COVID19 IN INDIA AND IT'S STATES: PERSPECTIVE FROM DEATHS

A PREPRINT

Siddhartha Mandal
For Niti Ayog
A work in progress

April 29, 2020

ABSTRACT

Highlights of the report:

1. Almost all models indicate a rising number of deaths till the first week of May.
2. The best fit model on global countrywise daily data indicates the total number of deaths in India to be around 1211 (95% CI: 755,2524).
3. Based on current trends, the largest number of cases will be in the middle age groups with around 3700 severe cases by the end of April.
4. The variable explaining variations in global death is a ratio of time before and after the strictest restrictions (eg: Lockdown).

1 Introduction

The current literature on progression of COVID19 focuses on projecting the number of cases, usually under the assumption of a compartmental model of susceptible, infected and recovered(or deaths), usually referred to as the SIR model in infectious disease epidemiology. The problems with modeing cases in an epidemic is fraught with problems as has been highlighted in multiple publications, resulting from the unknown number of infected people, long incubation period and varying degrees of testing in different locations. Deaths (and death rates), on the other hand, are more accurate to predict the progression of disease.

2 Methodology

We are using data on daily deaths from all countries along with disaggregated data wherever available. We model a) cumulative number of deaths and b) cumulative death rates against time using a sigmoidal curve (Gaussian error function) with three parameters and a nonlinear mixed effects modeling approach. In addition, we examined the effects of covariates, such as public health interventions, population density, proportion of urban population etc, on the outcome.

The function being used to model the cumulative outcomes is given by,

$$\frac{p}{2}\Big(1 + \frac{2}{\sqrt{\pi}} \int_0^{\alpha(t-\beta)} exp(-\tau^2) d\tau\Big) \tag{1}$$

where p indicates the maximum value of the outcome, α represents the growth parameter and β represents the inflection point of the curve. Under a nonlinear mixed model framework, we assume random effects and link functions to model the variation of the parameters between countries and the positive nature of specific parameters.

Data on deaths: We use data on daily deaths on 51 countries that recorded at least 25 deaths till 4th April 2020.

Covariates considered:

Table 1: Table showing model performances for the five models for cumulative deaths across countries

Model	AIC	BIC
Model 1	17999.52	18013.04
Model 2	17688.48	17703.93
Model 3	16215.87	16230.15
Model 4	17609.58	17626.97
Model 5	17753.56	17770.94
Model 6	16156.38	16170.65
Model 7	16041.22	16057.28

1. A stringency index has been developed by researchers at the Blavatnik School of Government, University of Oxford that tracks government response to the COVID19 situation. Apart from national lockdown, this index also incorporates policies such as school closures, travel bans, and financial indicators such as fiscal or monetary measures. (Reference: Hale, Thomas, Sam Webster, Anna Petherick, Toby Phillips, and Beatriz Kira (2020). Oxford COVID-19 Government Response Tracker. Data use policy: Creative Commons Attribution CC BY standard.). We use a weighted average of number of days when the index was below 25, between 25 and 50, between 50 and 75 and above 75 for each country, with highest weight attributed to days with strictest restrictions. This covariate is denoted by SI.
2. Ratio of the number of days under lockdown and number of days between the 10th case and lockdown implementation. This covariate is denoted by TR.
3. Days to lockdown from the 10th case.

The following model parameterizations were implemented:

1. All three parameters have random effects between countries (Model 1): $log(p_i) = \mu_1 + \eta_{1i}, log(\alpha_i) = \mu_2 + \eta_{2i}, \beta_i = \mu_3 + \eta_{3i}$
2. Maximum deaths depend on SI (Model 2): $log(p_i) = \mu_1 + \gamma * SI_i + \eta_{1i}, log(\alpha_i) = \mu_2 + \eta_{2i}, \beta_i = \mu_3 + \eta_{3i}$
3. Maximum deaths dependent on TR (Model 3): $log(p_i) = \mu_1 + \gamma * TR_i + \eta_{1i}$, $log(\alpha_i) = \mu_2 + \eta_{2i}$, $\beta_i = \mu_3 + \eta_{3i}$
4. Both maximum deaths and the inflection point depend on SI (Model 4): $log(p_i) = \mu_1 + \gamma * SI_i + \eta_{1i}$, $log(\alpha_i) = \mu_2 + \eta_{2i}$, $\beta_i = \mu_3 + \delta * SI_i + \eta_{3i}$
5. Inflection point dependent on SI and time to lockdown: (Model 5): $log(p_i) = \mu_1 + \eta_{1i}, log(\alpha_i) = \mu_2 + \eta_{2i}$, $\beta_i = \mu_3 + \delta_1 * SI_i + \delta_2 * TL_i + \eta_{3i}$
6. Inflection point dependent on TR: (Model 6): $log(p_i) = \mu_1 + \eta_{1i}, log(\alpha_i) = \mu_2 + \eta_{2i}, \beta_i = \mu_3 + \delta * TR_i + \eta_{3i}$
7. Maximum deaths and inflection point dependent on TR (Model 7): $log(p_i) = \mu_1 + \gamma * TR_i + \eta_{1i}, log(\alpha_i) = \mu_2 + \eta_{2i}, \beta_i = \mu_3 + \delta * TR_i + \eta_{3i}$

