March 30 Coronavirus Report

By Prof. Ken Comer, Computational Social Sciences Program, George Mason University

Perpetual optimism is a force multiplier. Colin Powell

Summary

The trajectory of COVID-19 cases and morbidity in both Italy and the United States continues to track along a logistic curve. Logistic curves go through exponential growth periods, reach an inflection point, and level off at a maximum. The maxima for the four fitted curves in question are not as large as most published estimates and will occur some time in late April. These observations are based on data through 30 March, and are essentially unchanged from the last report, 27 March.

Background

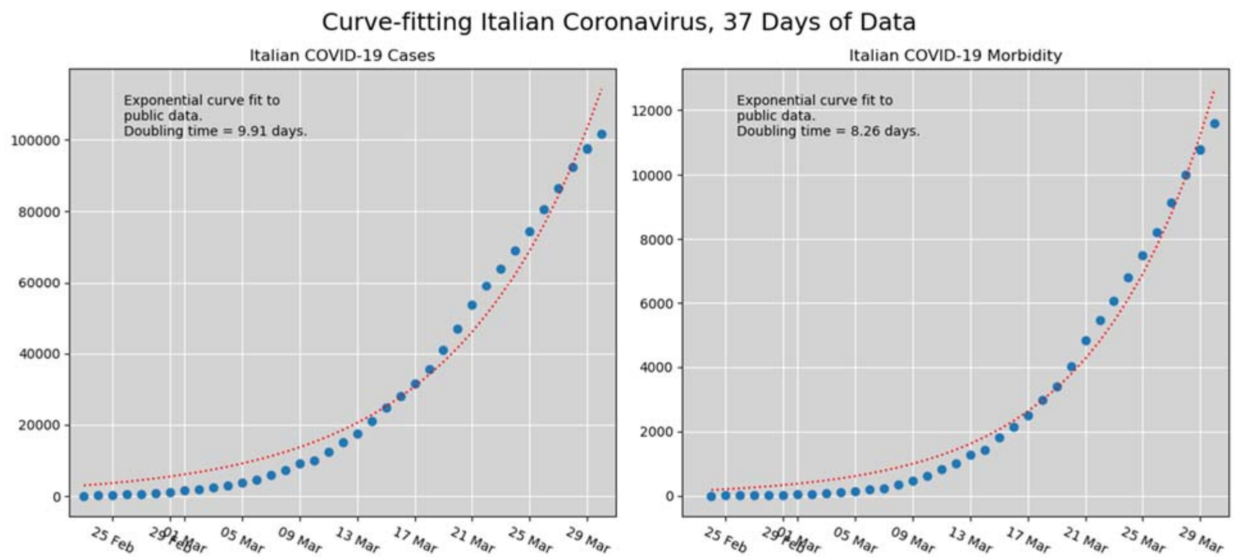
This is a *statistical* modeling effort. In essence, it is a curve-fitting exercise. As Newton said, "I make no hypothesis" as to why the data is following a particular curve. I just use a simple function in Python, the `scipy.curve_fit()` function, to find the parameters of a curve that causes it to deviate least from the known data points. You will see a red dashed curve that represents the best fit and blue data points. The data comes from worldometers.com.

The word "modeling" has come to mean much more. There is a process of *system* modeling in which the rules, relationships, boundaries, constants, inputs and outputs are hypothesized and then programmed into software. This *also* produces future output curves and error estimates. It's what I do the other 99% of my professional life. I will leave the presentation of my COVID-19 system modeling ideas for a different place and time. Here, we simply fit curves to data.

