COVID-19 PROJECTIONS

An alternative approach to modelling COVID-19 outbreaks for decision making by Governments

For the Govt. of Punjab Dated 06.06.2020

ABSTRACT

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has reached all continents of the world and is spreading rapidly in many regions. There is a growing need to understand and analyse the nature of the virus and recent trends in the rate and extent of its spread, symptoms, and mortality, to predict its future course and help plan necessary interventions.

To build an *accurate* model and provide insights, we attempt a comprehensive approach by combining the method of using Infection Fatality Rate to estimate extent of spread with trend analysis of recent detection data. We make four key observations:

1. Infection Fatality can be assumed constant for a given demographic for a large part of the initial period in the epidemic cycle of a disease
2. Rate and extent of spread is correlated with population density, government interventions and community behaviour – a bottom up approach to modelling gives a more accurate picture of spread than modelling for the entire state or nation
3. Examining deaths and hospitalizations provides a better estimate of the infection spread than detections
4. The first epidemic cycle is likely to peak when a certain share of the population has been infected

Based on these observations, we build a model to project infections, hospitalizations, and detections. We estimate the hospitalization, ICU, and ventilator requirements as on any day by cumulating requirements over rolling periods.

We arrive at ~2.8 million infections (~9% of the population or ~22% of the urban population) in Punjab by the end of the first epidemic cycle in September 2020. We assume that the epidemic cycle would be completed only in urban areas and that rural areas would see negligible spread. At the current detection rate, this would correspond to ~1,10,000+ detections and would translate to an infrastructure requirement of 11,200 hospital beds, 4,500 Oxygen beds, and 2,000 ICU beds at the peak which is estimated to be in the 3rd week of August. We plan to open-source both our model and raw data in due course.

TABLE OF CONTENTS

INTRODUCTION

The novel corona virus was identified as a cause of upper and lower respiratory tract infections in Wuhan (China) in December 2019. Since then cases have been reported in all continents and have been rising steadily in many countries. As per data from Johns Hopkins University[1], over 67.5 lakh cases had been detected across 188 countries as on 05th June, with the virus estimated to have caused more than 3.95 lakh deaths.

The epidemic appears to have progressed to varying degrees across geographies; there are broadly three categories of regions we see so far:

1. Regions which have witnessed a significant spike and subsequent tapering off – where we can assume the first epidemic cycle has passed / will soon pass
   E.g., New York City (USA), Lombardy (Italy), Madrid (Spain), England, etc.

2. Regions which witnessed an initial / few spike(s) but have been able to flatten the mortality curve and overall have lower deaths per million population so far
   E.g., Daegu (South Korea), Singapore, Qatar etc.

3. Regions which are currently witnessing a rising number of detected cases and deaths – which are in the early stages of the epidemic
   E.g., Various parts of India, Russia, Brazil, Mexico, etc.

A direct observation of the epidemic curve for these geographies throws up significant variations in patterns and growth rates as shown below:

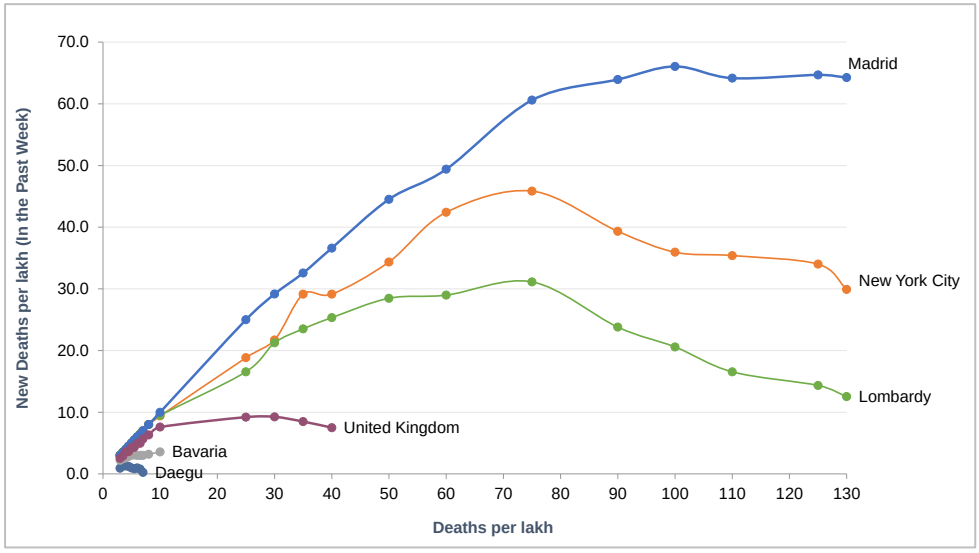


FIGURE 1: EPIDEMIC CYCLE PLOT FOR KEY GEOGRAPHIES [2,3,4,5,6,7]

While several studies have attempted to model this pandemic, we believe a comprehensive approach accounting for the nature of the disease and the behaviour of the peak is lacking. This study attempts

to bridge this gap by using statistical insights gleaned from regions where the first epidemic cycle has peaked and nearly completed, to build a practical model for India.

METHODOLOGY

Based on broad consensus among researchers and policy makers about the nature of SARS-CoV-2, we have made the following underlying assumptions:

1. SARS-CoV-2 has high infectivity[8]
2. Incubation period varies from 2 to 14 days[9]
3. A large share of those who are infected are asymptomatic[10]
4. Only a subset of those infected will be detected – large variations in detection rates may exist based on virus spread and the extent of testing done (Refer to Appendix 1)
5. Growth in infections for any period depends on:
    I. Infectivity of the virus – in the absence of strong evidence of infectivity of different strains of the virus, we assume that this is uniform
    II. Interactions between Infected and Susceptible (likely to be infected) – which further depend on:
        i. Government Interventions (like Lockdowns, Contact tracing and isolation)
        ii. Community Behaviour (compliance with lockdown, usage of masks, etc.)
        iii. Susceptible population size relative to Active Infected population
        iv. Population Density

Given the dynamic and inter-dependent nature of the abovementioned parameters, we base our modelling exercise on projecting the number of infections, using an estimated Infection Fatality Rate (IFR) and statistical trends observed in detections and deaths, adjusted for the impact of demographic and population density.

