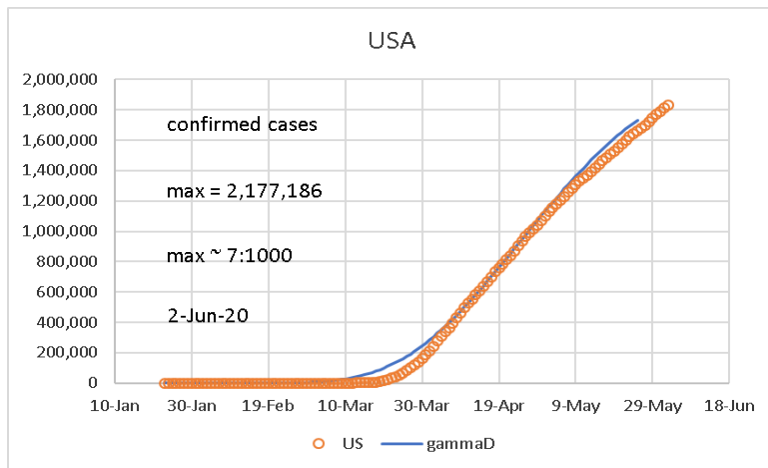
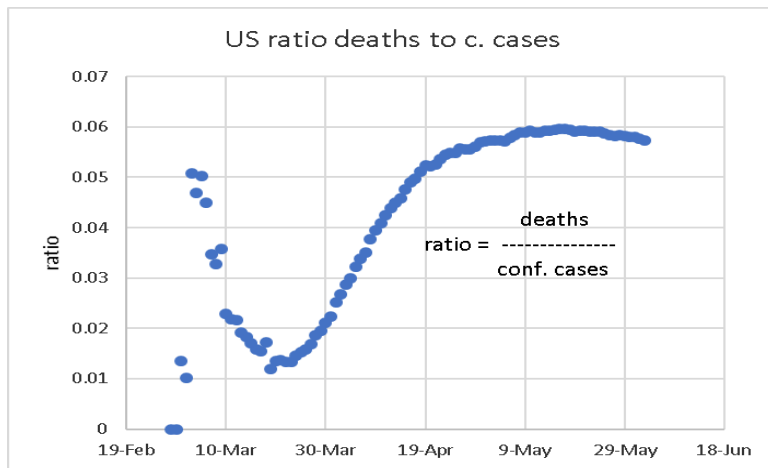


As of 03 June 2020, the US Covid-19 epidemic looks like this, cumulative:<sup>1</sup>



The ratio of deaths to confirmed cases looks like this:



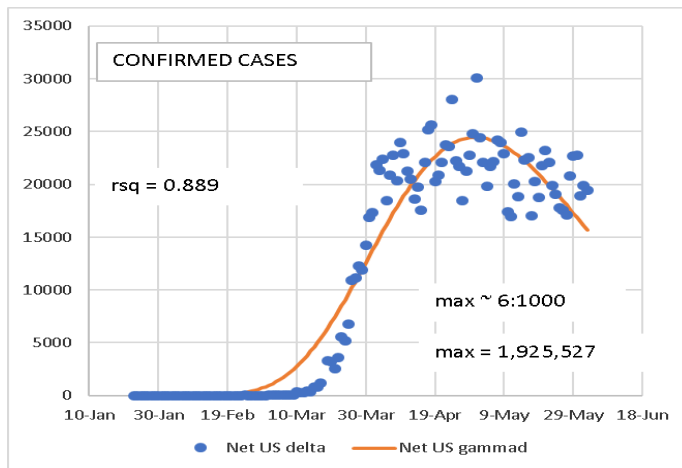
So, it appears that almost 6% of confirmed cases are classified as due to CV-19, which shows this is no common flu, regardless of discussions on how these are classified.<sup>2</sup>

The distribution of this data could be applied locally, assuming the US case applies to the local situation. It makes sense to subtract out New York City confirmed cases, the urban extreme (and about 20% of total US cases), to arrive at a more applicable distribution, although still predominantly urban. Since here in the Big Bend we live more spread out, and don't crowd together on rapid transit, our "wave" is probably something less than the NET US (that is, US confirmed cases minus NYC). It also probably lags in time the urban populations by quite a bit. But since an epidemic is nothing but a deadly diffusion phenomenon, sooner or later it reaches some saturation level that is hard to predict.