3 Results

3.1 Modeling crude cumulative deaths between countries

Model performance table:

From Table 1 showing the model fit criterion, we observed that Model 7 was providing the best fit with lowest AIC and BIC. The predictions from each model is shown in Figure 3.

4 From deaths to cases and severe cases: Age group wise projections

Over the month of April and May, based on our parameter estimates from Model 7, we projected the number of cases from the number of deaths using the case fatality rates in different age groups based on the case fatality rates reported from China and the age groupwise population for India and the age group wise distribution of deaths due to COVID19 in India (requires updated data).In addition, we provide estimates of severe cases based on age-group wise proportion of severe cases from a study on a subset of cases in China (Reference: https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30243-7/)

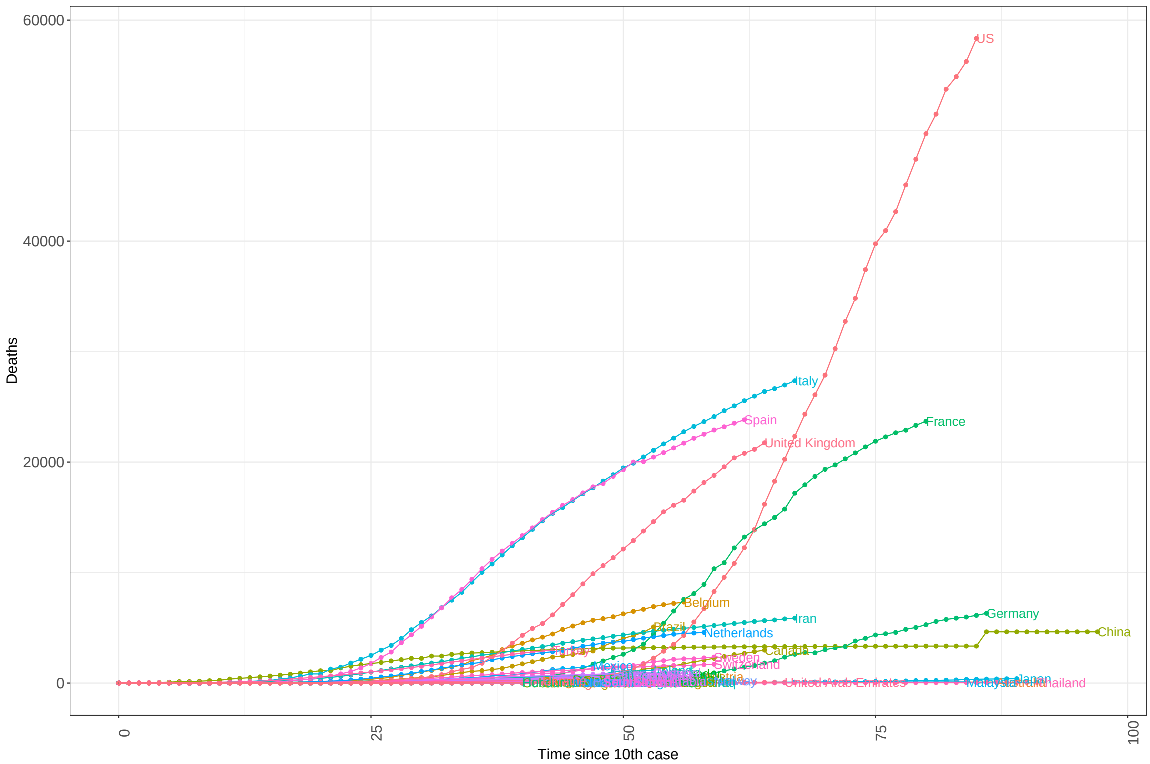


Figure 1: Cumulative number of deaths in countries with at least 25 deaths starting from time since 10th case