The SARS-CoV-2 virus started to spread in India only around late March and remains in its early stages in most parts of India. This allows us to study and apply learnings from the experiences of other regions globally, where the first pandemic cycle appears to have completed or is close to completion.

A summary of the insights from this examination are given below and elaborated subsequently:

1. Infection Fatality Ratio can be assumed constant for a given demographic
2. The pandemic has spread rapidly in population clusters, with the rate and extent of spread showing correlation with population density, Government interventions, and community behaviour (including compliance with social distancing)
3. Deaths and hospitalizations provide a better estimate of the true infection spread than detections due to wide variations in detection rates as well as testing regimes
4. The first epidemic cycle of the pandemic will complete once a finite share of population has been infected
5. Estimated Infection Fatality Rate (IFR) for India can be a reliable tool to estimate trends for India based on statistical trends from global cities

6. Hospitalization requirements will need to be matched against Infection trends and not Detection trends; they also need to be corrected for demographic (age structure) differences in population cohorts

Two snapshots of the only postulate used and methodology adopted for modelling is given below:

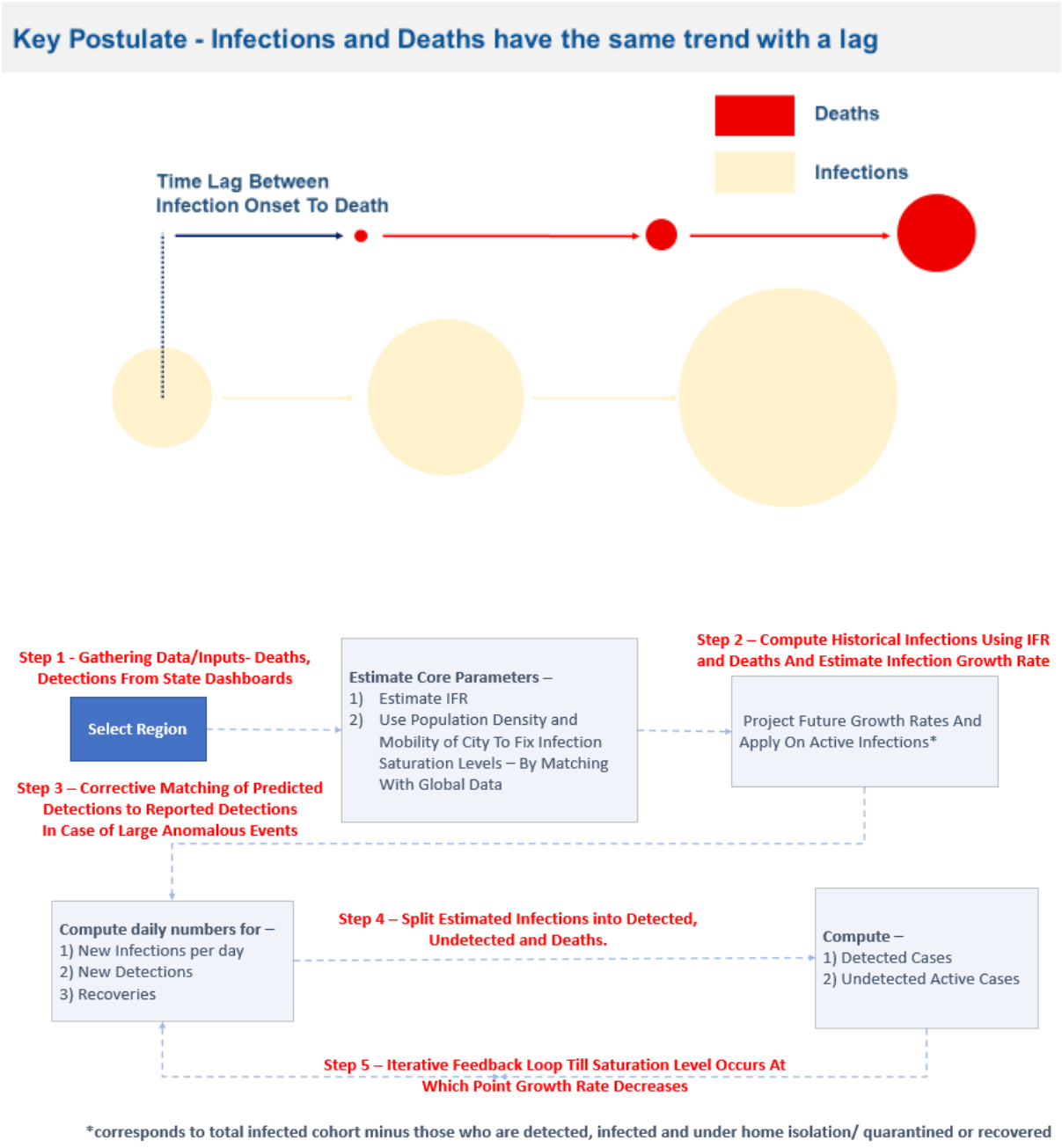


FIGURE 2: KEY POSTULATE AND FLOWCHART OF THE MODEL

Estimating Infection Fatality Rate (IFR) for India as on June 2020

We estimate the IFR for India to be between 0.11% and 0.25% (*0.15% assumed for modelling purposes*). We have arrived at this value by reviewing latest available literature on mortality trends by age-cohort for various countries and by applying five different methods:

#	METHOD	ESTIMATED IFR
1	Using estimates from Verity et. al. and Gangelt Based Serological Studies[11,12]	0.15%
2	Using estimates from Korean CDC and Time Correlated Study of the Same to Estimate IFR[13]	0.12-0.19%
3	Using age and censorship adjusted CFR numbers from Chinese Data[8]	0.12%
4	Using age-adjusted CFRs from Gangelt Data Directly	0.15%
5	Using CFR numbers of countries with high detection regimes like Qatar, Singapore and Iceland	0.15-0.5%

For details on the IFR calculations for each method, please refer to Appendix 2.