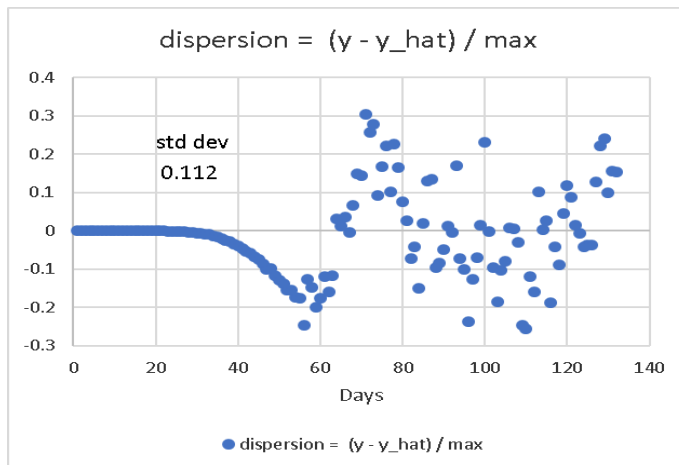
<sup>1</sup> <https://github.com/CSSEGISandData/COVID-19>

<sup>2</sup> A lot of useless discussion on actual vs. confirmed cases occurs, in my opinion, in order to minimize the deadliness of this virus. Why would you try to minimize the deadliness of something that is so obviously deadly?

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

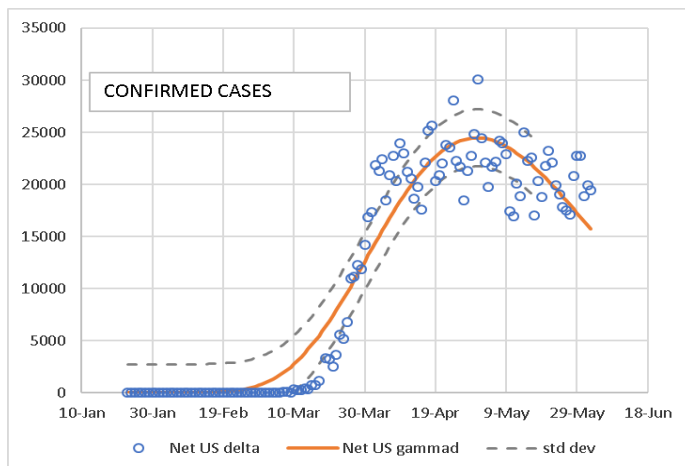


The curve fit is a gamma distribution, often used to characterize disease, population growth, and epidemics. A measure of dispersion is the standard deviation of the residuals scaled to a relevant constant, such as the maximum value of the curve, or  $(y - y_{\hat{}}) / y_{\hat{}}_{\max}$ .  $y$  is data recorded (in this case, the difference between each pair of successive values on the NET US version of the previously shown *cumulative* data),  $y_{\hat{}}$  (pronounced *wye hat*) is the predicted value at each date of the gamma distributed curve fit, and  $y_{\hat{}}_{\max}$  is the maximum value of the gamma distributed curve fit. The dispersion looks like this (below) and gives an idea of the variability of the data around the gamma distribution.  $R^2$  is for goodness of fit, and is equal to 0.889, 1.0 being a perfect fit. Notice that you can have a nearly perfect fit still with a lot of dispersion. Or vice versa.<sup>3</sup>



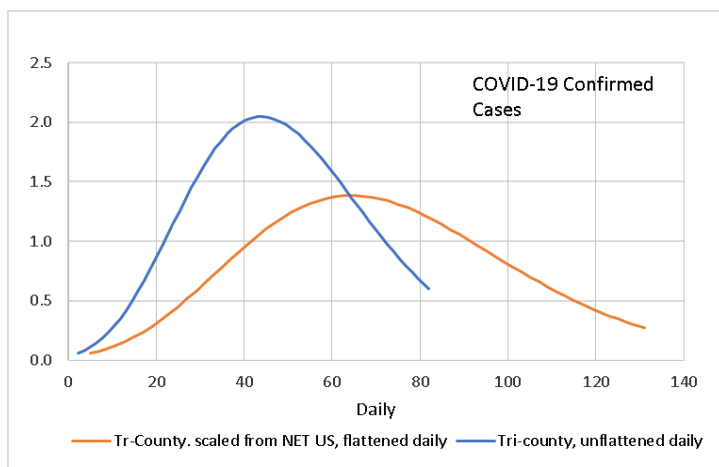
<sup>3</sup> This is an OK fit, but not great, but suitable for the following discussion. There is an alternative triangular distribution in the accompanying graphs.

