

Regional variations in the dynamics of the COVID-19 pandemic in the US

Author: Gal Almogy, PhD, Flurensics Inc. Tel Aviv, Israel

Abstract: Since identified as the pathogen responsible for an outbreak of severe respiratory distress in Wuhan, China, the 2019-nCoV coronavirus has caused over 30M cases and 1M deaths globally. Sporadic cases were identified in several states in the US from early January, and large-scale community transmission is believed to have started in late February, leading to a first spike in COVID-19 deaths and overall mortality in late April, and a second spike later in the summer. I show here that the dynamics of the pandemic were different in different regions of the US, showing a north-south pattern, with a first pandemic wave mainly in northern regions, followed by a second wave mainly in southern regions. Analysis of overall mortality data shows that the increase in mortality correlates well with COVID-19 incidence in most regions, and that from April through August COVID-19 deaths accounted for a substantial proportion of all deaths in all parts of the US.

Keywords: 2019-nCoV, COVID-19, excess mortality, All-cause deaths, Case Fatality Ratio, CFR, Epidemiology

Introduction

An outbreak of pneumonia cases of unknown etymology late in 2019, in Wuhan, China, led to the identification of a novel viral pathogen belonging to the coronavirus family, 2019-nCoV (1-2), or COVID-19. During January 2020 the pathogen was mostly limited to China (3-4), although sporadic cases, all of which believed to have been imported (5), were reported in several countries.

The first outbreak outside of China occurred onboard the Diamond Princess cruise ship (6). The attack-rate in the Diamond Princess outbreak, i.e proportion of population infected, was approx. 20%, with >700 of the 3600 crew and passengers infected. Although none of the crew became seriously ill, and many of the infected passengers showed only minimal, or no symptoms, 14 patients died, indicating a case fatality rate (CFR) of 2%, slightly lower than the rates reported in China (7), but still an order of magnitude greater than that of seasonal flu strains (8), as well as that of the 2009 pandemic strain (9).

Since the pandemic began, over 30M cases and 1M deaths had been reported worldwide, leading to a global CFR of 3%, yet there is still large variability across different countries, ranging from xx% in YY, to xx% in YY (10). Part of this variability reflects the different volume of testing, as well as the criteria used to categorize COVID-19 deaths in different countries.

COVID-19 cases data is limited by the number of tests, and by COVID-19 incidence - when the incidence is high but tests are limited, the case number will under represent the true number of cases, and conversely, even when incidence is lower a very large volume of tests can identify more cases. The COVID-19 death data may be more reliable in terms of covering most of the patient population, but it has two weaknesses of its own. First, even if tests were available for all patients in critical condition, the virus may not be detected due to false-negative test results (11), or even actually absent despite being the primary cause of death, thus under-estimating the number of COVID-19 deaths. On the other hand, it is possible that in some fatality cases in which COVID-19 was detected at the time of death, it was not in fact the primary cause, thus overestimating the number of deaths (12). Hence in reality it is nearly impossible to accurately determine the true number of infection cases, and fatalities caused by COVID-19; the numerator and denominator of the CFR, respectively.

Analysis of the number of “excess deaths” during the pandemic provides a complementary approach to understanding the fatality associated with COVID-19. Although it lacks the direct cause-and-effect logic that direct testing provides, mortality data is more comprehensive and definite than any test data, and can provide important insights into the true burden of the pandemic (13-14).

Here I examine the temporal relationship between COVID-19 cases, deaths, and all-cause deaths in the different regions of the US. I show that at the national level, the delay between new COVID-19 cases and deaths increases from one week in the period March-June (referred to here as the first wave), to 3 weeks in the period July-August (second wave). At the regional

level, the majority of cases and death during the first wave were recorded in the northern regions, and in the southern states in the second wave. The CFR also showed a regional pattern, peaking in the north during the first wave and in the south during the second.

Finally, using the per week ratio of COVID-19 deaths to all-cause mortality, I show that during the first pandemic peak COVID-19 deaths accounted for roughly 20% of all deaths in the US as a whole, and over 45% of deaths in the NC region.

Methods

Per week mortality data for the US (January 2020 to August 29, 2020) was retrieved from the CDC website on Sep 30 (15). Mortality data was broken down into states and regions of mainland US: North Coast (NC), East North Coast (ENC), South Atlantic (SA), Mid-West (MW), North-west (NW), South (S), and South-West (SW), as follows (in alphabetical order, state names abbreviated using standard two-letter code):

NC = CT, MA, ME, NH, NJ, NY, NYC, PA, VT
 ENC = IL, IN, MI, OH, WI
 SA = AL, FL, GA, KY, MD, MS, NC, SC, TN, VA, WV
 MW = IA, KS, MN, MO, NE
 NW = ID, MT, OR, WA
 S = AR, LA, OK, TX
 SW = AZ, CA, CO, NM, NV, UT

Data from HI, AS, DC, WY, ND and SD was not included, but adding it did not affect results.

Daily COVID-19 cases and COVID-19 deaths data (January to September) was retrieved from the CDC website on Sep 30 (16), and was arranged by weeks, starting from week 1 of 2020 (January 4), to match the weekly mortality data. Both mortality and COVID-19 per-week cases/deaths data were smoothed using a 1-week moving average.

Per week Case Fatality Ratio (CFR) was calculated as the number of COVID-19 deaths over the number of cases in the same week, i.e:

CFR on week w = new covid cases on week w over new covid deaths on the same week

Per week COVID-19 Death Ratio (CDR) was calculated as the number of new COVID-19 deaths in week w divided by the total number of deaths in the same week.

Results

COVID-19 cases and deaths at the national and regional levels

At the national level, there were two waves of COVID-19 cases and deaths in the US. The first wave begins in mid-March and reaches a peak late in April, and a second wave begins in June and reaches a higher case peak but lower deaths peak in late July (figure 1). It is possible that due to the low testing volume many COVID-19 cases went undiagnosed during March and April, and so the reported number of COVID-19 cases during the first peak may underestimate the true number of cases in the general population.

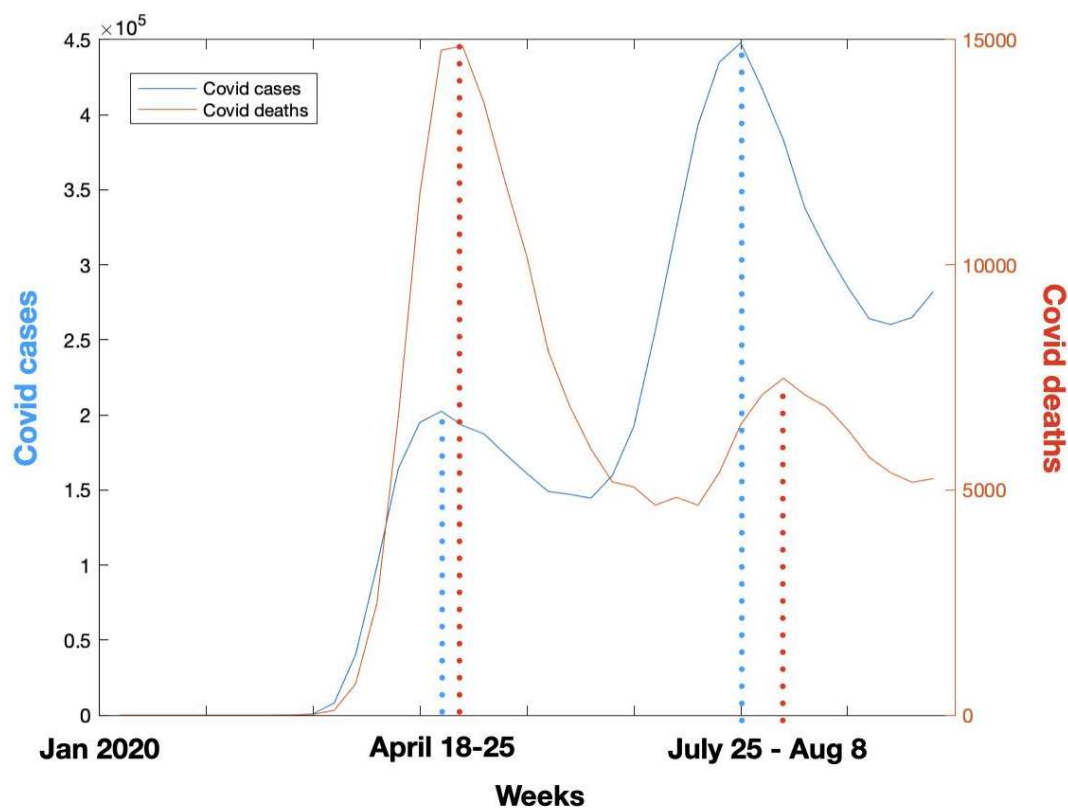


Figure 1. Legend: per week (x-axis) COVID-19 cases (left y-axis) and deaths (right y-axis) in the US. Dashed blue/red lines indicate peak week for COVID-19 cases/deaths, respectively.