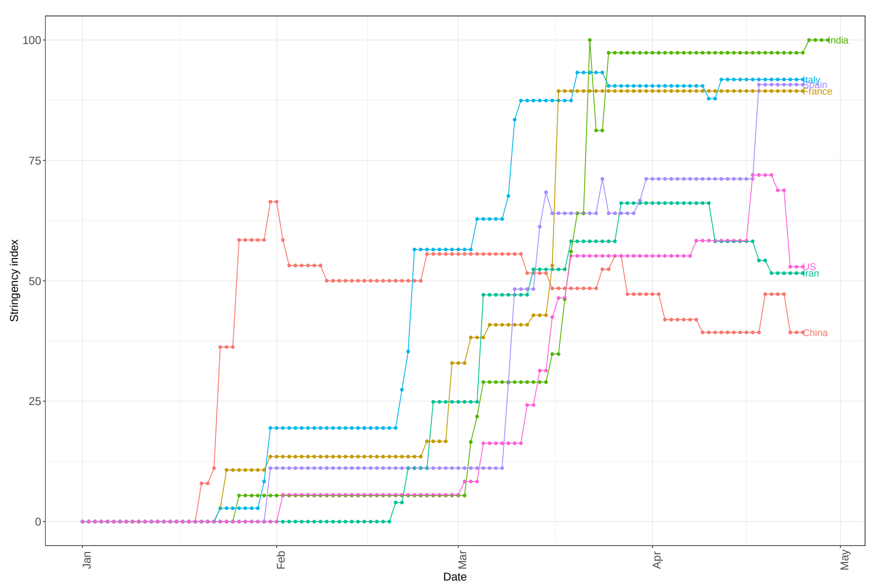


Figure 2: A global stringency index by country considering all measures taken by governments

These computations were carried out in R, specifically using the saemix package for Stochastic Approximation Expectation Maximization (SAEM) algorithm for nonlinear mixed effects modeling.

Figure 3: Predicted cumulative number of deaths in India based on different nonlinear mixed effect models.

