

COVID-19 Counts in Washington State and Select Other Locations

Nathan (Nat) Goodman

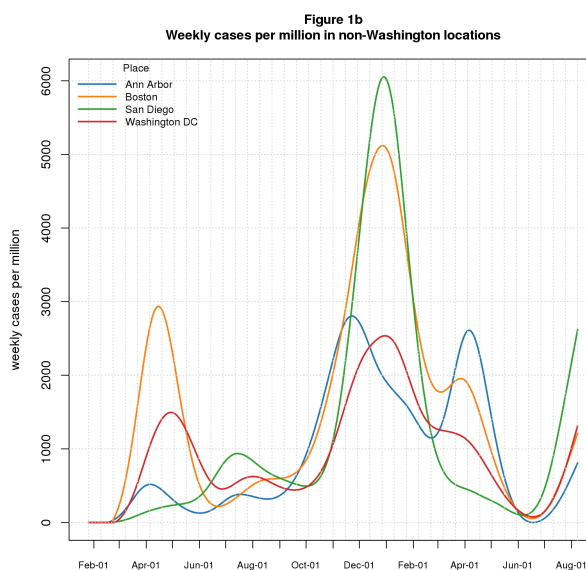
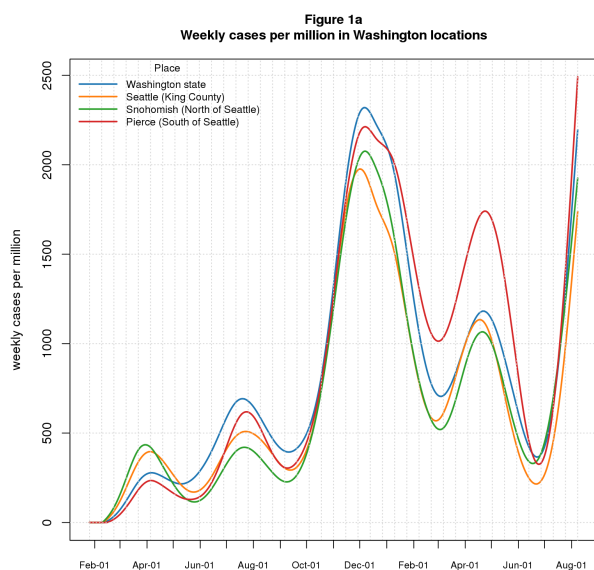
August 18, 2021

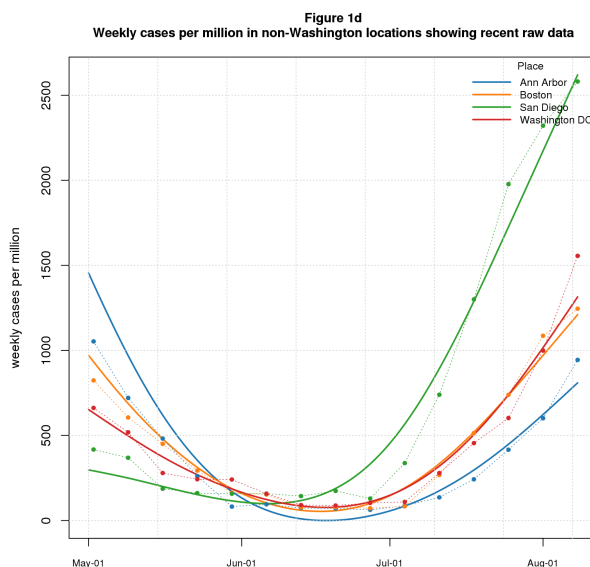
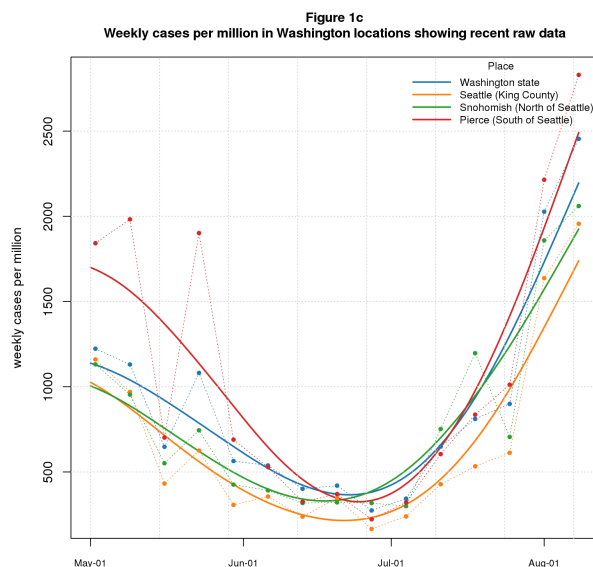
Counts of cases, hospital admissions, and deaths are key metrics of COVID-19 prevalence and burden, and are the basis for model-based estimates and predictions of these statistics. I present here graphs showing these metrics over time in Washington state and a few other USA locations of interest to me.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

The first several figures (1a-2d) show *case* and *death* counts per million for several Washington and non-Washington locations using data from Johns Hopkins Center for Systems Science and Engineering (JHU), described [below](#). 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.

