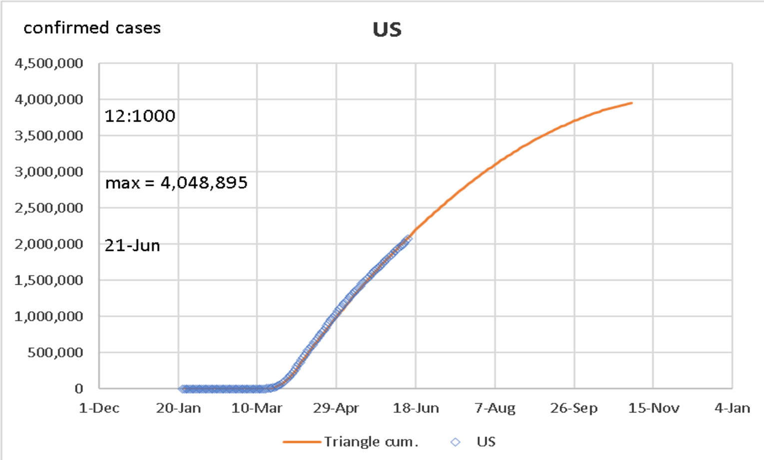
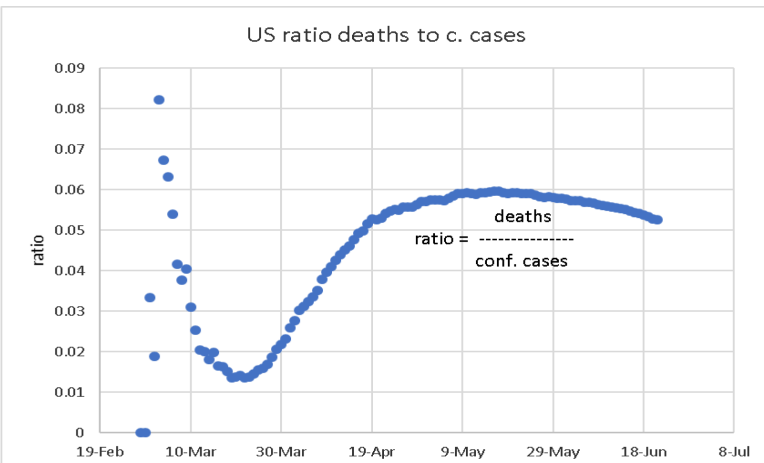
As of 22 June 2020, the US Covid-19 epidemic looks like this, cumulative:[1]



The ratio of deaths to confirmed cases looks like this:



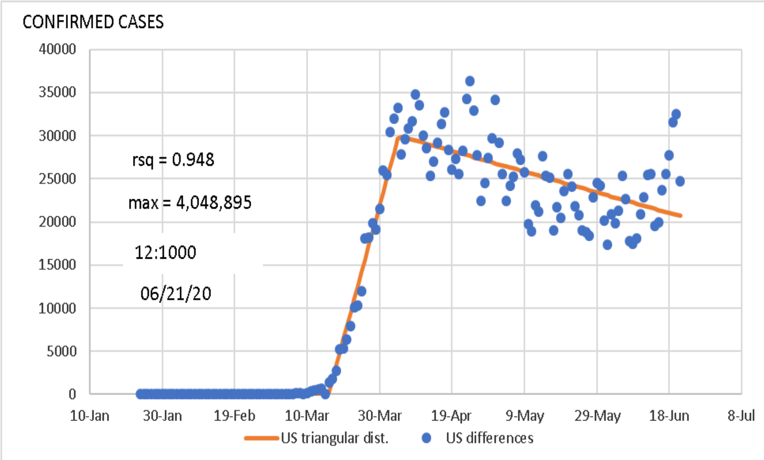
So, it appears that about 5.3% of confirmed cases are classified as due to CV-19, which shows this is no common flu, regardless of discussions on how these are classified.[2]

The distribution of this data could be applied locally, assuming the US case applies to the local situation in the Big Bend of Texas. This data has now been fitted with a triangular distribution, instead of gamma, since it fits better (so far). The Big Bend apparently lags in time the urban populations by quite a bit (see local graphs on this site). But since an epidemic is nothing but a deadly diffusion phenomenon, sooner or later it reaches some saturation level that is hard to predict, often claimed to be 60 to 70% of the population.