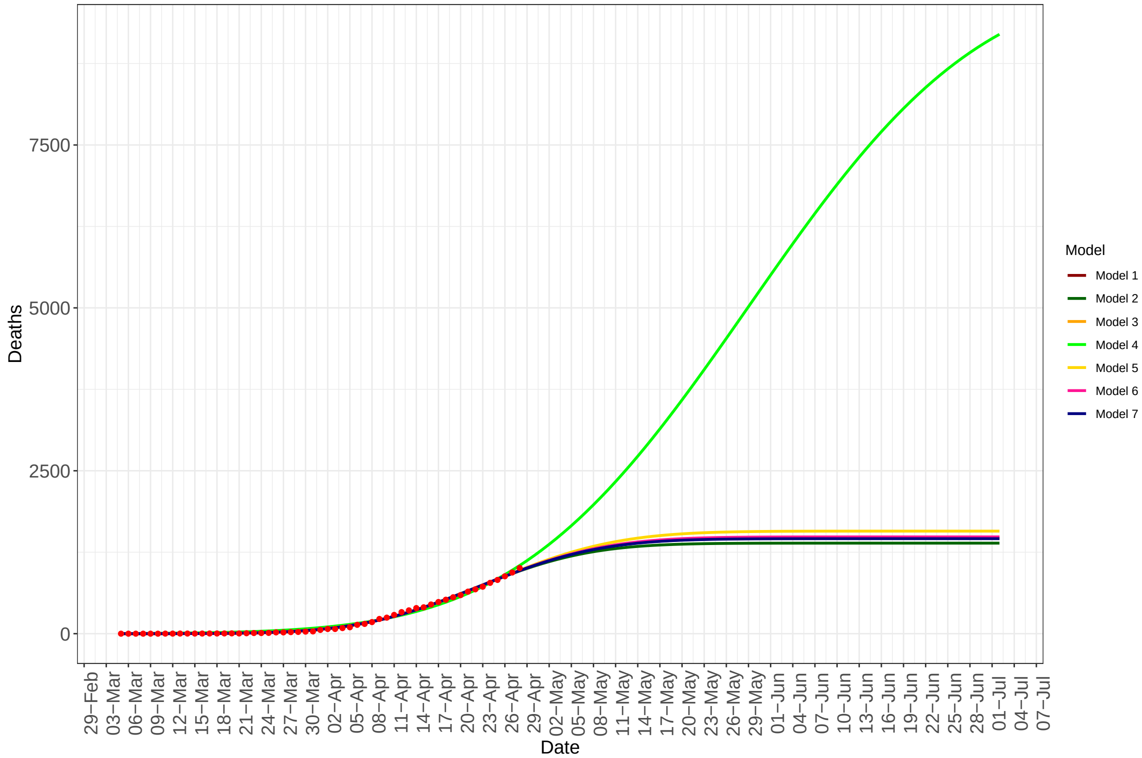


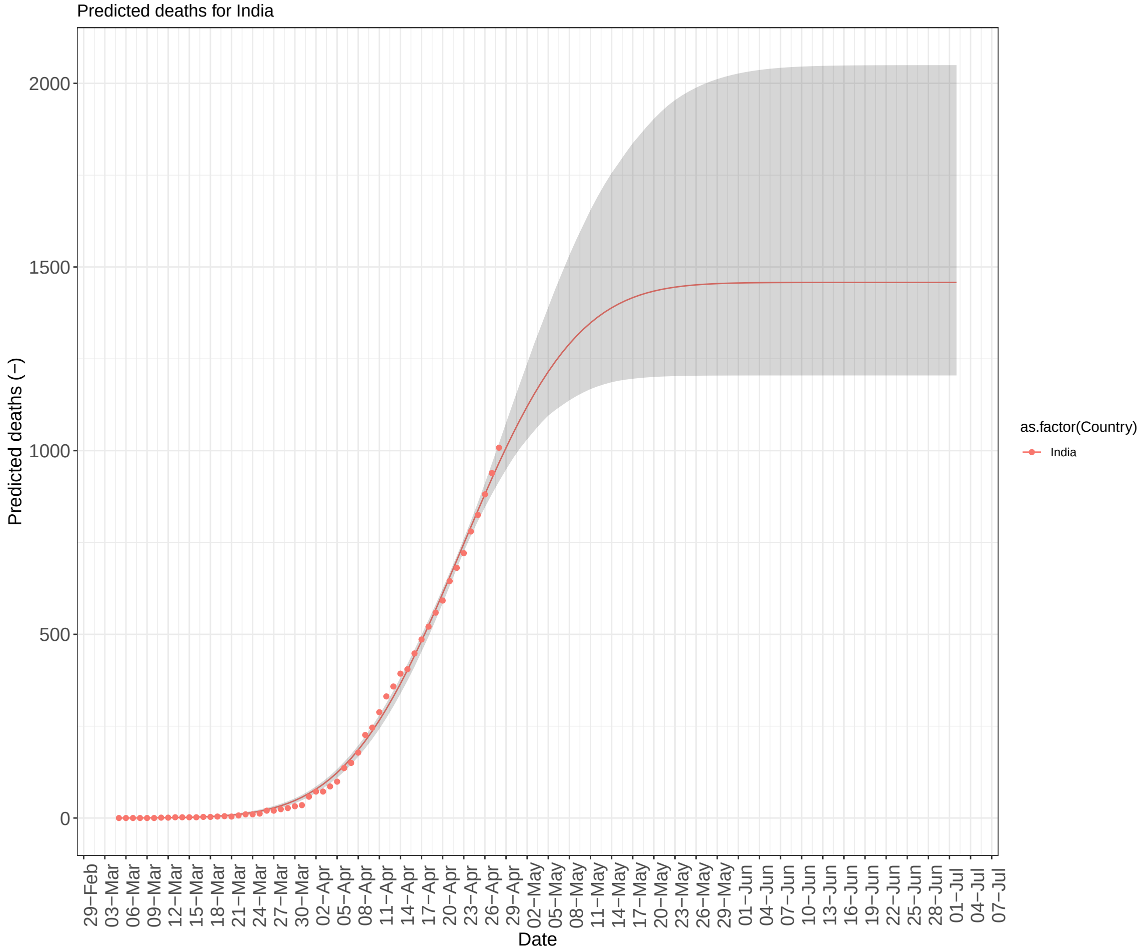
Table 2: Projected number of cases in India based on case fatality rates in China. This is most likely an overestimate. If we had the agewise distribution of deaths in India, this could be more precise.

Age group	Proportion	April 15	April 20	April 25	April 30	May 5	May 10	May 15
0-9	0.01	-	-	-	-	-	-	-
10-19	0.010	2200	3254	4137	4703	4981	5085	5115
20-29	0.010	2200	3254	4137	4703	4981	5085	5115
30-39	0.015	3300	4881	6205	7054	7471	7627	7672
40-49	0.125	13750	20336	25855	29394	31129	31781	31968
50-59	0.125	4231	6257	7955	9044	9578	9779	9836
60-69	0.235	2872	4248	5401	6140	6503	6639	6678
70-79	0.235	1292	1912	2430	2763	2926	2987	3005
80+	0.235	699	1033	1314	1494	1582	1615	1624

Table 3: Projected number of severe cases in India, based on a study in a subset of cases in China.

Age group	Proportion	April 15	April 20	April 25	April 30	May 5	May 10	May 15
0-9	0	-	-	-	-	-	-	-
10-19	0.010	1	1	2	2	2	2	2
20-29	0.010	23	34	43	49	52	53	53
30-39	0.015	113	167	213	242	256	262	263
40-49	0.125	584	864	1099	1249	1323	1351	1359
50-59	0.125	345	511	649	738	782	798	803
60-69	0.235	339	501	637	725	767	783	788
70-79	0.235	215	317	403	459	486	496	499
80+	0.235	129	190	242	275	291	297	299

Figure 4: Predicted cumulative number of deaths in India based on Model 7 along with 95% confidence intervals. The solid dots represent the observed data till 28th April 2020.



5 Modeling cumulative death rates in Indian states

Populations in countries and states vary by age composition, with some states and/ countries with a larger proportion of older individuals. This is particularly important in the case of COVID 19, where older age groups have been reported to have higher mortality. Hence we analyzed statewise cumulative death rates in the Indian states, considering the Italian population as a reference (choice of reference was mainly due to small size of country and availability of age wise distribution of deaths due to COVID 19).
We followed a similar methodology of nonlinear mixed effects modeling as before and modeled the cumulative death rates over time, using the national lockdown as the starting time point (t =0). Figure 5 shows the observed data till 28th April 2020 while Figure shows the predicted cumulative death rates for 10 Indian states, where we have had atleast 10 deaths and where agewise population distribution was available. Most of the states show a flattening of the curve 50-70 days post lockdown (mid May). Although the average population of India is young, special focus needs to be given to states that have 8-10% of population above the age of 60. Inclusion of more state specific covariates such as adherence to lockdown could provide more reliable estimates of these curves.

6 Ongoing work

1. Statewise analysis of age standardized cumulative death rates.
2. Including other covariates such as population density, travel information from Arogya Setu app or other sources.
3. Statewise implementation data on lockdown.

Figure 5: Observed cumualtive death rates in Indian states, standardized to the age groupwise death rates in the Italian population.

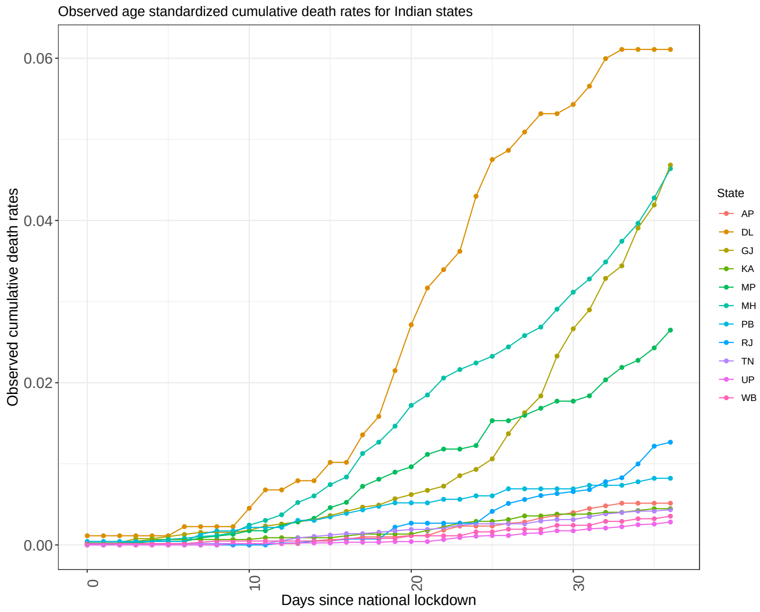


Figure 6: Predicted cumualtive death rates in ten Indian states, standardized to the age groupwise death rates in the Italian population. The curves are generated by simulating from the conditional distribution of the state specific parameters. The solid lines in each panel show the median and the grey shaded region represents the 95% confidence intervals.