Using IFR to find Infections and estimating Compound Growth Rates (CGRs)

Once we have an estimate of the IFR, we assume a starting point of actual detections as on a day 'T'. Average time taken to detect the infection from the day of contact is 10 days. We use IFR to compute the detection rate and apply this detection rate on actual detections as on day 'T', to compute Cumulative Infections as on the day 'T-10'. Using historical trends of the detected cases, we derive a Compound Growth Rate (CGR) which is adjusted and applied on the cumulative active infections as on 'T-10'.

We define Active Infections as Total Estimated Infections Less Total Detections till date multiplied by an Isolation multiple. The isolation multiple serves to attenuate the impact of better detection, contact tracing and isolation regimes wherein we observe that for every positive case, a group of primary and secondary contacts are typically isolated and quarantined for 14 days. We assume this multiple to be 2 conservatively, because at least one out of these primary contacts, on an average, would also be infected (even though they may not be tested) and thus, the effect of detecting one positive case would be to keep nearly two cases outside the Active Infections cohort.

An illustration of this exercise is given below:

Date	Detections	Actual Deaths	Implied Detection ratio	Implied Infections	CGR of Infections	CGR of Deaths
24 Apr	298	17	2.6%	11,461		
10 Apr	151	7	3.2%	4,719	**6.5%**	**6.5%**

Projecting Compound Growth Rates (CGRs) going forward

Going forward, the entire projection period is split into four distinct time periods:

1. Pre-lockdown period
2. Lockdown 1.0 (with strict restrictions),
3. Lockdown 2.0 & 3.0 (with fewer restrictions)
4. Post lockdown period

The infection growth rates for the Pre-lockdown and Lockdown 1.0 period are taken based on actual estimates from the historical detection data available. The growth rate for the most recent period is optimized periodically (usually a week) based on mortality, detections and variations in the detection rate to ensure accuracy. Lockdown 2.0 & 3.0 are combined as the restrictions put were of similar nature. Lockdown 4.0 is not considered in the model as the growth rate increases exponentially from Lockdown 2.0 & 3.0 to post lockdown period thereby effectively factoring in the Lockdown 4.0 by itself.

Determining Total Infections

We observe that the total estimated infections during the first epidemic cycle is likely to be between 20%-60% of the population based on population density. Refer to Appendix 3 for details on these assumptions. For this model we assume:

1. Total infections @ 50-60% of the population - For very high population density regions like Mumbai and Kolkata (on the lines of New York City and Madrid)

2. Total infections @ 30% of the population - For other urban areas / cities in the country (on the lines of Lombardy); also applicable for urban Punjab

3. Total infections @ 10% of the population for rural areas – however, we have assumed that the state can control spread to the villages and hence in the current projections have not assumed significant spread to the rural areas

Estimating the Peak and Deceleration Rates

An examination of the statistical trends across the globe allows us to estimate the shape of the infections curve by defining *Saturation Levels / Points* on the curve[14]:

1. **First Saturation Level:**

   Corresponds to an inflection point on the curve, where infection growth rates start declining – in theory this would correspond to the point where the susceptible population or the universe of interactions between active infections and susceptible cohorts falls below a certain minimum threshold.

We define this point to be:

a. 22.5% of total infections: For highest population density regions like Mumbai and Kolkata
b. 20% of total infections: For other urban areas / cities in the country
c. 12.5% of total infections: For rural areas

Once this point is reached, we assume a gradual decline in CGRs.

2. **Infections Peak:**

Corresponds to approximately 50-70% of total infections as observed for New York City (USA), Lombardy (Italy) and Madrid (Spain) (refer Appendix 3 for details). A deceleration rate of approximately 94-97% once the first saturation level is reached results in this peak being attained i.e. once the first saturation level is reached, subsequent CGRs are attenuated by 3-6% daily.

3. **Second Saturation Level:**

We observe that for smaller clusters i.e. densely populated cities, the decline of the curve is likely to be more rapid than the ascent for the first epidemic cycle. Accordingly, we accelerate the attenuation once a second saturation level of approximately 75% of the total estimated infections is attained.

Estimating Detections and Hospitalizations

For every projection date, we estimate the number of detections by applying the detection ratio on total estimated infections 10 days prior to the required date. The hospitalization, ICU and ventilator requirements on any day are projected by cumulating requirements over rolling time periods, using the following assumptions:

1. Hospitalization requirement is around 2% of infected population (refer to Appendix 3 for a derivation of this rate from global benchmarks) – this implies a hospitalization demand per detection rate of 20% to 40%, given a detection rate of 10% to 5%.

