COVID-19 Counts in Washington State and Select Other Locations

Nathan (Nat) Goodman

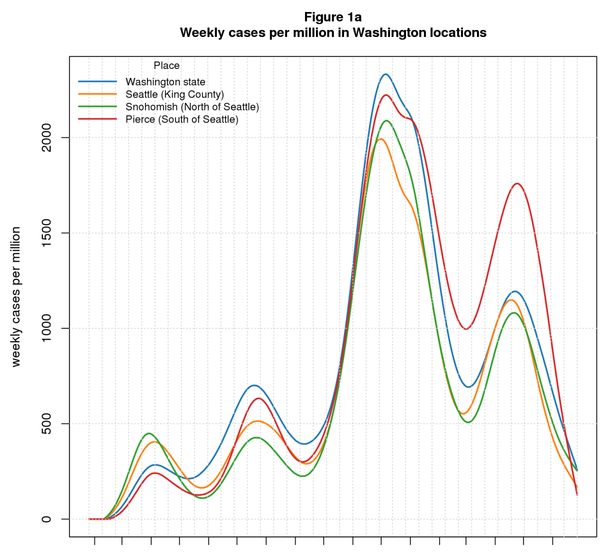
July 7, 2021

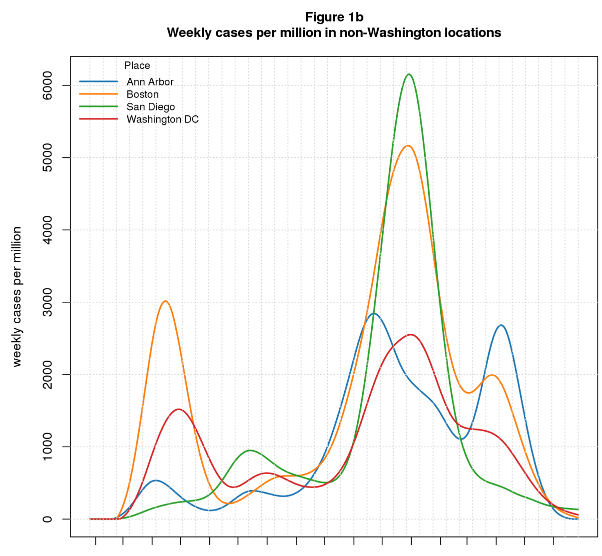
Counts of cases, hospital admissions, and deaths are key metrics of COVID-19 prevalence and burden, and are the basis for model-based estimates and predictions of these statistics. I present here graphs showing these metrics over time in Washington state and a few other USA locations of interest to me. 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. See below for details on the smoothing method.

Previous versions also showed raw data overlaid on the smoothed to illustrate variability. With *cases* and *deaths* so low, variability is no longer a big issue.

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 counts for Washington locations are about 150-250 per million; non-Washington counts are about 0-150 per million.





The graphs for Washington (Figure 1a) and non-Washington locations (Figure 1b) show that rates are declining in all locations. Trend analysis supports the decline.

Figures 2a-b show smoothed *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 0-7 per million; the non-Washington rates are 0-2 per million.

Figure 2a
Weekly deaths per million in Washington locations

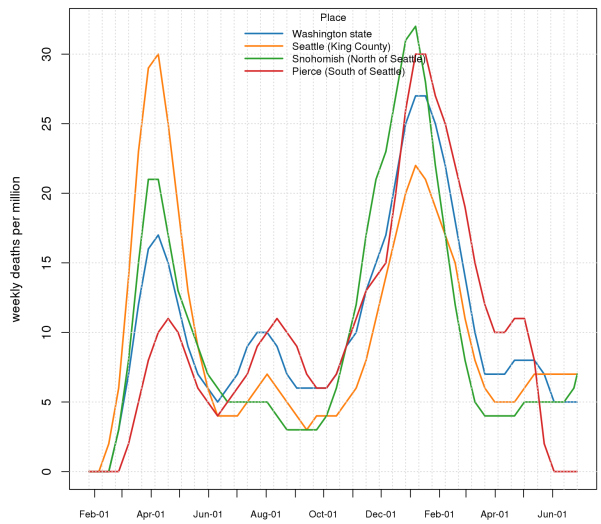
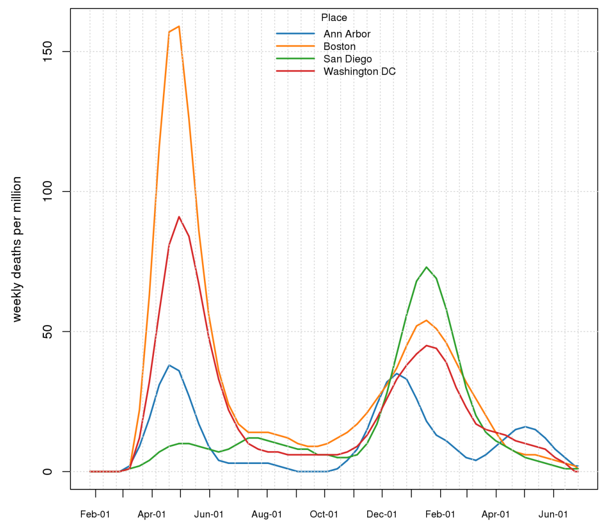


