Calculation of the corrected time evolution of COVID-19 cases in the USA

Ananth Dodabalapur

The University of Texas at Austin

Note: This is a living document and will be revised as additional information becomes available. This is expected to occur in a fast-moving situation.

Abstract

The time evolution of COVID-19 cases in the USA is highly inaccurate as a result of serious testing inefficiencies and biases, especially in the earlier stages. In this report, the corrected time evolution of such cases is reconstructed from death data, the average time between case recording and death in China, and the estimated overall death rate of COVID cases. The corrected time evolution in number of cases is compared to the officially reported case number evolution and a huge discrepancy exists between the two. The findings indicate that (i) the pandemic likely started earlier than though; (ii) the official number of cases is underreported by more than a factor of about 10; and (iii) the rate of increase in the calculated number of cases with time is slightly less compared to that of reported data.

Data Sources, Assumptions, and Calculation Procedure

The data on the number of deaths due to COVID 19 is understood to be relatively accurate due to the detailed diagnostics that are performed for all deaths. This data is obtained from Ref. 1. For cases that result in death, the average time duration between reporting of a case and death in the USA is not known accurately at this time, and will become available at a later stage. Such data from China is available from Ref. 2 and are used in these calculations. Such data may also be calculated from raw data in Ref. 3 for China. Data from S Korea is too noisy on account of the relatively small number of deaths. When USA data becomes available, a substitution can be made for greater accuracy. A key assumption is that the USA data will be similar to the China data that is used in the calculations reported in this report.

The overall death rate from COVID-19 is assumed to be between 0.5-3.0% based on various reports. When all cases, including asymptomatic cases, are included, the death rate is below or equal to 1% in countries such as S. Korea and Germany with extensive testing.

The number of COVID-19 cases is calculated by dividing the number of deaths by the death rate and displaced in time backwards by 9 days. The 9-day time displacement is reflective of the average time between case recoding and death in Ref. 2.

Results and Implications

The calculated results are shown below in Fig. 1 for a death rate of 1% and compared to the reported data in Ref. 1. The calculated number of cases is very much larger (by more than a factor of about 20) compared to the official data. The total number of cases in the USA exceeds 1 million on 26 March. The rate of growth in the number of cases is slightly less that that obtained from official data.

In Figure 2 is shown the ratio between calculated results of the total number of COVID-19 cases in the US and reported data from Ref. 1. This ratio is initially ~ 40, due to very inadequate testing and decreases to about 20 for later dates. This factor decreases to below 20 based on data extrapolated beyond 16 March. In Fig. 3, is shown extrapolated calculated number of COVID-19 cases for death rates of 0.5%, 1%, 2%, and 3%. These percentages range from the lowest reported death rates to 3%. Some countries, such as Italy, have reported higher death rates than 3%. This is likely due to lower levels of testing and possibly factors such as enhanced exposure of senior citizens to the virus resulting in more deaths. The total calculated number of cases decreases with increasing assumptions for the value of the death rate.

COVID Hospitalization – A test for this model:

New York state has the most accurate hospitalization rates among states and it is reported quickly. Governor Cuomo said the hospitalization numbers (and dates) are: March 18 (549), March 20 (1000), March 23 (2000), March 24 (3000), March 26 (5327). On March 26, NY had about half the cases in the US. Using this data to extrapolate to the US: There were approximately 2500 NEW hospitalizations on 26 March in the US. In my simple model (9 days from test to death, in case of death), I included an average of 4 days from test to hospitalization (for cases that result in hospitalization) and an average of 5 days from hospitalization to death (for cases that result in death), consistent with the overall average 9 days from testing to death (for cases of death).

Following the incremental set of people (who number 7500) who tested positive on 22 March all over the US, the model estimates that 2500 of these get admitted to hospital on 26 March (average) and of this set, 1000 will die on 31 March. 1000 is the extrapolated incremental death figure for 31 March for the US. According to the model, 75000-85000 people who have the virus were not tested on 22 March (for the 7500 who were).

The average death rate is 1-1.5%, and the hospitalization rate is about 2.7% as illustrated in Figure 4. These are average figures for the whole US. It is recognized that geographic and demographic variations will exist. The calculations are explained below:

Death Rate = 1000/(7500 + 85000)x100 % = 1.1%

Hospitalization rate = 2500/(7500 + 85000) = 2.7%

The 85000 number is the number of people who would have tested positive on 22 March if (hypothetically) everyone in the US was tested. In fact, only 7500 tested positive that day.

It must be noted that 7500 positive tests on 22 March and 2500 new hospitalizations on 26 March, are real numbers. The 1000 new deaths on 31 March is got from an extrapolation of trends. The 1-1.5% death rate is an important boundary condition that all experts agree on. For all these numbers to work, the set of key assumptions that were made in this report have to be valid. To recap, the key assumptions are that fewer than 10% of the actual positive cases are tested and that the death rate is about 1%.

Summary

In summary, the chance that someone in the US **with the COVID-19 virus** will end up in hospital is 2.7% and the chance that a person with COVID-19 will die is about 1%. The chance that they will be tested is currently less than 10%. The percentage tested will likely change in the future but the other two numbers should not change too much until an effective treatment is discovered.

Acknowledgments: The author thanks Xiao Wang for assistance in plotting data, several colleagues for discussions and encouragement, and Kelly Liang for web posting.

