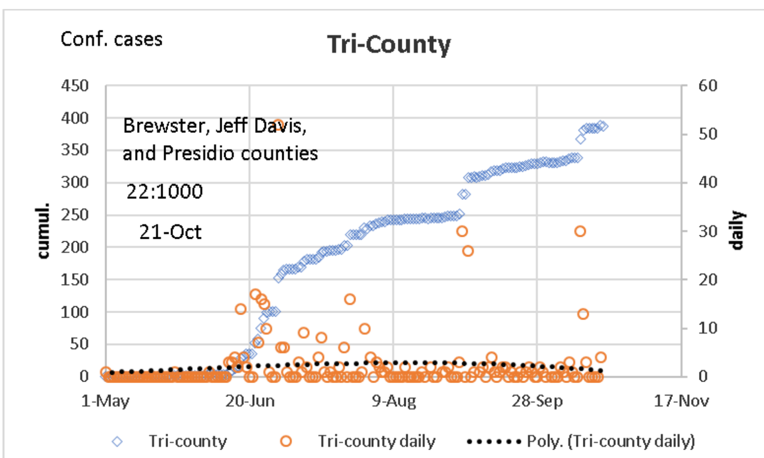
Tri-County (Brewster, Jeff Davis, and Presidio Counties) and the rest of the World, CV19

Please refer to accompanying two pdf files of graphs. The estimated number, and the related incidence relative to population, is based on maximums estimated from curve fits. They are subject to change of course, and should not be taken as precise projections, since there is no such thing. Another way to interpret them is as relative indices, that show a locale's progress relative to the others.[1]

US or Texas projections for confirmed cases scaled to Tri-County population:

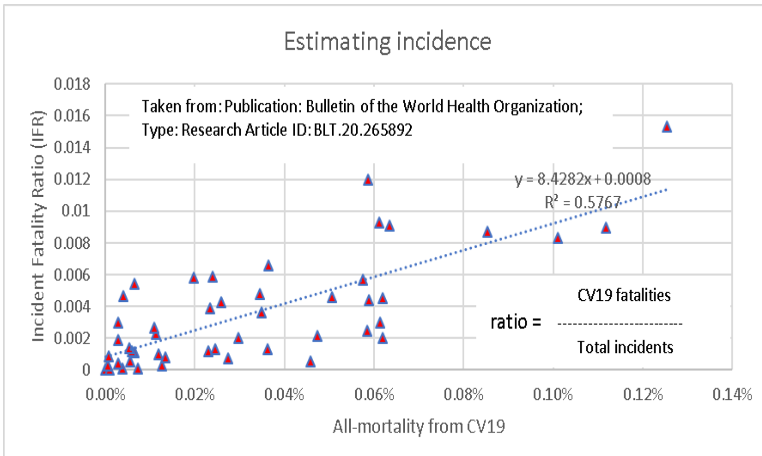
max US daily	63,762	persons
equiv. max Tri-County	3.5	persons
max TX daily	21,771	persons
scaled to Tri-County	14.0	persons
max US cumulative	23,737,951	persons
equiv. max Tri-County	1295	persons
incidence	72:1000	
max TX cumulative	818,426	persons
scaled to Tri-County	526	persons
incidence	29:1000	
Tri-County (current)	22:1000	10/21/2020



[1] Data all from publicly available Github site: https://github.com/CSSEGISandData/COVID-19

John Ionnadis, a highly respected epidemiologist, recently took a stab at estimating mortality rates; for example, infection fatality rate for deaths > 50:100,000 (e.g. USA) at 0.57%.[2] He is famous for his critique on research statistics released in 2005 which is available for free download on www.plosmedicine.org. I have taken a cross-plot of data he included in the paper cited below, to generate an estimate of the prevalence of COVID19 in the Tri-County area.

The idea is to use the following regression on his data, since from a known all-mortality figure one can generate an estimated IFR.



