



# Covid-19: One year on...

**Ron Fricker** assesses the impact of the pandemic in the United States by calculating the number of “excess deaths”

**T**he impact of the Covid-19 pandemic is unclear to many people. Some of this is due to the nature and newness of this disease, where our understanding of the SARS-CoV-2 virus and Covid-19 is evolving in real time, and some is due to an “infodemic” of misinformation. For example, at a point when the Covid-19 death toll had exceeded 180,000 in the United States, Donald Trump incorrectly claimed that only 6% of the deaths were actually caused by the virus (cnb.cx/2lZg2L0). Perhaps not surprisingly, a Cornell University study found that Trump “was likely the largest driver of the Covid-19 misinformation ‘infodemic’” (nyti.ms/3mvqzLU), but a lack of understanding of the pandemic’s impact was and is a worldwide problem (bit.ly/38e7TLj).<sup>1,2</sup>

There are many reasons for this. One is that social media has become a dominant source of information and within that communication ecosystem it is difficult for users to separate truth from fiction (bit.ly/3ph1DJV; bit.ly/3oYTNNK). In the United States, another is that some politicians and

broadcast media pundits have spread false or misleading facts, narratives, and explanations to further their self-interests (see, for example, the following *PolitiFact* articles: bit.ly/3r4FUXj; bit.ly/34paqBk; bit.ly/38vDtVh; bit.ly/34rCoMV). In addition, the SARS-CoV-2 virus spreads quickly but subtly and manifests in differential ways (bit.ly/3ak4LAl), making it hard to directly observe cause and effect and thus confounding people’s ability to accurately assess their risk of getting Covid-19 (bit.ly/3p86qgD). The cumulative effect is a populace overwhelmed by information yet unsure of what to believe or do.

## Illustrating the confusion

Consider the use of Covid-19 case counts as a way to characterise impact. Issues begin with a not uncommon misunderstanding of the definition of a “case”. According to the Merriam-Webster dictionary, a medical case is “an instance of disease or injury”, but Covid-19 case counts are typically *confirmed* case counts. That is, these counts are instances of the disease that have been

substantiated either by a test or medical professional. So, the actual case count must be estimated, a problem that has been exacerbated in the United States which has lagged in testing capability and uniform standards. Furthermore, random testing is necessary in order to accurately estimate the prevalence of Covid-19 (see bit.ly/3nrEbca and bit.ly/3nwvq0l for additional discussion). Yet, in the absence of random testing, the United States has had to rely on less desirable measures, such as the positivity rate, to try to understand the spread of Covid-19 (bit.ly/2WozX9s). Compounding this, the virus affects individuals in about the broadest way possible, meaning some contract the virus and have no symptoms at all and others end up in the hospital or die. Thus, to some, the notion of a case seems either ill-defined or, for asymptomatic cases, incorrectly identified.

As I write this in early January 2021, in the United States the number of confirmed Covid-19 cases currently exceeds 21 million and the number of deaths attributed to Covid-19 exceeds 360,000. While counting Covid-related deaths seems like it might be straightforward, it too has been challenged. When someone dies in the United States, the immediate cause of death, along with up to three underlying conditions that “initiated the events resulting in death”, is recorded on a death certificate by a medical professional (bit.ly/3nAvblg). Covid-19 is typically an underlying condition to an immediate cause of death such as pneumonia or acute respiratory distress syndrome (bit.ly/37rSO9U). Unfortunately, some have falsely alleged that medical facilities are incorrectly classifying deaths as Covid-related for financial gain (bit.ly/3gZld99). While not true, for some people it has raised doubts about the accuracy of the number of Covid deaths.

It is thus no wonder that a layperson can become confused about the true impact of the disease. But it is not necessary to appeal to Covid-19 case counts and death counts to get a sense of the magnitude of this pandemic. Instead, let us look at what is referred to as “all-cause” mortality counts, meaning the total number of deaths no matter what the cause.

## Mortality statistics

In the United States, death certificates are filed with local health departments which



**Dr Ronald D. Fricker Jr** is the interim dean of the Virginia Tech College of Science, and a professor in the Virginia Tech Department of Statistics.

then report them to the National Center for Health Statistics (NCHS). As part of the National Vital Statistics System, the NCHS uses this information to tabulate mortality statistics for each state and for the entire country. Once aggregated, the data is publicly available on the Center for Disease Control and Prevention (CDC) website ([bit.ly/37xnVB4](https://bit.ly/37xnVB4)). According to the CDC's data, there were 2.84 million deaths in the USA in 2018, 2.85 million in 2019, and as of 7 January 2021 an estimated 3.27 million deaths for 2020.