2. Hospitalizations if warranted take place on the same day as detections

3. For every detection [9,15,16,17,18]:
   a. 2% of the infections, wherein infection started 10 days ago will require hospitalization
   b. Of those requiring hospitalisation, ~ 40% will require oxygen support at some point
   c. Of those requiring hospitalization, 20% would require ICU stay
   d. Of those requiring ICU stay, around 50% would require ventilator support

4. The average length of stay (ALOS) for each category of patient is assumed as:

Category of Patient	Ward Bed ALOS	ICU Bed ALOS	Ventilator ALOS	Total ALOS
Requiring only ward stay; potentially oxygen	10	-	-	10
Requiring non-ventilated ICU care	2	8	-	10
Requiring ventilated ICU care	2	2	6	10

Limitations of the model

The projections are intended to serve as indicators on probable detections, hospitalizations, and deaths within a defined timeframe and are to be used to design strategies and take tactical decisions around infrastructure planning, testing and policy making. However, it is important to acknowledge that the growth in infections as well as detections are governed by multiple unpredictable factors and population behaviour.

This model has the following limitations:

1. It assumes that the detection rate remains constant throughout the epidemic – experiences of New York City (USA), Lombardy (Italy), Madrid (Spain), etc. suggests that once the epidemic cycle crosses a certain threshold, detection rates typically start falling given limitations on testing and tracing infrastructure

2. Excludes the possibilities of super spreader events which could speed up the epidemic depending on the number, geographical distribution and duration of such events

3. Should not be used for highly accurate day-by-day forecasts of detections and deaths

4. Does not factor in the specific impact of other interventions like mandatory use of masks, actions to protect the elderly (can potentially reduce hospitalization requirements by 60-70%), etc.; however, the model has the flexibility to handle changes in infectivity driven by these actions provided reasonable assumptions for these can be derived based on actual evidence

COMPARISON WITH OTHER MODELS

We provide a brief comparison of our model's methodology vis-à-vis other models in the table below:

	Current Model	ISB	Johns Hopkins (CCDEP)[19]	University of Michigan[20]
Cases	4,500-5,500 by Jun 30 1,10,000+ by Sep 30	3,50,000-20,00,000	N.A.	1,650-12,500 by May 31
Peak date	3rd Week of August	Long term peak in 2021	Only Provides Peak for India	Only Shows rise
Hospitalization needs	55,000 by Sep 30	3,000 by Aug-Nov 2020	1,40,000 (In Punjab @ 5% infections)	N.A.
State model basis	Bottom-up summation of urban and rural areas	State level directly (Treats Whole State as One Entity)	State level directly (Treats Whole State as One Entity)	State level directly (Treats Whole State as One Entity)
Population density	Accounted for	It is not accounted for here	It is not accounted for here	It is not accounted for here
Modelling basis	SIR* model with additional delay for 'exposed' factor Infection based modelling; detection lag accounted for	SEIR** modelling Infection based modelling; detection lag accounted for	Undisclosed; No detection lag considered	SIR* with additional delay for 'exposed' factor Detection based modelling
Infection Modelling	Calculates detection rate from IFR but is tuneable	Calculates detection rate from symptomatic and asymptomatic breakup	Assumes zero undetected cases (i.e. Infections = Detections)	Assumes zero undetected cases (i.e. Infections = Detections)
Mortality rates	IFR of 0.15% used; Computed using multiple methods including latest research using antibody studies	IFR of 0.67% used; old Imperial study based (Chinese IFR)	Does not differentiate between CFR & IFR	Does not differentiate between CFR & IFR
Peak modelling	Based on study of New York, Lombardy, – 50-70% at peak	Peak not factored in in projection period	Peak not mentioned in graph	Projected for short term hence, peak not factored