For example, all-mortality for Tri-County as of 21 Oct 2020 is 0.00066003 deaths per population (around 18,000 people). That predicts an IFR of 0.636% ± 0.23% (based on one standard deviation of scatter in this plot). The IFR is defined as:

$$IFR \equiv \frac{Fatality_{CV19}}{Incidents_{CV19}}$$

Likewise, the CFR is defined as

$$CFR \equiv \frac{Fatality_{CV19}}{Cases_{CV19}}$$

where *Cases* is the number of confirmed cases, known, always something less than the actual number of *Incidents*, unknown. The ratio of *CFR/IFR* is the same as the ratio of *Incidents* to *Cases*:

$$\frac{CFR}{IFR} = \frac{\dfrac{Fatality_{CV19}}{Cases_{CV19}}}{\dfrac{Fatality_{CV19}}{Incidents_{CV19}}} = \frac{Incidents_{CV19}}{Cases_{CV19}}$$

If you take the known average *daily* case incidence as a proxy for confirmed case prevalence (disregarding duration of infections), multiplying that number by the above factor results in an estimate for actual number of incidents. In this case, it's estimated at around five times the case incidence.

	Tri-County			
	all-mortality	IFR(all-mortality)		
	0.000660	0.006361	0.636%	IFR (calculated)
curve fit	8.42821478	0.000798	3.101%	CFR (known)

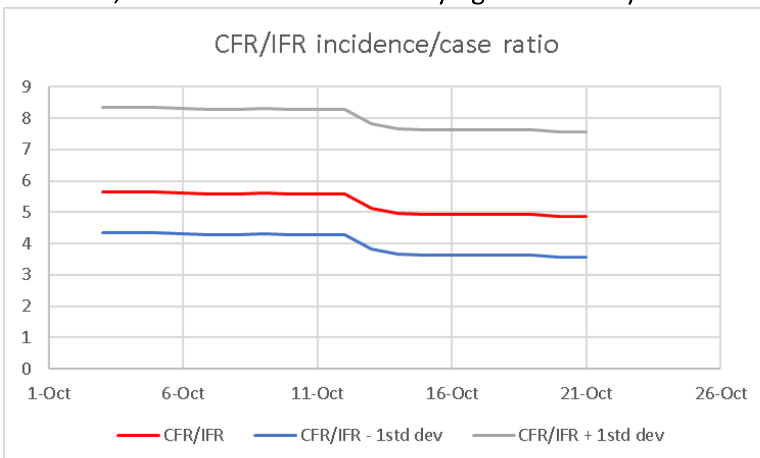
[2] https://www.who.int/bulletin/online_first/BLT.20.265892.pdf

Here is the range of ratios so generated:

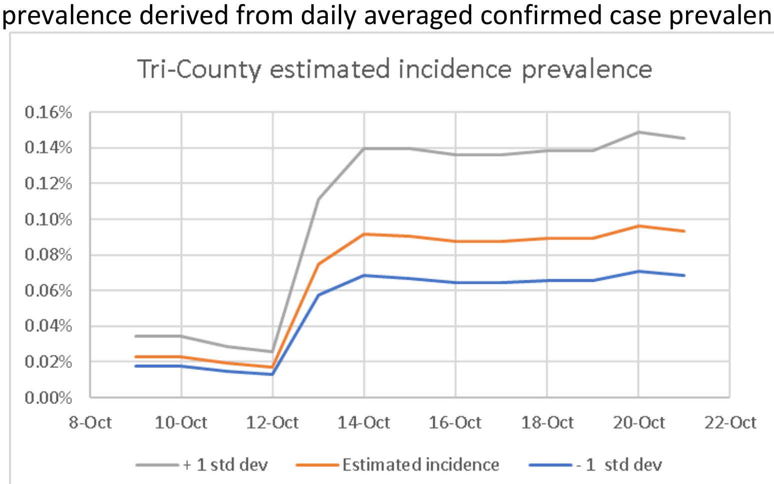
std dev

0.00226564		IFR	inc / conf. case ratio
+ 1 std dev	0.004095	0.41%	7.6
estimate	0.006361	0.64%	4.9
- 1 std dev	0.008627	0.86%	3.6

This creates a range of values, due to the scatter and varying all-mortality over time,