Heart disease is the leading cause of death in the United States, with an annual mortality rate of just over 647,000 deaths per year. The American Cancer Society estimates that in 2020 there were slightly more than 606,500 cancer-related deaths, the second leading cause. At more than 360,000 deaths, Covid is the third leading cause of death in the United States in 2020 as measured by total deaths. To put this in context, in a country of about 330 million people, there were 36,500 motor vehicle fatalities in 2018. In 2017, about 36,000 people died from unintentional falls and about 40,000 from firearms. So, the total number of Covid-related deaths thus far is more than one-half of the number of each of the two leading causes of death. But it is ten times the total annual number of deaths due to firearms, or unintentional falls, or motor vehicle accidents.

Covid-19 is now the leading cause of death in the United States as measured by the number of daily deaths.<sup>3</sup> For example, on 7 January 2021 the US exceeded 4,000 Covid-related deaths in a single day ([bit.ly/39LyNuK](https://bit.ly/39LyNuK)). At that rate, more people will die of Covid in 10 days than will die from automobile accidents in an entire year.

## Calculating excess deaths

A simple comparison of the total number of deaths illustrates the impact of Covid in the USA. With more than 2.8 million deaths in each of 2018 and 2019, 3.27 million deaths in 2020 corresponds to an increase of slightly more than 420,000 deaths compared to the previous two years. That is a 14.8% increase in one year. While the population of the United States has been increasing over the past three years, that increase is less than 1% per year (on average), so the nearly 15% increase in deaths in 2020 is a substantial jump, even accounting for population changes.

To do a more sophisticated analysis requires estimating what 2020 would have been like if the pandemic had never happened. The CDC actually does this using an algorithm based on a Poisson generalised linear model initially developed by Farrington *et al.*<sup>4</sup> and improved upon by Noufaily *et al.*<sup>5</sup> The model is fitted to historical mortality data, where more recent data is adjusted for reporting delays, and it is used to project the weekly mortality under “normal” conditions. (See “The Farrington algorithm”, page 15, for a more detailed description.) The difference between the estimated mortality counts from the model and upward deviations in the actual counts then represents the number of excess deaths.

Figure 1 (page 14) shows four years of weekly mortality counts for the United States, compiled by the CDC ([bit.ly/37xnVB4](https://bit.ly/37xnVB4)) from data submitted by states and the District of Columbia, from the week ending 14 January 2017 to the week ending 26 December 2020. The height of the bars is the weekly mortality count, where prior to 2020 it varied from about 50,000 deaths per week at the lowest point in the summer to about 60,000 deaths per week at the peak in the winter. The black curve is the expected weekly mortality count from the model and the grey curve is the upper bound of the 95% prediction interval for each week, the threshold above which the mortality count is considered to be significantly high.

A number of important aspects of US mortality are evident in the graph. First, a substantial increase in the number of deaths beginning in mid-March 2020 and continuing to the present is unmistakably visible, where the first confirmed Covid-19 case in the United States occurred on 20 January 2020.<sup>6</sup> Since March, mortality in the United States has increased by at least 11.6% compared to the past three years if we conservatively just look at deaths above the threshold, and it could be as much as 14.5% if we look at all deaths above the expected count. It does not take a sophisticated analysis to see that mortality has distinctly and substantially increased during the pandemic when compared to historical trends.

Second, the number of deaths above the threshold since March (the red part of the bars) sums to 328,900 and the number of deaths above the expected counts (the orange and red parts of the bars) sums to