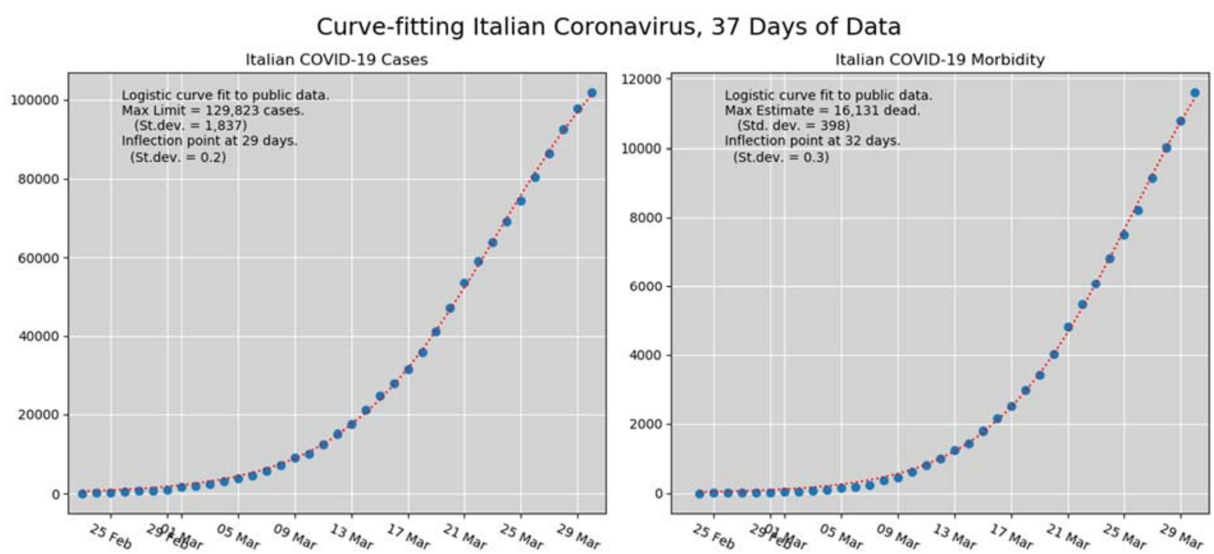
So the results you see have been selected by the Python code. That includes timing of the curve and maximum limits. The only subjective inputs I make are the selection of the data to use, the class of curves to fit (linear, exponential, logistic, etc.) and the choice of countries and variables to model. Case counts and mortality are the most stable for all data, so that choice is obvious. Epidemic spread theory points to a logistic curve as the form for cases and mortality. At the end I will discuss why I continue to focus on Italy and the US, and what we learn when we direct this analysis elsewhere.

Italian Analysis

Pandemic growth in Italy has clearly moved out of the exponential growth phase. The exponential growth best-fit curves have an increasingly large "white space" between the estimate and the data. Even the poorly-fitting exponential estimate curves have metrics that suggest the infection velocity is slowing down. The doubling time for cases was 8.32 days in the last report and is now 9.46. The doubling time for deaths was 6.96 and is now 7.91 days. It's an open question as to how much longer these metrics make sense and even whether I should present exponential growth curves.



There is also significant improvement of the values on the Italian logistic curve. Foremost, the logistic curve clearly fits the historical data for both cases and deaths. In the future it may move away from this, but it is clearly on an S-curve trajectory. The inflection point is clearly visible in the case curve (right).



For Italian cases, the expectation is to level off within a few thousand of 130K. As they are past 100K cases right now, this suggests they don't have very far to go. Over the past three days, the maximum case projections have been 117K, 121K, 125K, and 128K. The inflection point for cases started around 26 days, but has been at 29 days for two days now. This point was reached a week ago.

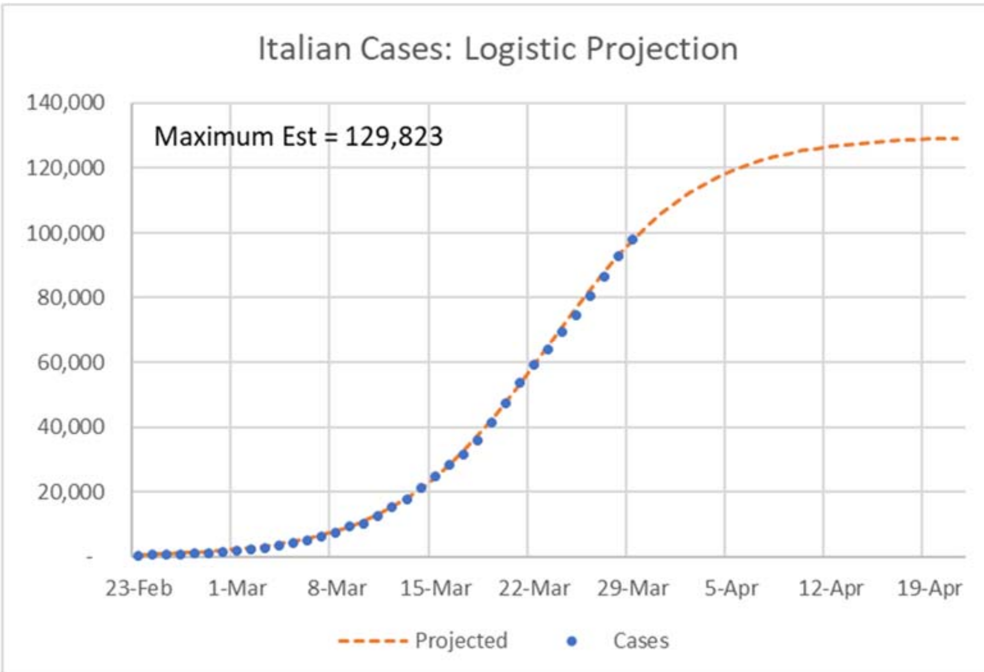
For mortality, a logistic curve fit to the Italian data tops off at 16,131, with a standard error of 398. Thus,