Figures 1a-b (the top row) show smoothed *case* counts (See [below](#) for details on the smoothing method). Figures 1c-d (the bottom row) overlay raw data onto the smoothed since May 1 to help explain recent trends.

When comparing the smoothed Washington and non-Washington graphs (Figures 1a-b), please note the difference in y-scale; the graphs with raw data (Figures 1c-d) have the same y-scale and are better for comparing Washington and non-Washington locations. The current raw counts for Washington locations are about 1900-2850 per million; non-Washington raw counts are about 950-2600 per million.

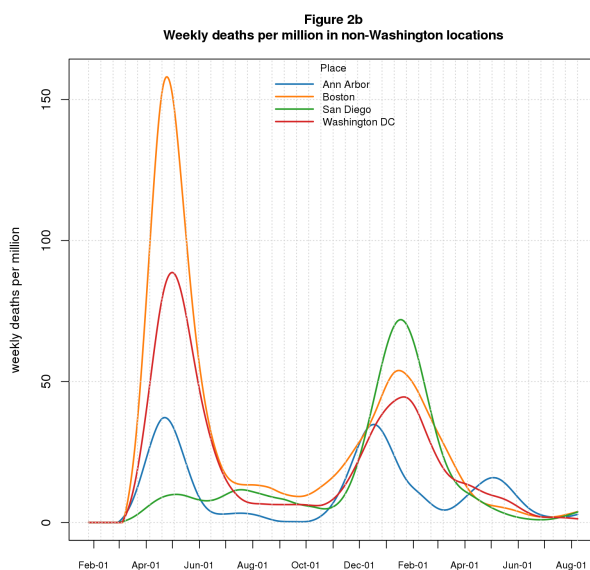
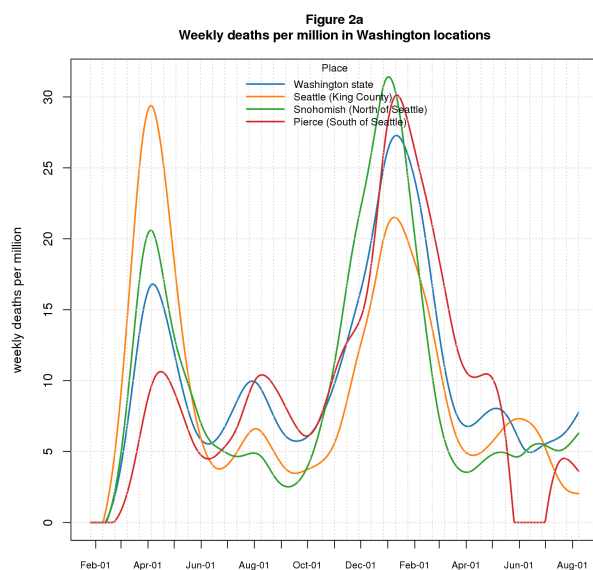


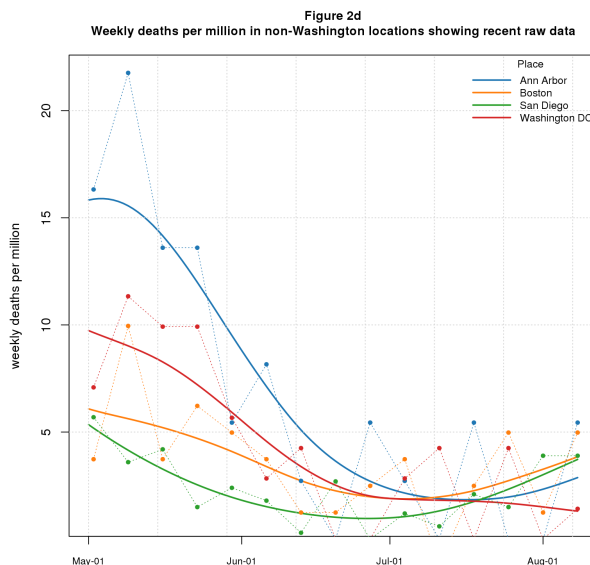
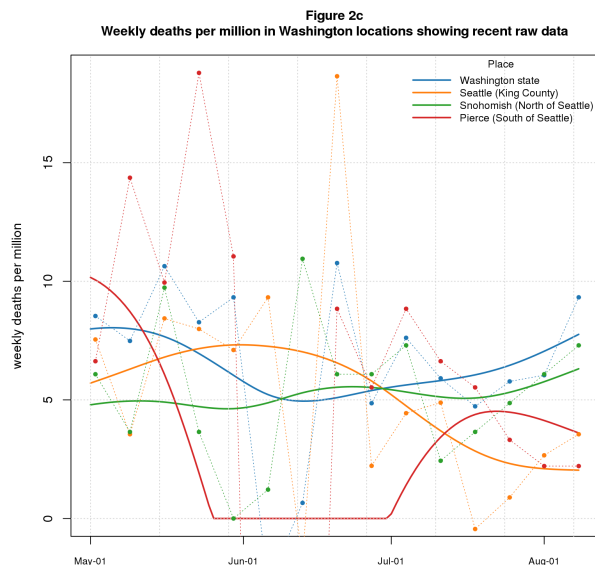


