COVID-19 Counts in Washington State and Select Other Locations

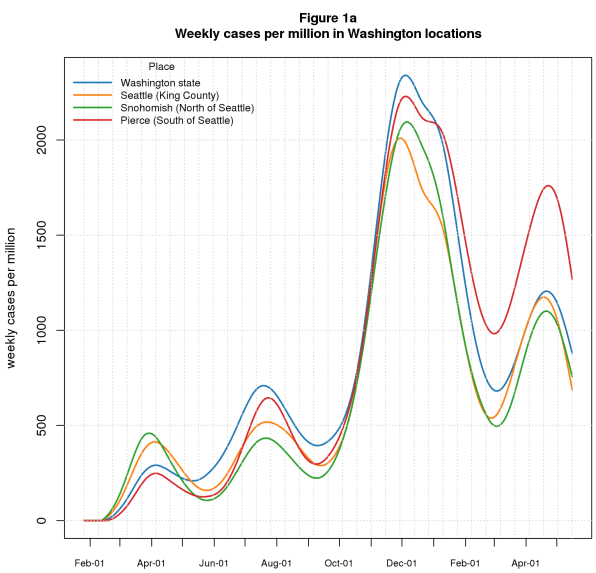
Nathan (Nat) Goodman

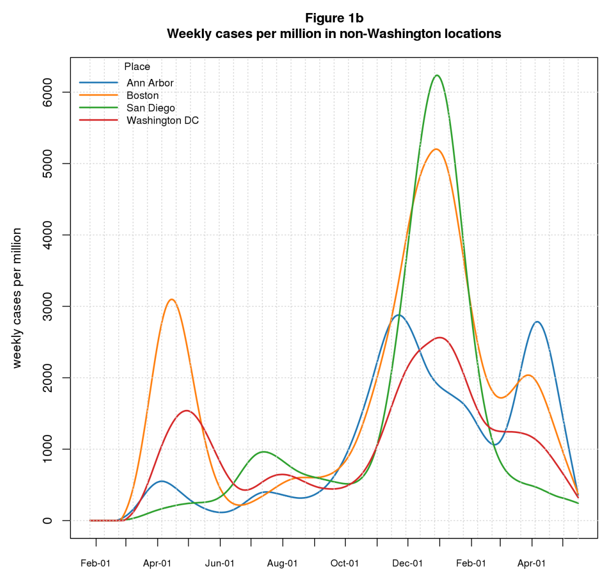
May 26, 2021

Counts of cases, hospital admissions, and deaths are key metrics of COVID-19 prevalence and burden, and are the basis for model-based estimates and predictions of these statistics. I present here graphs showing these metrics over time in Washington state and a few other USA locations of interest to me. I update the graphs and this write-up weekly. Previous versions are here. See below for caveats and technical details.

Figures 1a-b show smoothed *case* counts per million for several Washington and non-Washington locations. The Washington locations are the entire state, the Seattle area where I live, and the adjacent counties to the north and south (Snohomish and Pierce, resp.). The non-Washington locations are Ann Arbor, Boston, San Diego, and Washington DC.

The figures use data from Johns Hopkins Center for Systems Science and Engineering (JHU), described below. When comparing the Washington and non-Washington graphs, please note the difference in y-scale: the current Washington rates (about 700-1300 per million) are greater than the non-Washington rates (250-400 per million).





The Washington graph (Figure 1a) shows that rates are declining in all locations. Trend analysis suggests the decline is real everywhere except Pierce. The Pierce data is ambiguous: cases were increasing until the most recent data but declined considerably this last week. Let's hope this decline is real.

The picture for non-Washington locations (Figure 1b) is far better: cases are declining everywhere. Trend analysis confirms the decline everywhere except San Diego where cases continue to plateau.

Figures 2a-b show *deaths* per million for the same locations. When comparing the Washington and non-Washington graphs, again please note the difference in y-scale: the current Washington rates are 7-11 per million; the non-Washington rates are 3-18 per million.

Figure 2a
Weekly deaths per million in Washington locations

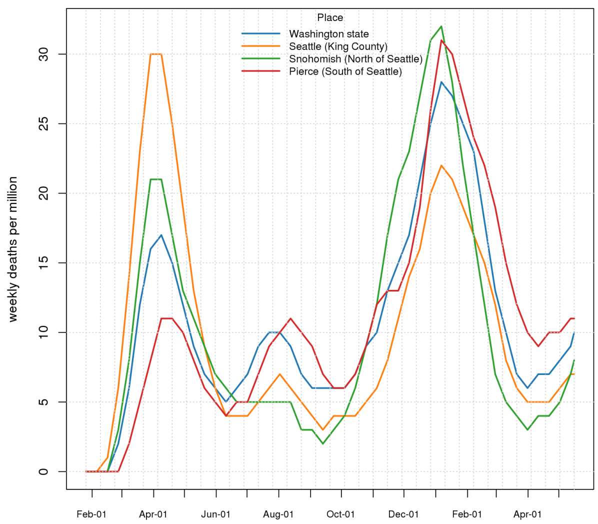
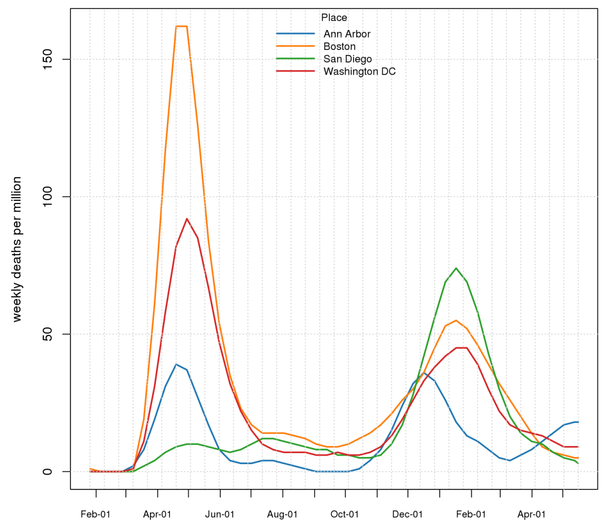