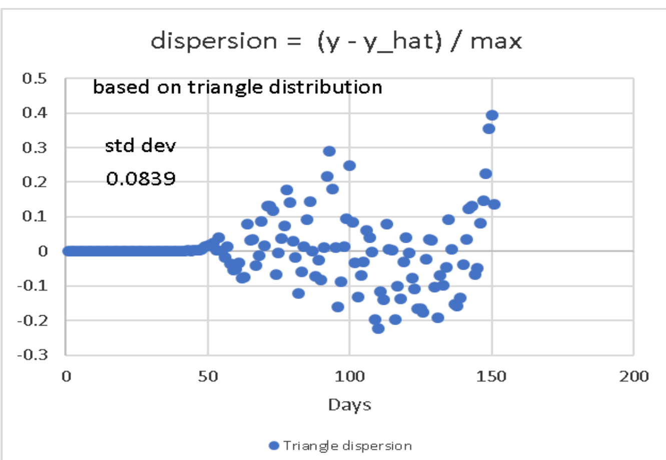
[1] https://github.com/CSSEGISandData/COVID-19

[2] The drop in the ratio could be attributed to more testing revealing more confirmed cases, though. Or better outcomes as treatments improve. Or both.

US appears like this, on a *daily* basis, not cumulative:



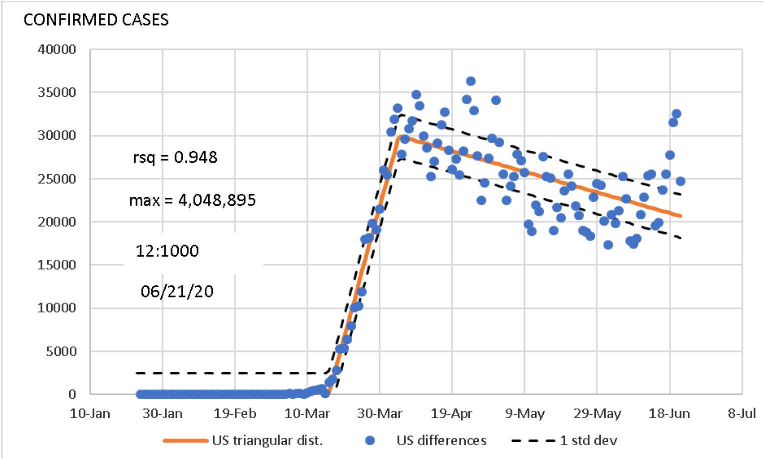
The curve fit is a triangular distribution, used to characterize disease, population growth, and epidemics. A measure of dispersion is the standard deviation of the residuals scaled to a relevant constant, such as the maximum value of the curve, or $(y - y\_hat) / y\_hat_{max}$. *y* is data recorded (in this case, the difference between each pair of successive values on the US version of the previously shown *cumulative* data), *y_hat* (pronounced *wye hat*) is the predicted value at each date of the curve fit, and y_hat_{max} is the maximum value of the distributed curve fit. The dispersion looks like this (below) and gives an idea of the variability of the data around the distribution. R^2 is for goodness of fit, shown as "rsq" on the graph, 1.0 being a perfect fit. Notice that you can have a nearly perfect fit still with a lot of dispersion. Or vice versa.[3]



