

A model for survival rate projection of Covid19 patients

Devika P. Madalli

DRTC, Indian Statistical Institute

Sagar Bhimrao Gajbe (✉ sagar@drtc.isibang.ac.in)

DRTC, Indian Statistical Institute <https://orcid.org/0000-0003-2924-4060>

Gopalji

DRTC, Indian Statistical Institute

Akhilesh Ajith Kumar

DRTC, Indian Statistical Institute

Research Article

Keywords: Virus infection, Corona Covid19, Prediction data modelling, survival rate

Posted Date: June 15th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-28560/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

A model for survival rate projection of Covid19 patients

Devika P. Madalli Sagar Bhimrao Gajbe Gopalji
Akhilesh Ajith Kumar

*Documentation Research and Training Center
Indian Statistical Institute, Bangalore*

Abstract

The novel coronavirus has been declared a worldwide pandemic. The pandemic has unleashed health as well as economic devastation across the world. In view of this, various governments, researchers and agencies are trying to work towards solutions to control the spread of the virus. There are many studies that have been initiated that consider the various parameters that contribute to the spread of the virus. We present a model to predict the survival rate of Covid19 infected patients. The study takes into consideration death rate (normal) and Covid infection death rate in India. Other factors considered are the number of infections and active infection numbers. We also incorporate a 'learning module' to learn from the observed error rate. We compute the moving average of error which is then deployed to minimise projection error in the model. The factor obtained from the learning is deployed into the survival rate computation to achieve best results.

Keywords: *Virus infection, Corona Covid19, Prediction data modelling, survival rate.*

1 Introduction

The novel coronavirus has unleashed unprecedented disruption in lives and livelihoods. The virus, though apparently has its origins in Asia, has spread to almost 215 countries/regions/territories across the world resulting in a large number of fatalities ([WHO, 2020](#)). The outbreak was first reported in China in December 2019, and it then began to spread to Asian countries, and then to European countries and American countries. Some of the worst affected countries at the end of April are the United States of America, Spain,

Italy, United Kingdom, and France in terms of lives lost and number of cases. Lockdowns are imposed by the governments either partially or completely to control the spread of the virus. The virus has put enormous strain on the emergency healthcare services in many of the countries so much that it is difficult to handle all Covid-19 cases. Many studies are being conducted to project the impact of the pandemic on the lives, deaths, economic, and healthcare facilities. The present study projects the daily survival rate of the people affected by the Covid-19 patients taking into account mortality rates in states of India. The survival rate and casualties projected will have an impact of preparedness of states in terms of infrastructure, hospital capacities and testing facilities.

2 Objective and scope

Our primary aim is to project the survival rate of the patients affected by the virus using covid-19 related data of the Ministry of Health and Family Welfare (MoHFW, India). The data contains statewise confirmed covid infected cases, discharges, deaths. Data is accessed from 30th March, 2020 on a daily basis. The data prior to this was published at a random time of the day so it does not have a time interval of 24 hours. Hence data from 30th March, 2020 is selected

3 Factors affecting the spread of virus

There are various factors that together contribute to the spread of the virus in communities. Some of the factors are distinct while others are dependent and related. The study of the effect of each factor is quite complicated, albeit necessary, for comprehensive results to gear up for the future disease ‘Y’. One evident factor for the spread of Covid19 is international travel that brings virus infected people to various places and thereby fuels the spread of the infection. According to [Qiu *et al.* \(2020\)](#), Covid-19 transmission depends on various social and economic factors such as mass gatherings, population flow between the cities and geographic proximity among others. [Florida \(2020\)](#) consider various factors which are responsible for the spread of the virus underlining that one of the major factors is population density. This is also observed in India especially in the cities of Delhi and Mumbai. [Zhan *et al.* \(2020\)](#), express population movement as a primary factor in the spread of Covid-19. [Araujo & Naimi \(2020\)](#), highlight the role of climate on spread of the virus. They opine that cold and warm climates are suitable for the spread

of viruses, and that arid and tropical conditions are less suitable. [Ebrahim & Memish \(2020\)](#), explain the effect of mass gatherings in Covid-19 spread. They considered some religious, cultural, sports and economic events that happened during the outbreak of Covid-19 and discussed how these events are correlated to spread of virus.

Consolidating the factors that are discussed widely and as evidenced from early studies and literature available on the topic, the list of factors that accelerate the spread of the virus and its impact maybe be enlisted as: population density, age of infected person, international travellers arriving from Covid affected countries, environmental factors like pollution, temperature, humidity and medical factors such as pre-existing health conditions like diabetes, kidney diseases, cardiac problems, respiratory problems etc., that make people more vulnerable to the virus infection. For our study we consider the death rate of states as a basis to compare with covid death rate. The death rate of a state or mortality is an indicator of the factors such as pre-existing health conditions, pollution, economic well being, medical facilities and infrastructure.

3.1 The Survival Rate Prediction Model

We use a dynamic projection approach in predicting the survival trend in the patients infected by Covid-19. The death rate data of the states to which the individual belongs and the current Covid-19 death number in that particular state are the key factors used for computation and the prediction. Death rate and death numbers at the national level are also analysed.

3.1.1 Explaining the selection of features

Covid-19, survival rate reduces if there is a pre-existing health condition in a patient. Another factor, we observed, that if mortality in general is high, then covid death rate is also high. For example, countries of Europe have high death (pre-Covid ratio to their population) and correspondingly we have seen that there, the death rate in Covid-19 patients is also high. Co-morbidity contributes significantly to fatalities and hence pre-existing health conditions of the populace is an important factor. We contend that it is fairly indicated by the mortality rate of a state in general. Hence the death rate factor is an important one.

Statewise Covid death cases number is the other factor taken into account. This indicates how the state is equipped to handle the cases and preparedness in terms of medical facilities, infrastructure, implementation of relief measures, efficiency and support. For example: Kerala vs Tamil Nadu

[contrast in infections and casualties]. Hence the second factor of individual death cases is vital and is included in the analysis.

3.1.2 Data collection

Data is collected from the Ministry of Health and Family Welfare, India [MoHFW] releases. The MoHFW provides statewise (different states of India) cases affected with Covid-19 on a daily basis. The data includes total number of Indian nationals infected, total number of foreign nationals infected, a combination of these two columns called confirmed cases, discharged cases, death cases. This data is updated everyday, multiple times at source.

