INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD • INDIA

Research and Publications

A Quick Estimate the Likely Total Infections and Deaths Due to COVID19 in Select Countries (Version April 1, 2020)

Sebastian Morris

W. P. No. 2020-04-01

April 2020

The main objective of the working paper series of the IIMA is to help faculty members, research staff and doctoral students to speedily share their research findings with professional colleagues and test their research findings at the pre-publication stage. IIMA is committed to maintain academic freedom. The opinion(s), view(s) and conclusion(s) expressed in the working paper are those of the authors and not that of IIMA.

INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD-380 015
INDIA

A Quick Estimate the Likely Total Infections and Deaths Due to COVID19 in Select Countries (Version April 1, 2020)

Sebastian Morris[1]

Abstract

Using a logistic model of cumulative cases and deaths it would soon become possible to give estimates to the final numbers of cases and deaths that are likely on account of COID19, for countries which have gone through about 60+ days since the first cases were recorded. Such estimates assume that the containment and preventive actions continue unabated. We also provide an upper bound to the final cases and deaths that are likely. Right now (with data up to April 1) the projections for Korea, China, Germany, Italy, Spain, Iran, UK have been made. We hope to update the same in due course as the disease progresses. In the UK the deaths are bound to increase, and in Italy the cases could rise further.

[1] Professor, Indian Institute of Management, Ahmedabad. Email: morris@iima.ac.in . Blogsite http://iima.ac.in/~morris/blog.php

INTRODUCTION

The COVID19 virus at breathtaking speed spreads across the globe. This is obvious to all, now that we have a media that is nearly instantaneous and with the capacity to quickly reach out to all. There may well have been other viruses of the same virulence in the past but their spread may not have been noticed the way this virus and its spread is being noted, counted, analysed and responded to. Collective response in crisis can have elements that could be dysfunctional and damaging to society. Thus, panic reaction of the markets can destroy value unnecessarily. Similarly, the fear of being infected can result in behavior like hoarding of medicines, masks etc., to the collective detriment of society. Since the scientific knowledge of the phenomenon is less than adequate for governments and institutions to manage the pandemic in a normal mode, the scope for rumours - soothsayers predicting doom, snake oil peddling, and governance collapse are all possible. They could adversely impact public systems, even when the functionaries and authorities are armed with emergency powers. More so where public systems are weak and the scope for rumors and mob like behavior is large

An estimate of the total number of infections and deaths, which it is unlikely to happen though the course of the pandemic before it comes under control, could help not only governments to gear up with a sense of what they are up to, but also people and markets to make the correct anticipatory choices and responses, and more importantly calm the public, and reduce dysfunctional behavior.

The spread of the disease at the micro level would depend on a whole host of factors that are difficult to incorporate in any model. Such an approach would also require much timely and detailed data. Studies of this kind are necessary for informing action at the local level and should have top priority among scientists and public managers analyzing and the spread.

THE LOGISTIC MODEL

Here we do a quick and very gross estimate of the likely total number of cases of infection, using an aggregate statistical approach, that builds on the well-known understanding that many phenomenon (including the spread of disease, of new species, and growth in a new advantageous environment, spread of pests, growth of organs, besides other human phenomenon like new product adoption, brand penetration) follow a logistic growth equation, with marginal differences. This is because unconstrained growth is exponential but all varieties of growth (other than probably aggregate economic growth) are constrained quickly.

Thus if Y_t is the population (in our case of those infected at any time in the past), then $(\frac{1}{Y})\frac{dY}{dt} = \alpha(1 - \beta Y)$ is the differential equation that underlies the population. In steady state (when the population is not growing), the maximum value of Y reached through a process of growth from small numbers i.e. Y^{max} is given by $\frac{1}{\beta}$. Thus, with data of mere COVID19 cases in a society which includes a period when the significant decline in the growth rates have happened, it is possible to estimate Y^{max}. Since societies improve their response with lockdowns and other methods to limit social contact, which affect the spread process, we can put an upper bound to the infections.[2] The method would be to estimate the following equation:

$(\frac{1}{Y})\frac{dY}{dt} = \ln\left(\frac{Y_t}{Y_{t-1}}\right) = \hat{\alpha} + \hat{\theta}Y_t + e_t$ where β is estimated by $-\hat{\theta}/\hat{\alpha}$. A very simple OLS can be used since the error terms are a priori known to be well behaved. Here we impose the idea of logistic growth on the phenomenon rather than let the data speak for itself[3]. The same is justified since we are sure that at the most abstract (birds eye view) any growth that starts has to be self-limiting in the biological world[4].

The data for the regressions are total number of confirmed COVID19 Cases to date, by absolute day, for nearly all countries as maintained and reported by "Our World in Data". The site has the merit that the data is organized as a readily used panel, for both deaths to date and cases to date. [5] We estimate

[2] Of course, this assumes that the efforts to contain the virus are not reduced from the present level. In societies with larger populations there could be pressures to prematurely relax the social distancing measures as the virus is contained, which then would find a fresh collection of individuals to attack. In societies with smaller populations where, no distinct sub-group has been left unaffected, relaxations may not bring back the spread if external restrictions continue. More importantly the spread of diseases has the characteristic of being a percolation phenomenon during normal times, leading to localization that is possible in percolation models. These require detailed data to be able to model. In a pandemic though a diffusion model with a rate reduction approach works best.

[3] Unlike what economists normally do. The contention here is that irrespective of the local factors that act, and the measures adopted to contain, the growth of infections would follow a logistic curve, with the parameters being depend on these measures and the variations in the potential to spread. No country's specific features can change the form of the curve.

[4] In animal populations that interact with each other, which is the reality in a complex ecosystem the population of organisms (animals, plants, bacteria and eve viruses) would vary with swings that in the abstract has been modeled as predator-prey, competition, generally giving rise to non-linear systems. Here stability of the population of a species is in the sense of Liapunov, in a dynamic non-linear system, i.e. as attractor points around which the population numbers would "orbit". Very early these systems were modeled by Volterra and Lotka. In the case of COVID19, since it is a new organism its population would rise (being paralleled by the rise in cases) to first reach a maximum after which the complex interactions would take place. Hence the validity in the use of the logistic or Verhulst equation.

[5] https://ourworldindata.org/coronavirus-data#confirmed-cases for Deaths; and https://ourworldindata.org/coronavirus-data#confirmed-cases for Confirmed cases.

the equation separately for each of the countries that have had significant days (after the first cases) over which the growth rate has begun to decline.

We report the analysis for a number of countries where the growth rate in the number of cases have begun to decline. Except for China and Korea, for the other countries it is still early days, so that the results are tentative. In most other countries the growth rates have not begun to decline (US, India notably). See Table 1.

Observe that only for Korea, Iran and China, are the estimates of β significant (the t values being -3.812, -4.85 and -5.844 respectively). For Italy it was -1.282 below 2 for 95% significance. For the others it was below 1 in magnitude. In another 5 days i.e. with 5 more points of data we expect the estimation to be possible for more countries.

INTERPRETING THE RESULTS

The forecast of the maximum number of cases for Korea, Iran and China are 9940, 33076 and 83432 respectively, a little above the level reached on 1st April. Thus, these countries if they continue to do what they have been doing and prevent further new imported cases, should be able to hold the level of infections without adding many more cases. The economic price of doing so can be stupendous since their efforts to date have come with closures, large restrictions on movement and massive efforts at making public places safe (especially China and Korea).

For Italy, Germany and Spain while the growth rates have begun to decline it is still too early to give a reliable estimate of the downward trend. Hence while the estimates are close to the maximum reached on April 1, (being under double the levels 127,000; 159,000; and 191,000 respectively), the most conservative (highest estimate) of cases are respectively 0.34, 0.67 and 0.94 millions). With more data being available and with no slippage (and expected further improvement in the efforts at containment), the upper bound of the maximum is likely to decline. For Spain the story is similar to that of Germany. In France, we don't see a steady retardant factor i.e. β so that the upper bound is in the range of 5 million. While it is a reason to examine the processes and measures in place in France, without a significant decline in the growth rate, the numbers of infections could reach much larger numbers than was the case for Korea or China. In another 5 days the model should give much better estimates for these countries. The US, India (being continental economies and in the same population size class along with China) these are still in the early days, but the pointers are to large number of cases.

China is the real positive surprise. Key to the containment, besides all the measures put in place in Wuhan, was the complete isolation of Wuhan itself, preventing any leakage of the virus out of Wuhan /Hubei. But China did take quite some number of days before the retardant factor becomes significant.

The contrast between South Korea and others who witnessed the initial few cases roughly within the space of a week show vastly different trajectories. In South Korea the cases have remained stagnant for over 18 days now, whereas in all others the cases are still rising since the retardant factor has not become large enough to quell the growth. Thus the early neglect in Europe and Iran may have been problematic and these countries would witness much higher numbers of cases. See Figure 1 (b).