The smoothed graphs for Washington and non-Washington locations (Figures 1a-b) show rates increasing in all locations. The raw data and trend analysis concur. Rates in all Washington locations have soared beyond all previous peaks except winter 20/21. Rates in non-Washington locations, though climbing rapidly, remain below their recent peaks.

Figures 2a-d show *deaths* per million for the same locations. As with the *cases* graphs, the top row (Figures 2a-b) show smoothed data (see details [below](#)) and the bottom row (Figures 2c-d) overlays raw data onto the smoothed since May 1.

As with the cases data, when comparing the Washington and non-Washington graphs, the graphs with raw data (Figures 2c-d) are probably more useful as they have similar y-scales. The current Washington rates are 2-9 per million and non-Washington rates are 1-5 per million.



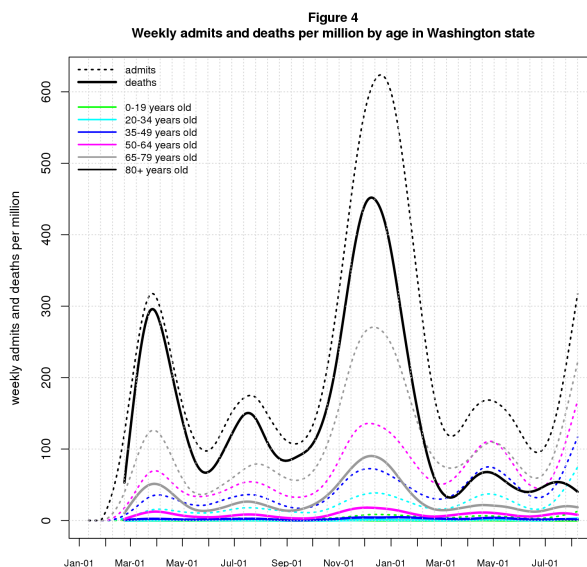
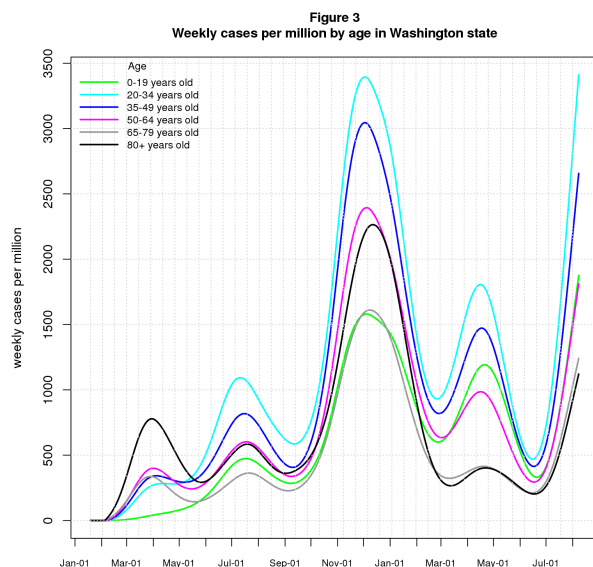


The smoothed graphs for Washington locations (Figure 2a) show some rates up and some down. Trend analysis and raw data (Figure 2c) suggest that Washington rates are heading up everywhere except Pierce; the Pierce downward trend may be artifacts of a few very high data points in early summer. The long flat line for Pierce in late May-June 2021 is due to negative counts for two weeks in early June: this arises when JHU discovers they overcounted previous weeks and are catching up; my software clamps the fit to zero reasoning that negative counts are meaningless.