3.1.3 Stages in the prediction model

The prediction model for survival rate is in three stages:

- i survival projection based on expected death which includes both the mortality rate and the deaths due to Covid-19;
- ii normalised learning from previous five-day errors and modification of the expected death number (obtained from the previous stage) for the next day; and
- iii based on modified expected death number from stage 2, modified survival is projected for the next day.

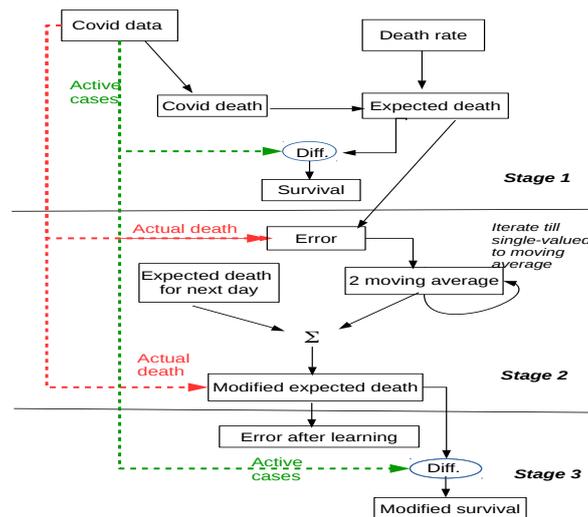


Figure 1: Survival rate prediction model flowchart

Stage 1

Survival projection is based on expected death which includes both the mortality rate and deaths due to Covid-19

$$E_D = \sum(A_{TS}(D_{RS} + D_{CS})) \quad (1)$$

$$S_R = A_{TN} - E_D \quad (2)$$

S_R = Survival Rate

E_D = Total expected death in the nation for next day

A_{TN} = Active Cases in the nation on a day

A_{TS} = Active Cases in a state on a day

D_{RS} = Death Rate (normal – non-Covid)

D_{CS} = Death due to Covid

A_{TN} - Active cases here means the number of patients currently affected by the Covid-19 in the nation. We remove the patients who are recovered and those deceased so that we are calculating based on the ‘current active’ cases only.

A_{TS} - Active cases in the state indicates the number of patients that are currently affected by the Covid-19 in the state. We remove the patients that are recovered and those deceased so that we are calculating based on the ‘current active’ cases only.

D_{RS} - State death rate is given per 1000 (GOI, 2019). To derive ‘per individual’, state death rate is divided by 1000.

D_{CS} - Death rate of a state pertaining to Covid-19 cases is calculated based on the total death due to Covid-19 in the particular state divided by the total cases of Covid-19 in the particular state.

$$D_{CS} = \frac{CovidDeaths/state}{CovidConfirmedCases/state} \quad (3)$$

Finally, total active cases in a particular state is calculated as a factor of the sum of the death rate (pre-Covid) of the state by the death rate of Covid-19 of that particular state.

The sum total indicates S_R - the number of death cases that might happen the next day (future) based on the assumption of today’s numbers. As the number keeps changing everyday so does the projected value. As seen from the data and visualizations, the projected survival rate values almost coincide with the actual.

Stage 2

Normalised learning from previous five-day errors and modification as per learning

At stage 1, we have estimated deaths in a state for the next day by adding normalized general death rate and death covid rate of the state. The results of death rate on the day are multiplied by the cumulative active cases of the state. Sum of all the expected deaths of states constitutes the total estimated death in the country for the next day.

If we observe the curve in figure 4a, the model has a growing difference in predicted death and actual death. We assert this difference is due to lack of error-learning. Consequently we integrate a learning function as follows in our model.

In order to integrate the error into the model, we calculate an error rate for the previous five days. We compute the ‘moving average of error’, with an iterative moving average of error for two days (the day of measurement and the day before). The cumulative predicted death is updated iteratively after every five days by adding the single-valued moving average of the preceding day. This is explained in the set of equations in the section below. The difference in expected death from actual death

$$e_{(\forall i, j=1)} = \begin{cases} A_D - E_D \\ A_D - \hat{E}_D \end{cases} \quad (4)$$

where $A_D - E_D$ is difference while learning, $A_D - \hat{E}_D$ is difference after learning

$$\begin{aligned} e_{(\forall i, j=1)} &= e \text{ for all values of } i \text{ and } j=1 \\ A_D &= \text{Total Actual Death} \\ E_D &= \text{Total Expected Death} \\ \hat{E}_D &= \text{Modified total expected death} \end{aligned}$$

Calculation of moving average

A diagonal matrix is constructed by calculating a two-day moving average lower diagonal matrix. In this lower diagonal matrix, the first column is the difference in expected death from actual death ($e_{(\forall i, j=1)}$). The rest of the element is calculated as shown below :

$$e_{(i, j)} = \frac{e_{(i, j-1)} + e_{(i-1, j-1)}}{2} \quad |_{i \geq j} \quad (5)$$

$$\begin{array}{cccccc}
e_{1,1} & 0 & 0 & 0 & \cdots & 0 \\
e_{2,1} & e_{2,2} & 0 & 0 & \cdots & 0 \\
e_{3,1} & e_{3,2} & e_{3,3} & 0 & \cdots & 0 \\
e_{4,1} & e_{4,2} & e_{4,3} & e_{4,4} & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
e_{i,1} & e_{i,2} & e_{i,3} & e_{i,4} & \cdots & e_{i,j}
\end{array}$$

$e_{(i,j)}$ for $i = j$, is a single-valued error after $(j-1)$ iterations of two day moving averages which is also the last value of the lower diagonal matrix. This value is considered as the error learnt. It is used to modify the total expected death for the next day by adding it to the total expected death as follows:

$$\hat{E}_{D_{(i+n)}} = E_{D_{(i+n)}} + e_{(i,j)} \Big|_{i=j} \quad (6)$$

where $i \geq 2$ and n takes value $1,2,3,4,\dots$ so on as the data for total expected death is populated.

