COVID-19 Counts in Washington State and Select Other Locations

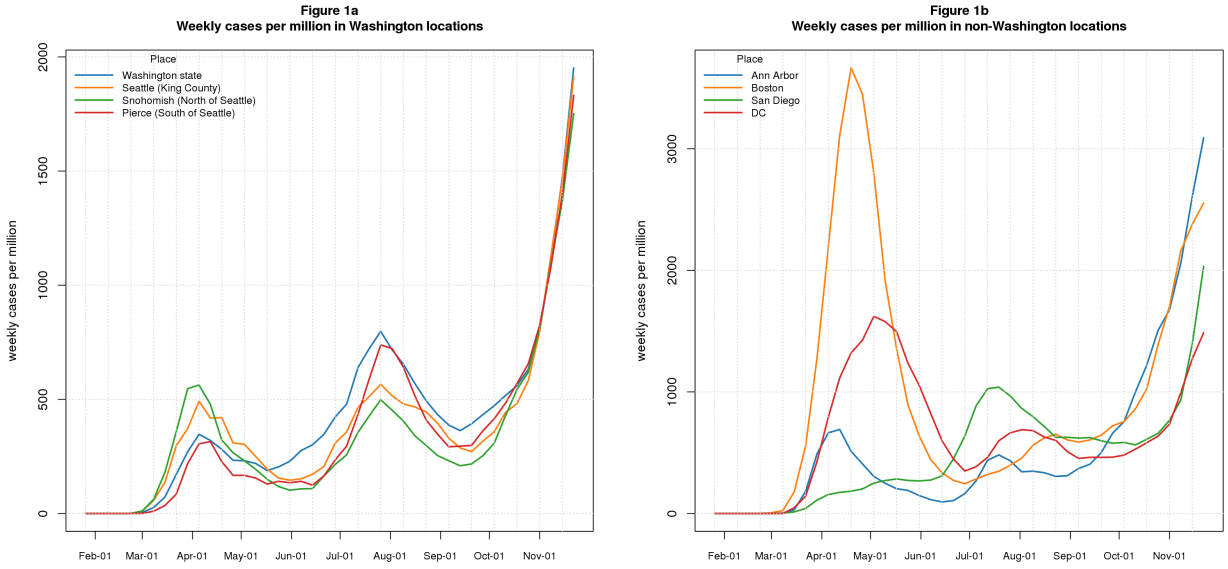
Nathan (Nat) Goodman

November 30, 2020

Counts of cases 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 hope to update the graphs weekly.

See below for caveats and details.

Figure 1a shows *case* counts per million for Washington state as a whole, the Seattle area where I live, and the adjacent counties to the north and south (Snohomish and Pierce, resp.) Figure 1b shows *case* counts for Ann Arbor, Boston, San Diego, and Washington DC. The figures use data from Johns Hopkins Center for Systems Science and Engineering (JHU), described below.



Cases are rapidly increasing everywhere. In Washington state (Figure 1a), case counts are at their highest ever and continuing to rise. What jumps out from the non-Washington data (Figure 1b) are the high early peaks in Boston and DC, followed by smaller peaks in most locations, and rapidly increasing counts now.

Figures 2a-b show *deaths* per million for the same locations.