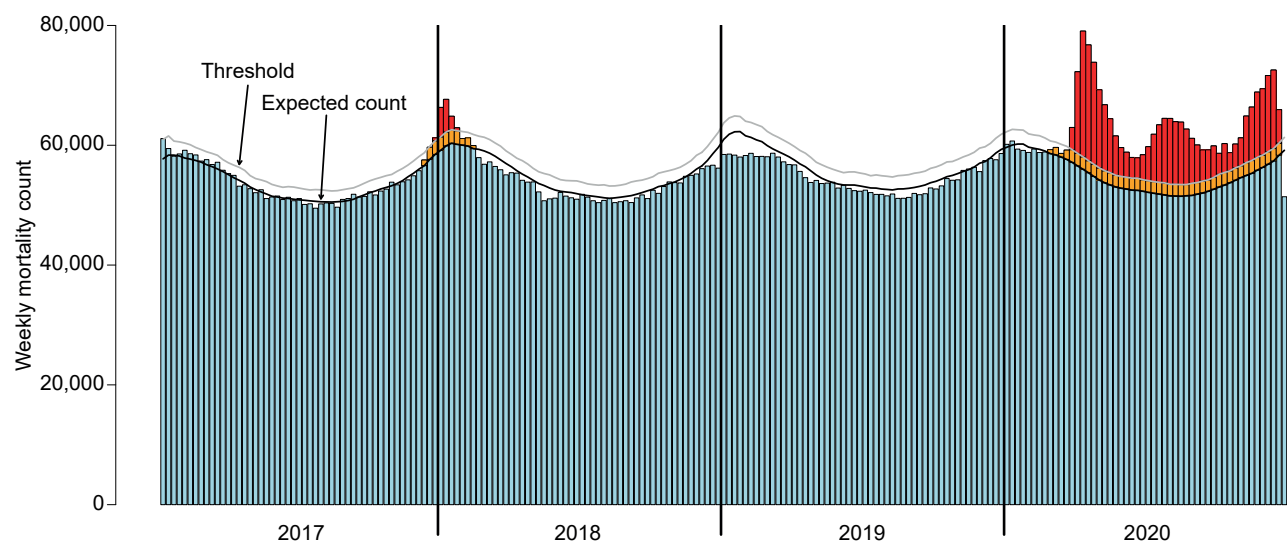
411,702. So, the number of additional deaths since the start of the pandemic in the United States is at least 329,000 and could be as large as 412,000. The number of deaths the CDC currently attributes to Covid-19 as of 6 January 2021 is 356,005 ([bit.ly/35uw73j](https://bit.ly/35uw73j)) which is at the low end of this range.

What is behind these additional deaths is not yet completely clear, where it may be that Covid-19 deaths are being undercounted. It also may be that public health measures taken during the pandemic have changed the baseline mortality, perhaps increasing the number of deaths due to other causes. It is likely some of both. For example, a recent study in the *Journal of the American Medical Association* found that 67% of US excess deaths between March and July 2020 documented Covid-19 as a cause of death.<sup>7</sup> But increased mortality from heart disease and two spikes for deaths related to Alzheimer's disease/dementia were also identified during that period. These may be due to delayed medical treatment, perhaps because of the impact of the pandemic on medical facilities, or perhaps because some people did not seek medical treatment to minimise their potential for exposure to Covid-19 ([bit.ly/3p6VLCW](https://bit.ly/3p6VLCW)).

Third, also visible in January 2018 is another period of excess deaths caused by an unusually virulent flu strain that winter. Comparing the two periods plainly shows that the mortality the United States is experiencing in this pandemic is much worse than the flu, even when compared to a year like 2018 in which an estimated 61,000 people died from influenza. Indeed, the weekly number of deaths during this spring and summer have frequently exceeded the peaks in mortality that tend to occur in the winter.

## Age and race/ethnicity

Looking a bit deeper into the data, there are differences by age and by race/ethnicity. Figure 2 (page 14) displays the number of deaths by age category for 2015 to 2020. The plot shows that mortality is up in 2020 in all age categories compared to 2015–2019, though note that for those under 25 the numbers were decreasing until 2020. Table 1 (page 14) shows the percentage increase for 2020 compared to 2019. While greater numbers of people are dying in the older age groups, which is natural, somewhat surprisingly given the media coverage, ►



**Figure 1:** Plot of weekly mortality counts in the United States from the week ending 14 January 2017 to the week ending 26 December 2020. The orange colour denotes the fraction of a weekly count that is above the expected count and below the threshold and the red colour denotes the fraction of a weekly count above the threshold. (The drop in the counts for the last two weeks of 2020 is probably due to data reporting lags exacerbated by the holiday period. Given the current high rate of daily Covid deaths in the United States, these bars will most likely be revised upwards once all the data is reported.)

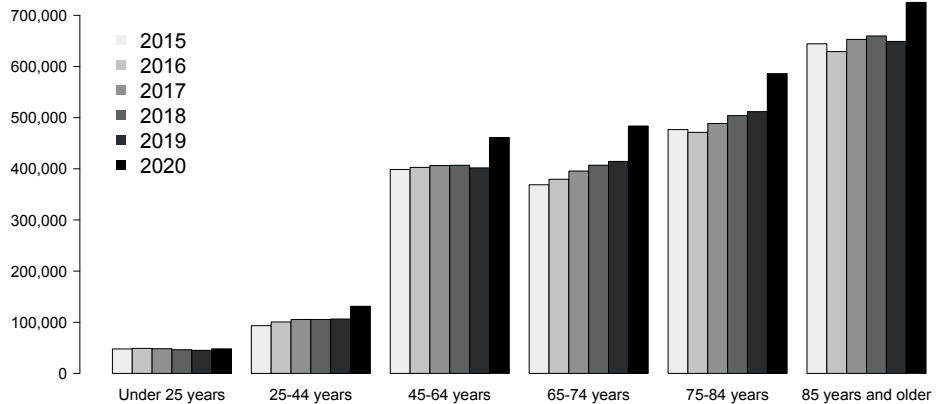
► Covid-19 has had the greatest percentage impact on those in the 25–44 age group. Figures 3 and 4 show the data by race and ethnicity, where Figure 3 shows increases across all categories, though they are hard to see in some groups because of the differences in magnitude. Figure 4 shows that, while total mortality is higher for non-Hispanic Whites simply due to the population size, the percentage increase across all minority groups from 2019 to 2020 is substantially greater – by a factor of nearly 2 and almost as much as 5 – compared to non-Hispanic Whites.