When examined at the regional level (regions indicated in figure 2a) it becomes apparent that the increase in COVID-19 cases during the first wave was much more pronounced in the northern regions of the US (figure 2b, blue curve) than in the southern regions, and vice versa during the second wave (red curve). As may be expected, most of the deaths during the first wave were reported in the north, whereas the majority of deaths during the second, smaller peak, were in the south (figure 2c).

The Case Fatality Ratio (CFR) for COVID-19, defined here simply as the number of deaths over the number of cases during the same week, rises sharply in the entire US and shows a less regional trend (17) during the first wave, peaking in late April in both the north and the southern regions (figure 2d), albeit reaching a higher maximal value in the north (~0.065) compared with the south (~0.045). Following the peak, CFR declines in both north and south until week 30, where there is a second increase in the CFR in the south, but a continued decline in the north, leading to a stable CFR in the US as a whole (yellow curve).

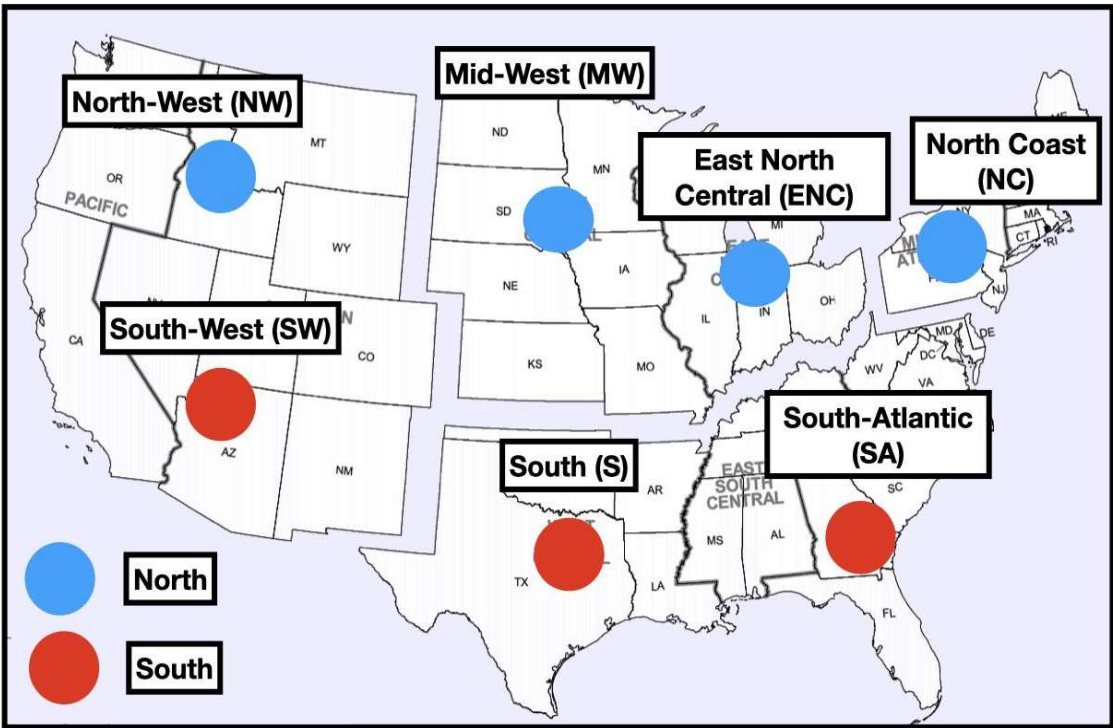


Figure 2a. Legend: the northern (blue) and southern (red) regions of the US. Map of the US was retrieved from the CDC website. The division into regions, and the description (i.e “South”) are based on, but not identical to the CDC definitions for convenience reasons.

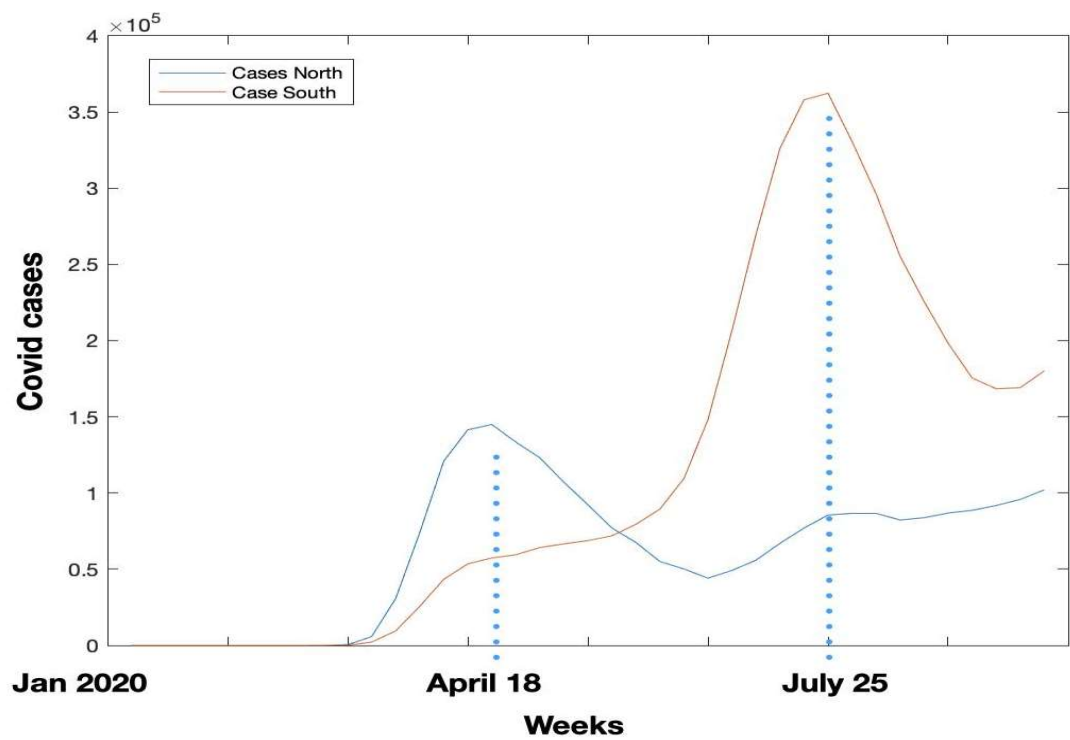


Figure 2b. Legend: per week (x-axis) COVID-19 cases (y-axis) in the northern and southern regions of the US. Dashed blue lines indicate peak COVID-19 cases week.

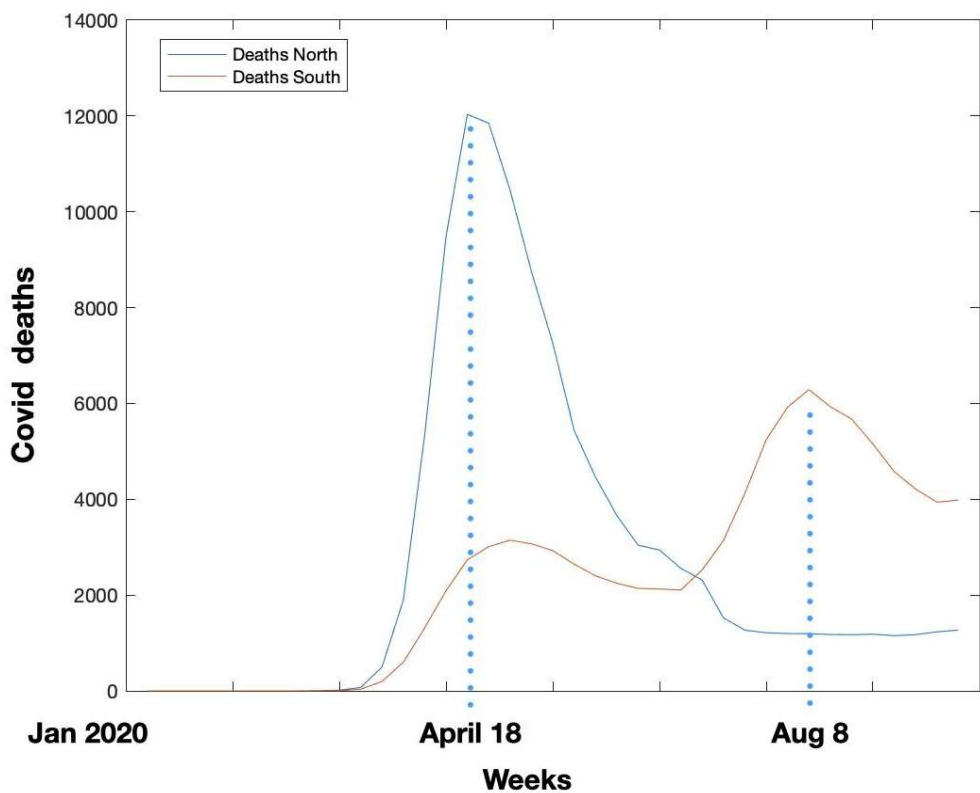


Figure 2c. Legend: per week (x-axis) COVID-19 deaths (y-axis) in the northern and southern regions of the US. Dashed blue lines indicate peak COVID-19 deaths week.