Figure 2a
Weekly deaths per million in Washington locations

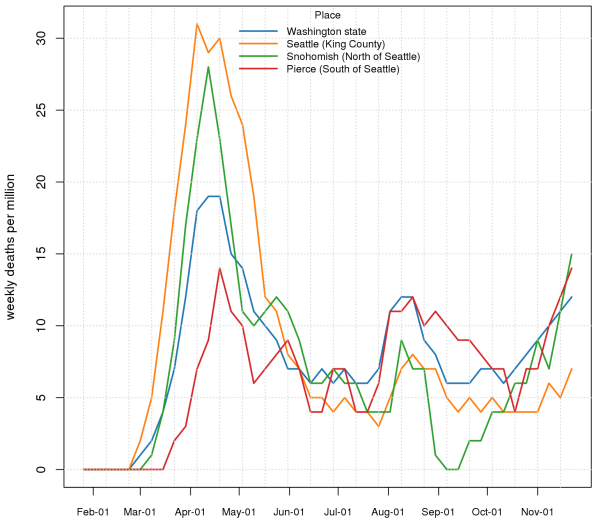
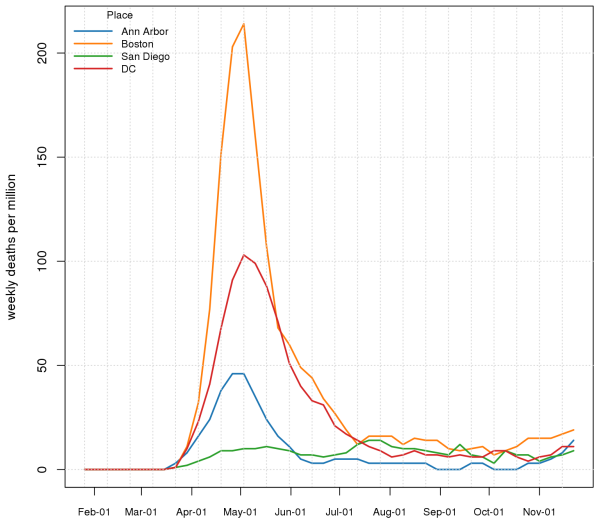


Figure 2b
Weekly deaths per million in non-Washington locations



The Washington data (Figure 2a) shows two waves with a third starting. The second peak was thankfully lower than the first; it remains to be seen how high the third wave will go. Early peaks dominate the non-Washington data (Figure 2b) making it hard to see details later on, but the rates are much lower than the early peaks and seem generally steady.

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 I extrapolate data for the last few weeks to compensate for systematic undercounting in the DOH weekly downloads. See below. The extrapolation is probably overshooting by as much as 25%. Inter-group comparisons are reasonable but the exact numbers may be too high.

Figures 3a-d are *cases*. The graphs are split into 20-year age ranges starting with 0-19, with a final group for 80+.