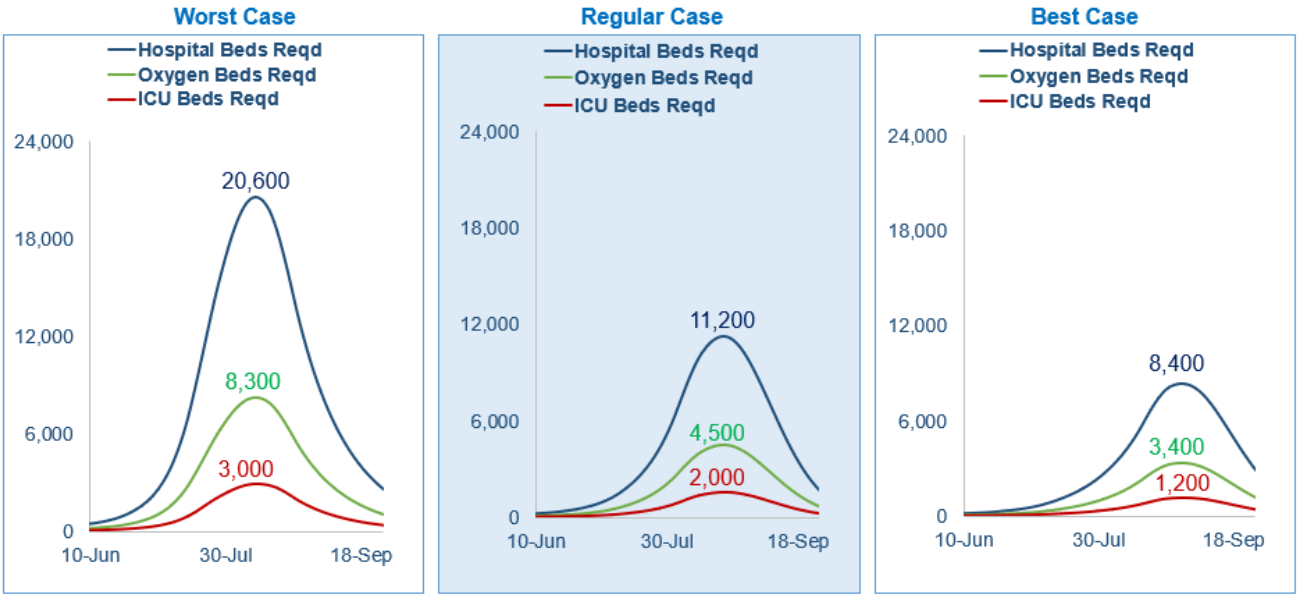
PROJECTIONS – PUNJAB

In Punjab, the first case was detected on 17th March. However, the start date 'T' has been considered as 30th March which had 41 detected cases. A Compound Growth Rate (CGR) of 13% is applied on the infections from 20th March ('T-10') to 1st April (Pre-lockdown Period). Thereafter, a lower CGR of 5.5% is considered factoring in the effective implementation of the lockdown by the state and cumulative detected cases which were restricted to 245 by 20th April. The post-lockdown growth rate is estimated at 10.5% to project active undetected infections. Given Punjab's urban population density is less than ~10,000 per sq. km, we assume the first epidemic cycle will result in a maximum of 30% of the total urban population being infected.

We observe that the Punjab reaches an infection population of ~9% (~22% of urban population) by 30th September which translates to ~2.8 million infections. This implies ~89,100 infections per million population (~222,500 per million urban population) and with the current detection rate, means ~3,550 detections per million population or ~8,860 detections per million urban population.

In terms of infrastructure requirements, the regular case scenario suggests that the state will require close to 11,200 hospital beds during the peak which is estimated to be in the 3rd week of August. The requirement for oxygen and ICU beds is estimated to be around 4,500 and 2,000, respectively.

However, in the pessimistic scenario the infra requirement for hospital beds, oxygen beds and ICU beds grow to 20600, 8300 and 3000, respectively. In case, proper mitigation steps are taken to stop the spread of this pandemic, the requirement can be curtailed to 8400, 3400 and 1200 for the hospital, oxygen and ICU beds.



The likely requirement of beds for key districts in the state are given below:

Regions	Oxygen Beds	ICU Beds	Ventilators
Amritsar	700-1,200	300-500	150-200
Jalandhar	700-1,200	300-500	150-200
Patiala	450-700	175-300	90-140
Ludhiana	1,250-2,000	500-900	250-400
Pathankot	100-150	40-70	20-40
Bathinda	350-550	120-240	60-100

MONITORING THE RIGHT METRICS

We believe the state should revamp its overarching strategy and policies to prioritize improved detections and lowering mortality. In order to do this, we propose the following framework for monitoring micro-zones:

- Measures to ensure accurate reporting and monitoring of COVID mortalities:
  - Classification of all SARI and ILI patient deaths as COVID likely – only if both RT-PCR and Antibody tests for such patients are negative should they be treated as non-COVID deaths
  - Testing of all other patient deaths which may be suspected to be from COVID
  - Classification of all COVID positive confirmed patient deaths as COVID deaths

- Measures to ensure tracking of the right metrics for every suspect / COVID patient:
  - Recording of date and time of symptoms
  - Recording of date and time of detection / testing
  - Recording of duration of hospitalization stay – in Ward (without O_2), in Ward (with O_2), in ICU (without Ventilator), in ICU (with Ventilator)
  - Recording of date and time of death

- Measures to ensure a minimum scale of testing is being undertaken:
  - At least 1% of high contact personnel (healthcare workers, delivery boys, grocers and vegetable vendors, pharmacists) are being tested every week
  - At least 80% of the test volumes undertaken the prior week are undertaken every week
  - At least 40% of the tests undertaken are on high risk cohorts (symptomatic patients, age above 55 years or with co-morbidities)

- Revamped Monitoring system for all micro-zones (*States, Districts, Talukas, Wards, Villages*) based on:

Sr. No.	Metrics	Desired trend	Optimum value	Acceptable Value
1	Average time from symptoms to detection	⇩	< 1 day	< 2 days
2	Average time from detection to death	⇧	> 6 days	> 4 days
3	% share of asymptomatic or mild symptom cases in active cases	⇧	> 90%	> 80%
4	Case Fatality Rate	⇩	< 1%	< 2%

OPENING UP STRATEGY

Several states have begun opening up their economies. It is critical to ensure that each region is prepared to tackle the pandemic in case of any acceleration of the outbreak. The state needs to be confident that it is capable of clamping down and managing the epidemic without stress in its healthcare infrastructure.

We recommend the following minimum criteria before opening up any district, city or part of the state:

Sr. No.	Parameters	Required Values
1	Testing Capacity per 1,000 Urban Pop'n	3-5 tests
2	Oxygen Bed Capacity per 1,000 Urban Pop'n	0.5-1 beds
3	ICU Bed Capacity per 1,000 Urban Pop'n	0.2 – 0.4 beds
4	Average time from detection to death in last month	> 7 days
5	PPE per Hospital Bed per day	2-3 PPE per day
6	Monitoring metrics	Above Acceptable Values

A region can open up without building up the above infrastructure only if it is estimated to –

a) Be COVID free i.e. there is no new infections in the past 21 days

OR

b) Cross the infection peak i.e. {(Total Deaths/IFR)/Population} > 20-40% (depending on the population density)

CONCLUSION

Several studies have attempted to provide an understanding of the SARS-CoV2 pandemic. These largely belong one of two camps (i) those trying to predict detections based on current and past detection trends and (ii) those trying to estimate the extent of spread after the first epidemic cycle has passed, by determining under ascertainment in testing using an estimated Infection Fatality Rate and detection data.