Figure 2b
Weekly deaths per million in non-Washington locations



The Washington data (Figure 2a) shows three waves. The second peak was thankfully lower than the first; the third wave exceeded the first in all areas except Seattle (King County). The graphs are well down from the first and third peaks, but have climbed from their recent valleys to levels near or above their summer 2020 peaks. The graphs are heading up in some locations and flat in others, but trend analysis indicates the data is too variable to be confident in the direction.

The non-Washington data (Figure 2b) shows early peaks in most locations, followed by a long trough, followed by a second wave starting in November 2020. The graphs are well down from their recent peaks. Deaths are increasing in Ann Arbor while continuing down elsewhere. Trend analysis indicates that the increase in Ann Arbor is slowing, and the recent data elsewhere may be flattening.

The next graphs show the Washington results broken down by age. This data is from Washington State Department of Health (DOH) weekly downloads, described below. An important caveat is that the DOH download systematically undercounts events in recent weeks due to manual curation. I extrapolate data for late time points as discussed below.

Figure 3 is *cases*. Figure 4 shows hospital admissions (*admits*) and *deaths*.

Figure 3
Weekly cases per million by age in Washington state

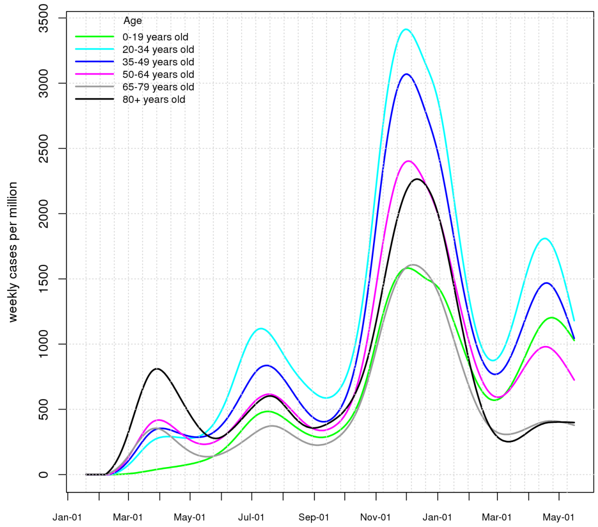
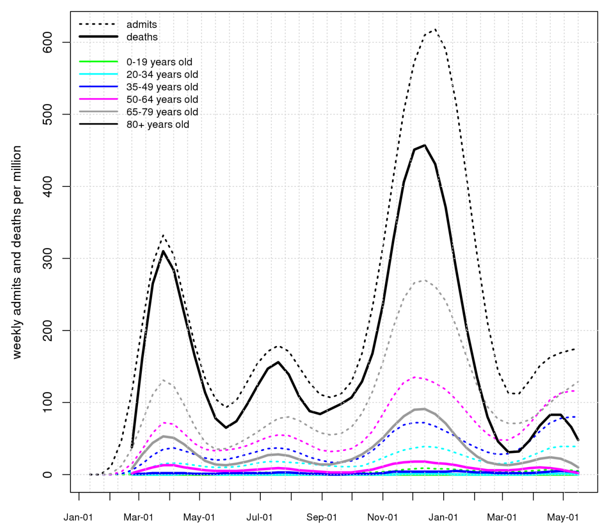


Figure 4
Weekly admits and deaths per million by age in Washington state



Early on, the pandemic struck older age groups most heavily. Over time, *cases* spread into all age groups, even the young. During the second wave, older groups did better with young adults and middle aged adults (20-34 and 35-49 years) becoming the most affected groups. The third wave swept into all age groups. As the wave grew, the oldest people (80+) were again strongly affected. The curves are heading down in all but the oldest age groups (65-79 and 80+) for whom the rates are flat.

The shocking *death* rate of the 80+ age group jumps off the page in Figure 4. The *death* rate in this age group shows four waves. The third wave reached its peak in December 2020. The fourth wave peaked in April 2021 at a level well below the earlier peaks. Throughout the pandemic, the death rate for the 80+ group has been much higher than for the 65-79 group; this discrepancy continues even as rates subside.