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

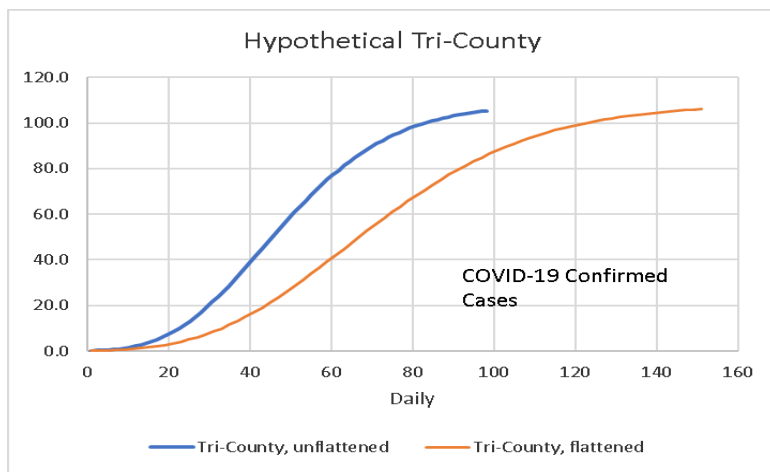


This data is for a population of 320 million (US population less NYC's), but why not scale it to a Tri-County population of 16,000, in order to have a better idea of what to expect here? You certainly wouldn't expect the Tri-County to be *worse* than this.

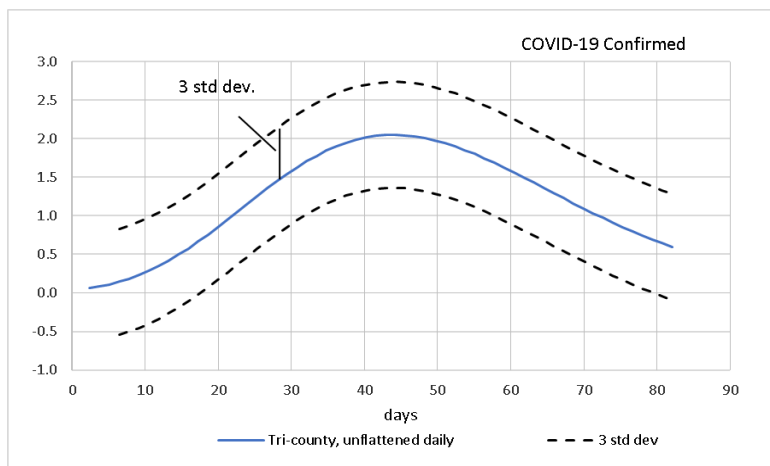
First, assume the above data represents a *flattened* curve—that is, all the restrictive measures that were put in place succeeded in lengthening its duration and reducing its peak (but not significantly changing the total number of infections, the area under the curve). This was the original explanation 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 this distribution, to see what the time compression does to it. Notice it doesn't change the final, cumulative value, which is equivalent to 6 in 1,000 being a confirmed case.<sup>4</sup>



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<sup>5</sup>, to avoid sending any virus sufferers 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.