---

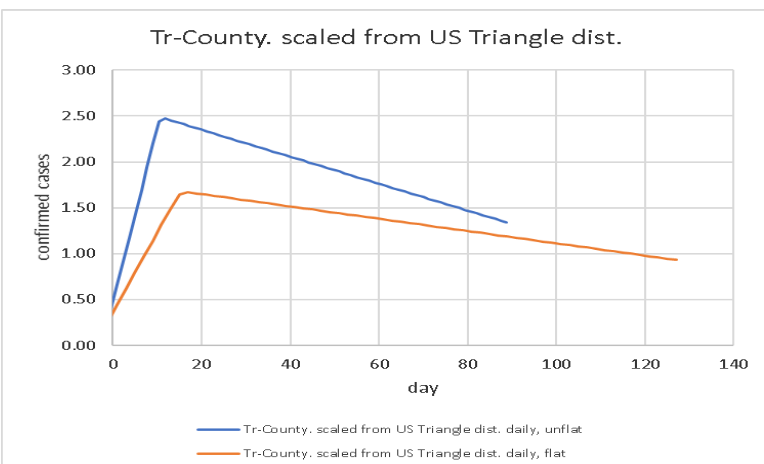
[3] As mentioned before, the gamma distribution was not fitting very well. That is why the data was switched to triangular. Also see Iran in accompanying graphs for a striking example of what can happen to these distributions; it's easy to fall prey to projections like these.

The standard deviation of the dispersion values can be used to construct a 1 standard deviation boundary around the curve fit:

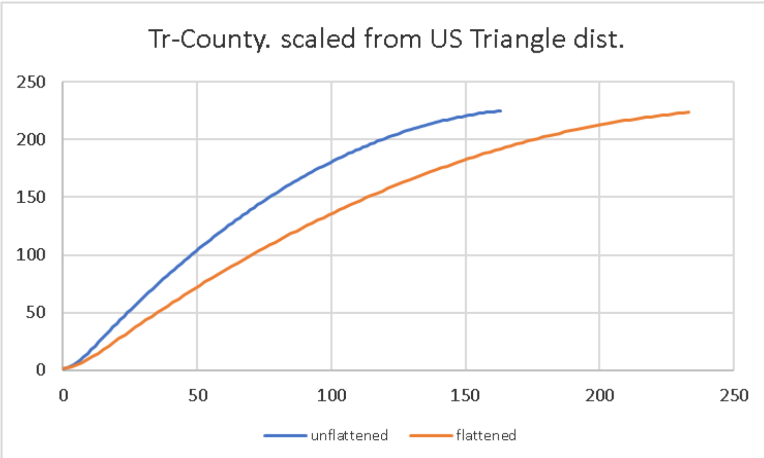


This data is for a population of 330 million but why not scale it to a Tri-County population of 18,000, in order to have a better idea of what to expect here? You certainly wouldn't expect the Tri-County to be *worse* than this, although time will tell.

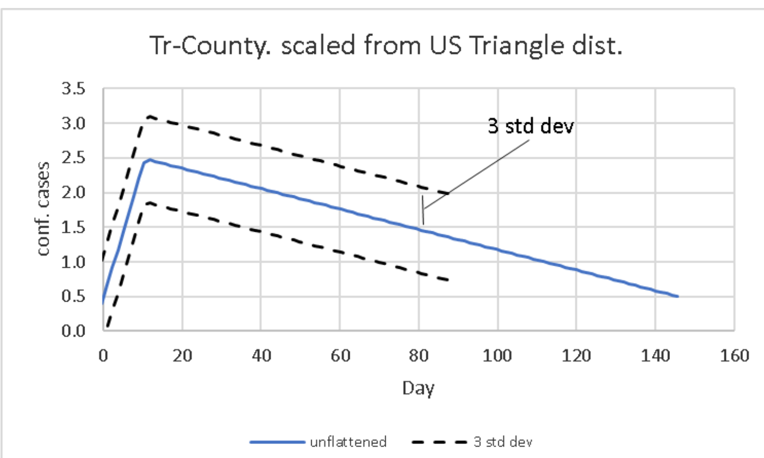
First, assume the above data represents a *flattened* curve—that is, all the restrictive measures that were put in place succeeded in lengthening its duration and reducing its peak (but not significantly changing the total number of infections, the area under the curve). This was the original explanation given for the restrictive measures, to alleviate the burden on the medical facilities, and it made sense. It can be scaled by a simple population ratio to fit the Tri-County area. The resulting curve can also have its time scale compressed, to simulate an unflattened curve, presumably what would happen here without the restrictions. This is compressed by a factor 1.5; it's impossible to verify the real extent of compression, so this seemed like a reasonable assumption. Here is what that looks like:



The area under each curve is the same, there is no change in total cases. It's also worth looking at the cumulative version of the same data, to see what the time compression does to it. Notice it doesn't change the final, cumulative value.[4]



Secondly, taking the unflattened curve (blue), and applying the dispersion factor previously created, we can apply a three standard deviation upper bound—this could be a conservative design criteria to use for county-supplied COVID-19 treatment centers[5], to avoid sending any virus suffers to the regional hospital. Three standard deviations should take care of 95% of the possibilities. The counties could provide something less than this, and still do a lot of good, of course.



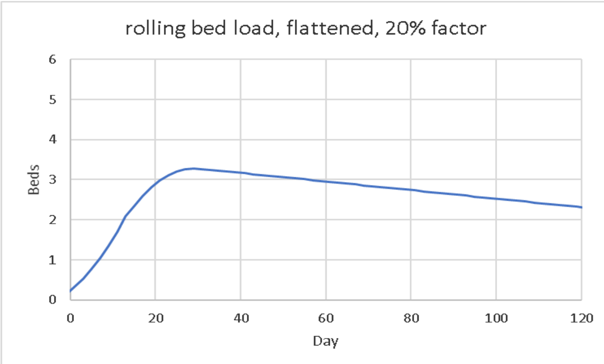
On first glance, it appears that *four* (rounding up) confirmed cases are what to expect for the maximum daily cases in the Tri-County during a virus wave. But there will already be cases ongoing, with an average length of 14 days. A rolling sum of the two (new cases, plus already existing cases lasting 14 days) can be constructed.

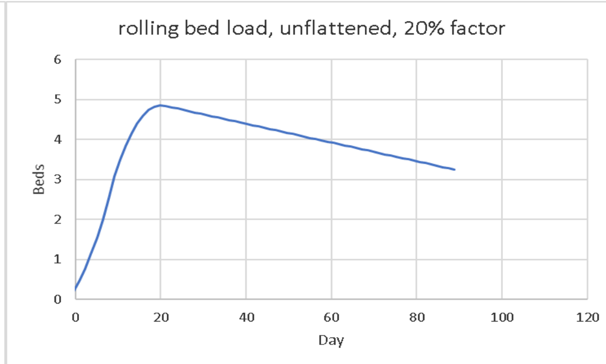
---

[4] If one assumes the Tri-County is a sample of the US population, scaling it this way will likely increase its variability, since that is the nature of small samples. The rolling sum probably smooths that out enough that it is not significant, but that is not addressed here.

[5] Call them M.A.S.H. units, if you wish.

To find the impact on medical care one needs to apply a 20% factor to this confirmed case estimate, since only about 10 to 20% of confirmed cases require medical care or hospitalization:





The unflattened case peaks at 5, or about 2 beds per county, or to be apportioned by each counties' respective populations. A similar calculation can be done for the flattened curve case, where the expected facility load is 4, or about one to two beds per county.

Either case is one that the regional hospital could probably handle, or each county could undertake on their own, with separate medical facilities. It's worth reviewing the spread of both SARS-Covid-1 and SARS-Covid-2 (the current one) to see the key role of hospitals in initially *spreading* the disease, another good reason to *not* use the hospital if possible. Looked at this way, with life and economic restrictions, it would require one to two beds per county, while unrestricted liberty and economics costs two beds per county, for a net cost of zero to one bed per county to preserve personal and economic liberty.