Figure 3a
Weekly cases per million by age in Washington state

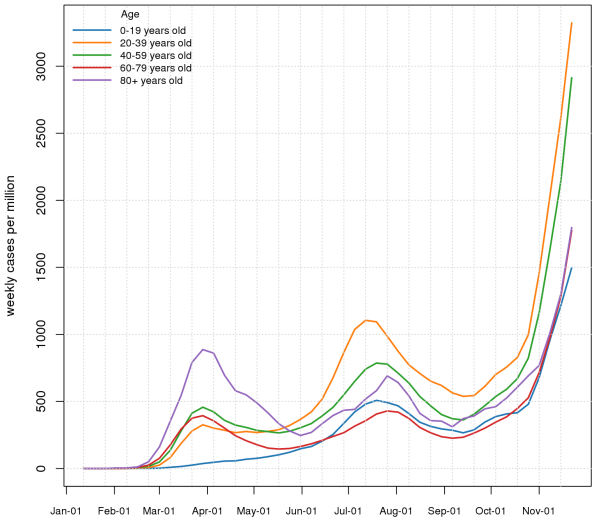
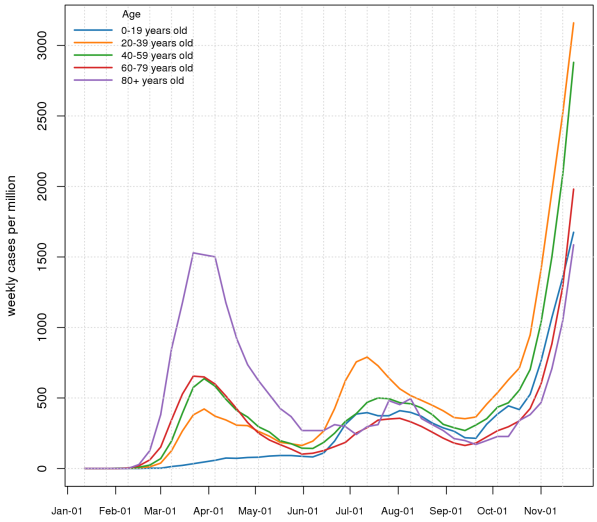
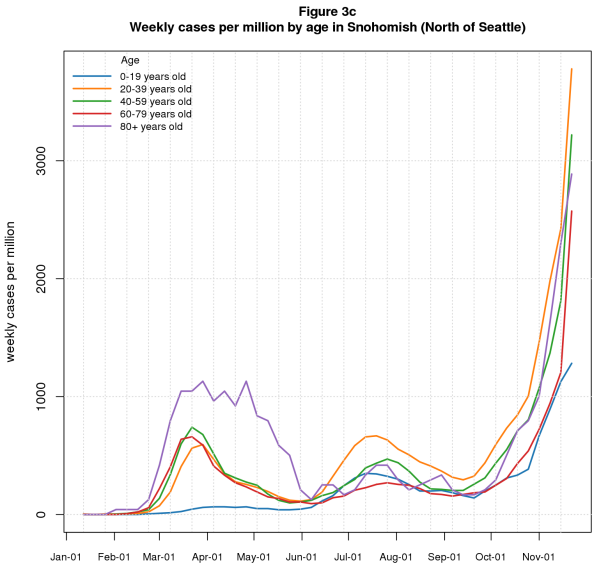
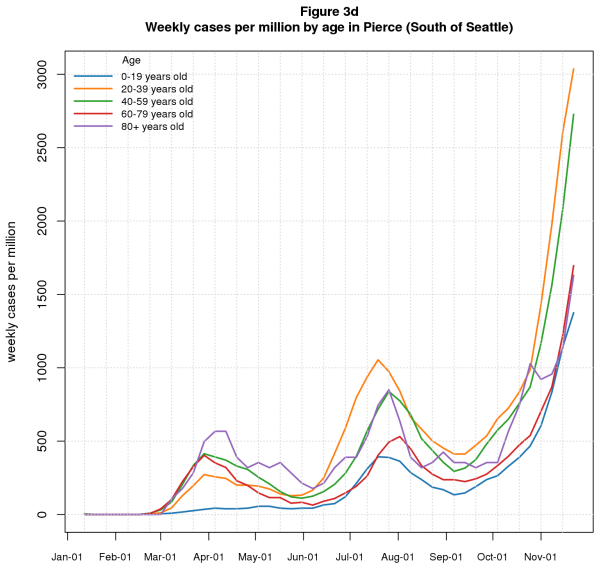


Figure 3b
Weekly cases per million by age in Seattle (King County)