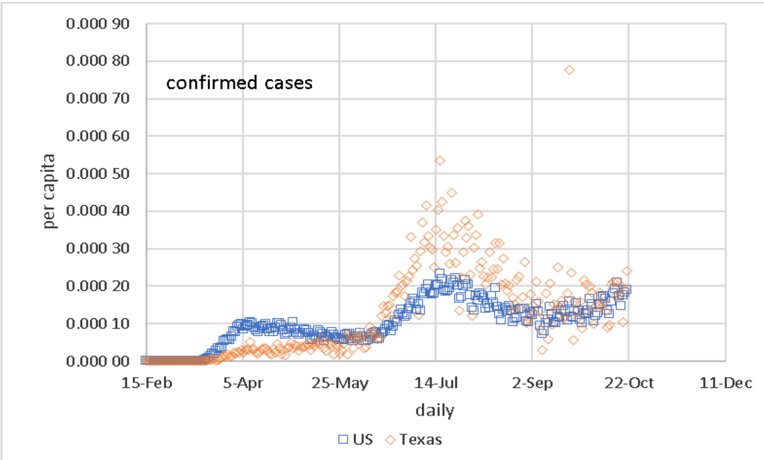
Estimated incident prevalence derived from daily averaged confirmed case prevalence,



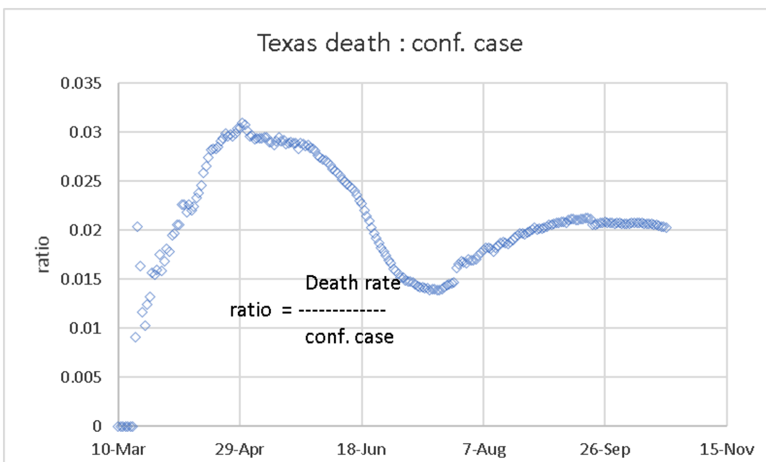
This is just an estimate, obviously, and the original source lists a whole bunch of caveats that go with it. But as you can see, having some idea of what the actual incidence is, when it can't be directly measured, has a whole lot of uses. Read discussion in "Details on Ro you may or may not be interested in" document on the effect of the actual prevalence on false positives for CV19 testing, for example. Another valuable use is correctly assessing risk, so that stupid things aren't done.

Comments

If Texas and the USA confirmed cases are plotted as daily data concurrently, on a per capita basis, here is what you see, along with the apparent *third* wave in US confirmed cases:



Mortality looks in a good place (in the sense it could have been a lot worse). Arizona has a similar curve, although no doubt they have a higher proportion of retired, elderly people that don't handle CV-19 very well. Confirmed cases are taking a turn for the worse, as Fall progresses.



What's to be done? Your mom taught you to cover your mouth when you coughed as a act of consideration for others so you wouldn't spread germs. That is all wearing a mask does, except more effectively, since this virus is stronger than the usual germs. What is so hard to understand about that?