The area under the curve is the *same* in either flattened or unflattened case, which implies shutting everyone in *doesn't* reduce total cases. The restrictions were to help avoid inundating health facilities at the peak. Recently there are claims found in the news media and made by politicians that lockdowns save lives by reducing infections, but it's very hard to prove. It's really an attempt to lower R_o, but how much CV-19 is affected by quarantine and isolation is not known well enough, and helps explain why the original introduction of lockdowns a couple of months ago only claimed to change duration and reduce maximum daily peak, not total infections[6]. Some have made the case that the drop in cases from April to May is evidence that lockdowns reduced them, but that's without understanding epidemics usually follow a bell shape curve, with a peak at some point.[7] That peak was in April for the US, as can be seen in the curves above, so the decrease unfortunately proves nothing, since it is expected.

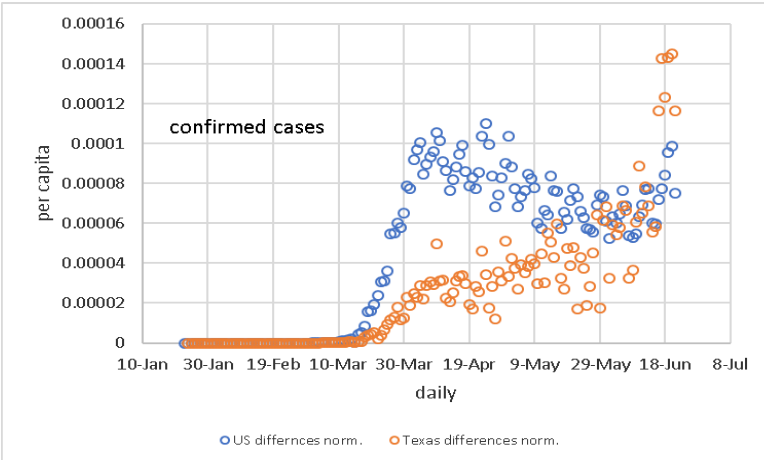
Lockdowns undoubtedly delay the rate of virus infections, and they certainly can have a benefit for relieving medical facilities of their load, which surely saves lives, but it's not known to what level lockdowns have to be done in order to actually reduce cases, and leads to the question of what other more effective and less damaging actions can we take to reduce the threat. We surely will have a second, and maybe a third wave, and it's not clear if we can survive any more lockdowns.

[6] Basic Reproduction Rate. R_o is a measure of transmissibility: $R_o < 1$, disease disappears; $R_o = 1$, it's endemic; $R_o > 1$, epidemic.

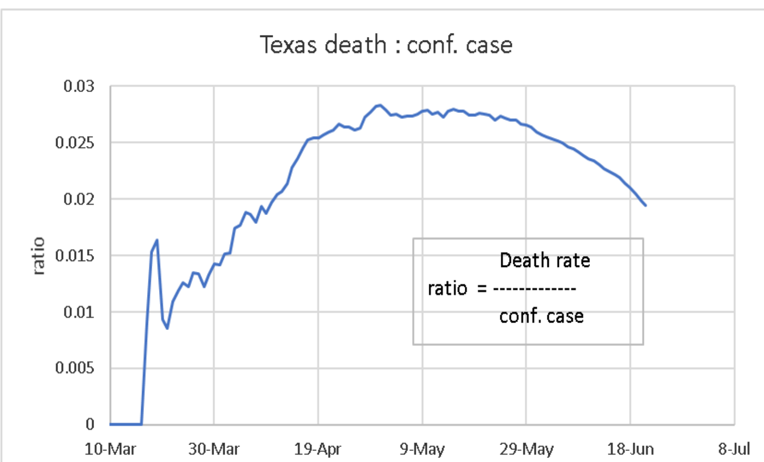
[7] There is a uptick since about the beginning of June, that is concerning a lot of people, though. See US daily data above and you can see it. There is also a zoomed view of it in accompanying graphs. It's particularly pronounced in Texas.