The smoothed graphs for non-Washington locations (Figure 2b) show rates going up everywhere except Washington DC. Trend analysis and raw data (Figure 2d) suggest that the data are variable, but the trends remain flat.

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*.



Early on, the pandemic struck older age groups most heavily, but *cases* quickly spread into all age groups, even the young. Rates are heading up again in all age groups. Young adults (20-34 years) are leading the way, with the next older group (35-49 years) close behind. Though not apparent in the graph, the youngest group (0-19) is also increasing rapidly and is now at about the same level as pre-seniors (50-64). The oldest age groups (65-79 and 80+ years) are doing best, perhaps due to the high vaccination coverage in these people, but they too are climbing fast.

Throughout the pandemic, the death rate for the 80+ group was much higher than any other group, but thankfully deaths even in this group have dropped to near zero.

Admits are increasing for all age groups, even the young, and are now above their spring 2021 levels and near or above their spring 2020 levels. *Admit* rates vary with age groups exactly as one would expect: lowest in the young and increasing with age.

It remains to be seen whether the increasing *case* and *admit* rates will lead to higher *deaths* in the coming weeks.

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 data by week before graphing.
6. I truncate 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. Now that *death* counts are low, the smoothing method seems not to be working well.
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 most versions before March 17, 2021, I showed counts broken down by age (as in Figures 3 and 4) for each Washington location. Now I only show the statewide graphs: the other locations are similar.
5. 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.

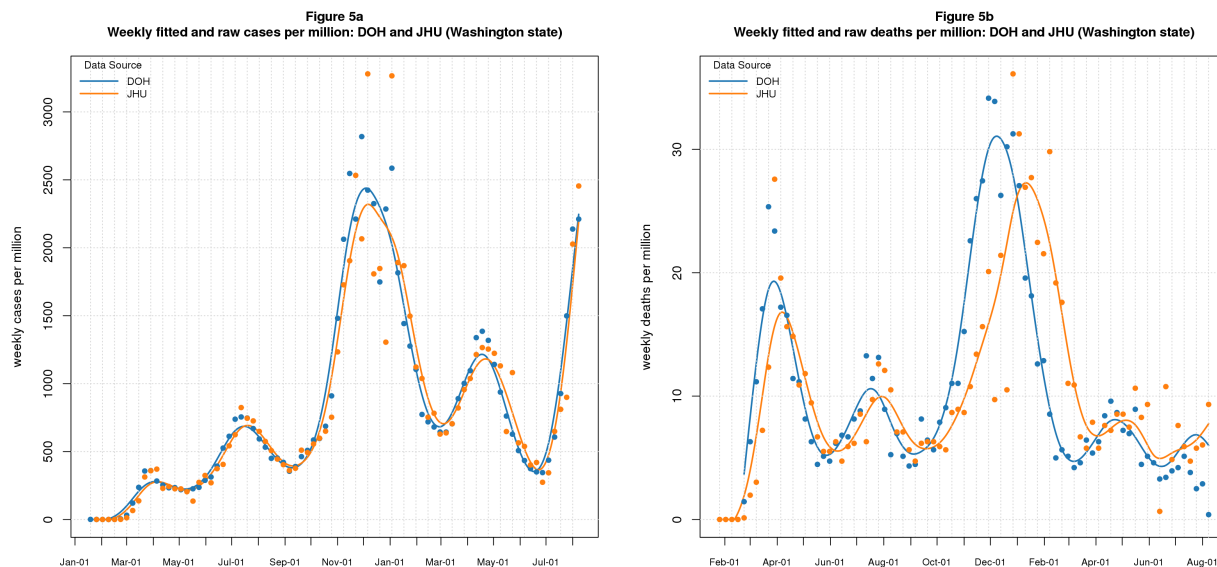
1. [Washington Disease Reporting System \(WDRS\)](#) provides daily “hot off the presses” results for use by public health officials, health care providers, and qualified researchers. It is not available to the general public, including yours truly.
2. [COVID-19 Data Dashboard](#) provides a web graphical 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+.

Issues with DOH Undercounting

Figures 5a-b compare DOH and JHU *cases* and *deaths* for Washington state to illustrate the undercount in the raw DOH data. The *cases* data matches well except for a few periods spanning several weeks, including

most of May 2020. The *deaths* data matches less well and is presently much lower than JHU. I believe the discrepancy in the *deaths* data reflects the consistent undercount of recent DOH data.



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

Copyright (c) 2020-2021 Nathan Goodman

The software is **open source and free**, released under the [MIT License](#). The documentation is **open access**, released under the [Creative Commons Attribution 4.0 International License](#).