Stage 3

Based on modified expected death from stage 2, modified survival is projected for next day

$$S_R = A_{TN} - \hat{E}_D \quad (7)$$

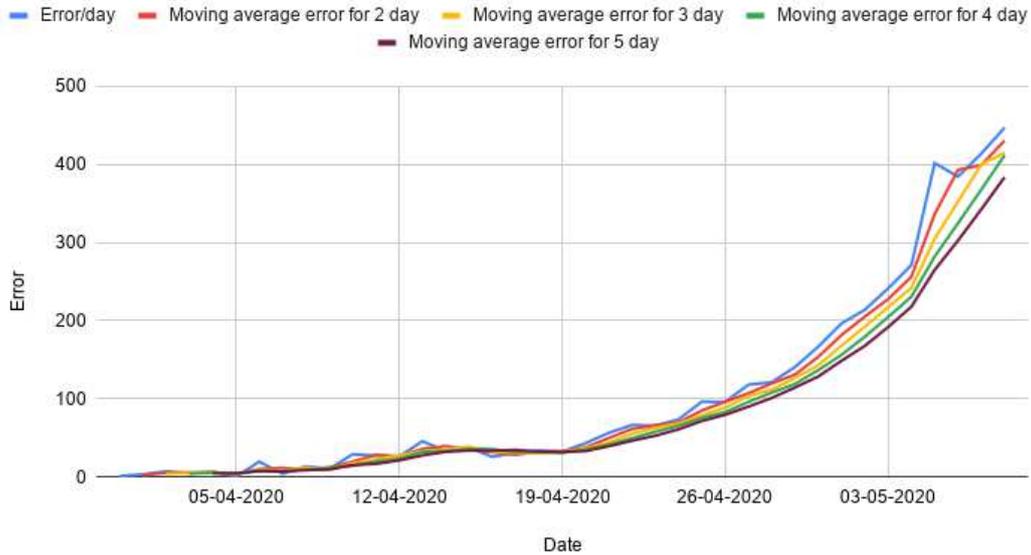
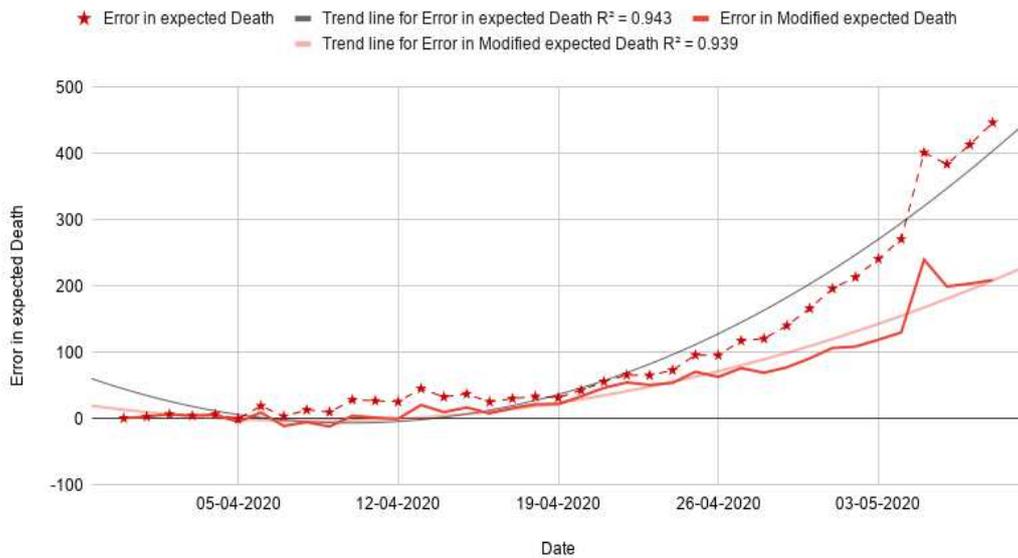


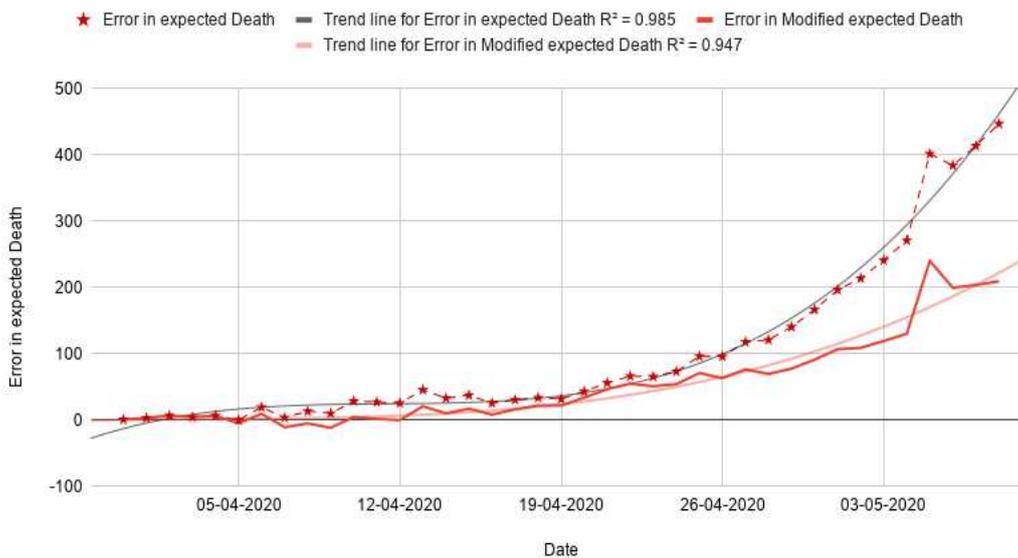
Figure 2: Two, three, four, and five moving averages of error and the error in expected death w.r.t. actual death

4 Findings

In stage 1, we found that the expected death for the next day was projected correctly with only slight error but with time when other factors came into existence, such as testing speed, reporting, etc, the error in next day's ex-



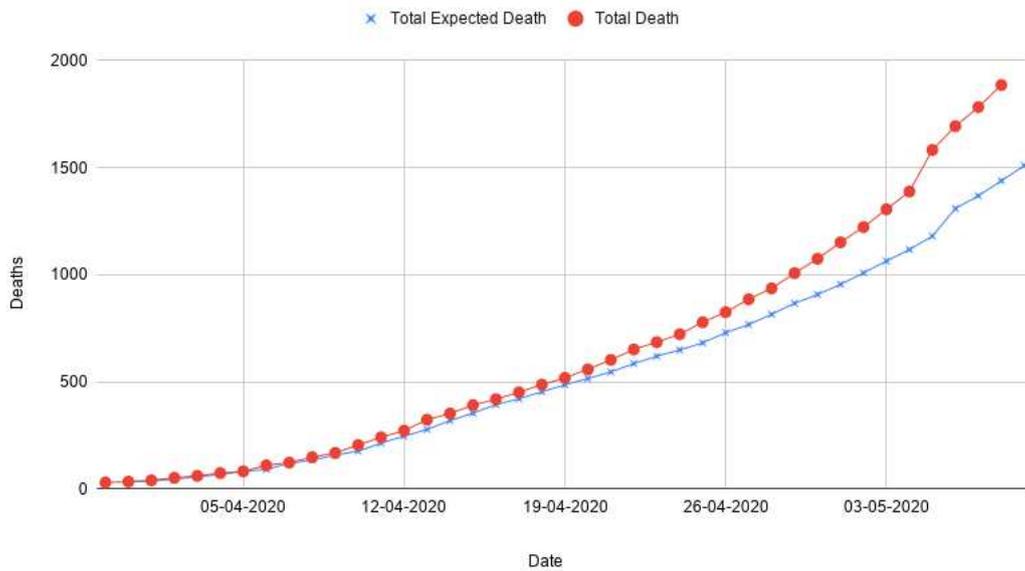
(a) Trend line with 2nd degree polynomial curve fitting