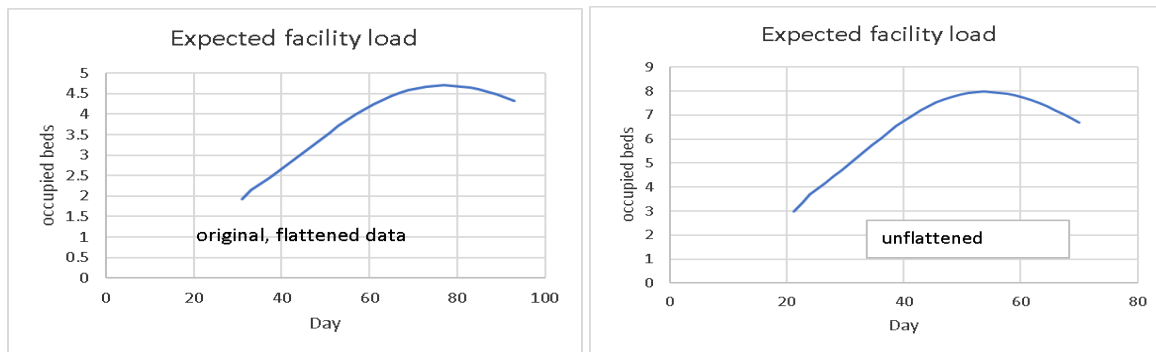


On first glance, it appears that *three* (rounding up) confirmed cases are what to expect for the maximum daily cases in the Tri-County during a virus wave. But, there will already be cases ongoing, with an average length of 14 days. A rolling sum of the two (new cases, plus already existing cases lasting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

<sup>4</sup> If one assumes the Tri-County is a sample of the US NET population, scaling it this way will likely increase its dispersion, 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 rigorously addressed here.

<sup>5</sup> 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 the confirmed case estimate, since only about 10 to 20% of confirmed cases require medical care or hospitalization:



This peaks at 8, or about 3 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 5, or about 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 two beds per county, while unrestricted liberty and economics costs three beds per county, for a net cost of one bed per county to preserve personal and economic liberty.

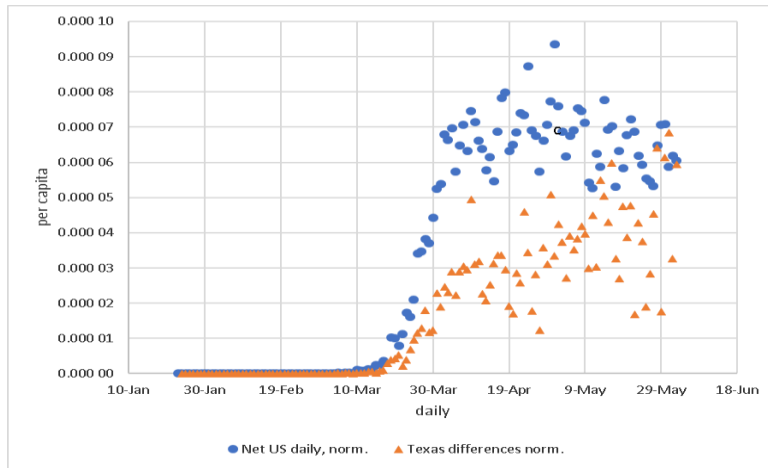
The area under the curve is the *same* in either flattened or unflattened case, which implies shutting everyone in *doesn't* reduce total cases. The restrictions were to help avoid inundating health facilities. Recently there are claims found in the news media and made by politicians that lockdowns save lives, but it's very hard to prove. It's really an attempt to lower  $R_0$ , 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<sup>6</sup>. 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. That peak was probably in the latter part of April for the US, as can be seen in the curves above, so the decrease unfortunately proves nothing.

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.

<sup>6</sup>  $R_0$  is a measure of transmissibility:  $R_0 < 1$ , disease disappears;  $R_0 = 1$ , it's endemic;  $R_0 > 1$ , epidemic.

## Postscript

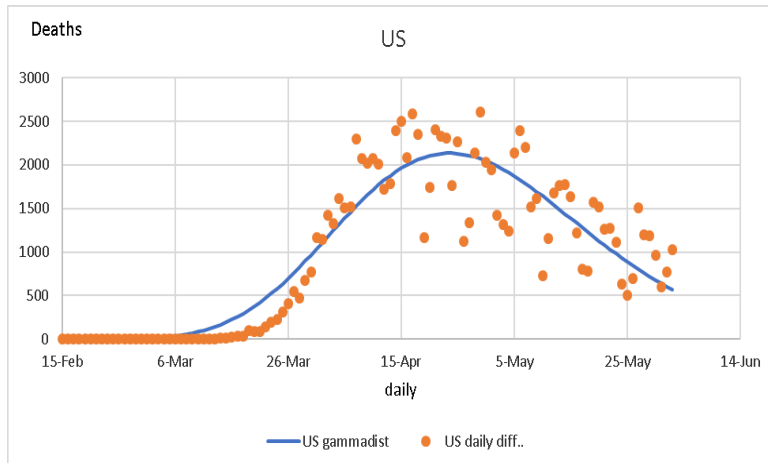
If Texas and the NET USA data are plotted as daily data concurrently, on a per capita basis, here is what you see:



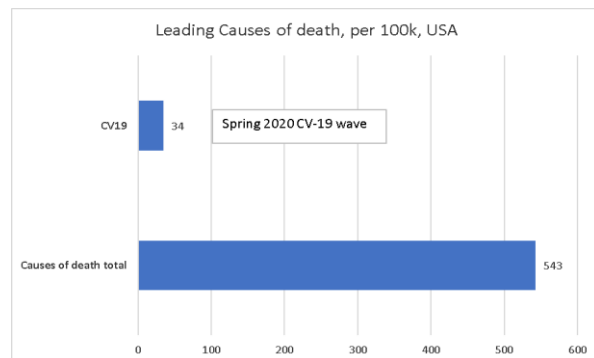
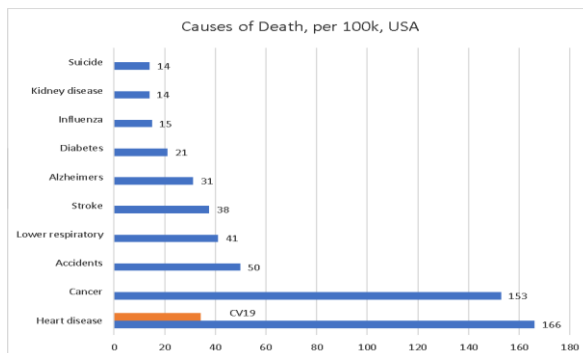
It appears Texas lags the NET USA a little, but even more significantly, the per capita basis looks about two thirds to half of the NET USA, although it now appears Texas is trying to catch up to the rest of the country, part of the lag effect. Texas would probably be a better dataset to scale the Tri-County from, since we're in Texas. But USA is more complete. Also, USA represents a more conservative assumption.

## 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 gamma distribution curve fit of USA deaths that generates a cumulative maximum for a completed wave.



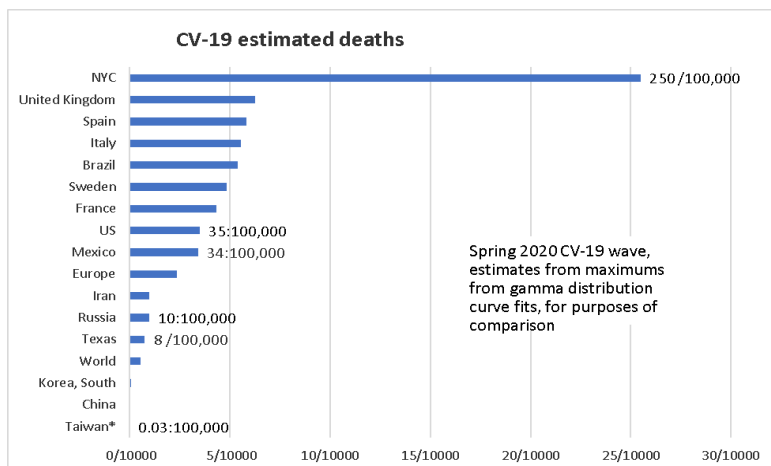
This distribution predicts 115,000 deaths, once this particular wave is over. We may get one or two more waves this year. This value is plotted as deaths per 100,000 with the WSJ data.



It's right up there with the major killers, although it's clear 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 250 deaths per 100,000 in this wave, while Texas only will only have 7 per 100,000. That is an estimate based on the maximums predicted by their respective gamma distributed 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, 8 per 100,000 is 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.

Using the estimator (the maximum value parameter from the gamma distribution curve 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 and are indistinguishable from those that did. I put NYC in here to show how much it skews the US data, since it is a fifth of all the cases. Not sufficient data to prove anything, particularly with Mexico still in the early stages of this wave (and it later went to lockdowns), 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. 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:

