γ : recovery rate.
- δ : decrement to case fatality ratio.
- δ_0 : death rate coefficient.

Parameter Estimation

Distribution of $i(t)$

To ascertain the Likelihood, we divide t into h small intervals of length, $\Delta t = \frac{1}{h}$

Thus, the probability of a *susceptible* coming in contact with an *infected* (with infection age τ) in each small interval and getting infected is given by

$$\frac{1}{h} C \frac{i(t-1)}{N}$$

Where,

$$C = (\beta_0 + \frac{\beta}{I(t-1)^\lambda})$$

Annotations

- C : non-linear contact rate, with a negative exponential λ term to adjust for the effect of lockdowns.

Probability of a *susceptible* escaping an infection is given by,

$$(1 - \frac{1}{h} C \frac{i(t-1)}{N})^h$$

$$\Rightarrow \lim_{h \rightarrow \infty} (1 - \frac{1}{h} C \frac{i(t-1)}{N})^h = \exp(-C \frac{i(t-1)}{N})$$

Let $p(t)$ be the probability that the *susceptible* does not escape an infection, i.e., gets infected.

$$\Rightarrow p(t) = 1 - \exp(-C \frac{i(t-1)}{N})$$

$$\Rightarrow p(t) = 1 - \exp(-\frac{i(t-1)}{N} (\beta_0 + \frac{\beta}{I(t-1)^\lambda}))$$

Linear Approximation of $p(t)$

$$\Rightarrow p(t) \approx \frac{i(t-1)}{N} (\beta_0 + \frac{\beta}{I(t-1)^\lambda})$$

Appendix 1

Thus, $i(t)$ can be assumed to follow Binomial Distribution, with parameters,

$$i(t) \sim \text{Binomial}(S(t-1), p(t))$$

$$i(t) = 0, \text{ for all } t < 0.$$

Poisson Approximation of $i(t)$,

$$i(t) \sim \text{Pois}(S(t-1) \times p(t))$$

Appendix 2

So, the expected ΔS can be given by :

$$\Rightarrow E[i(t)] = \frac{S(t-1)}{N} i(t-1) (\beta_0 + \beta e^{-\lambda \ln(I(t-1))})$$

This can further be simplified to

$$E[i(t)] = S(t-1) \frac{i(t-1)}{N} (\beta_0 + \frac{\beta}{I(t-1)^\lambda})$$

It is reasonable straightforward to optimise the parameters by minimising the Mean Absolute Errors (MAE).

Given the multitude of parameters to be computed, this approach is relatively simpler than the Maximum Likelihood approach.

Distribution of $r(t)$

We assume the probability of recovery to be proportional to the concerned subinterval of time.

During any time interval $(t, t + \frac{1}{h})$, an infected Individual is supposed to recover from infection with probability

$$\gamma \left(\frac{1}{h} \right)$$

The probability of an infected person recovering in a day is given by,

$$\lim_{h \rightarrow \infty} \left(1 - \frac{\gamma}{h} \right)^h = e^{-\gamma}$$

$$\Rightarrow q_\gamma = 1 - e^{-\gamma}$$

Through linear approximation we get,

$$\Rightarrow 1 - e^{-\gamma} \approx \gamma$$

$$\Rightarrow q_\gamma \approx \gamma$$

Appendix 1

Thus, $r(t)$ can also be assumed to follow Binomial Distribution :

$$r(t) \sim \text{Binomial}(I(t-1), q_\gamma)$$

So, the expected ΔR can be given by :

$$\Rightarrow E[r(t)] = I(t-1) \times q_\gamma$$

Where,

$$\Rightarrow \text{Infection Period} = \frac{1}{1 - e^{-\gamma}} \approx \frac{1}{\gamma}$$

Poisson Approximation of $r(t)$

$$r(t) \sim \text{Pois}(I(t-1) \times q_\gamma)$$

Appendix 2

Maximum Likelihood Estimate of γ

$$L(\gamma) = \prod_{t=0}^{\hat{t}} P(r(t) / r(1), r(2), \dots, r(t-1))$$

$$\Rightarrow L(\gamma) = \prod_{t=t_0}^{\hat{t}} \frac{\phi^{r(t)} e^{-\phi}}{r(t)!}$$

Where,

- $\phi = I(t-1) \times q_\gamma$
- $q_\gamma = 1 - e^{-\gamma}$
- $\hat{t}$: Time until which observations are available.