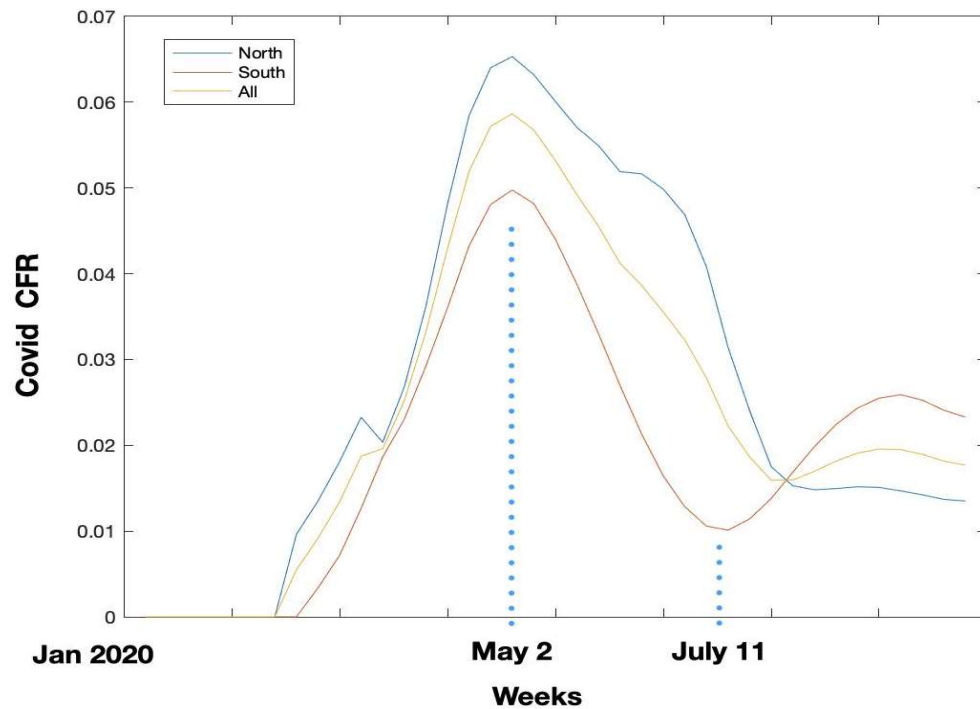


Figure 2d. Legend: per week (x-axis) COVID-19 CFR (y-axis) in the northern and southern regions of the US. Dashed blue lines indicate peak week for CFR.

Lag between COVID-19 cases and deaths

Clinical studies suggest that there is a period of approx. 2-3 weeks between patients developing symptoms and testing positive for COVID-19, and patient death, in fatal cases (18-19). This delay between infection and death on the patient level translates into a lag between COVID-19 cases and deaths in the US (figure 2b) - a one week lag during the first wave, and a 3 week lag in the second.

At the regional level there is a similar overall trend, albeit with several notable differences. The NC is the only region where there is a single wave of cases, closely followed by a wave of COVID-19 deaths (blue and red curves, respectively in figure 3a). In the ENC the death curve qualitatively matches the curves in the NC region, yet the death curve in the ENC peaks a week before the peak in case number (blue and red dashed lines, figure 3b), and surprisingly, continues to decline while the number of new cases, after a transient dip, increases from July onward.

In the S, SW and SA regions the number of cases continues to increase but does not peak during the first wave, yet the number of COVID-19 deaths does peak, and curiously, approx. 2 weeks after the NC and ENC regions (figure 3d-e). In the second wave the rise in case numbers precedes the increase in COVID-19 deaths by 1-3 weeks.

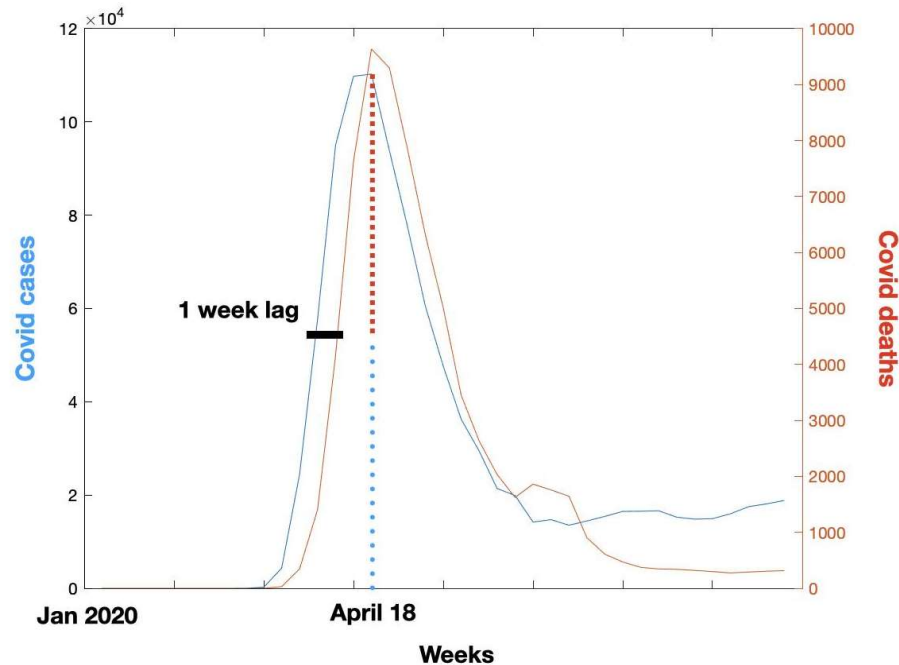


Figure 3a. Legend: per week (x-axis) COVID-19 cases (left y-axis) and COVID-19 deaths (right y-axis) in the NC region. Dashed blue/red lines indicate peak week for cases/deaths, respectively.

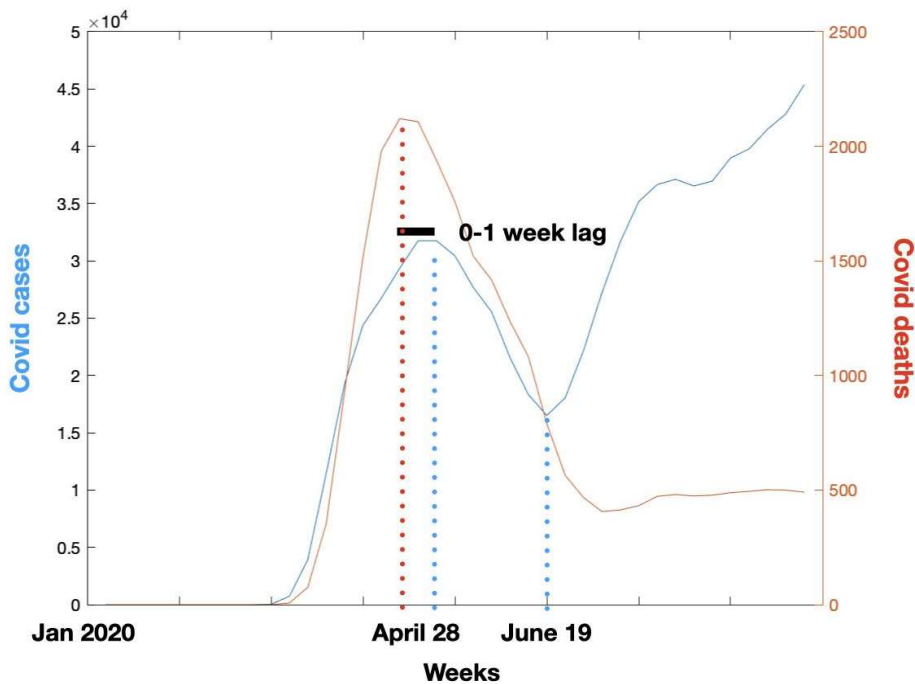


Figure 3b. Legend: per week (x-axis) COVID-19 cases (left y-axis) and COVID-19 deaths (right y-axis) in the ENC region. Dashed blue/red lines indicate peak week for cases/deaths, respectively.