Conclusions

Using excess deaths as a measure, the impact of the Covid-19 pandemic on the United States should be much clearer. It has resulted in a substantial increase in mortality across all age groups and races/ethnicities, although with a disproportionately greater impact on non-White populations. Also, while much of the reporting and discussion has been on increased mortality among older populations, the greatest percentage increase in mortality is in the 25–44 age group. Assessing the effects of the pandemic using excess mortality sidesteps the sometimes contentious issues related to whether any particular death was caused by Covid-19. Moreover, excess mortality is useful as an overall measure of the pandemic’s impact.

For example, as previously mentioned, Woolf *et al.* found increases in heart disease and Alzheimer’s disease/dementia-related mortality coincident with the spring surge in Covid-19 cases in the United States.<sup>7</sup> They also found increases in Alzheimer’s disease/dementia-related mortality coincident with the summer Covid-19 surge in sunbelt states. These increases may not be directly attributable to Covid-19 but they could be the result of pandemic-related impacts on the health-care

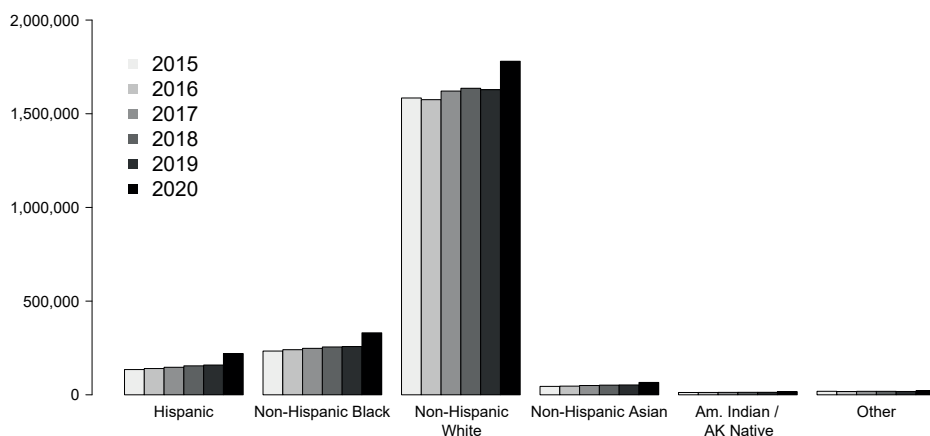
system and/or unintended side effects of policies to slow the spread of Covid-19. That said, the number of deaths attributed to Covid-19 is within the range of excess deaths and, in fact, it is at the lower end of that range. This suggests that the number of deaths currently attributed to Covid may also be an undercount of the actual number of Covid-related deaths. As a final sobering note, as of 7 January 2021 the Institute for Health Metrics and



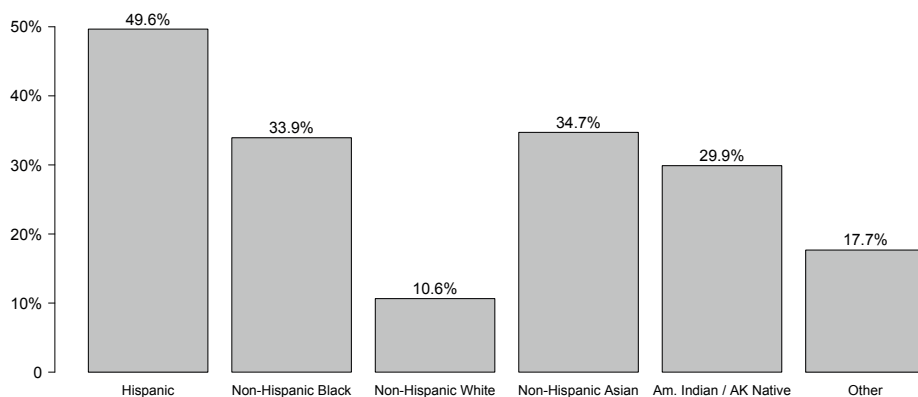
**Figure 2:** Total US mortality by age category and year.