Thus,

$$\Rightarrow \hat{\gamma} = -\log \left(1 - \frac{\sum_{t=0}^{\hat{t}} r(t)}{\sum_{t=0}^{\hat{t}} I(t-1)} \right)$$

Simulation of $I(t)$

λ : Extent to which Lockdowns assist in the containment of the virus

Figure 02 : Simulating values of $I(t)$ based on 100%, 105% & 95% of λ

As discussed earlier, as $\lambda \rightarrow 0$ the transmission rate increases, and when $\lambda = 0$, the effective transmission rate is reduced to $\beta_0 + \beta$. This implies that there is a negative relationship between the value of λ and the number of infectives, i.e., the higher the value of λ the faster virus will spread, and more persons would be affected in a shorter period of time.

γ : Rate of Infection

Figure 03 : Simulating values of $I(t)$ based on 100%, 50% & 10% of γ

As, $\gamma \rightarrow 0$, the recovery period also increases, and hence the time it takes for an infected person to recover from the disease increases. This implies that a greater number of persons are infected at any time (as they recover slower too). Thus, we can assume γ to directly affect the longevity of an epidemic. The recovery rate also has an effect on the accumulated number of infectives. Thus, as the recovery period increases, the total number of infected persons at any given time also increases.

β & β_0 : Rate of Transmission

Figure 04 : Simulating values of $I(t)$ based on 100%, 110% & 90% of β

Figure 05 : Simulating values of $I(t)$ based on 100%, 101% & 99% of β_0

As, $\beta \rightarrow 0$, the transmission rate also decreases. A similar effect can be seen in the case of β_0 . A transmission rate of 1.1 for a population with 1000 infectives, implies there to be 1100 infected persons on the very next day. Thus, a reduction implies lesser infectives at any given time. Thus, establishing a positive relationship between effective transition rate and the number of infectives. Activities such as social distancing, wearing face masks etc assist in reducing its value, and hence reducing the totality of infections. Smaller values of β and β_0 enables the delay of the epidemic, and also reduction in the totality of infectives at any given time. The difference is that $I(t)$ more sensitivity to change in β_0 .

Numerical Simulation

$$S(t) = N - \text{confirmed}$$

$$I(t) = \text{confirmed} - \text{recovered} - \text{dead}$$

$$R(t) = \text{recovered} + \text{dead}$$

The Model has been fitted to data^[10] until 19 May 2020. The remainder of the dates (till 05 June, 2020) have been used to check the model's accuracy. The following parameters were used to fit the model to the Cases Data mentioned above.

- $N = 136.26 \text{ crores}$ ^[9] : total population.
- $\beta = 0.43527$: transmission coefficient.
- $\beta_0 = 0.89934$: base transmission coefficient.
- $\lambda = 0.10394$: effects of lockdown on transmission rate.
- $\gamma = 0.05012$: recovery rate.

Comparison

1. $S(t)$: Susceptible

Date	Actual	Estimated
01-03-2020	1,36,25,99,997	1,36,25,99,997
02-03-2020	1,36,25,99,995	1,36,25,99,994
03-03-2020	1,36,25,99,994	1,36,25,99,990
04-03-2020	1,36,25,99,972	1,36,25,99,985
05-03-2020	1,36,25,99,970	1,36,25,99,979
...	...	...
18-04-2020	1,36,25,85,208	1,36,25,86,903
19-04-2020	1,36,25,83,884	1,36,25,85,756
20-04-2020	1,36,25,82,344	1,36,25,84,531
21-04-2020	1,36,25,81,015	1,36,25,83,225
...	...	...
02-06-2020	1,36,24,01,294	1,36,23,99,483
03-06-2020	1,36,23,92,385	1,36,23,90,688
04-06-2020	1,36,23,83,081	1,36,23,81,629
05-06-2020	1,36,23,73,230	1,36,23,72,301

Table 01 : $S(t)$: Susceptible

Figure 06 : $S(t)$: Susceptible

1. $I(t)$: Infective

Date	Actual	Estimated
01-03-2020	3	3
02-03-2020	5	6
03-03-2020	3	10
04-03-2020	25	14
05-03-2020	27	20
...	...	...
18-04-2020	12,289	9,091
19-04-2020	13,295	9,827
20-04-2020	14,255	10,608
21-04-2020	15,122	11,434
...	15,859	12,307
02-06-2020	...	...
03-06-2020	93,322	1,00,806
04-06-2020	97,581	1,04,593
05-06-2020	1,01,497	1,08,461

Table 02 : $I(t)$: Infective

Figure 07 : $I(t)$: Infective

2. $R(t)$: Removed

RMSE = 17.7042

Date	Actual	Estimated
01-03-2020	-	-
02-03-2020	-	-
03-03-2020	3	-
04-03-2020	3	-
05-03-2020	3	1
...	...	...
18-04-2020	2,503	4,006
19-04-2020	2,821	4,416
20-04-2020	3,401	4,861
21-04-2020	3,863	5,341
...	4,612	5,860
02-06-2020	...	...
03-06-2020	97,213	91,177
04-06-2020	1,01,125	95,923
05-06-2020	1,06,118	1,00,851