29 March	15,498
28 March	14,773
27 March	13,471
26 March	12,701

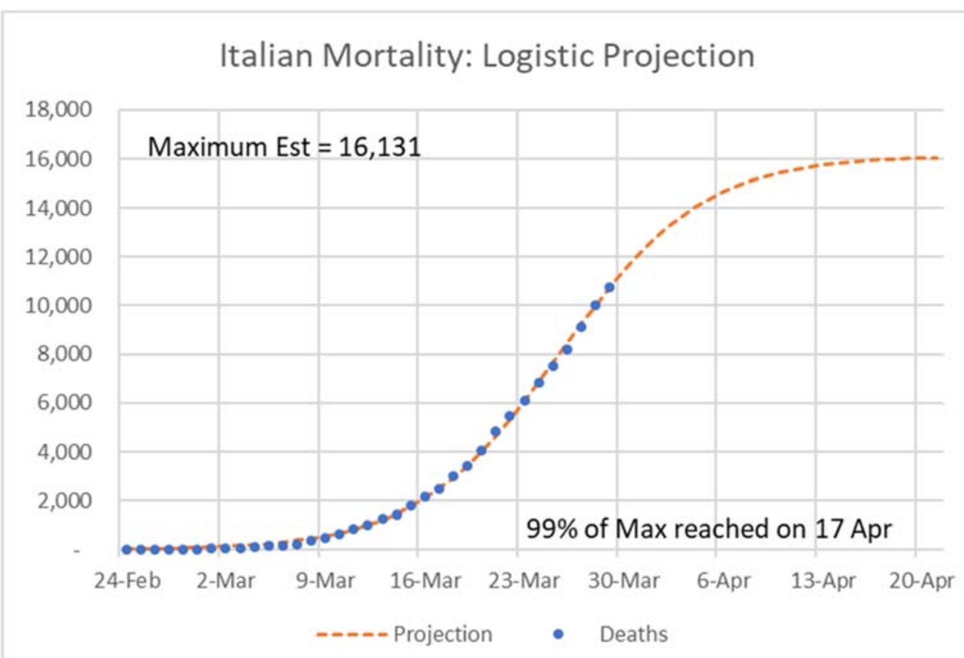
the 95% confidence interval would be 15,351 to 16,911. This number has been drifting up over the past few days, but it is unlikely to reach 20,000. As I noted in earlier reports, Italian leaders have noted the tendency not to include all of the victims of the epidemic. It is assumed that this underreporting behavior will continue for the rest of the epidemic.

As before, I can place the best-fit curve parameters on a projected trajectory to show where and when it is likely to level off. As you can see, it can be expected that the Italian epidemic will be mostly over by Easter.

For cases:



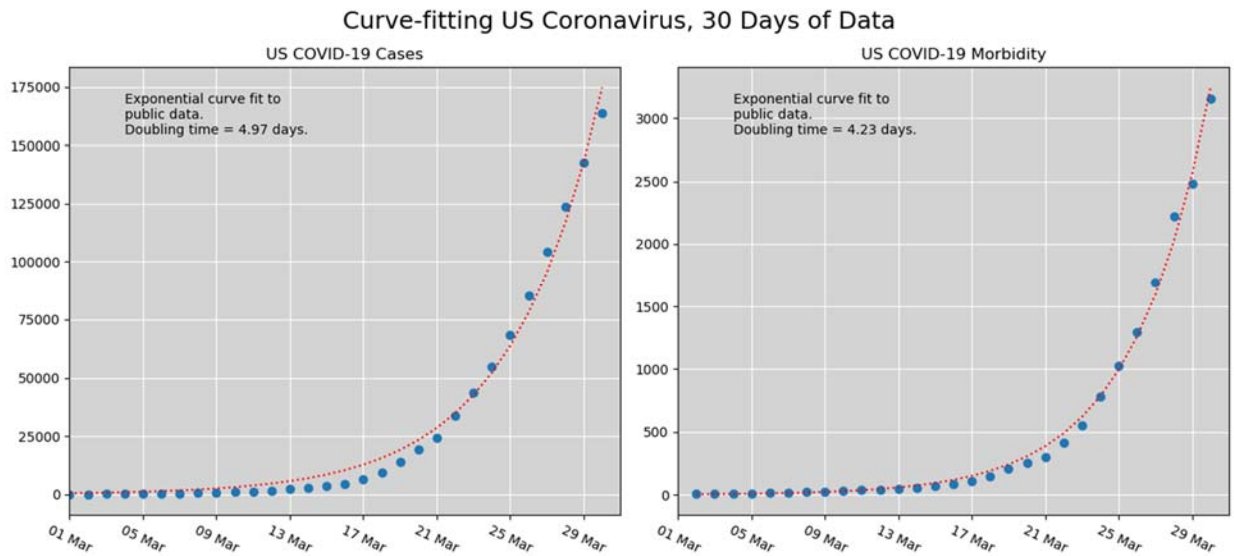
And for mortality:



These are all based on the assumption that the future data falls on the same curve that fits the past data. But, things would have to change quite radically at this point for the Italians to experience an outcome substantially different than this.

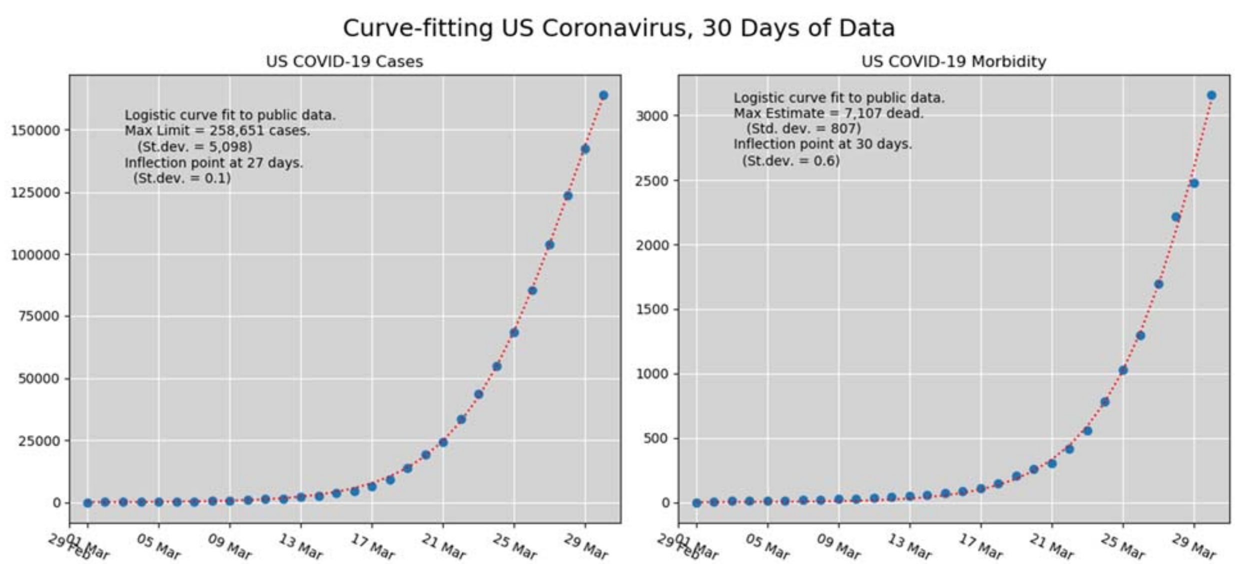
US Analysis

The US trajectory for COVID-19 is also becoming clearer, although the visual cues are a bit more subtle. First, the exponential phase seems to be coming to an end in the US as well:



Doubling time continues to grow for both these curves, as does the mismatch in curvature.