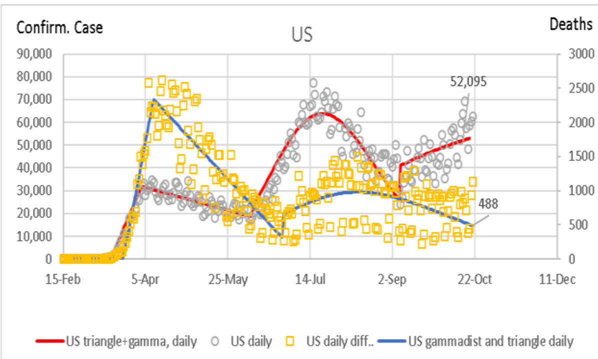
Another good thing to do is avoid enclosed areas with low or no ventilation where there are lots of people. This would include most restaurants and bars, and probably schools, too. Or step up the air changes in such places. But instead, we gripe about the government telling us what to do, or criticize it for not doing anything, not sure which is more irritating. For better or for worse, Germany is now taking action on fresh air in buildings.[3]

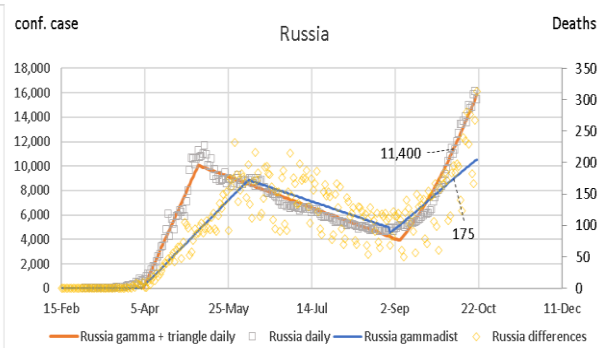
[3] https://www.bbc.co.uk/news/world-europe-54599593

COVID-19 deaths

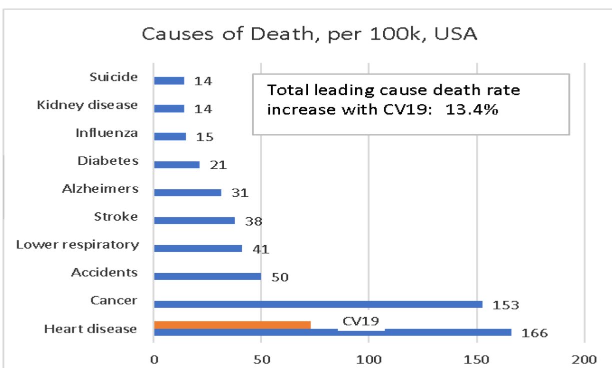
The USA CV-19 death statistic is based on a triangular distribution, combined with a later gamma distribution curve fit of US deaths that generates a cumulative maximum for the two combined (minus their overlap). See “Experimental Page” for a look at ratios of the curve fits, rather than the data itself. It appears the ratio of deaths to cases is gradually decreasing:

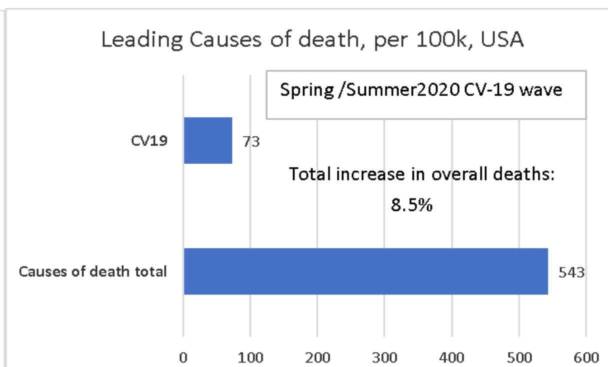
> A point of comparison:
>
> Around 1920, the population of the US was 105 million. It is estimated that 500,000 people in the US died from the Spanish Flu epidemic in those years. This is a death rate of around 480 per 100k. Most deaths occurred during the Fall of 1918.