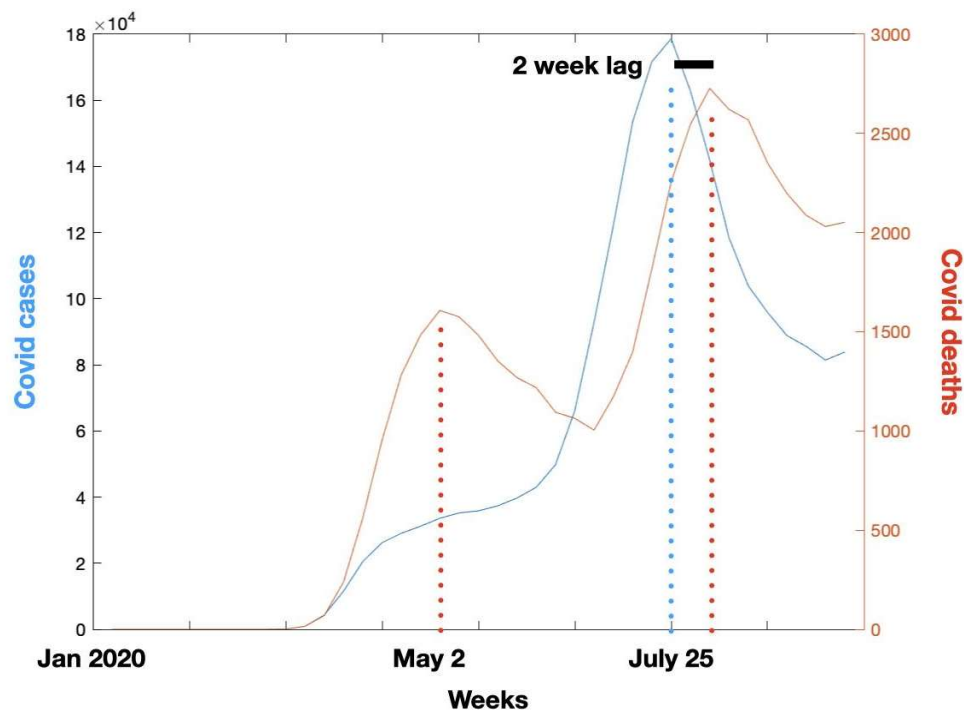


Figure 3c. Legend: per week (x-axis) COVID-19 cases (left y-axis) and COVID-19 deaths (right y-axis) in the SA region. Dashed blue/red lines indicate peak week for cases/deaths, respectively.

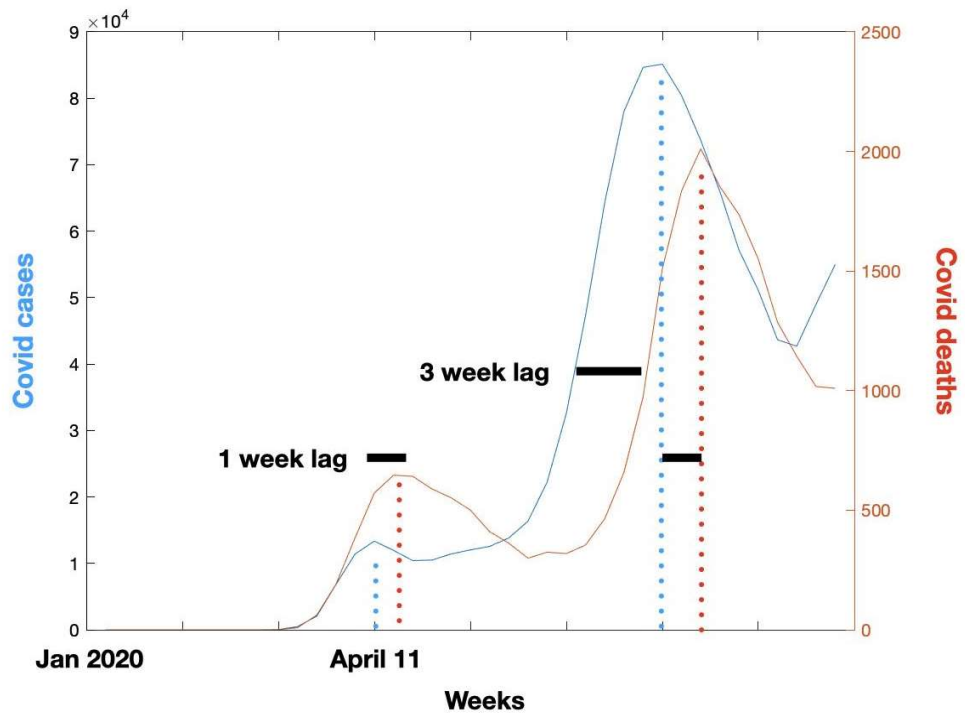


Figure 3d. Legend: per week (x-axis) COVID-19 cases (left y-axis) and COVID-19 deaths (right y-axis) in the S region. Dashed blue/red lines indicate peak week for cases/deaths, respectively.

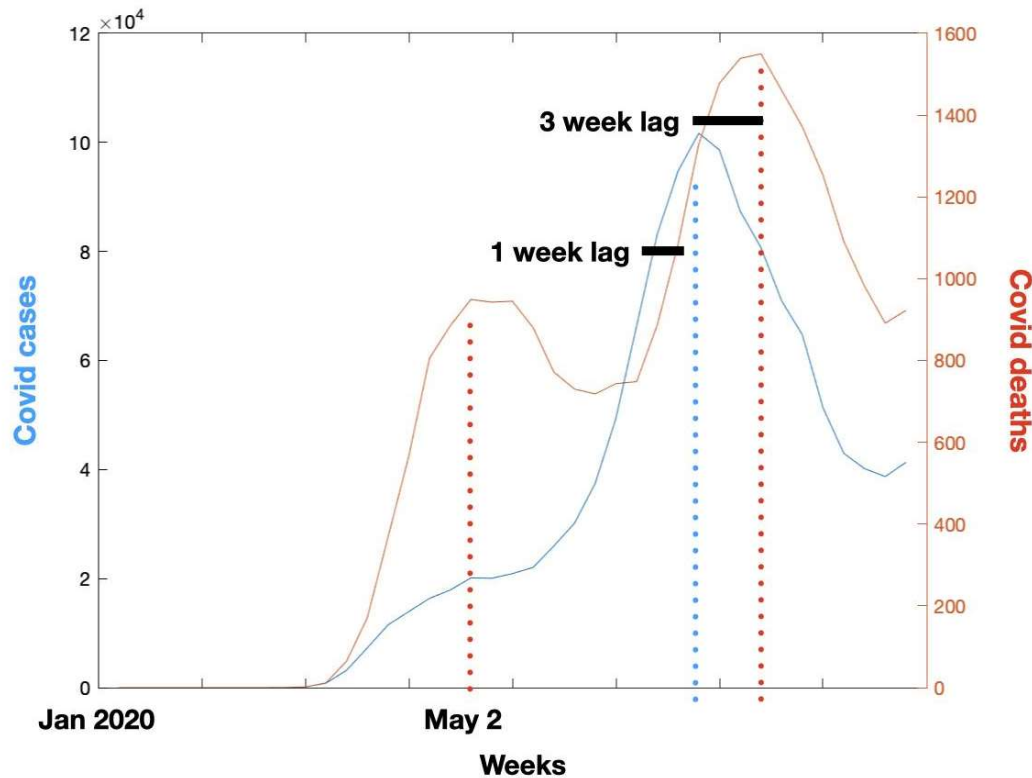


Figure 3e. Legend: per week (x-axis) COVID-19 cases (left y-axis) and COVID-19 deaths (right y-axis) in the SW region. Dashed blue/red lines indicate peak week for cases/deaths, respectively.

COVID-19 cases, deaths, and all-cause mortality

A clear cut indication of an ongoing pandemic is an unusual rise in mortality following the increase in the incidence of the pathogen (20-21). At the national level, there is a sharp increase in overall mortality (all-cause) beginning around week 11 of 2020, almost exactly at the same time COVID-19 case numbers begin to increase (figure 4a). Mortality in the US reaches a peak around April 18, declines until week 25, and reaches a second peak in August, albeit a much smaller than during the first wave, despite the larger number of cases.