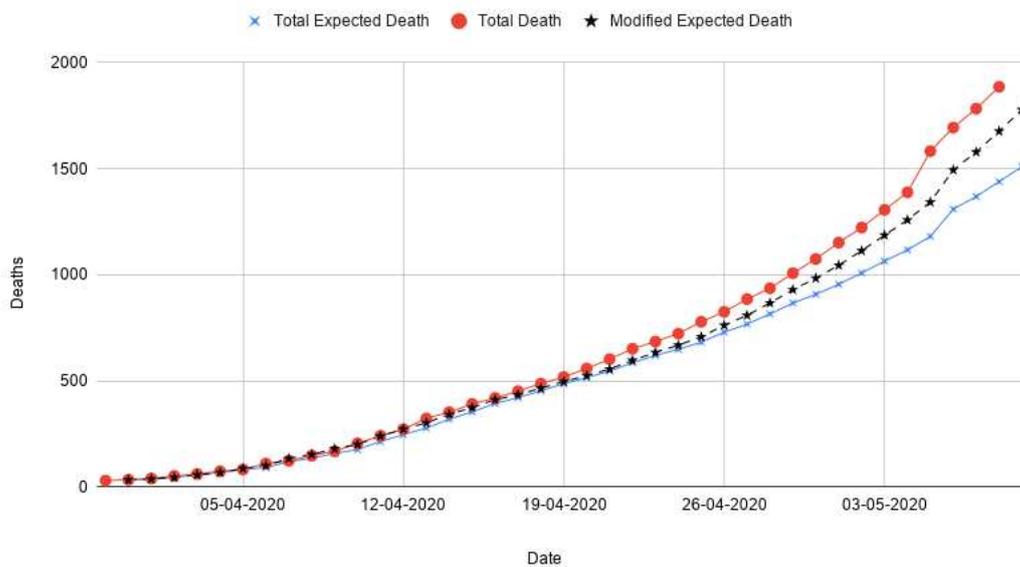
(b) Trend line with 3rd degree polynomial curve fitting

Figure 3: Error in expected and modified expected death for each day

pected death was increasing as shown in graph 2. The graph 2 shows the moving average error in the expected death per day. This graph also shows the moving average trend of error for 2, 3, 4, and 5 days.



(a) Expected death vs Actual death



(b) Expected death vs Actual death vs Modified expected death

Figure 4: Comparison plots between deaths with respect to date

The graph 3a and 3b show the variation in error for the projected death per day with a second and third degree polynomial curve fitting line with

R^2 value 0.943 and 0.985 respectively. However, the statistical measure for representing the proportion of the variance in error in the expected death from actual death for every day is a measure of R^2 value for trend line. It is also visible by comparing modified expected error in both the graphs that their R^2 is the same which means it is more stable as compared with error in expected death at first stage. After learning from the error of the previous 5 days, the proposed model found a significant difference in the expected modified death relative to the expected death.

We have estimated the expected death for all states and union territories with the suggested method in the stage 1 and these projected deaths for the next day are compared with the exact death that occurred that day. In addition, such a relation has been shown in the figure 4a which shows that as the data is populated, the cumulative death is more or less similar to the actual cumulative death but the error continues to increase. This is where the model attempts to learn from the error, thereby reducing the gap (modified expected death curve in figure 4b) and making it more stable.

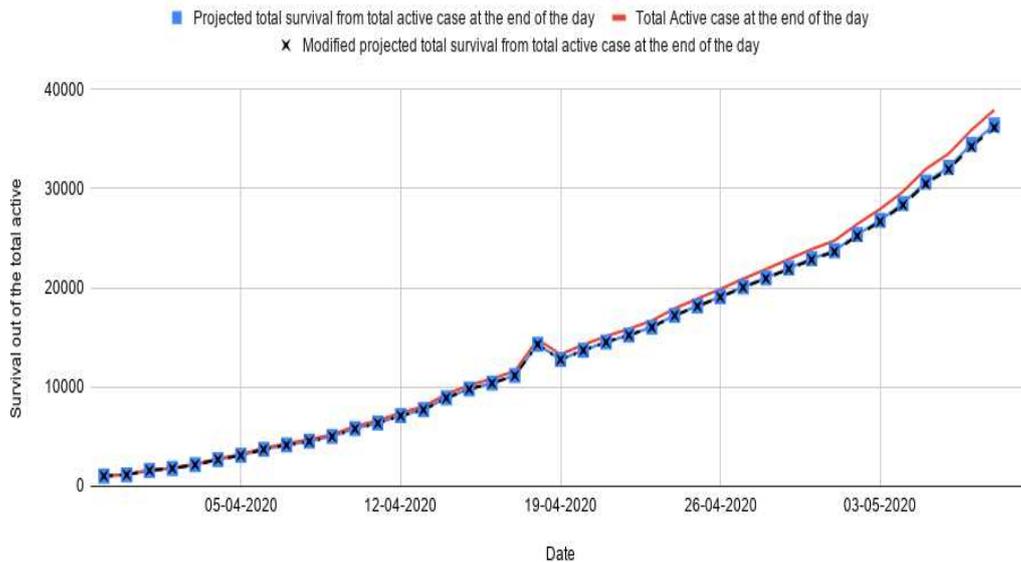


Figure 5: Survival Out from active cases in India by expected deaths and modified expected deaths

Similarly when the modified death projection improved, the survival projection out of total active cases also improved. This improvement is shown in the graph 5 of survival rate. This change in expected death is less significant for survival out of active cases as a whole but when counted for number of survival, it becomes significant.