Table 03 : $R(t)$: Removed

Figure 08 : $R(t)$: Removed

3. $i(t)$: Newly Infected

RMSE = 6.5700

Date	Actual	Estimated
01-03-2020	-	3
02-03-2020	2	4
03-03-2020	1	5
04-03-2020	22	6
05-03-2020	2	7
...	...	...
18-04-2020	957	1,147
19-04-2020	1,324	1,225
20-04-2020	1,540	1,306
21-04-2020	1,329	1,392
...	1,486	1,481
02-06-2020	...	...
03-06-2020	8,392	8,534
04-06-2020	8,171	8,795
05-06-2020	8,909	9,060

Table 04 : $i(t)$: Newly Infected

Figure 09 : $i(t)$: Newly Infected

4. $r(t)$: Newly Removed

RMSE = 7.0476

Date	Actual	Estimated
01-03-2020	-	-
02-03-2020	-	-
03-03-2020	3	-
04-03-2020	-	-
05-03-2020	-	1
...	...	...
18-04-2020	284	410
19-04-2020	318	444
20-04-2020	580	480
21-04-2020	462	519
...	749	559
02-06-2020	...	...
03-06-2020	5,065	4,746
04-06-2020	3,912	4,927
05-06-2020	4,776	5,113

Table 05 : $r(t)$: Newly Removed

Figure 10 : $r(t)$: Newly Removed

Gauging Contagious Infections

Case Fatality Rate (CFR)

Case Fatality Rate (CFR), in epidemiology, refers to the proportion of persons who die from a specified disease amongst all concluded cases. CFR is typically used as a measure to ascertain the disease severity and is often used for prognosis. It also can be used to evaluate the effect of new forms of treatments.

CFR can conceivably be a function of virulence of any disease, and so was used to ascertain the number of deaths in the above model.

Typically,

$$CFR = \frac{Deaths}{Deaths + Recoveries}$$

But, to model the number of deaths caused by Covid-19, the following approximation of CFR is used :

$$CFR_t = \delta_0 e^{-\delta \ln(S(t))}$$

This can further be simplified to

$$CFR_t = \frac{\delta_0}{S(t)^\delta}$$

Since, CFR is not constant, it is a function of time. Seeing the current trend, it is assumed to be decreasing at a decreasing rate. Also, it is assumed to be proportionate to the number of confirmed cases at any given time t .

The aforementioned rate comprises of the following components :

- The base CFR (δ_0) is the death rate coefficient.
- The overall death rate reduces at a reducing rate. A negative exponential term is inscribed to allow for this effect.
- CFR is proportionate to the logarithmic scale of $S(t)$. The number of confirmed cases is scaled down to respond to skewness towards its large values. δ represents the proportionality constant.

The initial empirical values of the CFR are found to be as large as 28.53%. Such variations are a consequence of deficient data points, ergo the substantial sampling errors. As additional data is made available, the accuracy of the estimates is enhanced (assuming no bias). Henceforth, a higher weightage is accorded to the values at later times. The aforementioned reasoning helps in reaffirming this assumption.

Figure 11 : Case Fatality Ratio

Figure 12 : Change in CFR

Estimated values of the parameters :

- $\delta = 0.72128$: decrement to case fatality ratio.
- $\delta_0 = 101.425$: death rate coefficient.

R_0 : The Basic Reproductive Number

The basic reproduction number (R_0) is defined as the average number of secondary cases caused by a single infectious individual in a wholly susceptible population. It is a threshold parameter, intended to quantify the contagiousness of any disease: if $R_0 < 1$, the disease dies out, whereas if $R_0 > 1$, the disease persists. Also, the higher the value of R_0 , the more contagious the infection.

The basic reproductive number is affected by several factors:

- The rate of contacts in the host population.
- The probability of infection being transmitted during contact.
- The duration of infectiousness.

There are multiple methods to estimate the value of R_0 . Two of those theoretics are discussed below,

- The Next Generation Method

Diekmann et al. (1999)

$$\Rightarrow R_0 = \sqrt[\tau]{\prod_{t=1}^{\tau} \frac{i(t)}{r(t)}}$$

Appendix 3

This does not produce a unique result, *Jing Li et al. (2011)*. Furthermore, the next-generation method does not return the number of persons infected by a single human if there is an intermediate host, but rather the geometric mean of the number of infections per generation (R_t). But this method is still widely used due to its simplicity.

- S_∞ Approach

Dietz K. (1993)

$$R_0 = N \frac{\log \left(\frac{S_0}{S_\infty} \right)}{S_0 - S_\infty}$$

Appendix 3

Here, the percentage of the population that eventually gets infected (S_∞) is linked to the basic reproductive number.

The infection spread halts during an epidemic because the replacement number R_t becomes less than one, when $S(t)$ becomes small. However, it is important to note that the final susceptible population S_∞ is not zero.

Based on early transmission dynamics R_0 was computed to be 1.605. Whereas, R_t was computed to be 4.880 as on 19 March, 2.756 as on 19 April, and 1.995 as on 19 May. Hence, an overall depreciating trend can be seen.

The implication of the aforementioned finding is that a person who is infected by the disease will (on average) transmit it to 2 others. This replication will continue if no one has been vaccinated against the disease or is already immune to it (through recovery or death).

Figure 13 : R_t : Effective Reproductive Number

A population is said to have herd immunity for a disease if enough people are immune so that the disease would not spread if it were suddenly introduced somewhere in the population.

Appendix 4

The proportion of population that needs to be immunized (through inoculation, recovery, or death) to halt the infection spread is estimated to be 37.69%, ergo, the Herd Immunity Threshold (*Hethcote, 1989*) is estimated to be 51.36 crore recoveries, if R_t remains below 2.

Recovery Rate (γ)

γ is a biological parameter. So, while estimating such a parameter, preference is given to its biological factors over the statistical ones. This is because the virulence (and fecundity) of any pathogen is predicated on this critical threshold.

Figure 14 : Distribution of Infection Age Across the concluded cases in India

Using Individuals Details Dataset^[11], the average infectious period (*after* being diagnosed) was computed to be 13.23 days.

Two approaches have been used to estimate the value of the recovery rate γ . The first is the Maximum Likelihood Estimate (MLE), the other is the minimisation of the Mean Absolute Error (MAE).

Total duration of infectiousness was computed to be 27.36 days (MLE) and 20.46 days (MAE).

Thus, the difference between the total infection period and the infection period *after* being diagnosed must return the average period of infectiousness *before* being diagnosed.

The incubation period of the virus is the time between the exposure and the display of symptoms. And common-sense dictates that a person would only go to a doctor after he/she starts experiencing the symptoms. And so, the average infectious period (*before* being diagnosed) can be considered as the incubation period.

The incubation period for Covid-19 has been found to be anywhere between 1 to 12 days (with a median duration of 5-7 days) ^[13]. This period is computed to be 14.13 days (MLE) and 7.23 days (MAE).

From this, we infer that γ in the MLE Model has an unrealistic value. Ergo, in the long run, Model MAE returns more accurate results due to the practicality of its estimates.

Furthermore, if an individual remains infected for a longer period of time, there is a higher change of him/her to spread the infection. This leads to a slower growth in the number of persons in the *removed* class as compared to growth of persons in the *infected* class. The implication of this inaccuracy is an initial extenuation of the projections, followed by major overestimations at later times.

The number of confirmed cases as on 5 June 2020 was projected to be 2,14,040 (MLE) and 2,27,699 (MAE). The actual number of confirmed cases on the aforementioned date was 2,26,770. The number of recovered persons was projected to be 83,809 (MLE) and 1,05,017 (MAE). The actual number of recoveries was 1,09,462.

Hence, this outcome reaffirms our belief that prioritising biological factors over statistical ones enhances the accuracy of the model projections.

These results also suggest that :

- Groundless assumptions lead to peculiar predicaments. They may not be necessary in every proposition, but they are necessary in the arrangement of a syllogism. They fail to extricate the flaws of our hypothesis.
- Mindless adhesion to the crude values can also be problematic as this supposition leads to a chain of unprecedented outcomes.

Model Checking

Computing the following values as

$$confirmed = N - S(t)$$

$$active = I(t)$$

$$concluded = R(t)$$

$$deaths = CFR_t \times concluded$$

$$recovered = (1 - CFR_t) \times concluded$$

Comparison

1. Confirmed Cases

Date	Actual	Estimated
01-03-2020	3	3
02-03-2020	5	6
03-03-2020	6	10
04-03-2020	28	15
05-03-2020	30	21
...	...	...
18-04-2020	14,792	13,097
19-04-2020	16,116	14,244
20-04-2020	17,656	15,469
21-04-2020	18,985	16,775
...	...	...
02-06-2020	1,98,706	2,00,517
03-06-2020	2,07,615	2,09,312
04-06-2020	2,16,919	2,18,371
05-06-2020	2,26,770	2,27,699

Table 06 : Confirmed Cases

Figure 15 : Confirmed Cases

2. Active Cases

Date	Actual	Estimated
01-03-2020	3	3
02-03-2020	5	6
03-03-2020	3	10
04-03-2020	25	14
05-03-2020	27	20
...	...	...
18-04-2020	12,289	9,091
19-04-2020	13,295	9,827
20-04-2020	14,255	10,608
21-04-2020	15,122	11,434
...	15,859	12,307
02-06-2020	...	...
03-06-2020	93,322	1,00,806
04-06-2020	97,581	1,04,593
05-06-2020	1,01,497	1,08,461

Table 07 : Active Cases

Figure 16 : Active Cases

3. Concluded Cases

RMSE = 17.7042

Date	Actual	Estimated
01-03-2020	-	-
02-03-2020	-	-
03-03-2020	3	-
04-03-2020	3	-
05-03-2020	3	1
...	...	...
18-04-2020	2,503	4,006
19-04-2020	2,821	4,416
20-04-2020	3,401	4,861
21-04-2020	3,863	5,341
...	4,612	5,860
02-06-2020	...	...
03-06-2020	97,213	91,177
04-06-2020	1,01,125	95,923
05-06-2020	1,06,118	1,00,851

Table 08 : Concluded Cases

Figure 17 : Concluded Cases

4. Recoveries

RMSE = 18.6304

Date	Actual	Estimated
01-03-2020	-	-
02-03-2020	-	-
03-03-2020	3	-
04-03-2020	3	-
05-03-2020	3	1
...	...	...
18-04-2020	2,015	3,553
19-04-2020	2,302	3,927
20-04-2020	2,842	4,332
21-04-2020	3,260	4,771
...	...	...
02-06-2020	95,527	90,204
03-06-2020	1,00,303	94,898
04-06-2020	1,04,107	99,771
05-06-2020	1,09,462	1,04,826

Table 09 : Recoveries

Figure 18 : Recoveries

5. Deaths

RMSE = 3.2498

Date	Actual	Estimated
01-03-2020	-	-
02-03-2020	-	-
03-03-2020	-	-
04-03-2020	-	-
05-03-2020	-	-
...	...	...
18-04-2020	488	453
19-04-2020	519	490
20-04-2020	559	529
21-04-2020	603	570
...	...	...
02-06-2020	5,598	5,719
03-06-2020	5,815	5,953
04-06-2020	6,075	6,192
05-06-2020	6,348	6,438

Table 10 : Deaths

Figure 19 : Deaths

Projected Number of Cases

Date	Confirmed	Active	Concluded	Deaths	Recovered	New Cases
05-06-2020	2,27,699	1,16,434	1,11,265	6,438	1,04,826	9,600
06-06-2020	2,37,299	1,20,539	1,16,759	6,691	1,10,068	9,874
07-06-2020	2,47,173	1,24,722	1,22,451	6,950	1,15,501	10,153
08-06-2020	2,57,326	1,28,983	1,28,343	7,216	1,21,127	10,434
09-06-2020	2,67,759	1,33,320	1,34,439	7,488	1,26,951	10,718
10-06-2020	2,78,477	1,37,734	1,40,744	7,767	1,32,977	11,005
11-06-2020	2,89,483	1,42,222	1,47,260	8,053	1,39,208	11,295
12-06-2020	3,00,778	1,46,785	1,53,993	8,345	1,45,648	11,588
13-06-2020	3,12,366	1,51,421	1,60,945	8,644	1,52,300	11,884
14-06-2020	3,24,250	1,56,130	1,68,119	8,951	1,59,169	12,182
15-06-2020	3,36,431	1,60,910	1,75,521	9,264	1,66,257	12,482
16-06-2020	3,48,914	1,65,761	1,83,153	9,584	1,73,569	12,785
17-06-2020	3,61,699	1,70,681	1,91,018	9,911	1,81,107	13,091
18-06-2020	3,74,790	1,75,669	1,99,120	10,245	1,88,875	13,398
19-06-2020	3,88,188	1,80,725	2,07,463	10,587	1,96,877	13,708
20-06-2020	4,01,896	1,85,846	2,16,050	10,935	2,05,115	14,019
21-06-2020	4,15,915	1,91,031	2,24,884	11,290	2,13,593	14,333
22-06-2020	4,30,248	1,96,280	2,33,968	11,653	2,22,315	14,648
23-06-2020	4,44,896	2,01,591	2,43,305	12,023	2,31,283	14,965
24-06-2020	4,59,862	2,06,962	2,52,900	12,400	2,40,500	15,284
25-06-2020	4,75,146	2,12,392	2,62,753	12,784	2,49,970	15,604
26-06-2020	4,90,750	2,17,880	2,72,870	13,175	2,59,694	15,926
27-06-2020	5,06,676	2,23,424	2,83,251	13,574	2,69,678	16,249
28-06-2020	5,22,925	2,29,023	2,93,901	13,980	2,79,922	16,573
29-06-2020	5,39,498	2,34,675	3,04,822	14,393	2,90,429	16,898
30-06-2020	5,56,396	2,40,379	3,16,017	14,813	3,01,204	17,224
01-07-2020	5,73,620	2,46,132	3,27,488	15,241	3,12,247	17,552
02-07-2020	5,91,172	2,51,934	3,39,238	15,676	3,23,562	17,880
03-07-2020	6,09,052	2,57,783	3,51,269	16,118	3,35,150	18,208
04-07-2020	6,27,260	2,63,677	3,63,583	16,568	3,47,015	18,538
05-07-2020	6,45,798	2,69,614	3,76,184	17,025	3,59,159	18,868
06-07-2020	6,64,665	2,75,593	3,89,072	17,489	3,71,583	19,198
07-07-2020	6,83,863	2,81,612	4,02,251	17,960	3,84,291	19,528
08-07-2020	7,03,392	2,87,670	4,15,722	18,439	3,97,283	19,859
09-07-2020	7,23,251	2,93,764	4,29,487	18,925	4,10,563	20,190

Table 11 : Number of Cases 01 June 2020 to 05 July 2020

Conclusion

Mathematical modelling of any epidemic plays a crucial role in quantifying the impact of such pathogens on public health, control strategies, development of treatments, and other measures taken by the Government. It helps discern the severity of the situation, and disseminate what would transpire (given the current growth rates).

An SIR Model helps in the development of a broader cognizance of the core features of any epidemic.

1. R_0 : Basic Reproductive Number.

A valuable insight derived from an SIR Model is R_0 . The pathogen persistence is predicated on this critical threshold. Airborne diseases have a relatively higher R_0 , in contrast to a contagion transmitted through close contact only. Based on the estimates, it could be established that on average a single person transmits the contagion to 2 others.

2. Inadequate Health Infrastructure.

Inadequate health infrastructure (i.e., medical staff, equipment & medicines, etc) has a major implication on the number of recoveries per day. An SIR Model provides an expedient method to estimate the number of new patients (on any given day). Necessary arrangements (acquisition of additional medication, equipment etc) can be made to alleviate the situation by coping with the welter of newly infected individuals. For instance, the above model suggests there to be 5,56,396 cases confirmed by end of June. Accordingly, necessary preparations can be made by such medical institutions.

3. "Worst-hit" Date

Such a model also helps in ascertaining a tentative date the country would be "worst-hit" by the epidemic. The deferral of this date prolongs the epidemic. This is not necessarily inimical, as it provides researchers and professionals in the field additional time to develop new forms of treatment or vaccine for the pathogen. The model simulated above suggested a delay in the "worst-hit" date. This delay was a consequence of the Lockdown. This can be considered as a major finding, as it basically suggests that the Lockdowns were able to contain the spread of the virus to a great extent, and thus, were a success.

The aforementioned reasoning helps in syllogising the imperative nature of SIR modelling in Epidemiology. Though, it is still not without its flaws.

1. Long-Term Stationary Distribution

Firstly, a major drawback lies in its long-term stationary distribution. Almost all persons who were in originally in the *susceptible* category, end up transitioning to the *removed* category. So, according to this model, after sufficient time has passed, almost the whole population would end up moving to the *removed* class, leaving a small proportion (S_∞) in the *susceptible* class.

This drawback can be adjusted if the theory of *Herd Immunity* is taken into consideration. This approach argues that if a certain proportion of persons move to the *removed* class the pandemic ends as its transmissibility reduces. This proportion is evaluated to be 37.69% of the population of India. Based on these estimates, if 51.36 crore persons were to attain immunity from Covid-19 (either by means of vaccination, recovery or death), only then will the pandemic will come to a halt (assuming the whole population is exposed to the risk of getting infected at all times).

Also, it is estimated there to be approximately 55 lakh cases by early December. Once this critical threshold is achieved a steady decline is observed in the number of new cases. The surge is hindered due to the exceedance of the recoveries over the number of newly infected (in the succeeding days). This abatement is a consequence of the number of persons in the *susceptible* class to fall below a certain number, thus limiting the infection spread.

2. All other things remain constant

Secondly, the SIR model mentioned in this paper works under the precedence – *ceteris paribus*, i.e., "all other things remain constant". This implies that unprecedented occurrences such as the development of new forms of treatments or vaccines, mass gatherings (such as protests, for any apparent reason), or amendments to the rules of the Lockdown (or other forms of Government interventions), cause the actual values to deviate from the estimates of the model. Variables can be placed for the same, but the inclusion of such parameters may overfit the model (and over-complicate it). Regardless of the number of new parameters incorporated, there might still be other external forces that aberrate the predicted numbers. Also, it would be impossible to account for all possible factors, and impractical to actually attempt it.

3. Waves

Furthermore, most outbreaks are characterised by waves of spread. This is caused by the exhibition of seasonality (in the ensuing stages of the epidemic). The lack of seasonality assumption of the SIR Model disregards this characteristic. This more often than not leads to overestimation of results.

4. Simplifying Assumptions

An objection often made to these types of mathematical models is that they involve too many simplifying assumptions, but it would be unwise to reject such models merely on the basis of their simplicity. One such supposition involves uniformly varying the age distribution of its population (between 0 and life expectancy). This extricates the age factor from the transmissibility. This assumption can be debated to be impractical as persons in different age groups interact with other persons of the population at different rates, thus varying the contact rate for each group. The demography of India, 45% of the population lies in the age group of 0-19, 36.3% in the age group 20-44, and remaining (18.7%) above the age of 45 [\[12\]](#), advocates this assertion.

Contrarily, this assumption can be substantiated due to the prolonged continuance of the lockdown. Age as a determinant for transmissibility is extricated as the entirety of the population is expected to stay confined to their homes (regardless of their age).

5. Sampling Error

Furthermore, the data used to estimate parameters only includes the "confirmed cases". Usually whenever a person is considered as an "Active Case", they are required to socially distance themselves from the remaining population. There could be other infectives who have not been reported, and so they're also spreading the infection. This too increases the infection rate. These "hidden" infections may affect the accuracy of the estimates in the long run (due to the accumulation of sampling errors).

6. Effective Transmission Rate

The effective transmission rate is the cumulative of a base rate and a transmission coefficient which is depreciated at a decreasing rate, thus allows for the change in the number of susceptibles ($-\Delta S$) to decrease at a decreasing rate. This is analogous to a scenario where the fear of getting infected increases over time. The theoretical implication of this is the reduction of the effective contact rate due to the "fear". Contrarily, the prolonged lockdown has led to unrest amongst the people. This "impatience" is resulting in people breaking the rules of the lockdown (and ultimately increasing the progression rate). This effect is contrary to the "fear" assumption. Also, it was assumed that as the number of infected persons increases the lockdowns will get stricter thus limiting the contact between persons. This does not seem to be the case, as it can be seen Lockdown 3.0 has relatively eased rules in comparison to Lockdown 2.0. The net effect deviates the projections from the actual values, thus exposing a flaw in the underlying hypothesis.

To extricate this flaw, another parameter could be incorporated to be analogous to this increasing "impatience". But the quantification and estimation of such a parameter would be cumbersome. Also, it would be almost impossible to verify its authenticity, given the nature of the factor (human behaviour) affecting the value of the parameter. This enforces the belief that simplified models perform better in the long run.

Conclusively, majority of the model predicaments have been extricated. This was done so by justifying the welter of parameters and substantiating the suppositions made.

Appendix

1. Linear Approximation

For, $x \ll 1$

$$\begin{aligned}1 - e^{-x} &= 1 - \left(1 - \frac{x^1}{1!} + \frac{x^2}{2!} - \frac{x^3}{3!} + \dots \infty\right) \\&\Rightarrow 1 - e^{-x} = x + \text{hots} \\&\Rightarrow 1 - e^{-x} \approx x\end{aligned}$$

Since, the higher order terms (hots) ($x^2, x^3, \dots$) ≈ 0

2. Poisson Approximation

For any random variable X, following Binomial Distribution,

$$X \sim \text{Binomial}(n, p)$$

We can assume X to approximately follow Poisson distribution if n is sufficiently large and p is very small.

$$X \sim \text{Poisson}(np)$$

3. R_0 : Basic Reproductive Number

Let, σ denote the Average Contact Rate.

Hethcote (1989), If the initial replacement number is greater than one, then an epidemic occurs since the prevalence (the infective fraction) increases to a peak and then decreases to zero. Otherwise, there is no epidemic since the prevalence decreases to zero.

$$\text{Let, } S(0) = S_0 > 0 ; I(0) = I_0 > 0$$

$$\text{If } \frac{\sigma S_0}{N} \leq 1$$

Then, as $t \rightarrow \infty$, $I(t)$ decreases to 0.

$$\text{If } \frac{\sigma S_0}{N} > 1$$

Then, as $t \rightarrow \infty$, $I(t)$ initially rises to a maximum value (I_{max}), then then decreases to 0. Where,

$$I_{max} = 1 - R_0 - \left(\frac{1 + \log\left(\frac{\sigma S_0}{S_\infty}\right)}{\sigma}\right)$$

The susceptible fraction $S(t)$ is a decreasing function and the limiting value S_∞ is the unique root in $(0, \frac{1}{\sigma})$ of the equation :

$$1 - R_0 - \frac{S_\infty}{N} + \frac{\log\left(\frac{S_\infty}{S_0}\right)}{\sigma} = 0$$

The infection spread stops during an epidemic because the replacement number $S(t)$ becomes less than one when $S(t)$ becomes small. However, the final susceptible population S_∞ is not zero.