**Table 1:** Percentage change in 2020 mortality from 2019 by age category.

Under 25	25–44	45–64	65–74	75–84	85 plus
6.7%	23.4%	14.8%	16.7%	14.5%	11.7%



**Figure 3:** Total US mortality by race/ethnicity and year.



**Figure 4:** Percentage increase in 2020 mortality compared to the average mortality from 2015 to 2019 by race/ethnicity.

Evaluation projects more than 567,000 Covid-19 deaths in the United States by 1 April 2021 ([bit.ly/3mzsQWa](https://bit.ly/3mzsQWa)). That puts the number of deaths in the United States due to this pandemic on track to equal or exceed the number of deaths from the 1918 Spanish flu pandemic. While the US population in 1918 was one-third of what it is today, it is no less of a horrifying comparison. Furthermore, it is a tragedy that, in an age when scientists were able to create multiple highly effective Covid vaccines in less than a year, a misinformation “infodemic” has caused the death toll to be far greater than it should have been. ■

#### Disclosure statement

The author declares no conflicts of interest.

#### References

1. Editorial (2020) The COVID-19 infodemic. *The Lancet Infectious Diseases*, **20**(8), 875.
2. Galvão, J. (2020) COVID-19: The deadly threat of misinformation. *The Lancet Infectious Diseases*. doi: 10.1016/S1473-3099(20)30721-0.
3. Woolf, S. H., Chapman, D. A. and Lee, J. H. (2020) COVID-19 as the leading cause of death in the United States. *Journal of the American Medical Association*. doi: 10.1001/jama.2020.24865
4. Farrington, C. P., Andrews, N. J., Beale, A. D. and Catchpole, M. A. (1996) A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society, Series A*, **159**, 547–563.
5. Noufaily, A., Enki, D. G., Farrington, P., Garthwaite, P., Andrews, N. and Charlett, A. (2012) An Improved algorithm for outbreak detection in multiple surveillance systems. *Statistics in Medicine*, **32**(7), 1206–1222.
6. Holshue, M. L., DeBolt, C., Lindquist, S., Lofy, K. H., Wiesman, J., Bruce, H., et al. (2020) First case of 2019 novel coronavirus in the United States. *New England Journal of Medicine*, **382**, 929–936.
7. Woolf, S. H., Chapman, D. A., Sabo, R. T., Weinberger, D. M., Hill, L. and Taylor, D. D. H. (2020) Excess deaths from COVID-19 and other causes, March–July 2020. *Journal of the American Medical Association*, **324**(15), 1562–1564.
8. Fricker, R. D., Jr (2013) *Introduction to Statistical Methods for Biosurveillance, with an Emphasis on Syndromic Surveillance*. Cambridge: Cambridge University Press.
9. Salmon, M., Schumacher, D., and Höhle, M. (2016) Monitoring count time series in R: Aberration detection in public health surveillance. *Journal of Statistical Software*, **70**(10), 1–35.

## The Farrington algorithm

In health surveillance, algorithms such as Farrington’s are used to predict a disease’s expected or “normal” state using historical data. Then a substantial increase above what is expected is taken as evidence of a possible outbreak or, in this case, an unusual increase in mortality. Critical in the implementation of any surveillance algorithm is calibrating it to maximise the probability of detecting an increase while constraining the number of false positive signals. These trade off much like Type I and Type II errors do in classical hypothesis testing.<sup>8</sup>

The original Farrington algorithm<sup>4</sup> and its improved version<sup>5</sup> are based on an overdispersed Poisson generalised linear model with spline terms to account for trends such as seasonality in the mortality counts and then to assess deviations in the observed count from the prediction. The algorithm also incorporates logic to address issues of missing data and the presence of a linear trend. The *surveillance* package in R is used by the CDC to implement the Farrington algorithms.<sup>9</sup>

When the Farrington algorithm is used for surveillance, an increase is taken as statistically significant if the observed count exceeds a threshold calculated as a one-sided 95% prediction interval for the next week’s mortality count. As employed here, the threshold is used to establish a lower bound on the number of excess deaths, which is the sum of excess counts exceeding the threshold from February to 26 December 2020. In comparison, the sum of the differences between the expected counts predicted by the model and the observed counts is taken as the likely or best estimate of the number of excess deaths.