5 Discussion

The covid pandemic has been devastating taking into account the number of countries infected, the number of infections and casualties in a very short period. As mentioned earlier, there are various factors that affect the spread of the virus including socio-economic, medical, environmental and biophysical factors among others. As an immediate response to the pandemic, it is necessary to study the most important and potent of these factors. The availability of open data sets corresponding to the various factors helps accelerate the research and forge collaboration (Schwalbe, 2020).

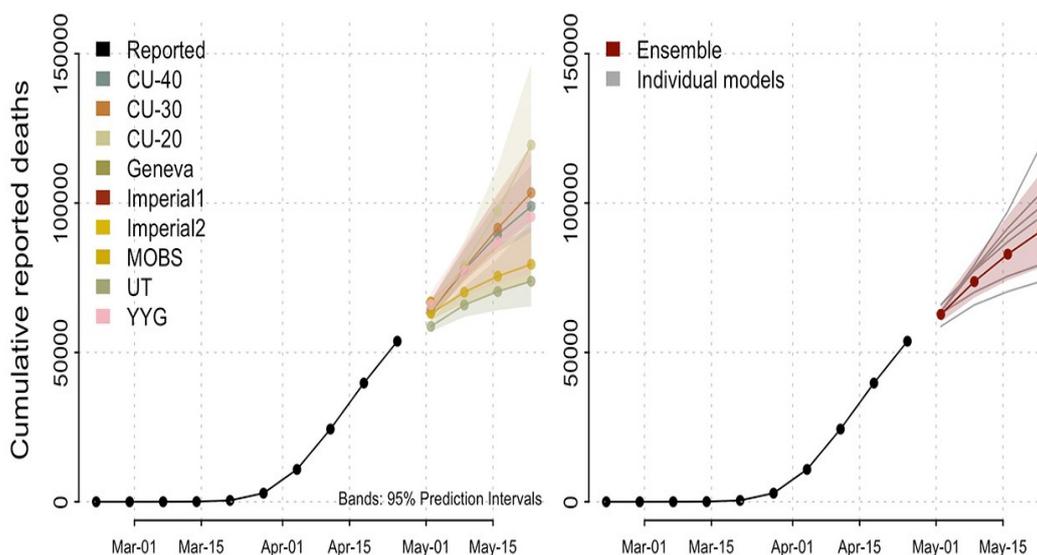


Figure 6: USA National Forecast of death in epidemic by different universities (CDC, 2020)

Many studies compare and contrast a set of factors such as state preparedness in terms of quarantine facilities, medical testing facilities, hospital capacities etc. Others studied the environmental factors such as pollution and sanitation facilities. Yet other studies took into account deaths due to Covid and age of people, location and other demographic details (Henderson & Keiding, 2005; Hollnagel, 1999; Soyiri & Reidpath, 2013; Zeegers *et al.*, 2016). There are studies and discussions that point to comorbidity as a significant factor for the number of casualties (Guan *et al.*, 2020; Wang *et al.*, 2020). Without considering comorbidity, fatalities may be mistakenly read as

covid deaths. There are several universities in the United States of America which have done projections for deaths. One such study is done at Columbia University (CU), CDC (2020) where they used ‘death’ as an independent variable and made a projection based on social distancing factor using the metapopulation SEIR model. The study takes several degrees of social distancing, like 20%, 30%, and 40% contact reduction, among the population shown in figure 6. The CU forecast is weekly for the next one and half month whereas our model projects deaths for the next day by learning from previous days’ errors in death projection.

We also need to focus on designing more robust mathematical models that will not just accommodate the above mentioned factors or features but also parameters like economic affluence, education level and society strata into account, so we could plan our health action plans and strategy better in tackling covid or ‘disease Y’ in future. We, however, have taken into account the general mortality rates of the states/provinces in the nation before pandemic under consideration. In our case, data was available state wise and we took into consideration the death rate of the states. We would like to propose that the death rate is a good indicator of other factors such as age profile of communities, pollution, underlying health conditions, economic conditions and others that generally contribute to mortality rate. It can be concluded that death rates for a nation shows significant effect in projecting covid deaths and hence survival rates. The model used has a dynamic learning component. As data and the results of studies including all factors that impact the spread of virus are made available, they could be ingested in the proposed model to predict the magnitude of spread of infection, casualties and survival which would contribute significantly to future preparedness.

References

- (2019). Sample Registration System Bulletin. Retrieved from http://www.censusindia.gov.in/vital_statistics/SRS_Bulletins/Bulletins.html.
- Araujo, M. B., & Naimi, B. (2020). Spread of SARS-CoV-2 Coronavirus likely to be constrained by climate. *medRxiv*. Retrieved from <http://dx.doi.org/10.1101/2020.03.12.20034728>.
- CDC (2020). COVID-19 Forecasts. Retrieved from <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html>.
- Ebrahim, S. H., & Memish, Z. A. (2020). COVID-19—the role of mass gath-

- erings. *Travel medicine and infectious disease*. Retrieved from <http://dx.doi.org/10.1016/j.tmaid.2020.101617>.
- Florida, R. (2020). What do we know so far about the types of places that are more susceptible to the spread of Covid-19? In the U.S., density is just the beginning of the story. Retrieved from <https://www.citylab.com/equity/2020/04/coronavirus-spread-map-city-urban-density-suburbs-rural-data/609394/>.
- Guan, W.-j., Liang, W.-h., Zhao, Y., Liang, H.-r., Chen, Z.-s., Li, Y.-m., Liu, X.-q., Chen, R.-c., Tang, C.-l., Wang, T., *et al.* (2020). Comorbidity and its impact on 1590 patients with Covid-19 in China: A Nationwide Analysis. *European Respiratory Journal*. Retrieved from <http://dx.doi.org/10.1183/13993003.00547-2020>.
- Henderson, R., & Keiding, N. (2005). Individual survival time prediction using statistical models. *Journal of Medical Ethics*, *31*(12), 703–706. Retrieved from <http://dx.doi.org/10.1136/jme.2005.012427>.
- Hollnagel, H. (1999). Explaining risk factors to patients during a general practice consultation: conveying group-based epidemiological knowledge to individual patients. *Scandinavian journal of primary health care*, *17*(1), 3–5. Retrieved from <http://dx.doi.org/10.1080/028134399750002809>.
- Qiu, Y., Chen, X., & Shi, W. (2020). Impacts of social and economic factors on the transmission of coronavirus disease (COVID-19) in China. *medRxiv*. Retrieved from <http://dx.doi.org/10.1101/2020.03.13.20035238>.
- Schwalbe, N. (2020). We could be vastly overestimating the death rate for COVID-19. Here's why. Retrieved from <https://www.weforum.org/agenda/2020/04/we-could-be-vastly-overestimating-the-death-rate-for-covid-19-heres-why/>.
- Soyiri, I. N., & Reidpath, D. D. (2013). An overview of health forecasting. *Environmental health and preventive medicine*, *18*(1), 1. Retrieved from <http://dx.doi.org/10.1007/s12199-012-0294-6>.
- Wang, B., Li, R., Lu, Z., & Huang, Y. (2020). Does comorbidity increase the risk of patients with COVID-19: evidence from meta-analysis. *Aging (Albany NY)*, *12*(7), 6049. Retrieved from <http://dx.doi.org/10.18632/aging.103000>.