A few months ago the Wall Street Journal published data on annual causes of death which is plotted with the USA CV-19 death statistics. The possible death value is plotted as deaths per 100,000 with WSJ data of annual averages of leading causes:

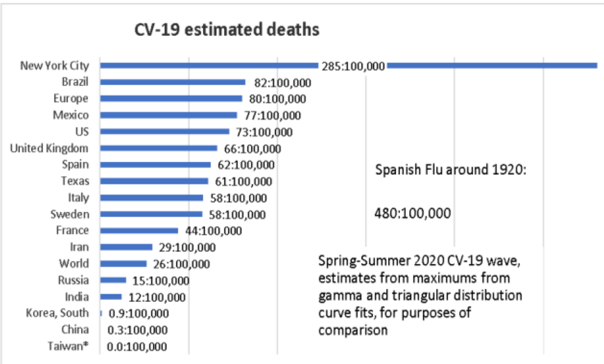


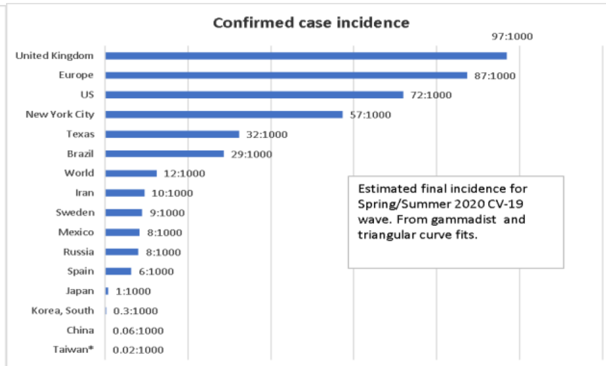


It’s right up there with the major killers, although it’s known many who died from CV-19 already had these potentially fatal issues.[4] It’s not clear where dying from old age fits here, either. (It’s strange how nobody dies of old age anymore—it’s always some other cause.) It’s interesting to read off the mortality table the IRS uses that the expected age of death is 77.3 yrs (median is 80.3). The probability of dying within age bin 77 to 78 is 0.026, and the 77 to 78 age bin proportion of population is 2.7%, so fraction dying is 15:100,000. Similarly, the probability of dying in each year age, from 77 on up, is 62%, with population proportion 29.9%, so fraction dying in that group is 257:100,000. Total deaths are around 860:100,000 every year, so old people are a significant chunk of that, not surprisingly. See table towards end of this document.*

[4] On the right chart above, about 13% increase in overall annual deaths is attributed to CV19. Or, put another way, a 13% decrease in all those existing causes of death would completely cancel this year’s projected CV19 deaths. Or, just an 23% decrease in cancer and heart disease deaths cancels all CV19 for the year.

What follows are some comparative tabulations of these estimates. As you can see, NYC had about half the death rate as the Spanish Flu disaster in the 1920s. But that doesn't make the current situation good. Texas finally reached a daily maximum, which affected the estimate in a positive way. Good news is Iran hit a turnaround point, which also drastically changed its death estimate, downward. Japan has been struggling as of late, but a lot better than we are, nonetheless.





It's also worth reminding yourself of basic arithmetic—a confirmed case incidence of 32:1000 is big on this ranking, but still is only 3.2% of the population! Most of those are now recovered.

Recently, a lot of discussion has centered around excess deaths. The CDC has taken a stab at this, while recognizing the difficulties of getting accurate death certificate information (e.g. did this person die *from* CV-19, or die while *having* CV-19, etc.). They took weighted data from the previous three years and averaged it. They also created a threshold above the average (I didn't quite understand that, so just took the simple three year average as a basis of comparison here). It's also weighted, presumably against change in population over the three year period. Those are points you can check out on your own, as they're not that relevant to this discussion. Here are the results:

USA Excess Deaths (from CDC data):[5]

Summary Table	All Causes Observed - Expected	All Causes , excl. CV19 Observed - Expected	CV19 Difference
2017-2019 avg	-5:100,000	-5:100,000	0:100,000
2020 (annualized) Week 1 thru Week 39	87:100,000	28:100,000	59:100,000

USA Average Deaths (from CDC data):