An epidemic enters a population as one or very few cases, $I(0)$ is negligibly small so $S(0) = N - R(0)$

Then the above equation can be solved for σ to obtain :

$$\sigma = N \frac{\log(\frac{S_0}{S_\infty})}{S_0 - S_\infty}$$

Dietz K. (1993), under the S_∞ Approach :

$$R_0 = N \frac{\log(\frac{S_0}{S_\infty})}{S_0 - S_\infty}$$

Diekmann et al. (1999), under the Next Generation Method :

$$\frac{dI}{dt} = i(t) - r(t)$$

$$\text{Let, } \frac{dI}{dt} = 0$$

$$\Rightarrow R_t = \frac{i(t)}{r(t)} = \frac{-\Delta S(t)}{\Delta R(t)}$$

R_0 is the geometric mean of R_t

$$\Rightarrow R_0 = \sqrt[t]{\prod_{t=1}^t \frac{i(t)}{r(t)}}$$

4. Herd Immunity & Vaccination

A population is said to have herd immunity for a disease if enough people are immune so that the disease would not spread if it were suddenly introduced somewhere in the population. If the population is homogeneously mixing and the immune people are distributed uniformly in the population, then herd immunity will be obtained if a large enough uniformly distributed fraction is immune. In order to prevent the spread of infection from an infective, enough people must be immune so that the replacement number must be less than 1. That is, the susceptible fraction must be small enough so that the average infective infects less than one person during the infectious period.

Herd immunity in a population is achieved by vaccination of susceptibles in the population. If H is the fraction of population which is immune due to vaccination or recovery.

Since,

$$S = 1 - R \quad \& \quad I = 0$$

Herd Immunity is achieved by

$$H > 1 - 1/R_0$$

Hethcote (1989)

For example, if the contact number (R_0) is 5, at least 80% must be "Removed" to achieve herd immunity.

Also,

$$R_0 = N \frac{\log(\frac{S_0}{S_\infty})}{S_0 - S_\infty}$$

$$H = 1 - \frac{1}{R_0}$$

This, implies that at least H percent of the population must be immune (either through vaccination, recovery or death) for the pandemic to end.

References

- [01] Hethcote H.W. (1989). *Three Basic Epidemiological Models*. In: Levin S.A., Hallam T.G., Gross L.J. (eds) Applied Mathematical Ecology. Biomathematics, vol 18. Springer, Berlin, Heidelberg.
- [02] Changguo Li (2018). *Parameter Estimation on a Stochastic SIR Model with Media Coverage*. Yongzhen Pei, Meixia Zhu, Yue Deng Hindawi, Discrete Dynamics in Nature and Society, Article ID 3187807.
- [03] Dietz K. (1993). *The estimation of the basic reproduction number for infectious diseases*. Statistical Methods in Medical Research.
- [04] Diekmann, O. and H. Heesterbeek (2000). *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. New York: Wiley.
- [05] R. Fierro, V. Leiva, and N. Balakrishnan (2015). *Statistical Inference on a Stochastic Epidemic Model*. Communications in Statistics—Simulation and Computation, vol. 44, no. 9.
- [06] Hethcote, H. W. (2000). *The Mathematics of Infectious Diseases*. SIAM Review.
- [07] Kermack WO, McKendrick AG (1927). *A Contribution to the Mathematical Theory of Epidemics*. Proc. Roy. Soc.
- [08] Begon, M., Bennett, M., Bowers, R., French, N., Hazel, S., & Turner, J. (2002). *A Clarification of Transmission Terms in Host-Microparasite Models: Numbers, Densities & Areas*. Epidemiology & Infection.
- [09] Wu JT, Leung K, Leung GM. (2020). *Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China*. Lancet.
- [10] Li, J., Blakeley, D., & Smith, R. J. (2011). *The failure of R0*. Computational and mathematical methods in medicine, 527610.
- [11] Frank, S. A. (1996). *Models of parasite virulence*. Quarterly Review of Biology 71 (1).
- [12] Murthy S, Gomersall CD, Fowler RA (2020). *Care for Critically Ill Patients With COVID-19*. JAMA Insights.
- [13] Niels Becker (1979). *The Uses of Epidemic Models*. Vol. 35, No. 1, Perspectives in Biometry, pp. 295-305. International Biometric Society.
- [14] Coburn, B.J., Wagner, B.G. & Blower, S. (2009). *Modelling influenza epidemics and pandemics: insights into the future of swine flu (H1N1)*. BMC Med 7.