WHO, W. H. O. (2020). Coronavirus disease (COVID-19) Pandemic. Retrieved from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.

Zeegers, M., Bours, M., & Freeman, M. (2016). Methods used in forensic epidemiologic analysis. In *Forensic Epidemiology*, (pp. 71–110). Elsevier. Retrieved from <http://dx.doi.org/10.1016/b978-0-12-404584-2.00003-3>.

Zhan, C., Tse, C., Fu, Y., Lai, Z., & Zhang, H. (2020). Modelling and prediction of the 2019 Coronavirus Disease spreading in China incorporating human migration data. *Available at SSRN 3546051*. Retrieved from <http://dx.doi.org/10.2139/ssrn.3546051>.

* This reference list is based on *Publication Manual of the American Psychological Association*, 6th ed. (2010).

Declarations:

Competing Interests: The authors declare no competing interests.

Figures

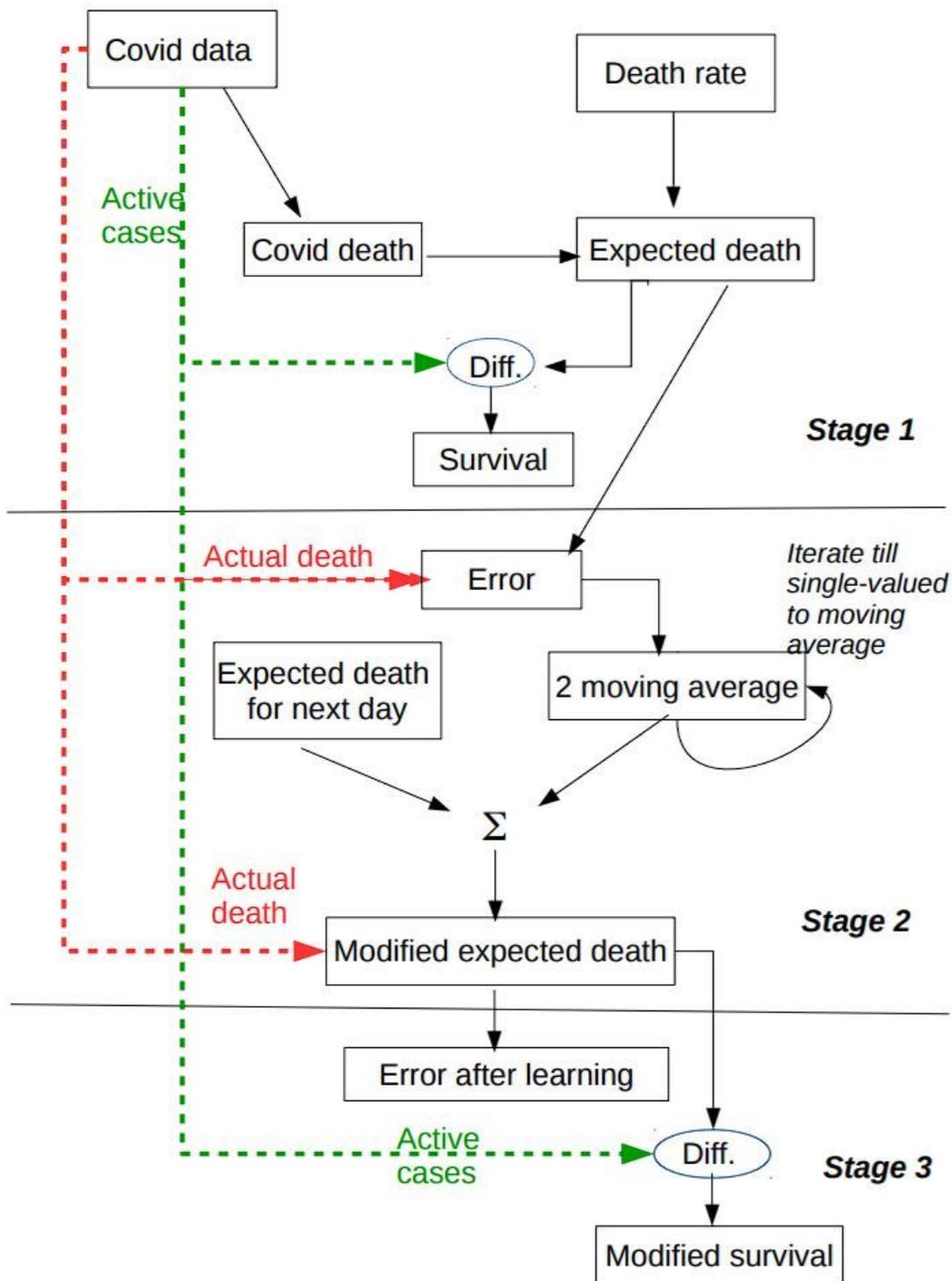


Figure 1

Survival rate prediction model flowchart

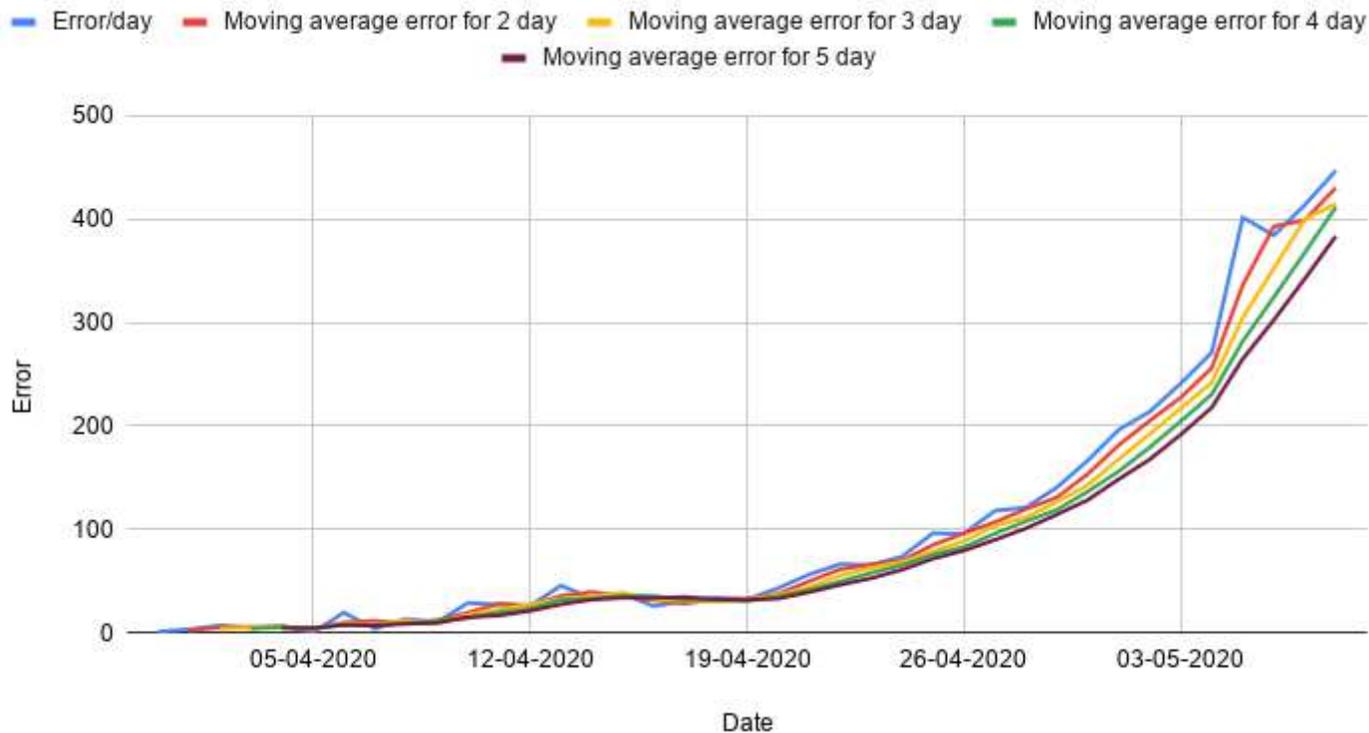


Figure 2

Two, three, four, and five moving averages of error and the error in expected death w.r.t. actual death

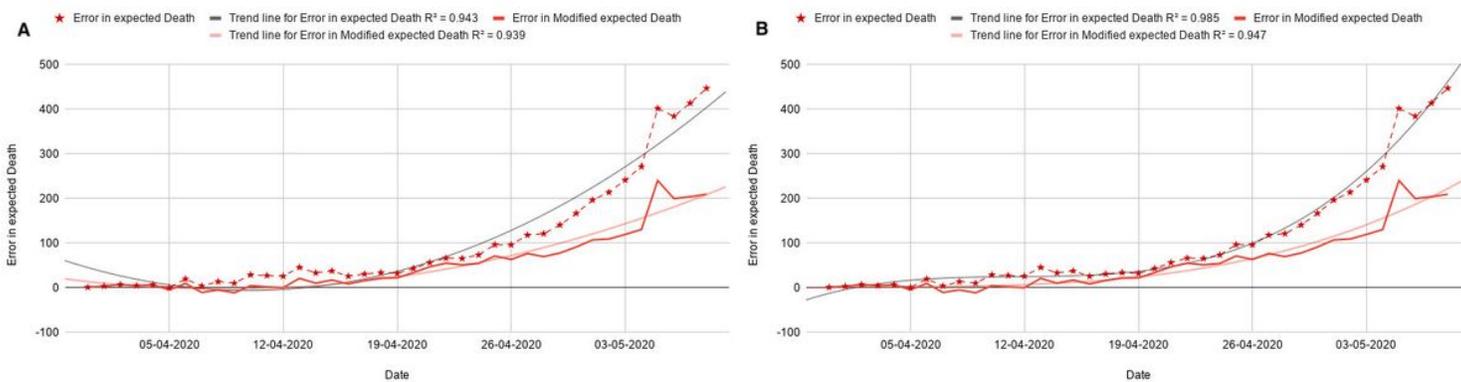


Figure 3

(a) Trend line with 2nd degree polynomial curve fitting (b) Trend line with 3rd degree polynomial curve fitting

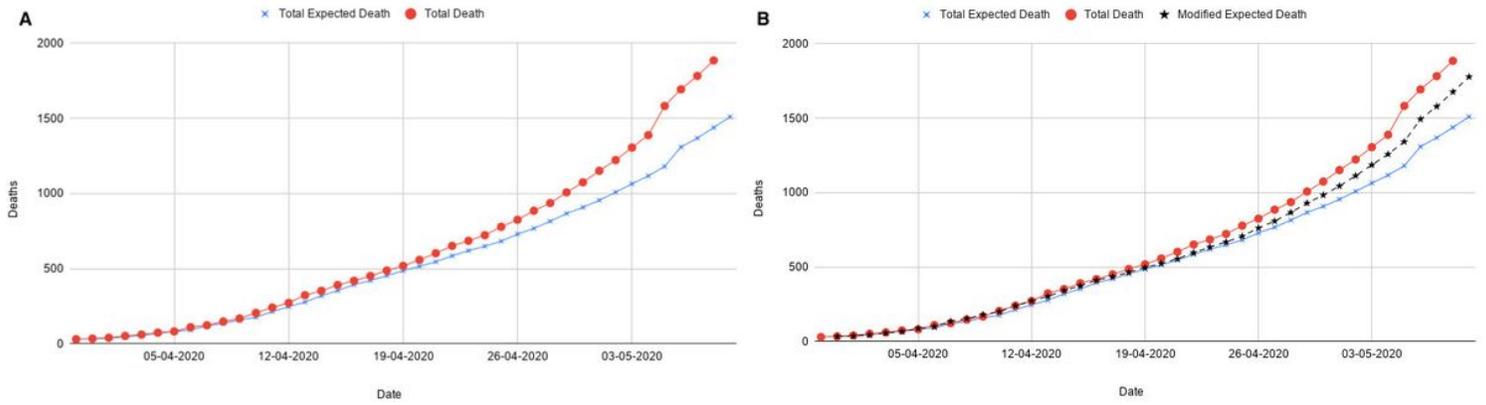


Figure 4

(a) Expected death vs Actual death (b) Expected death vs Actual death vs Modified expected death