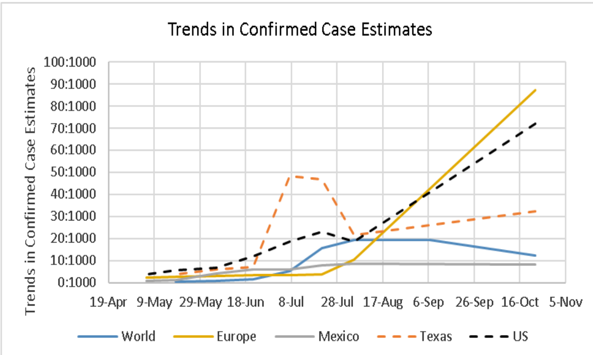
2017-2019 avg	864:100,000

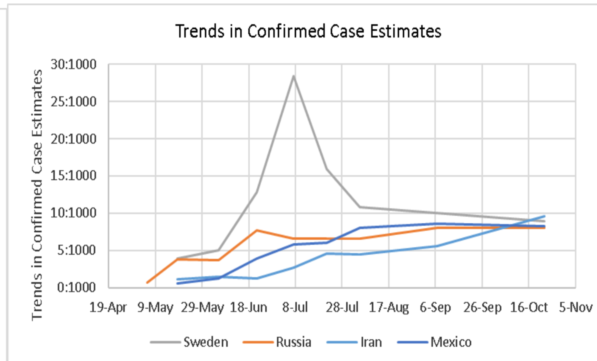
Trends in Confirmed Case Estimates

The confirmed case estimates are based on the maximum value parameter of a curve fit for a *completed* epidemic wave. A way to look at this is this estimate is what it *would* be, if the data followed a gamma (or other distribution). Historically that's what they tend to do, but you can see plenty of examples here

[5] https://data.cdc.gov/NCHS/Excess-Deaths-Associated-with-COVID-19/xkkf-xrst/data

where the data suddenly diverged (e.g. US, World, Texas, etc.). Still, it's a useful way to characterize it. The curve fit has been based on data that was in the process of being completed for the Spring/Summer 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; UK, US, and Europe look particularly bad:





---

*This is an approximation of deaths by age out of 100k, normalized to an average 860:100,000 each year:

Age range	Probability of dying in age range (IRS tables)	Fraction of US population	Fraction dying
0 - 5	0.01474	6.62%	11:100,000
5 - 10	0.00079	6.38%	2:100,000
10 - 15	0.00099	6.37%	2:100,000
15 - 20	0.00334	6.48%	7:100,000
20 - 25	0.00462	6.86%	10:100,000
25 - 30	0.00452	7.14%	10:100,000
30 - 35	0.00551	6.77%	11:100,000
35 - 40	0.0078	6.46%	15:100,000
40 - 45	0.01151	6.11%	22:100,000
45 - 50	0.01677	6.34%	33:100,000
50 - 55	0.0238	6.48%	48:100,000
55 - 60	0.03616	6.65%	74:100,000
60 - 65	0.05371	6.07%	100:100,000
65 - 70	0.0743	5.05%	115:100,000
70 - 75	0.10233	3.87%	121:100,000
75 - 80	0.13742	2.67%	112:100,000
80 - 85	0.16348	1.77%	89:100,000
85 - 90	0.15999	1.11%	55:100,000
90 - 95	0.11601	0.56%	21:100,000
95 - 100	0.05394	0.18%	3:100,000
100+	0.00827	0.03%	0:100,000
Total	1	100.00%	860:100,000

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. 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, for a completed curve. Each locale's "wave" 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 anything, but nonetheless that is how most of this data is usually presented to the public.

Cases are described here as incidents per thousandths; deaths are described as incidents per hundred thousandths. In this way, one can get a better feel for the relative performance of locales, as long as the current population figure is available. It's also a way to gauge one's own personal risk. 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 (at least the first section), which also changed the comparative statistic. It should be thought a qualitative measure, since it's impossible to put error bounds on it, at least in its earlier stages before the inflection point.

R_o

Basic Reproduction Rate (said as R-sub-0, or just R-0). R_o is a measure of transmissibility: $R_o < 1$, disease disappears; $R_o = 1$, it's endemic; $R_o > 1$, epidemic. R_o is mentioned a lot during this epidemic, along with flattening of curves, with not a lot of understanding or relevance. The real trick is figuring out the *effective* reproduction rate, R_e. More on that in the "*Details on R_o you may or may not be interested in.doc*" found elsewhere on this site.