Postscript

If Texas and the USA data are plotted as daily data concurrently, on a per capita basis, here is what you see:

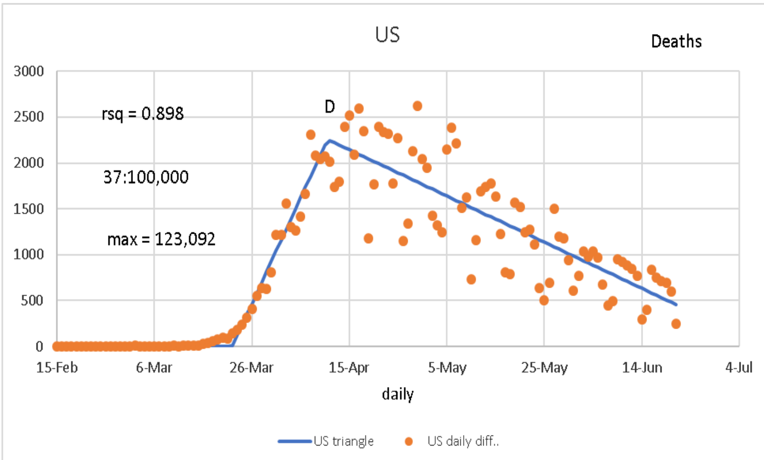


Texas is closing in on the national total of 12:1000 cases predicted by the triangular distribution. In fact, confirmed case estimate for a completed wave for Texas is 8:1000, and still rising, at least there are consecutive spikes the last few days. It's interesting that the Texas ratio of deaths to confirmed cases is dropping, and always has been significantly less than the US numbers, which could mean better treatment of a growing number of confirmed cases, or more testing finding more cases, or some combination of that. The good news, in a world of bad news, is a drop in the ratio might decrease the fear and panic factor and lead to more rational decisions by people in authority.

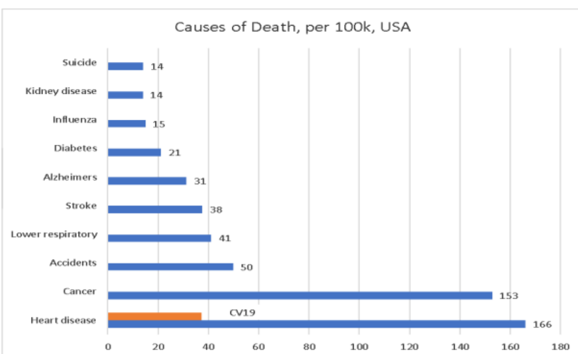


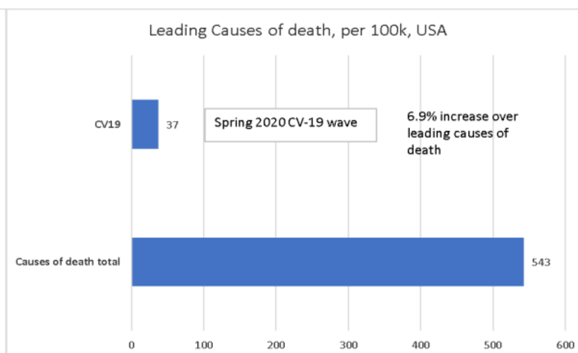
P.S.S.

A few weeks ago the Wall Street Journal published data on annual causes of death which I have plotted with the USA CV-19 death statistics. The USA CV-19 death statistic is based on a triangular distribution curve fit of USA deaths that generates a cumulative maximum for a completed wave.



This distribution predicts 123,000 deaths, once this particular wave is over.[8] We may get one or two more waves this year. This value is plotted as deaths per 100,000 with the WSJ data:



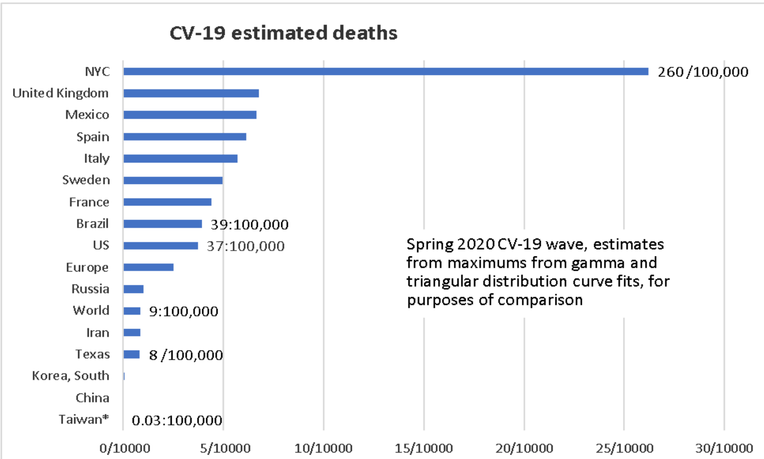


It's right up there with the major killers, although it's known many who died from CV-19 already had these potentially fatal issues. It's not clear where dying from old age fits in here, though. I suspect old age deaths would deduct from all these more or less equally, so the relative size remains unchanged.

NYC will suffer 260 deaths per 100,000 in this wave, while Texas only will only have 8 per 100,000. Those are estimates based on the maximums predicted by their respective curve fits. It's something to keep in mind when considering the desperation found in NYC and the northeast United States; they have every reason to be desperate. On the other hand, Texas' incidence is still less than any of the categories listed above. It's still a serious issue for Texas, no doubt, but two orders of magnitude less than what NYC is experiencing, and one order of magnitude less than the US average. Certainly don't need to treat Texas like NYC.

[8] Can't help but wonder why the news media ignores this and focusses on the recent pronounced rise in confirmed cases, instead.

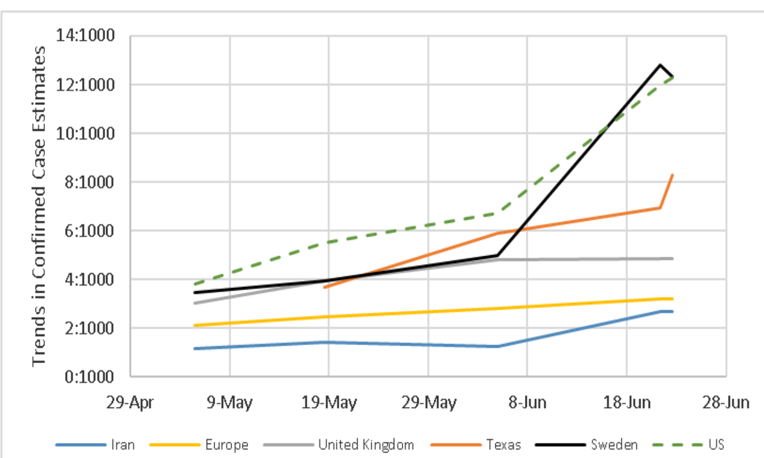
Using the estimator (the maximum value parameter from the gamma distribution curve fit, or the maximum cumulative value for the triangular fit), it's possible to do a bit of comparison of deaths around the world. Interestingly, two countries that did not do lockdowns (Sweden and Mexico) are mixed up in here. Mexico is not doing well at all, so far. I put NYC in here to show how much it skews the US data, since it is almost a fifth of all the cases. Not sufficient data to prove anything, but it does make one wonder, or I hope it makes one wonder.



For example, lockdowns are putting incredible stresses on people, losing jobs, being cooped up, maybe drinking too much. All of that stresses the immune system, and that makes people more susceptible to infection. In that case, lockdowns would not only *not* decrease the eventual number of infections, but might even increase them, due to the compromised immune systems.