The *admit* rate shows the same four waves, and seems to be increasing slowly in some age groups. Given the declining death rates now, it seems reasonable to be optimistic that deaths will remain low in the coming weeks.

In versions before April 28, 2021, Figure 4 only showed *deaths*, but the growing impact on young people combined with the high vaccination rate of the old, make admissions a useful indicator of pandemic status. These earlier versions aggregated 0-64 years into a single group, since the *death* rate in these ages is near 0. Now that the graph includes *admits*, it's useful to keep the age groups separate.

Caveats

1. The term *case* means a person with a detected COVID infection. In some data sources, this includes "confirmed cases", meaning people with positive molecular COVID tests, as well as "probable cases". I believe JHU only includes "confirmed cases" based on the name of the file I download.
2. Detected *cases* undercount actual cases by an unknown amount. When testing volume is higher, it's reasonable to expect the detected count to get closer to the actual count. Modelers attempt to correct for this. I don't include any such corrections here.
3. The same issues apply to *deaths* to a lesser extent, except perhaps early in the pandemic.
4. The geographic granularity in the underlying data is *state* or *county*. I refer to locations by city names reasoning that readers are more likely to know "Seattle" or "Ann Arbor" than "King" or "Washtenaw".

5. The date granularity in the graphs is weekly. The underlying JHU data is daily; I sum the data by week before graphing.
6. I truncate the data to the last full week prior to the week reported here.

Technical Details

1. I smooth the graphs using a smoothing spline (R's `smooth.spline`) for visual appeal. This is especially important for the *deaths* graphs where the counts are so low that unsmoothed week-to-week variation makes the graphs hard to read.
2. The Washington DOH data (used in Figures 3 and 4 to show counts broken down by age) systematically undercounts events in recent weeks due to manual curation. I attempt to correct this undercount through a linear extrapolation function (using R's `lm`). I have tweaked the extrapolation repeatedly, even turning it off for a few weeks. The current version uses a model that combines date and recentness effects. In past, I created models for each Washington location and age group but had to change when DOH changed its age groups on March 14, 2021 (see below). I now create a single model for the state as a whole and all age groups summed together, then blithely apply that model to all locations and ages.
3. The trend analysis computes a linear regression (using R's `lm`) over the most recent four, six, or eight weeks of data and reports the computed slope and the p-value for the slope. In essence, this compares the trend to the null hypothesis that the true counts are constant and the observed points are randomly selected from a normal distribution. After looking at trend results across the entire time series, I determined that p-values below 0.1 indicate convincing trends; this cutoff is arbitrary, of course.
4. In previous versions of the document, I've shown plots that overlay raw *case* and *death* counts onto the smoothed graphs in Figures 1 and 2. Recent data is much less variable than in past making these plots superfluous.
5. In most versions before March 17, 2021, I showed counts broken down by age (as in Figures 3 and 4) graphs for each Washington location. Now I only show the statewide graphs: the other locations are similar.
6. In past, DOH reported results in 20 year age groups starting with 0-19, with a final group for 80+. As of the March 14, 2021 data release (corresponding to document version March 17), DOH changed age groups. The new groups are 0-19, followed by 15 year ranges (20-34, 35-49, 50-64, 65-79) with a final group for 80+.

Data Sources

Washington State Department of Health (DOH)

DOH provides three COVID data streams.

1. Washington Disease Reporting System (WDRS) provides daily "hot off the presses" results for use by public health officials, health care providers, and qualified researchers. It is not available to the general public, including yours truly.
2. COVID-19 Data Dashboard provides a web graphical user interface to summary data from WDRS for the general public. (At least, I think the data is from WDRS - they don't actually say).
3. Weekly data downloads (available from the Data Dashboard web page) of data curated by DOH staff. The curation corrects errors in the daily feed, such as, duplicate reports, multiple test results for the same incident (e.g., initial and confirmation tests for the same individual), incorrect reporting dates, incorrect county assignments (e.g., when an individual crosses county lines to get tested).

The weekly DOH download reports data by age group. In past, the groups were 20-year ranges starting with 0-19, with a final group for 80+. As of March 14, 2021, they changed the groups to an initial 20-year range (0-19), then several 15-year ranges (20-34, 35-49, 50-64, 65-79), with a final group for 80+.

Johns Hopkins Center for Systems Science and Engineering (JHU)

JHU CSSE has created an impressive portal for COVID data and analysis. They provide their data to the public through a GitHub repository. The data I use is from the csse_covid_19_data/csse_covid_19_time_series directory: time_series_covid19_confirmed_US.csv for *cases* and time_series_covid19_deaths_US.csv for *deaths*.

JHU updates the data daily. I download the data the same day as the DOH data (now Tuesdays) for operational convenience.

Other Data Sources

The population data used for the per capita calculations is from Census Reporter. The file connecting Census Reporter *geoids* to counties is the Census Bureau Gazetteer.

Comments Please!

Please post comments on Twitter or Facebook, or contact me by email natg@shore.net.

Copyright & License