We attempt a comprehensive approach of modelling the pandemic in India by combining these approaches and accounting for government interventions. We used the broad consensus among researchers about the nature of SARS-CoV-2 such as infectivity, incubation period, and large number of asymptomatic cases as assumptions, listed out below:

1. Infection Fatality can be assumed constant for a given demographic for a large part of the initial period in the epidemic cycle of a disease

2. Rate and extent of spread is correlated with population density, government interventions and community behaviour

3. Examining deaths and hospitalizations provides a better estimate of the infection spread than detections

4. The first pandemic cycle is likely to peak when a certain share of the population has been infected

We used the resulting model to project infections, hospitalizations, and detections. We estimate the hospitalization, oxygen beds, and ICU requirements as on any day by cumulating requirements over rolling periods.

We arrive at the key conclusions mentioned below:

1. ~2.8 million infections (~9% of the population; ~22% of the urban population) projected in Punjab by the end of the first epidemic cycle in September 2020

2. We assume that the epidemic cycle would be completed only in urban areas and that rural areas would see negligible spread

3. This implies ~89,100 infections per million population (~222,500 infections per million urban population) and with the current detection rate, means ~3,550 detections per million population or ~8,860 per million urban population (total of 110,000+ detections)

4. This suggests a peak infrastructure requirement of 11,200 hospital beds, 4,500 oxygen beds and 2,000 ICU beds at the peak which is estimated to be in the 3rd week of August

We will release our model online in due course. We hope that the findings in this work and our model will inspire new research directions, ideas, and rigorous and more accurate models to understand the SARS-CoV-2 virus. We also hope that these forecasts assist in scientific planning of interventions and infrastructure

TEAM PROFILES

A group of enthusiastic individuals, with varied experience, have worked together on this report and modelling exercise. The core team for this exercise consisted of:

1. **Akshay Ravi**

   Associate Principal @ Areté Advisors | IIM Ahmedabad

2. **Lakshya Sharma**

   Associate @ Invest India | BITS Pilani

3. **Nipun Sawhney**

   Ph.D. @ University of Cambridge | IIT Roorkee

4. **Raghav Gupta**

   Investment Specialist @ Invest India | Ex Qualcomm | Carnegie Mellon

5. **Vipul Kedia**

   Chief Data Officer @ Affle | Ex-IBM | IIM Ahmedabad | IIIT Hyderabad

6. **Vivek Gala**

   Associate Consultant @ Areté Advisors | IIM Lucknow | ICAI

We thank our friends Abhijit Basak, Ravishankar Iyer and Satyam Mehra for guiding our thought. We are also thankful to Dr. Arun Singh, Dr. Ajay Bakshi, Dr Anurag Agrawal, Dr. John Jacob, Dr. Venkata Raghava and Dr. Sujata Baveja for their feedback and guidance. The team has also been actively supported by volunteers and experts, without whose support, the finalization of the model would not have been possible. We acknowledge especially the support of Nitin M. Pai, Prajval Somani, Rithika Ardeshir, Shashank Mudlapur, Shrenik Bohra and Srishti Gupta.

We would also like to acknowledge the voluntary contribution of Areté Advisors LLP to this exercise, to which they lent their resources, research data, experience, and expertise.

APPENDIX

Appendix 1: The COVID Universe

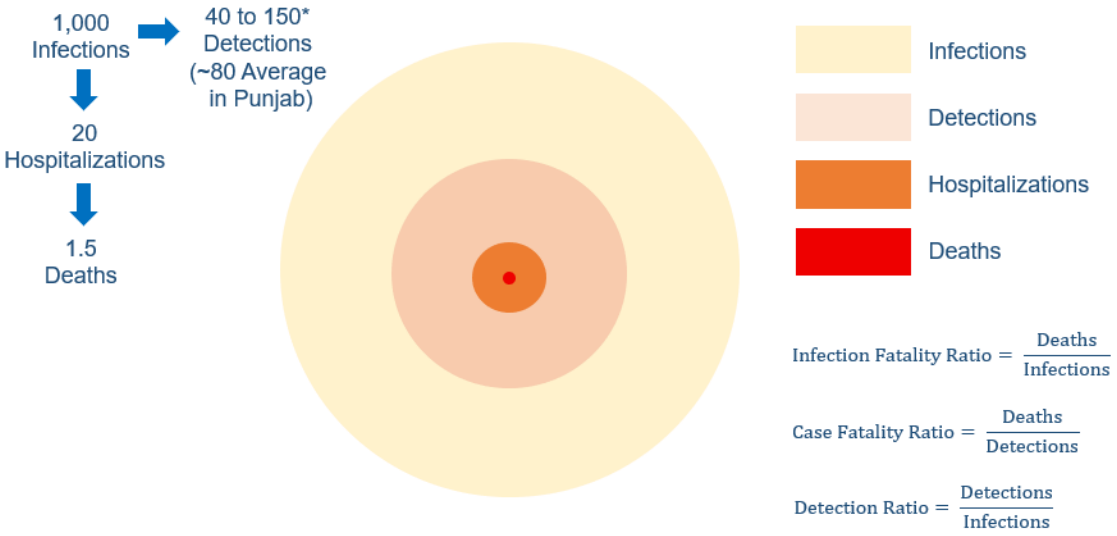


FIGURE 3: THE COVID UNIVERSE

Appendix 2: Estimating Infection Fatality Rates for India

Method 1: Using estimates Verity et. al. and Gangelt Based Serological Studies

1. Using age-cohort stratified IFRs from Verity, et. al. and applying to the Indian Demographic, we get an IFR of 0.36% for the Indian Demographic.

2. However, IFRs from Verity, et. al. depends largely on their estimations of under ascertainment based on PCR testing. To check for error or any further under ascertainment, we use these ratios to estimate IFR of approximately 0.8% in Gangelt. On the other hand, Serological testing, which can give statistical estimates of spread of infection, shows that Gangelt has a true IFR maximum of 0.37%.