Figure 2b
Weekly deaths per million in non-Washington locations



The graphs for Washington (Figure 2a) show rates increasing in Snohomish (North of Seattle) and flat everywhere else. Trend analysis suggests that rates are actually flat everywhere; the apparent increase in Snohomish is probably an artifact of data variability and fitting.

The graphs for non-Washington locations (Figure 2b) are declining everywhere. Trend analysis suggests that rates are actually flat; counts are so low that further decline is meaningless.

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* and figure 4 shows hospital admissions (*admits*) and *deaths*.

Figure 3
Weekly cases per million by age in Washington state

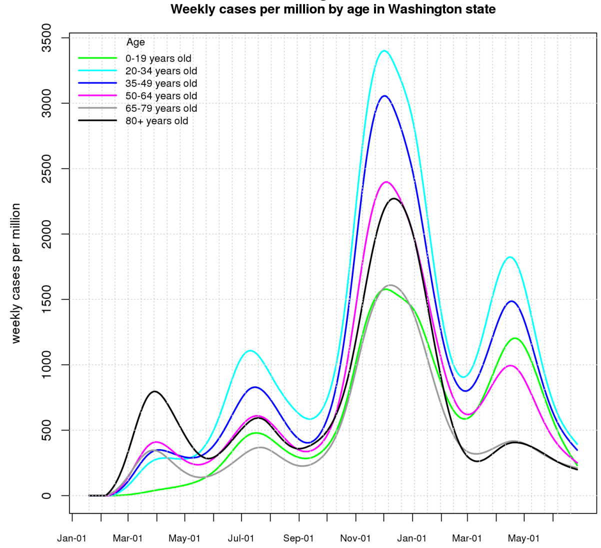
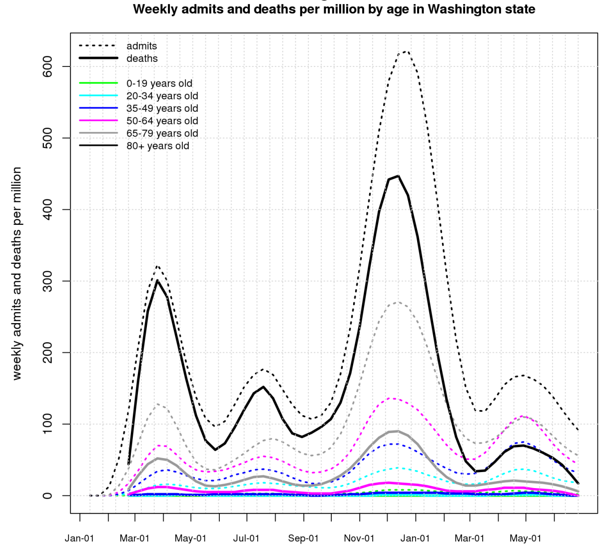


Figure 4
Weekly admits and deaths per million by age in Washington state



Early on, the pandemic struck older age groups most heavily, but *cases* quickly spread into all age groups, even the young. Thankfully, rates are now declining in all age groups. The oldest age groups (65-79 and 80+) are doing best, perhaps due to the high vaccination coverage in these people.

The shocking *death* rate of the 80+ age group jumps off the page in Figure 4. The *death* rate for 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 was much higher than any other group, but it now appears that deaths even in this group have dropped to near zero. Given the declining *case* rates, it seems reasonable to be optimistic that deaths will remain low in the coming weeks.

The *admit* rate shows the same four waves, and seems to be decreasing in all groups.

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. I also compute a regression using daily data over the most recent 7-42 days. 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 0.15 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