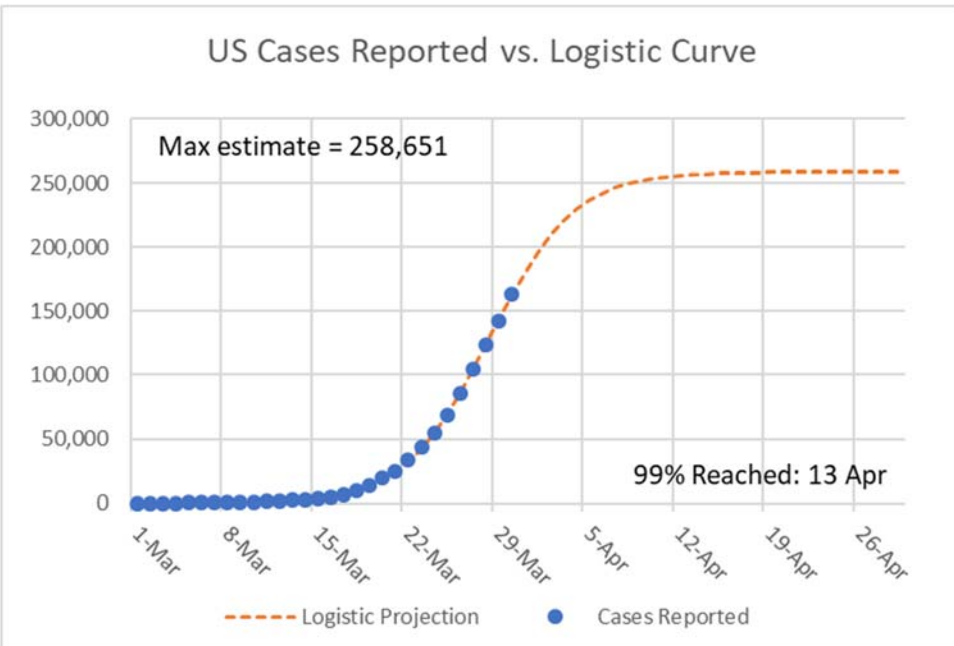
And, for the US logistic curves, the picture is much clearer than on 27 March. In the same way as Italy, a logistic curve much more closely matches the US data. This is confirmed by the small errors in the parameter estimates (curve max and inflection point), three of which fall within 2% of their values. The estimate of maximum morbidity for the US is still quite broad at 11% of the estimate.



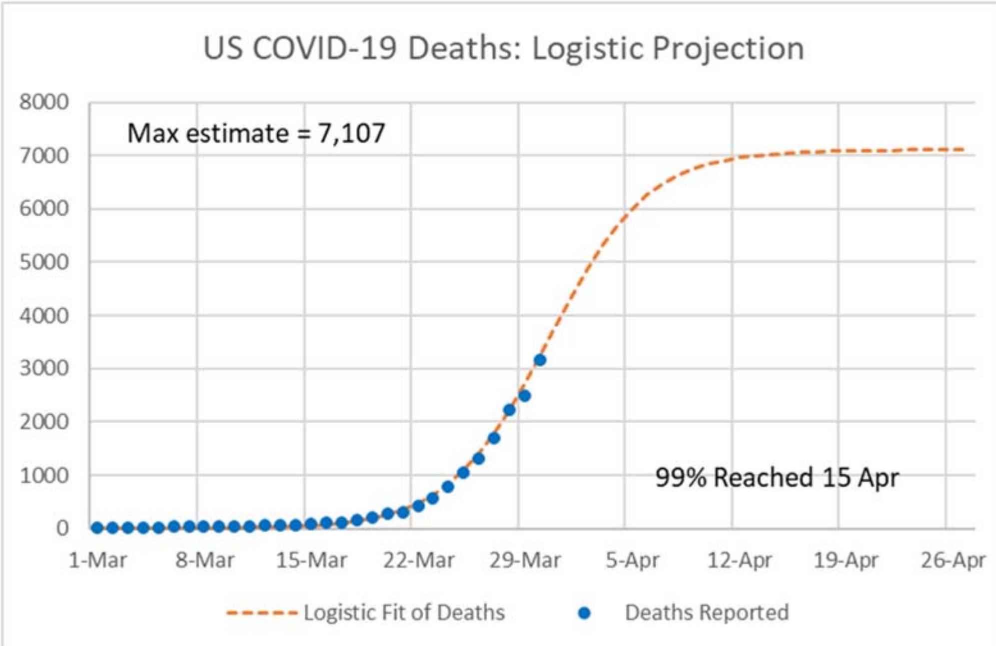
Date	Est of Max Cases	Est of Max Deaths
26 Mar	195,605	62,908,991
27 Mar	227,149	37,342
28 Mar	242,110	47,141
29 Mar	244,077	5,540
30 Mar	258,651	7,701