3. Based on the above estimates from Verity et. Al, the IFR is over-estimated by 56%. We therefore revise the Indian IFR from 0.36% to 0.15%.

Method 2: Using estimates from Korean CDC and Time Correlated Data to Estimate IFR

1. Using age-cohort stratified CFRs from CDC Korea and applying to the Indian Demographic, we get a CFR of 0.58% for the Indian Demographic, if detection rate in India = detection rate in Korea.

2. This detection rate can be found from the time correlation paper that estimates the IFR of Korea to be 0.4-0.7% giving a detection rate around 30%

3. Since detection rate in India = detection rate in Korea for above CFR, detection rate of India= 30%

4. Since CFR India calculated above assumes the same detection rate, IFR India, 30% * 0.58% = 0.19%, with a range of 0.12-0.19%

Method 3: Using age and censorship adjusted CFR numbers from Chinese Data

1. Using age-cohort stratified CFRs from China adjusted for censorship and applying to the Indian Demographic, we get a CFR of 0.8% for the Indian Demographic, if detection rate in India = detection rate in China.

2. The Japanese Diamond Princess Cruise ship example when applied to China gives an IFR of 0.55% in China, this implies detection in China = IFR China/ CFR China = 0.55%/3.3%

3. Since detection rate in India = detection rate in China for above CFR, detection rate of India= 16%

4. This Implies IFR India = 16%*0.8% = 0.12%

Method 4: Using age-adjusted CFRs from Gangelt Data Directly

1. Using age-cohort stratified CFRs from Gangelt Study (assuming all infected were detected either by RT-PCR or through the Serological Tests) and Demographics of Infected Population/ Detected Population and applying to the Indian Demographic, we get an IFR of 0.15% for the Indian Demographic

Method 5: Using actual case fatality rates from regions with high detections as proxy

1. Countries such as Singapore, Qatar and Iceland have exhibited amongst the lowest Case Fatality Rates even though detected cases have crossed 10,000 cases

2. Even if we assume these countries have detected nearly all infected cases in their regions and thus Case Fatality Rate = Infection Fatality Rate, the IFR comes to be between 0.15% to 0.5%

Region	Total Detections (31.05.2020)	Total Deaths (31.05.2020)	Case Fatality Ratio
Singapore[21]	34,366	23	0.07%
Qatar[22]	55,262	36	0.07%
Iceland[23]	1,806	10	0.56%

Appendix 3: Learnings from nearly complete Epidemic Cycles

We investigate key statistical trends from the epidemic cycles of four regions – New York City, Lombardy, UK, and Madrid. The detection and mortality time trends are shown below:

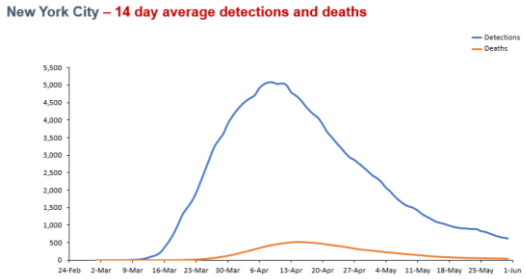


FIGURE 4: NEW YORK CITY EPIDEMIC CURVE

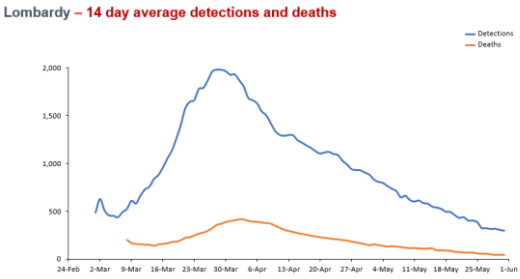


FIGURE 5: LOMBARDY EPIDEMIC CURVE

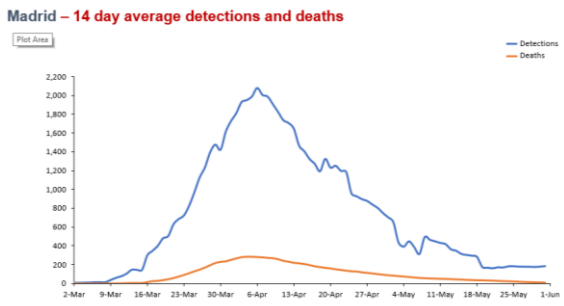


FIGURE 6: MADRID EPIDEMIC CURVE

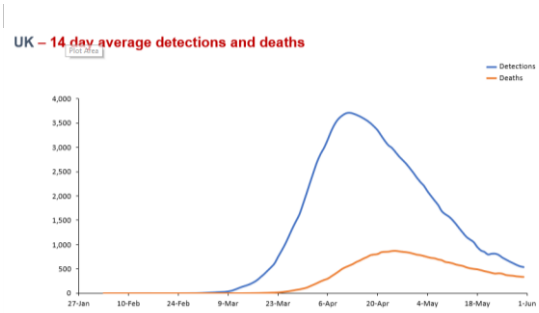


FIGURE 7: UK EPIDEMIC CURVE

The following statistical trends can be derived from the above curves:

Region	Peak date / period	Area under the curve at peak	Inflection point date	Area under the curve at inflection point
New York City	April 16	69%	April 2	27%
Lombardy State	March 28	44%	March 17	19%
United Kingdom (UK)	April 11	49%	March 29	18%
Madrid City	April 6	58%	March 23	15%

Based on the above data, the first saturation level has been defined as:

1. 22.5% of total infections: For highest population density regions like Mumbai and Kolkata
2. 20% of total infections: For other urban areas / cities in the country
3. 12.5% of total infections: For rural areas