Trends in Confirmed Case Estimates

The confirmed case estimates are based on the maximum value parameter of a curve fit for a *completed* epidemic wave. The US was changed to a triangular distribution recently, which predicts more confirmed cases than before. The curve fit has been based on data that was in the process of being completed for the Spring 2020 CV-19 wave, so it's to be expected that the estimates might change. Here are confirmed case trends for a few select countries:

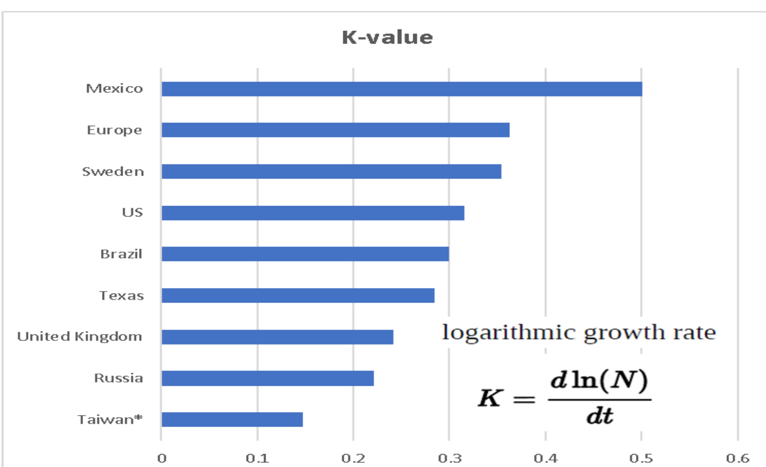


Explanation of statistic used for comparative purposes:

Often the generation of confirmed cases or deaths tend to be in a gamma distribution, which is a slightly skewed bell-shaped curve. Other times a triangular distribution better describes what's happening. A least squares procedure is used to get the closest fit to the data. Generally, the data is provided in a *cumulative* format, so the number increases each day, until the wave is over, at which point it is at its maximum and no longer changes. This can be converted to a daily format just by finding the difference between each pair of successive days. In mathematical terms, the daily format is the time derivative of the cumulative format, or the cumulative format is the anti-derivative (integral) of the daily format. In any case, the area under the daily curve is identical to the last, largest and unchanging value in the cumulative format. This is the value that is used here as a comparative statistic, on a per capita basis. Each data set has a different beginning, develops at a different rate, so comparison of values on a particular date among any given datasets doesn't tell you much.

Cases are usually described as incidents per thousandths; deaths are described as deaths per hundred thousandths. In this way, one can get a better feel for the relative performance of countries, continents, cities, states, counties, as long as the current population figure is available. It's also a way to gauge one's own personal risk—for example, the rate for Texas of 8:1000 confirmed cases over several months just isn't a very high risk for anything that's not fatal, despite being worse than average.[9] Keep in mind this statistic can and will change over time, not as a direct result of increasing cases or deaths, but because of the possible change in shape of the distribution. In some cases, particularly with the US, the distribution was changed from a gamma to a triangular because it fit better, which also changed the comparative statistic.

R_o is a pretty subjective parameter, but much discussed. If one can find the exponential part of cumulative growth (a short section near the beginning of the cumulative curve, but loses validity if less than N = 100 people), one can find the growth rate constant for an exponential, which is $K = d \ln(N) / dt$. $R_o = e^{K\tau}$ where τ is the average number of days for someone to infect R_o individuals. This is a simplified explanation. Here are some K values:



[9] As far as fatalities goes, 8:100,000 mortality estimated for Texas is less than the 11:100,000 national rate for car fatalities. You have to ask yourself if you stay up at night worrying about getting killed in a car crash. Of course, those are all unconditional probabilities—given you're over 65 with significant health problems, it's a whole different story.