In terms of timing the peaks in all-cause mortality and COVID-19 cases overlap during both waves, but new cases precede the rise in all-cause mortality by approx. 1 week during the second wave. When examined at the regional level there are again stark differences between the north and the south, as expected given the north-south pattern of COVID-19 deaths (figure 3). In the NE region (figure 4b) there is a nearly perfect overlap of the cases and mortality curves during the first wave, both reaching a peak around Apr 18, same week as COVID-19 deaths, and there is no second wave of COVID-19 deaths or mortality in the region.

In the ENC region there is a single peak in mortality, as in the NC region, but unlike in the NC, peak mortality precedes the peaks in both COVID-19 cases and deaths by 1-2 weeks (figure 4c), which is a rather unexpected result given COVID-19 deaths are believed to be the cause of the unusual increase in the number of deaths.

In the SA and south regions there is a very good fit between the mortality, COVID-19 cases and COVID-19 deaths curves, albeit at an unusual order of events, with the increase in overall mortality following COVID-19 cases but preceding COVID-19 deaths during both waves, even more so during the second wave in the south region (figure 4 f-g). In the SW region the data is similar except for the mortality curve preceding COVID-19 deaths by ~2 weeks during both waves.

In the MW the first peak in mortality precedes the peaks in COVID-19 cases and deaths by 1 and 3 weeks, respectively, but overlaps the second wave in cases until early August, where overall mortality begins to decline, despite a continued rise in COVID-19 cases and deaths. In the NW region mortality peaks earlier than in the rest of the US (early April), and overlaps the peak in COVID-19 cases. The second peak in mortality follows the peak in COVID-19 cases by 2 weeks and mostly overlaps the COVID-19 death curve, as would be expected (data not shown for NW and MW).

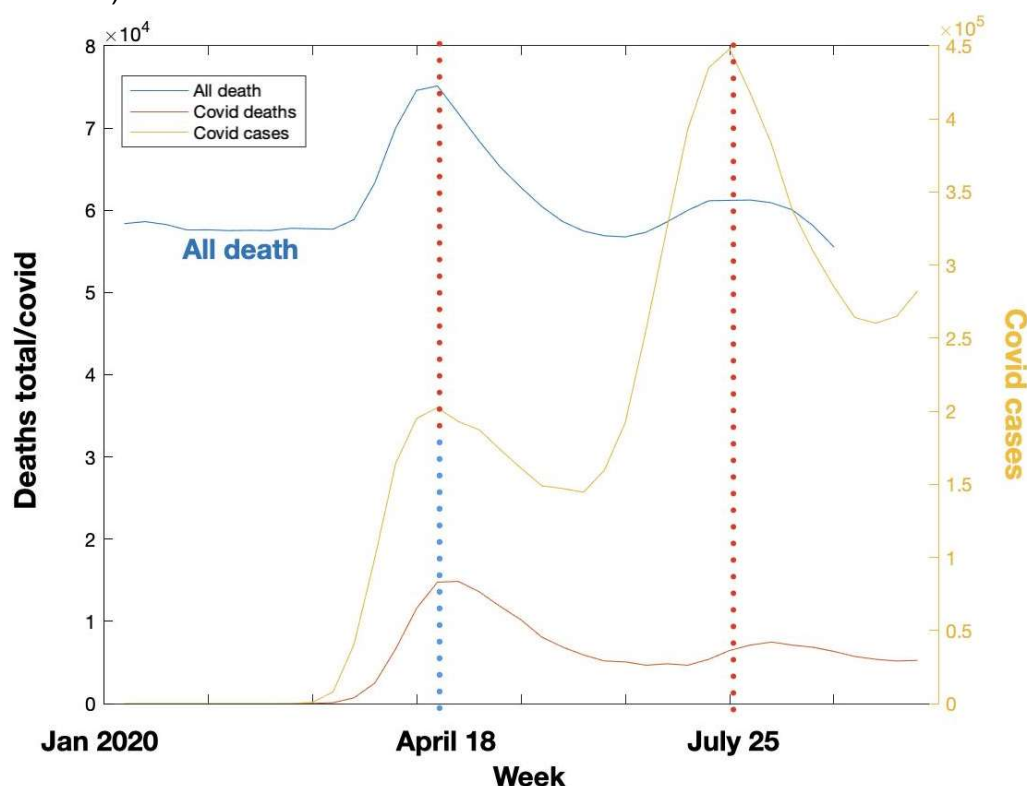


Figure 4a. Legend: Overall mortality (blue), COVID-19 deaths (red, left y-axis) and COVID-19 cases (yellow, right y-axis) in the entire US. Dashed blue/red lines indicate peak week for overall mortality and COVID-19 cases, respectively.

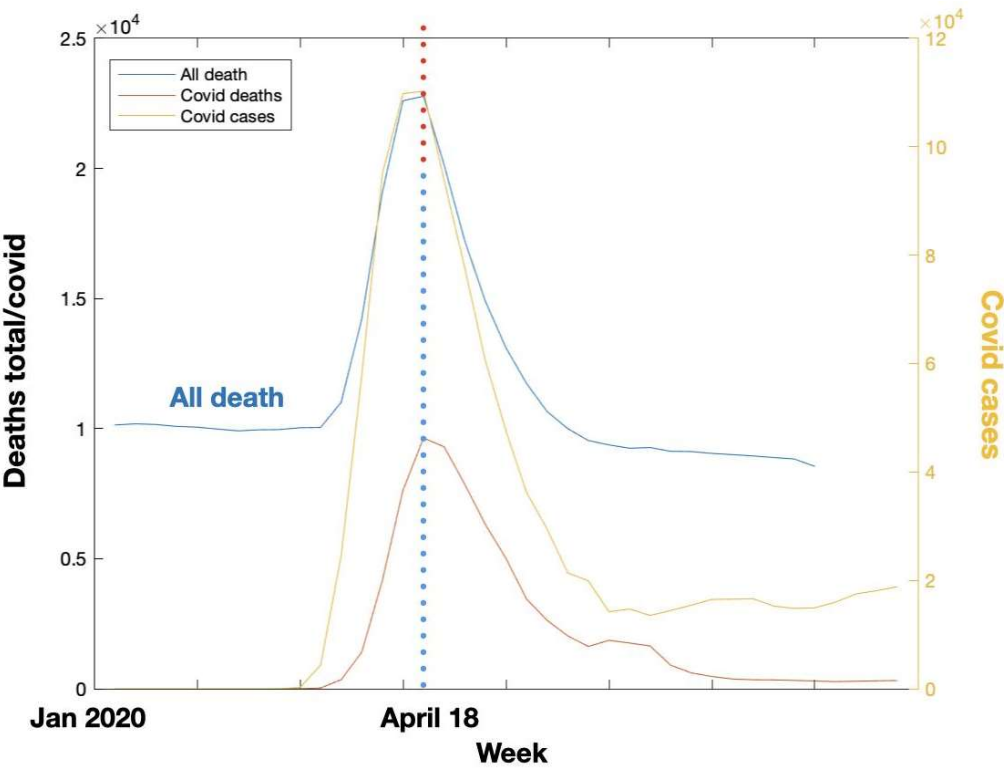


Figure 4b. Legend: Overall mortality (blue), COVID-19 deaths (red, left y-axis) and COVID-19 cases (yellow, left y-axis) in the NC region. Dashed blue/red lines indicate peak week for cases/deaths, respectively.

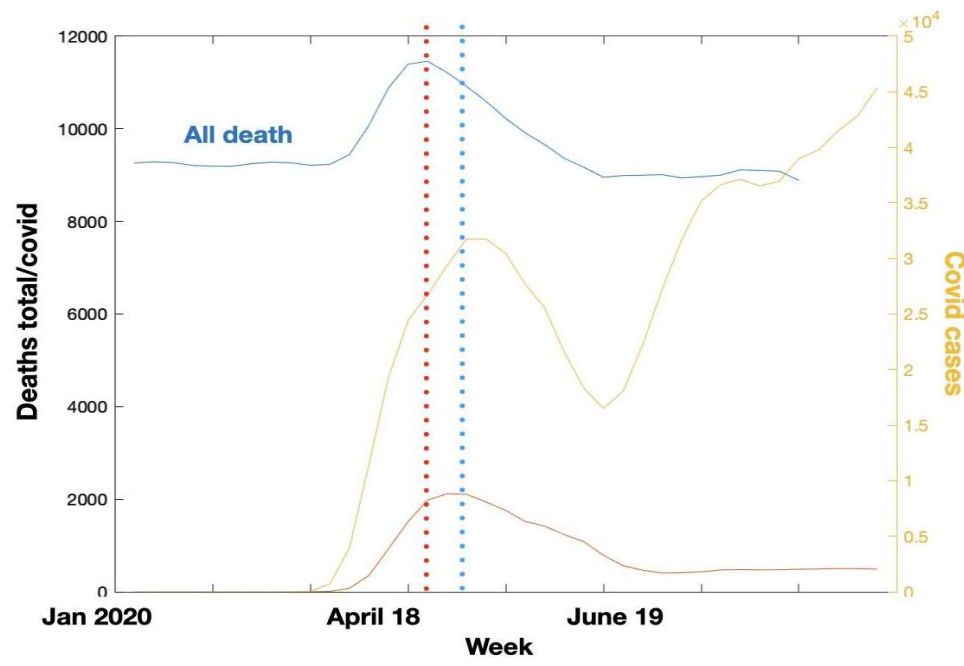


Figure 4c. Legend: Overall mortality (blue), COVID-19 deaths (red, left y-axis) and COVID-19 cases (yellow, left y-axis) in the ENC region. Dashed blue/red lines indicate peak week for cases/deaths, respectively.