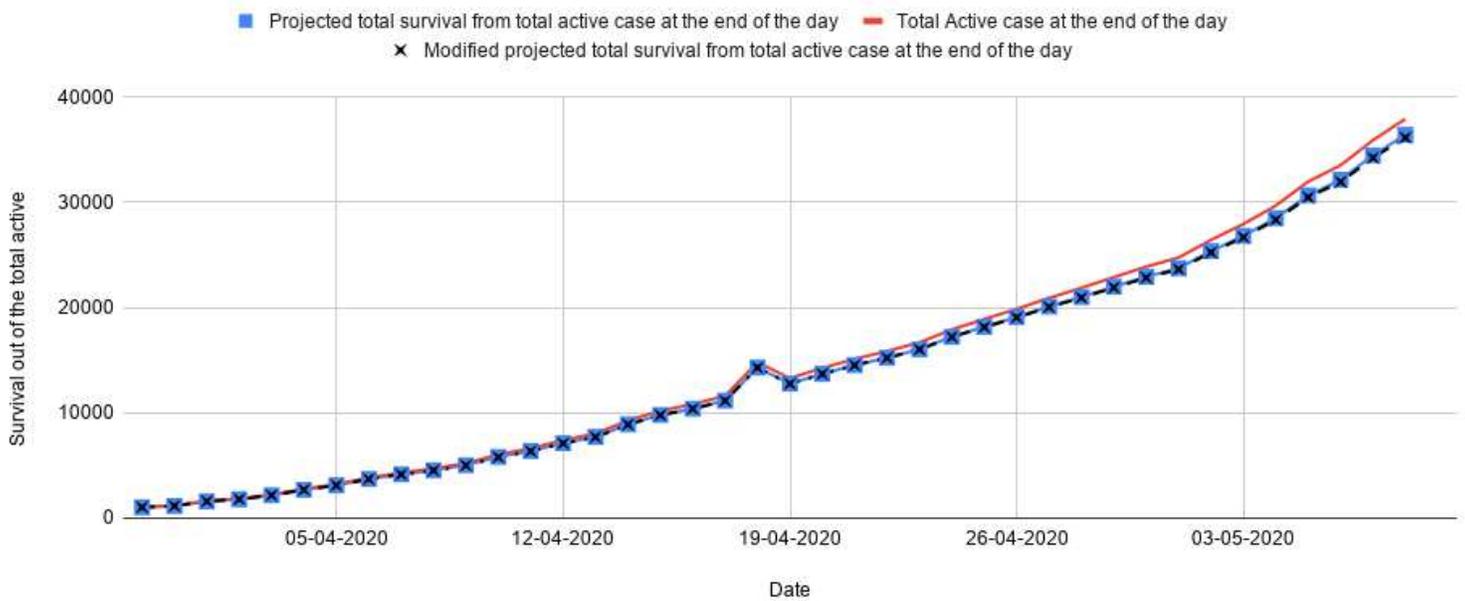


Figure 5

Survival Out from active cases in India by expected deaths and modified expected deaths

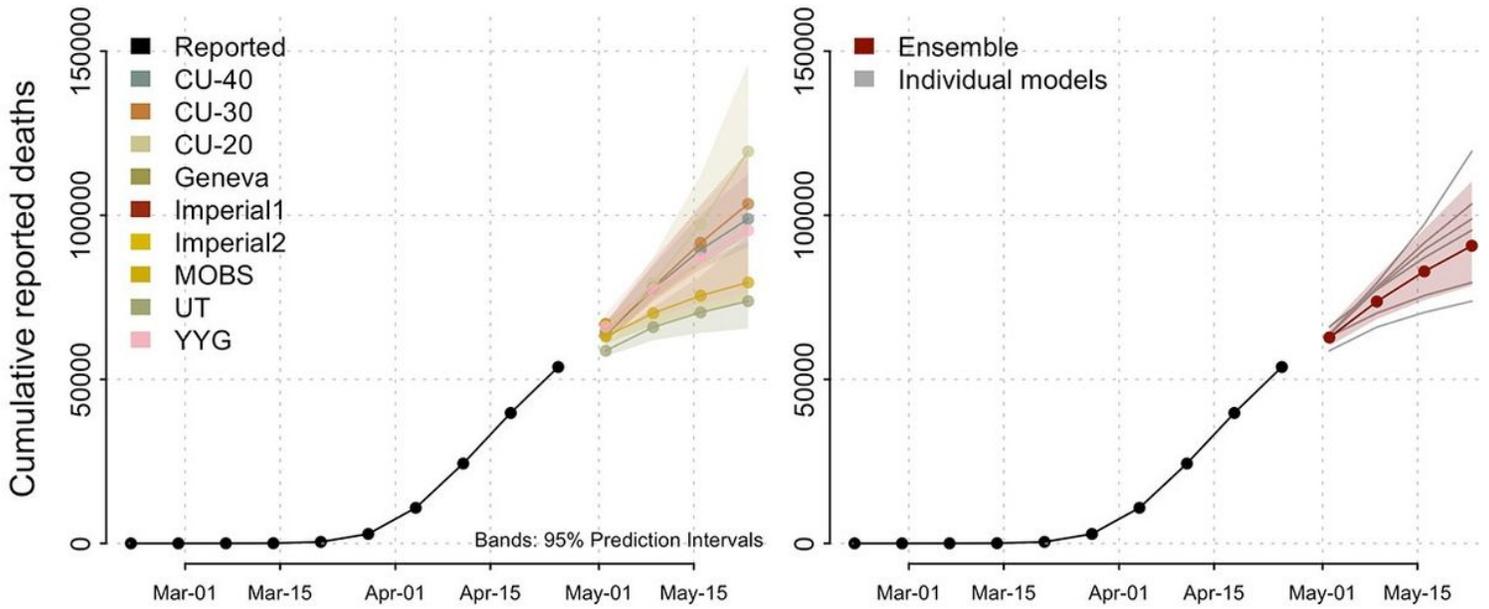


Figure 6

USA National Forecast of death in epidemic by different universities (CDC, 2020)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [equationonly.pdf](#)