From Fig 1(a) notice that Italy and Iran have had many more days over which the growth rate is still positive (continues as on date) though at a declining rate. For France, Germany and Spain the decline in the growth rates have just begun to happen. France is particularly to be noted for the fact that the growth in the number of cases even after 50 days is not significantly being attenuated to give numbers in the low tens of thousands. Right now with its β having high standard error, one hopes that within a few days the data would show a significant and large retardant.

DEATHS

The deaths per case, given the very rapid spread of the disease and the significant incubation and illness period in the range of a month before death takes place of those who are unfortunate, would fall initially then rise marginally, or remain stagnant only to fall quite dramatically as the morbidity naturally reduces due to better treatment, and some expected reduction in the intrinsic biological fatality falls. Deaths per case are likely to be very different across different countries not only because the identification of COVID19 cases could vary much given the different approaches to testing, but also because of variations in the effect of the virus due to genetic differences, differences in age profile of the population (since there is a known distinct higher fatality among older people), the prior diseases and state of health of the population. In addition, possible variation in resistance arising not only due to age and genetic differences but also because of prior history of inoculations, with some like BCG having a small positive spillover[6]. We would not hazard a prediction on the reduction right now since the rate would ultimately fall.

[6] Newman, Tim "COVID-19: Could TB vaccine offer protection?", April 2, 2020, *Medical News Today* https://www.medicalnewstoday.com/articles/covid-19-could-tb-vaccine-offer-protection

See Figures 2 for deaths per case for some countries. Notice the wide variations to start with though with more cases and deaths there is some kind of convergence to a levels over a narrower range.

Figure 3 brings out the log of deaths for a few countries. Deaths unlike cases are more objective since in most countries the cause of death is actually recorded. The measurement of cases is problematic since it would depend on the testing protocols in each country, which vary a great deal.

While early deaths may not have been recorded as being due to COVID19, some deaths may have been missed. With COVID19 now having become a major concern, it would be difficult even in poor countries to not investigate COVID19/pneumonia deaths to distinguish between the two. At this juncture, Korean, Iranian and Chinese death data can be fitted into a logistic model with much predictability to estimate the ultimate number of deaths, which would be very close to the actual numbers on 1 April. We have considered only those countries among the set chosen for which the t-values of the coefficient θ have a magnitude larger than 1.

However, in another week in a number of countries the logistic model should yield results to be able to get broad estimates on the number of deaths in the other countries in Figure 1(a). Meanwhile we report the upper bound at 95% significance of the cumulative deaths that are likely along with the "best" estimates of the same.

Again Korea and China show control over deaths to a far higher degree than the other countries. In the UK deaths would increase substantially, while in Italy and Spain if the efforts are slacked there could be a resurgence. See Table 2.

CONCLUSIONS

Using a logistic model of cumulative cases and deaths it would soon become possible to give estimates to the final numbers of cases and deaths that are likely on account of COID19, for countries which have gone through about 60+ days since the first cases were recorded. Such estimates assume that the containment and preventive actions continue unabated. We also provide an upper bound to the final cases and deaths that are likely. Right now (with data up to April 1) the projections for Korea, China, Germany, Italy, Spain, Iran, UK have been made. We hope to update the same in due course as the disease progresses. In the UK the deaths are bound to increase, and in Italy the cases could rise further. For the world as a whole, the model for Cases is just beginning to be meaningful. The forecast is about 1.25 million cases, with a maximum of 3.4 million, when the cases as on 1st April stood at 0.85 million. Deaths for the world are forecasted (with little confidence) at 72000 with an upper bound of 227,000, while the deaths till 1st April have been 41,887.

Figure 1 (a)

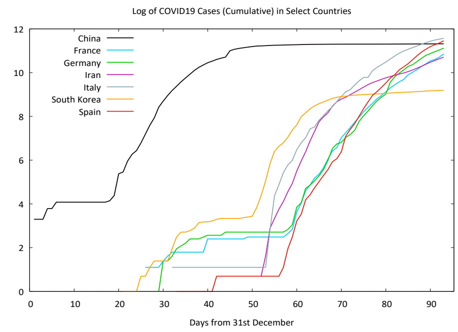

Fig 1 (b)

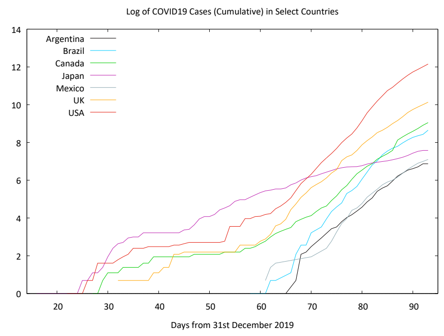

Figure 1 (c)

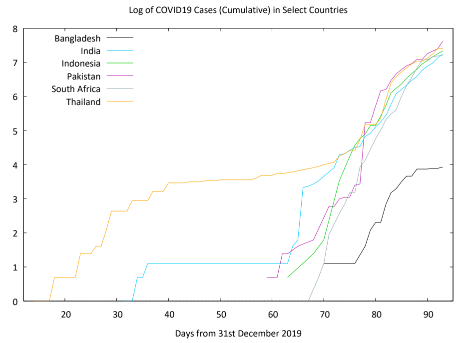

Figure 2(a)

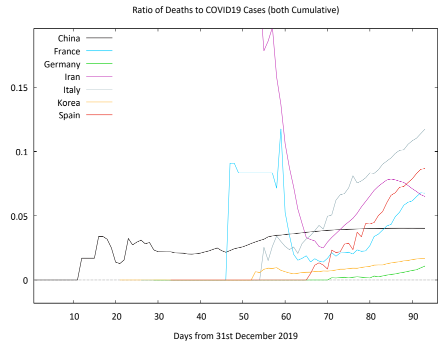

Figure 2 (b)

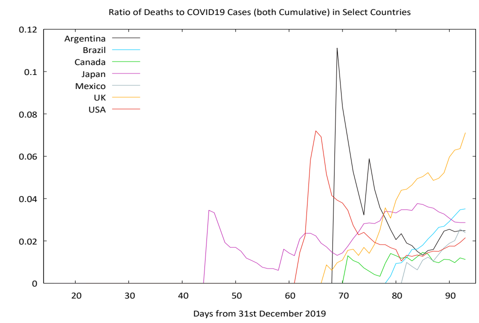

Figure 3

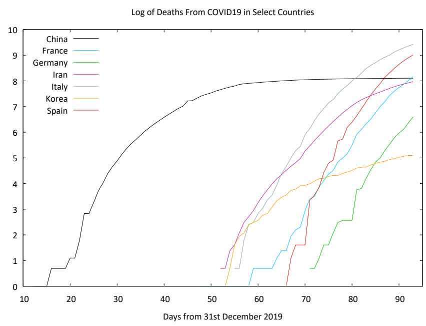

Table: Regression results of growth rate in total COVID19 cases to date on COVID19 cases to date; and estimates of maximum number of cases							
	Korea	Iran	Italy	China	France	Germany	Spain
Constant (α)	0.201987	0.399701	0.203382	0.194349	0.147712	0.183609	0.20377
Coefficient of Y (θ)	-2.03E-05	-1.21E-05	-1.60E-06	-2.32E-06	-3.28E-07	-1.52E-06	-1.06E-06
Std_error (α)	0.0287557***	0.0452762***	0.0461223***	0.02332***	0.228504***	0.0318127***	0.0309379***
Std_error (θ)	5.33E-06***	2.49E-06***	1.25E-06	3.99E-07***	1.69E-06	1.66E-06	1.16E-06
R-squared	0.171898	0.376231	0.027107	0.275078	0.00579	0.00773	0.014234
F	14.53 (1,70)	23.52 (1,39)	1.64 (1,59)	34.15 (1,90)	0.04 (1,65)	0.48 (1,62)	0.84 (1,58)
Growth rate (% per day)	20.2%	40.0%	20.3%	19.4%	14.8%	18.4%	20.4%
Retardant factor (β) (per unit of Cases)	1.006E-04	3.023E-06	7.871E-06	1.199E-05	2.224E-06	6.275E-06	5.223E-06
Estimated Y (max) (1/β)	9940	33076	127054	83432	449728	159354	191448
Error in Y (max)	6005	15749	208876	35292	4.69E+06	467430	427486
Y (upper bound)	15945	48825	335931	118724	5.14E+06	626784	618934
Y(2nd April)	9786	44606	105792	82298	5.21E+04	67366	94417

Table 2: Regression of growth rate in total COVID19 deaths to date to COID19 deaths to date							
	Korea	Iran	Italy	China	UK	Spain	World (very tentative)
t-value of coeff of Deaths D (θ)	-3.955	-5.035	-3.418	-6.298	-1.243	-1.839	-0.9413
Expected cumulative deaths	137	2953	12375	3125	3199	8665	71586
Maximum level of expected cumulative deaths at 95% significance	221	4244	20253	4365	8497	18813	227124
Deaths cumulative as on 1st April	163	2898	12430	3310	1789	8189	41887