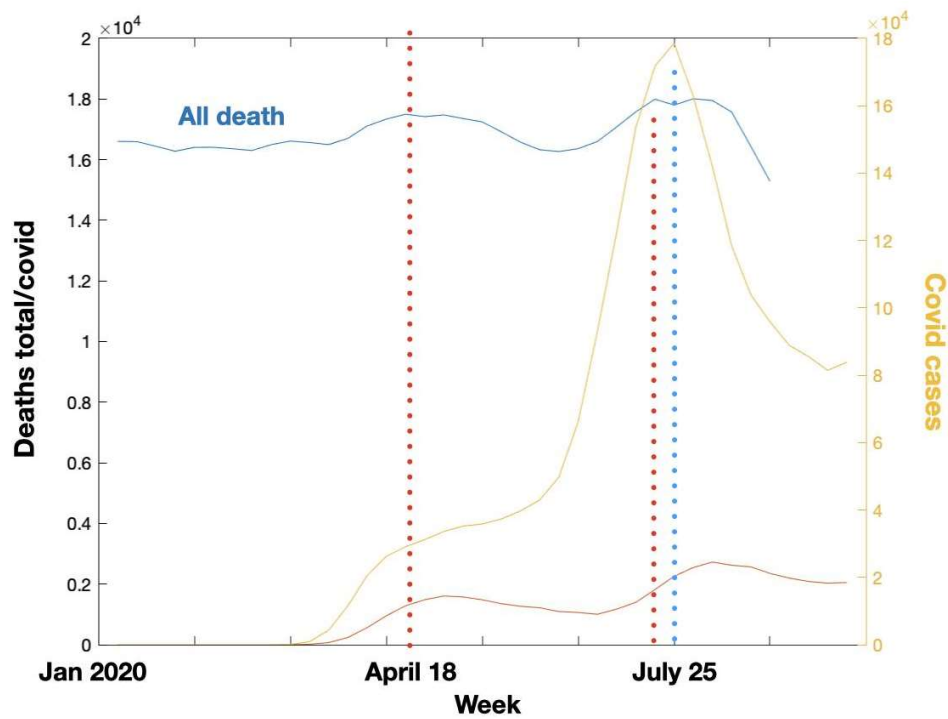


Figure 4d. Legend: Overall mortality (blue), COVID-19 deaths (red, left y-axis) and COVID-19 cases (yellow, left y-axis) in the SA region. Dashed blue/red lines indicate peak week for COVID-19 cases and overall mortality, respectively.

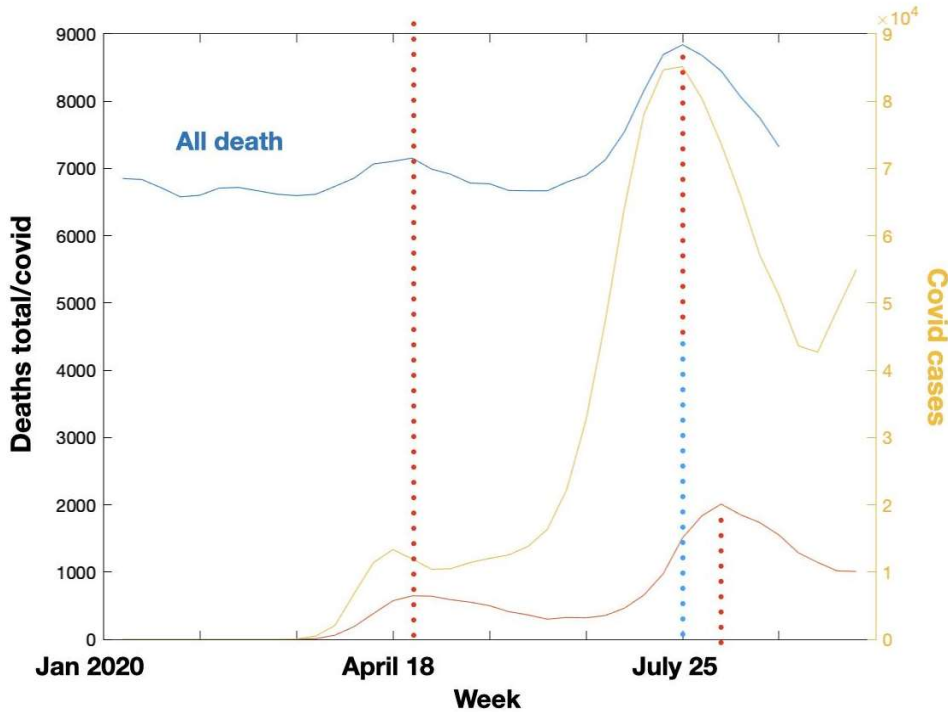


Figure 4e. Legend: Overall mortality (blue), COVID-19 deaths (red, left y-axis) and COVID-19 cases (yellow, left y-axis) in the S region. Dashed blue/red lines indicate peak week for COVID-19 cases and overall mortality, respectively.

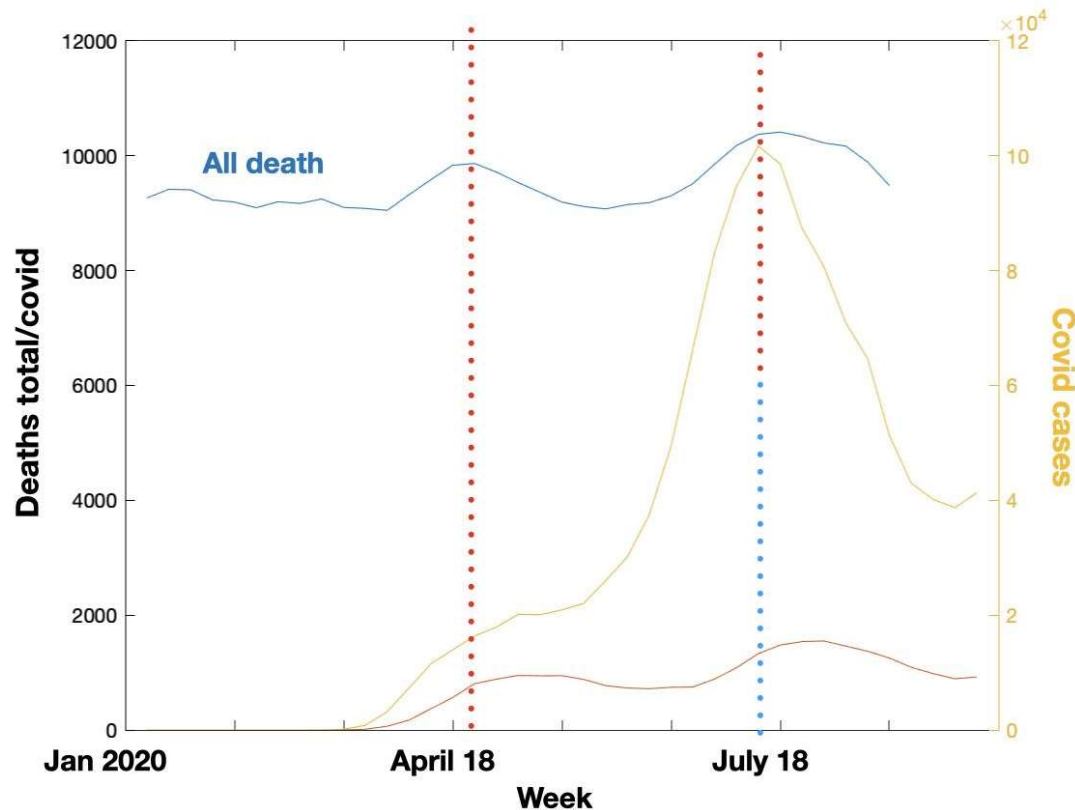


Figure 4f. Legend: Overall mortality (blue), COVID-19 deaths (red, left y-axis) and COVID-19 cases (yellow, left y-axis) in the SW region. Dashed blue/red lines indicate peak week for COVID-19 cases and overall mortality, respectively.