The curve-top estimates for the ultimate values of cases and deaths have been trending down. Morbidity, which could not be realistically estimated four days ago has seen a truly dramatic reduction. The large estimate on 26 March was the first time any actual number could be deduced. Prior to that the curve estimate was on the order of 10^{180}, so the last three additional points have been extremely important to the final estimate. The estimate of max cases has been trending up, but this can be expected to settle in in the same way the Italian numbers have.

The timing of the US curves also seems to have resolved. Both inflection points appear to have been reached. As the logistic curve is symmetrical around the inflection point, the top of the curves can be expected to occur in late April.



For the first time, there are enough parametric estimates to draw the US morbidity curve:

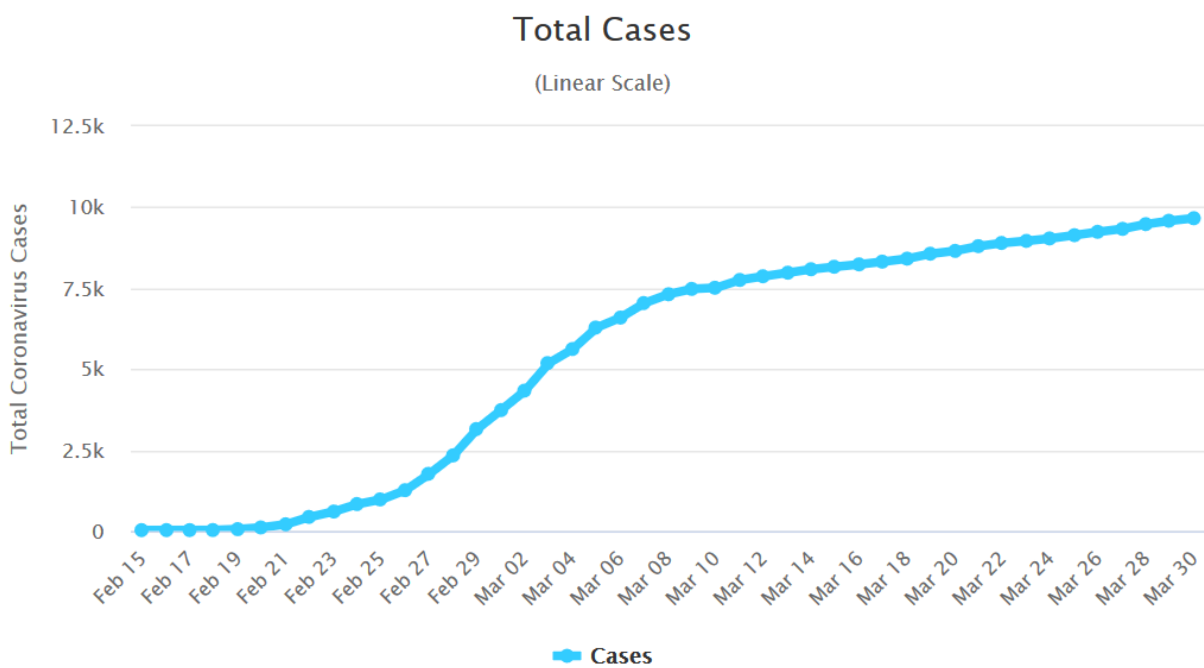


There are important similarities between the US and Italian curves. The inflection point for morbidity is settling in at three days after the inflection point for cases. While the ultimate number is drifting up, the inflection points are moving very slowly. And, all of the standard errors in the estimates are coming down.