Once this point is reached, we assume a gradual decline in CGRs

Total infections in these regions for the first epidemic cycle can also be derived using their respective age adjusted IFRs as shown below:

Region	Age-adjusted IFR	Case Fatality Rate	Implied Detection Rate	Total Detections (31.05.2020)	Est. Total Infections (31.05.2020)	% of Population infected
New York City	0.2-0.4%	8.3%	2-4%	200,730	50-60 lakhs	60-70%
Lombardy State	0.5-0.7%	18.1%	2-4%	88,970	25-35 lakhs	25-35%
United Kingdom (UK)	0.4-0.6%	14.2%	2-4%	274,760	80-100 lakhs	10-15%
Madrid City	0.4-0.6%	13.1%	3-5%	68,830	16-20 lakhs	50-60%

Peak Infections and subsequent deaths have thus, been assumed corresponding to approximately 50-70% of total infections as observed for New York City, Lombardy and Madrid. A deceleration rate of approximately 94-97% once the first saturation level is reached results in this peak being attained i.e. once the first saturation level is reached, subsequent CGRs are attenuated by 3-6% daily.

The hospitalization rate across these regions when looked at in relation to the estimated total infections is estimated to be between 0.8% to 3%. Given the absence of age-specific hospitalization rates in these regions and given India has a younger population, we assume a hospitalization rate of 2% per infection.

Region	Est. Total Infections (31.05.2020)	Hospitalized Cases	Hospitalized Cases as % of Est. Total Infections
New York City	50-60 lakhs	52,000	0.8-1.2%
Lombardy State	25-35 lakhs	50,101	1.4-2.0%
Madrid City	16-20 lakhs	42,041	2.1-2.6%

REFERENCES

[1] COVID-19 Map. (n.d.). Retrieved from https://coronavirus.jhu.edu/map.html

[2] UK data: Government Digital Service, UK Government. Retrieved from https://coronavirus.data.gov.uk/

[3] Germany data: Das Robert Koch-Institute. Retrieved from https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Fallzahlen.html

[4] Spain data: Carlos III Health Institute (ISCIII). Retrieved from https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Paginas/InformesCOVID-19.aspx

[5] New York City data: NYC Department of Health and Mental Hygiene. Retrieved from https://www1.nyc.gov/site/doh/covid/covid-19-data-archive.page

[6] South Korea data: Korean Centre for Disease Control and Prevention (KCDC). Retrieved from https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030

[7] Italy data: Ministero della Salute. Retrieved from http://www.salute.gov.it/portale/nuovocoronavirus/archivioNotizieNuovoCoronavirus.jsp?lingua=italiano&menu=notizie&p=dalministero&area=nuovocoronavirus&notizie.page=0

[8] Wu, Z. and McGoogan, J.M., 2020. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. Jama – https://jamanetwork.com/journals/jama/fullarticle/2762130

[9] Ferguson, N. (2020, March 16). Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Retrieved from https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf

[10] Mallapaty, S. (2020, March 26). What the cruise-ship outbreaks reveal about COVID-19. Retrieved from https://www.nature.com/articles/d41586-020-00885-w

[11] Verity, R., Okell, L. C., & Dorigatti, I. (2020, March 30). Estimates of the severity of coronavirus disease 2019: a model-based analysis. Retrieved from https://www.sciencedirect.com/science/article/pii/S1473309920302437

[12] Regalado, A. (2020, April 10). Blood tests show 14% of people are now immune to covid-19 in one town in Germany. Retrieved from https://www.technologyreview.com/2020/04/09/999015/blood-tests-show-15-of-people-are-now-immune-to-covid-19-in-one-town-in-germany/

[13] Juan, Manuel, V., & Rivas Santos. (2020, April 4). Estimating the infection fatality rate of COVID 19 in South Korea by using time series correlations. Retrieved from https://figshare.com/articles/Estimating_the_infection_fatality_rate_of_COVID_19_in_South_Korea_by_using_time_series_correlations/12083322

[14] Verity, R., et al., 2020. Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet Infectious Diseases (2020) – https://www.sciencedirect.com/science/article/pii/S1473309920302437

[15] Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days & deaths by US state in the next 4 months. Retrieved from http://www.healthdata.org/sites/default/files/files/research_articles/2020/covid_paper_MEDRXIV-2020-043752v1-Murray.pdf

[16] Graham, J. (2020, April 11). What Recovery From COVID-19 Looks Like. Retrieved from https://www.scientificamerican.com/article/what-recovery-from-covid-19-looks-like/

[17] Walker, M. (2020, April 27). High Mortality Rate in Intubated COVID-19 Patients in NYC. Retrieved from https://www.medpagetoday.com/infectiousdisease/covid19/86101

[18] Expert reaction to ICNARC report on the first reported 775 patients critically ill with COVID-19. (2020, March 29). Retrieved from https://www.sciencemediacentre.org/expert-reaction-to-icnarc-report-on-the-first-reported-775-patients-critically-ill-with-covid-19/

[19] Modeling the Spread and Prevention of COVID-19. (n.d.). Retrieved from https://cddep.org/covid-19/

[20] Salvatore, M., Rix, A., Kleinsasser, M., & Mukherjee, B. (2020, May 5). COVID-19 Outbreak in India. Retrieved from https://umich-biostatistics.shinyapps.io/covid19/

[21] Singapore Data: Ministry of Health, Singapore. Retrieved from https://www.moh.gov.sg/covid-19/situation-report

[22] Qatar Data: Ministry of Public Health, Qatar. Retrieved from https://www.moph.gov.qa/english/Pages/Coronavirus2019.aspx#

[23] Iceland Data: Iceland Review. Retrieved from https://www.icelandreview.com/ask-ir/whats-the-status-of-covid-19-in-iceland/