COVID-19 deaths as a proportion of all death

To better understand the relationship between reported COVID-19 deaths and all other deaths, I calculated the per week proportion of COVID-19 deaths to the overall mortality, termed here COVID-19 death ratio, or CDR (see methods). CDR values in the US (figure 5a) are consistent with previous reports (12, 14, 17), with COVID-19 deaths accounting for up to 20% of all deaths in the US, >50% of all deaths in places like MA, NJ, and >90% of deaths in NYC in late April (not shown).

As expected the CDR in the NC and ENC regions increased rapidly during the first wave and steadily declined after peaking in late April. Interestingly, despite some decline from peak levels, the impact of COVID-19 deaths in these regions was not negligible during the summer months, accounting for 5 and 10% of all deaths in the ENC and the NC regions, respectively. In the southern regions CDR was lower during the first wave at around 10%, but reached roughly 15% in the southern states in August (figure 5b). Both at the national and the north-south levels, there is a remarkable correlation between the CDR and the total number of COVID-19 deaths (figure 5c).

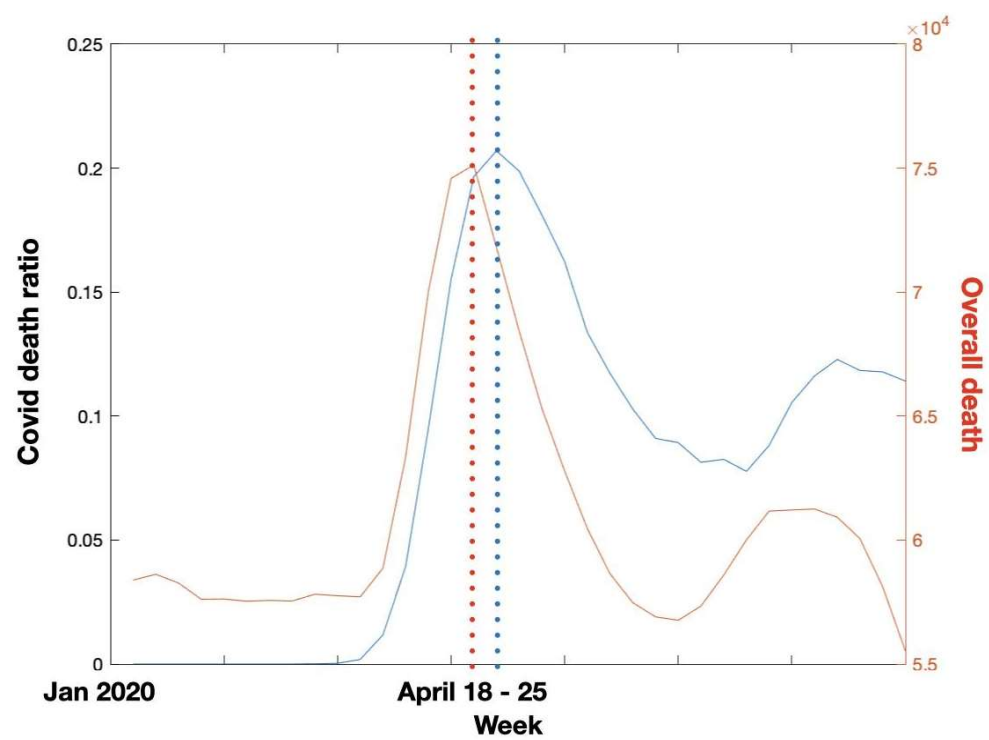


Figure 5a. Legend: COVID-19 death ratio (blue, left y-axis) and overall mortality (red, right y-axis) in the entire US, over time (x-axis). Dashed blue/red lines indicate peak week for CDR and overall mortality, respectively.

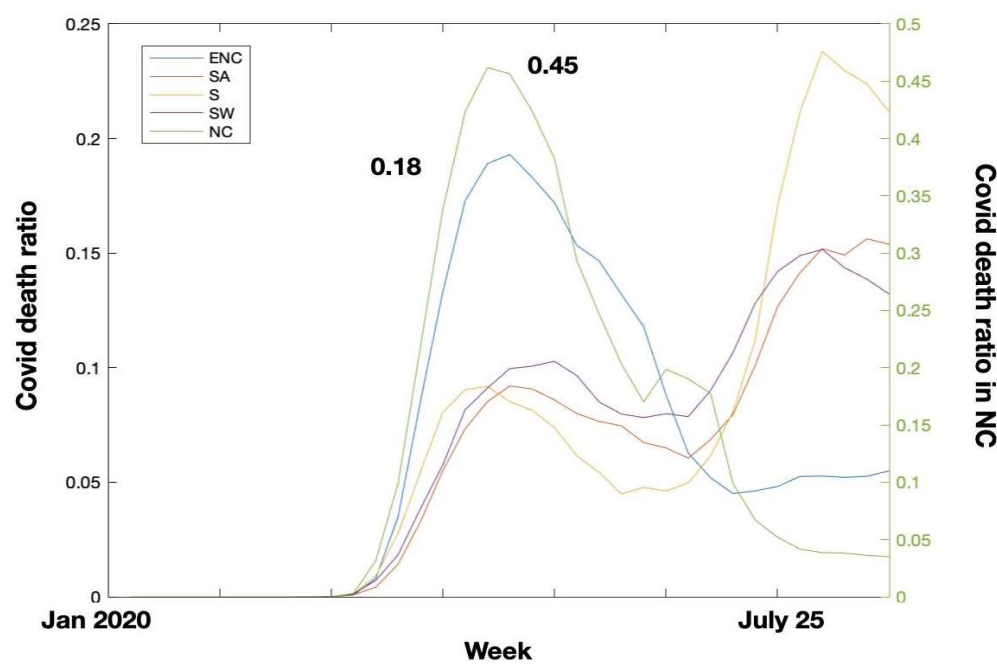


Figure 5b. Legend: COVID-19 death ratio (CDR) in the US, by region. During the first wave, the CDR in the NC region (green) reaches a peak (0.45) more than twice as high as in any other state (plotted on the right axis for comparison purposes). Note that the maximal CDR in the southern regions was <0.1 in the first wave, but higher in the second wave (<0.15 in the SA and

SW regions, <0.25 in the S region). Post-peak CDR in NC and ENC regions declined to 0.04-0.05, i.e during the summer months COVID-19 still accounted for 4-5% of all deaths in those regions

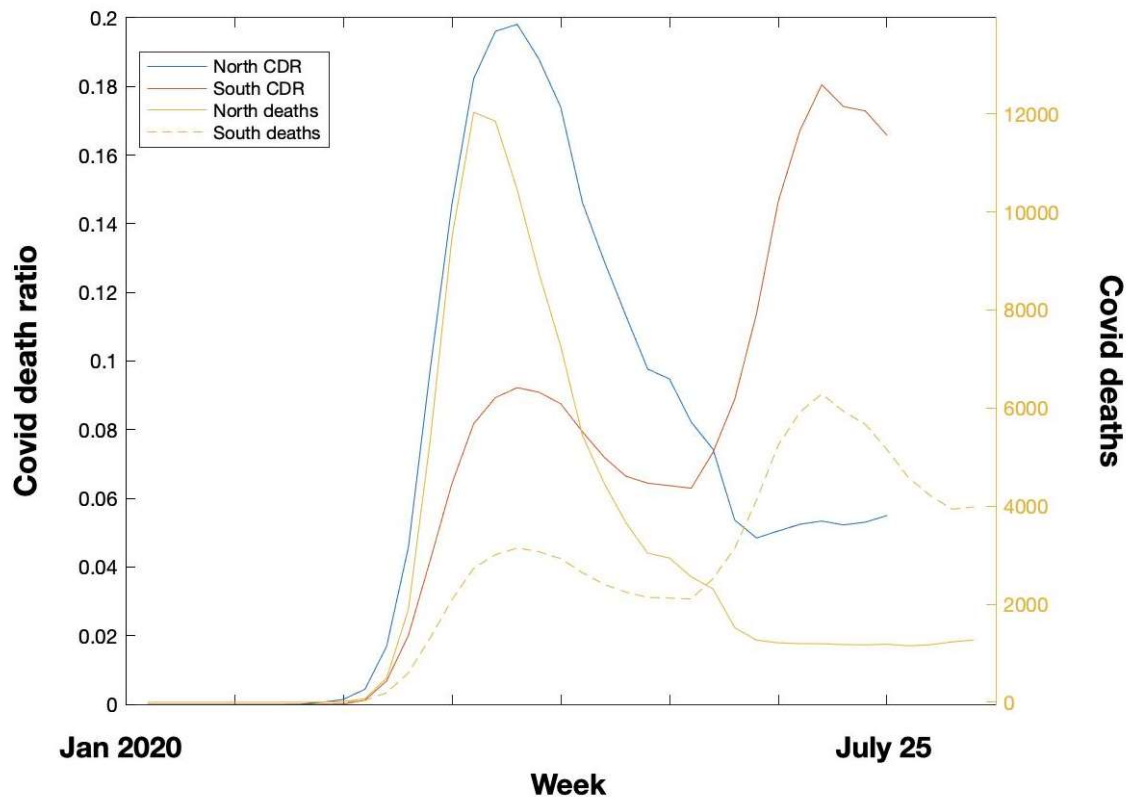


Figure 5c. Legend: COVID-19 death ratio (right y-axis) and COVID-19 deaths (left y-axis) in the north (NC, ENC, MW, NW) and south (SA, S, SW) of the US.