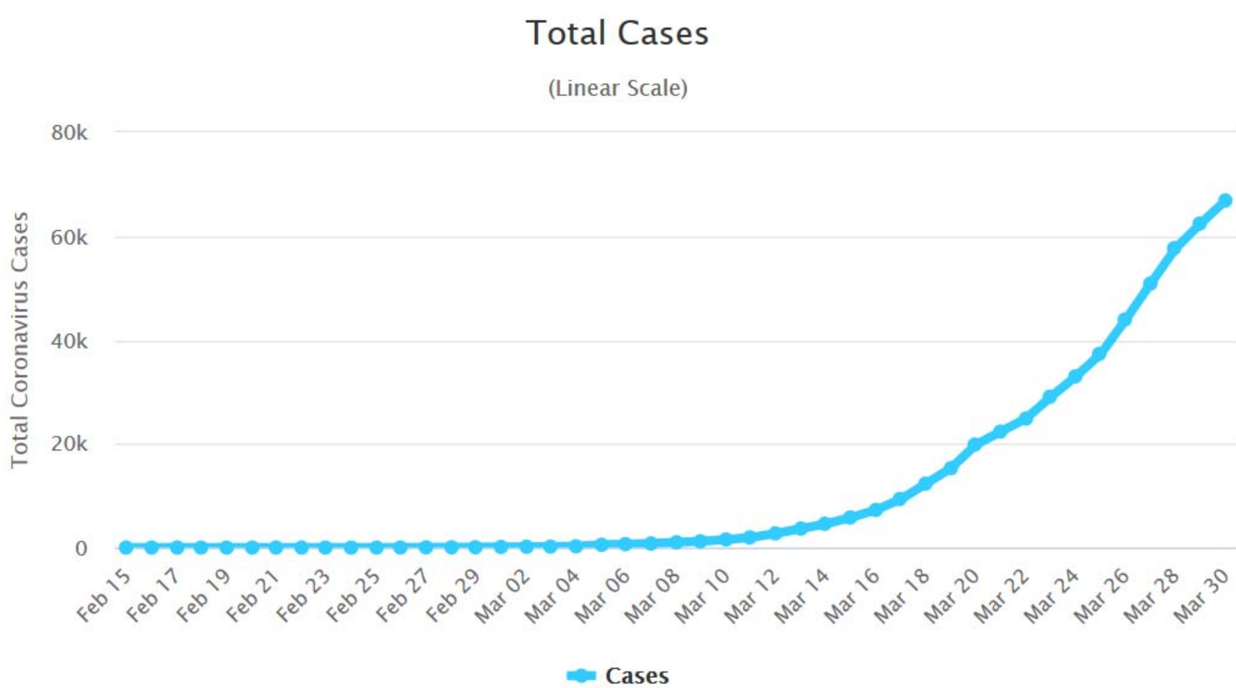
Context

I chose Italy and the US because their curves appeared the most regular. A week later they still can be fit to a single, simple function. Other countries' curves shift somewhat in their shapes. Germany and South Korea, for example, shift *down*.

South Korea (below) shifted in early March to what is essentially linear growth:



Germany (below) is a bit more subtle, but there is a drop in this curve away from a smooth function around 20 March:



Conclusions

It’s important to realize that, for Italy and the US, we are seeing the results of radical reduction in public mixing. Italy shut down in early March and the US did mostly telework around March 16. If the curves were on a different, sharper trajectory in those early days they would have introduced only small errors. These curves are mostly driven by the data from the past two weeks.

This analysis is all conducted under the assumption that the curve patterns will persist into April. Certainly, the circumstances may change and the numbers accelerate. For those countries that have deviated from the logistic curve, however, it is more likely that they shift down (South Korea and Germany) instead of up. The most prominent “late growth” country is Hong Kong, but that is only in terms of reported cases. As of this date there have only been four deaths in Hong Kong from COVID-19, according to worldometer.com.

We’ve been this way before: There is precedent for substantially overestimating disease spread. In the summer and fall of 2014 great concern arose about the threat from Ebola in West Africa, where it had not previously been seen. The US CDC, based on computer models, estimated in September 2014 that there would be 1.4 million cases in Sierra Leone and Liberia by January 2015. In reality, the final case count for the West African epidemic was 30,000. (King et al. 2015; Althaus 2016)

References

Althaus, Christian L. 2016. “Modeling the Ebola Epidemic : Challenges and Lessons for the Future.” Application/pdf. https://doi.org/10.7892/BORIS.91589.

King, Aaron A., Matthieu Domenech de Cellès, Felicia M. G. Magpantay, and Pejman Rohani. 2015. “Avoidable Errors in the Modelling of Outbreaks of Emerging Pathogens, with Special Reference to Ebola.” *Proceedings of the Royal Society B: Biological Sciences* 282 (1806): 20150347. https://doi.org/10.1098/rspb.2015.0347.