References

(1) https://www.worldometers.info/coronavirus/country/us/

(2) https://csblab.github.io/novoCoronavirus-Analysis/

(3) https://ourworldindata.org/coronavirus

(4) https://www.cdc.gov/mmwr/volumes/69/wr/mm6912e2.htm

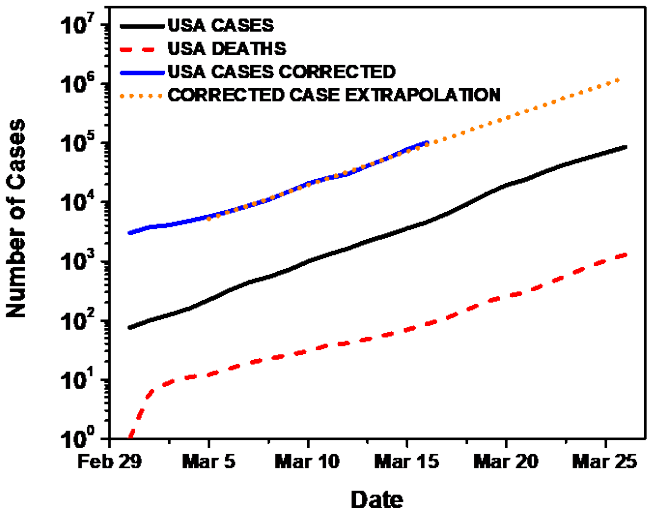

Figure 1. The number of deaths due to COVID-19 in the USA [1] along with the reported number of cases [1] and the calculated number of cases on various dates between 1 March and 27 March 2020. The calculated number of cases is extrapolated till 26 March 2020.

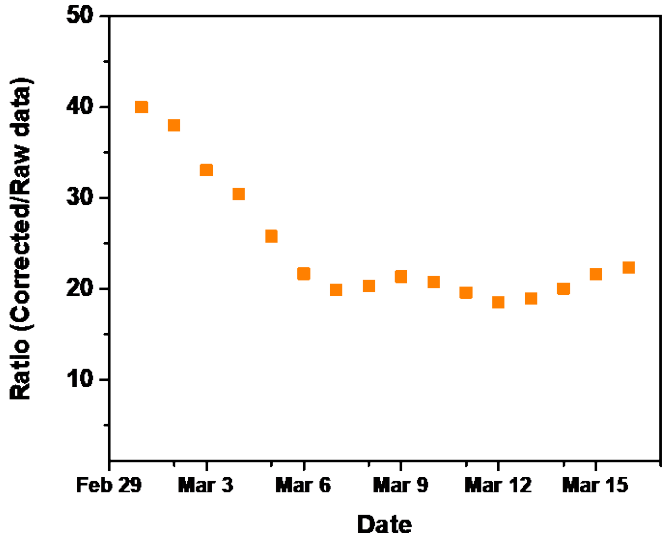

Figure 2. The ratio between calculated COVID-19 case numbers and the reported data in [1] for various dates.

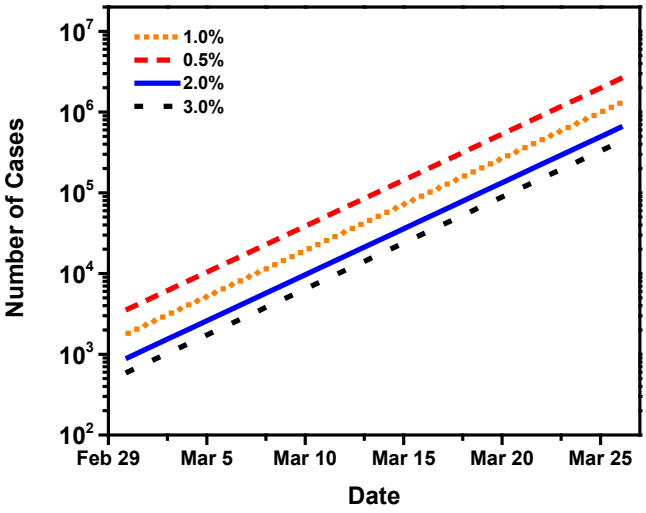

Figure 3. Extrapolated calculated number of COVID-19 cases in the USA for various death rates in the range 0.5%-3%. The growth rates are exponential.

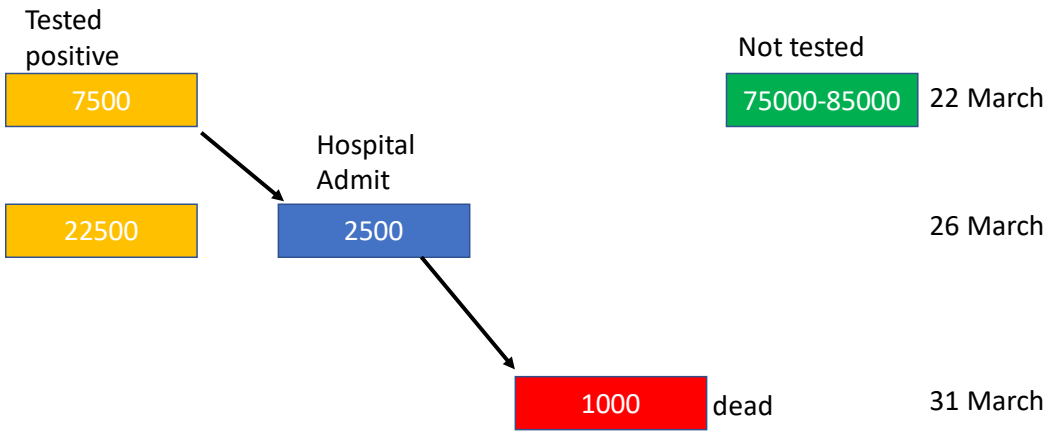

Fig. 4: Incremental analysis that follows 92500 people in the US. If all of the US were tested these 92500 will have tested positive on 22 March. In actuality only 7500 people tested positive that day due to insufficient testing. Of these 7500 people who tested positive on 22 March, 2500 will have been admitted to hospital on 26 March and of these 1000 will die on 31 March. These are average numbers and correspond to a death rate of 1.1% and hospitalization rate of 2.7% overall.