Conclusions

The potential danger to a population posed by a pathogen may be estimated from its rate of transmission, and the probability that an infection will lead to severe disease and death, that is, the total number of deaths the pathogen is expected to cause.

In the context of the COVID-19 pandemic these parameters were entirely unknown and had to be calculated from surveillance and clinical data. Therefore the accuracy of estimates of the virus transmission and fatality rates are limited by the volume and distribution of clinical tests performed in the patients as well as the general population, and the accuracy and reliability with which the virus can be determined to be the primary cause of death when identified in patients.

A complementary estimate of the human cost of the pandemic is the number of unexpected deaths, i.e excess deaths reported during the outbreak, which presumably were caused by the pathogen. The limitation here is that some degree of variation in the number of deaths per week exists regardless, and therefore even when the number of deaths is significantly elevated when

compared to past periods, it does not offer a quantitatively exact measure of the deaths directly caused by a pandemic.

Here I examine the relationship between COVID-19 cases and deaths, and the impact of the pandemic in terms of all-cause mortality, at the national level and in different geographic regions of the mainland US.

I show that the two waves of the pandemic in the US show a strong north-south pattern. Compared with the second wave, there were fewer cases, but more deaths recorded during the first wave, and it was far more pronounced in the northern regions of the US, especially the North, and East North Coast regions (figure 2).

The higher CFR during the first wave, as well as the time lag between cases and deaths during the two waves suggest that many cases during the first wave were not identified due to low volume of testing, and that the increase in cases may have been earlier than assumed, leading a shorter lag between cases and deaths during the first wave. These results apply at the national, as well as the regional levels, with several curious aberrations. For example, in the ENC region the peak number of deaths preceded the peak in cases during the first wave, and even more curiously, the increase in cases during the second wave was not followed by a corresponding increase in death (figure 3b). In the southern regions there is a steady increase in cases during the first wave, but the number of COVID-19 deaths peaks approx. a week after the northern regions, and declines until the end of June - i.e from May through June the number of COVID-19 deaths decreased while the number of cases increased (figure 3c-e).

When contrasted with all-cause mortality there are two waves of increased mortality, corresponding to the two pandemic waves. Surprisingly, the peaks in mortality coincided with COVID-19 cases, rather than COVID-19 deaths, which followed overall mortality by 1-3 weeks at the national level (figure 4a), as well as the regional level, with the exception of the NC region, where COVID-19 cases and deaths, and overall mortality all peaked on the week of April 18 (figure 4b).

As may be expected, the proportion of all-cause deaths that were caused by COVID-19, referred to here as the CDR, also followed a strong regional pattern, and correlated nearly perfectly in all regions with the number of COVID-19 deaths. Interestingly, although the number of COVID-19 deaths fell dramatically from peak levels in the NC and ENC regions, COVID-19 deaths remained substantial through the summer months, accounting for roughly 5% of weekly deaths in those regions (figure 5b)

The changes in the COVID-19 CFR over time are consistent with increased testing volume (the denominator when calculating CFR), and the greater number of deaths during the first wave may simply be a reflection of the far larger number of deaths in the NC region, especially in NY and NJ, [which itself may have had more to do with the decisions of health authorities in these states than it does with the virus.] The same may be applied to explain the shorter lag between

COVID-19 cases and deaths at the beginning of the outbreak as compared to the second wave of cases during the summer (figure 3).

The 1-3 week lag between increases in overall mortality and COVID-19 deaths is surprising and difficult to explain as a result of the increasing testing volume, as the lag is in fact longer during the second wave. For similar reasons it is also unlikely that this discrepancy results from differences in the way dates of “regular deaths” and COVID-19 deaths are reported. Moreover, there was no delay between the peak in overall mortality and peak in COVID-19 deaths in the NC region, further suggesting a reporting or data error is unlikely the cause.

In conclusion, COVID-19 cases and deaths in the US show a north-south pattern, with a first wave affecting mainly northern regions, and a second wave mostly in the southern regions of the US. As expected, the increase in overall mortality during the pandemic period shows a corresponding trend, but surprisingly, it appears to overlap the number of cases rather than COVID-19 deaths. Finally, at the peak of the pandemic COVID-19 deaths accounted for roughly 20% of all US deaths, and COVID-19 remains a significant cause of death even in all regions of the US.

References

- 1 Wu, Fan, et al. "A new coronavirus associated with human respiratory disease in China." *Nature* 579.7798 (2020): 265-269.
- 2 Zhou, Peng, et al. "A pneumonia outbreak associated with a new coronavirus of probable bat origin." *nature* 579.7798 (2020): 270-273.
- 3 Chinazzi, Matteo, et al. "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak." *Science* 368.6489 (2020): 395-400.
- 4 Kinross, Pete, et al. "Rapidly increasing cumulative incidence of coronavirus disease (COVID-19) in the European Union/European Economic Area and the United Kingdom, 1 January to 15 March 2020." *Eurosurveillance* 25.11 (2020): 2000285.
- 5 Jorden, Michelle A., et al. "Evidence for limited early spread of COVID-19 within the United States, January–February 2020." (2020).
- 6 Mizumoto, Kenji, et al. "Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020." *Eurosurveillance* 25.10 (2020): 2000180.
- 7 Huang, Chaolin, et al. "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China." *The lancet* 395.10223 (2020): 497-506.
- 8 Abubakar, I. I., T. Tillmann, and A. Banerjee. "Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013." *Lancet* 385.9963 (2015): 117-171.
- 9 Dawood, Fatimah S., et al. "Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study." *The Lancet infectious diseases* 12.9 (2012): 687-695.
- 10 Rajgor, Dimple D., et al. "The many estimates of the COVID-19 case fatality rate." *The Lancet Infectious Diseases* 20.7 (2020): 776-777.
- 11 Xiao, Ai Tang, Yi Xin Tong, and Sheng Zhang. "False-negative of RT-PCR and prolonged nucleic acid conversion in COVID-19: rather than recurrence." *Journal of medical virology* (2020).
- 12 Ioannidis, John PA, Cathrine Axfors, and Despina G. Contopoulos-Ioannidis. "Population-level COVID-19 mortality risk for non-elderly individuals overall and for non-elderly individuals without underlying diseases in pandemic epicenters." *medRxiv* (2020).

13 Almog, Gal. "A Pandemic since When?." (2020).

14 Weinberger, Daniel M., et al. "Estimation of excess deaths associated with the COVID-19 pandemic in the United States, March to May 2020." JAMA Internal Medicine (2020).

15 https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm (retrieved Sep 30).

16 <https://www.cdc.gov/nchs/nvss/vsrr/COVID19/index.htm> (retrieved Sep 30).

17 COVID, Team CDC, et al. "Geographic Differences in COVID-19 Cases, Deaths, and Incidence-United States, February 12-April 7, 2020." MMWR Morbidity and mortality weekly report 69 (2020).

18 Baud, David, et al. "Real estimates of mortality following COVID-19 infection." The Lancet infectious diseases (2020).

19 Wilson, Nick, et al. "Case-fatality risk estimates for COVID-19 calculated by using a lag time for fatality." Emerging infectious diseases 26.6 (2020): 1339.

20 Viboud, Cécile, et al. "Global mortality impact of the 1957–1959 influenza pandemic." The Journal of infectious diseases 213.5 (2016): 738-745.

21 Simonsen, Lone, et al. "A review of the 1918 herald pandemic wave: importance for contemporary pandemic response strategies." Annals of epidemiology 28.5 (2018): 281-288.