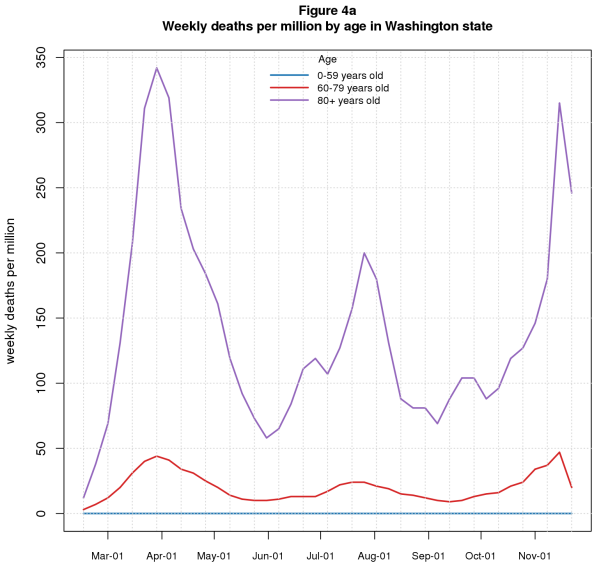


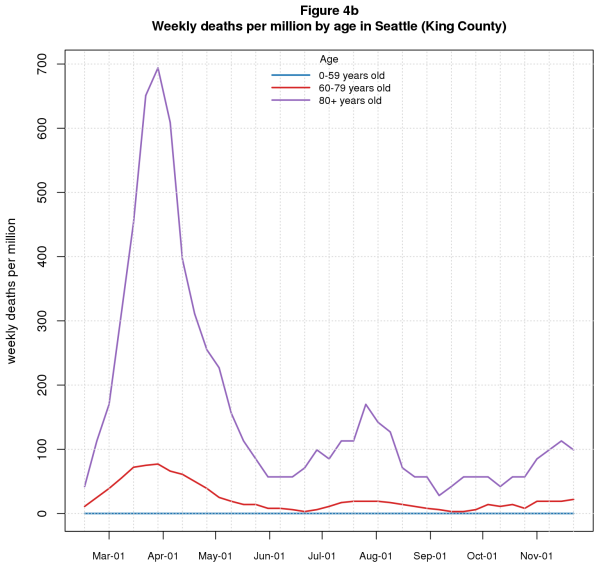


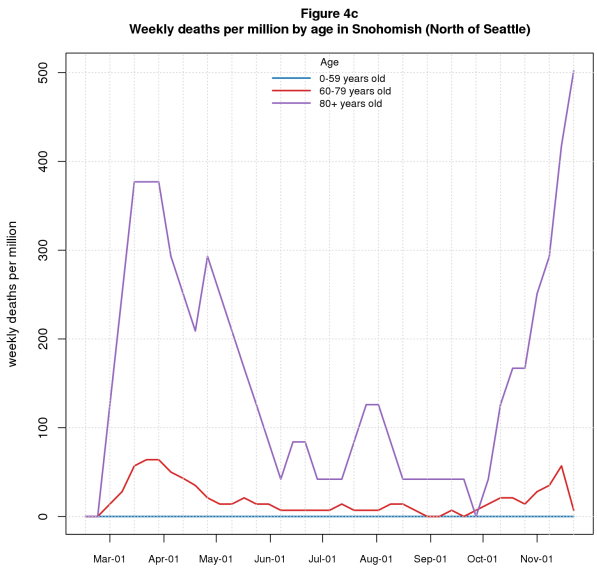


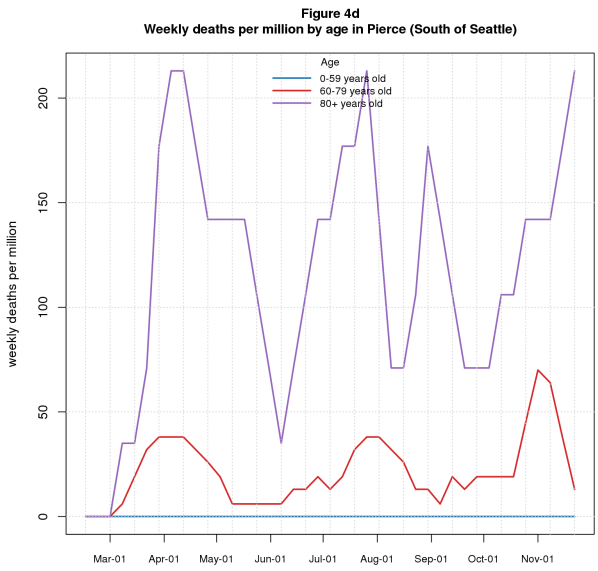
Early on, the pandemic struck older age groups most heavily. Over time, *cases* spread into all age groups, even the young. During the second wave, older groups did better in most locations, but now counts are going up rapidly in all age groups and locations.

Figures 4a-d are *deaths*. These graphs aggregate 0-59 into a single group, since the death rate in these ages is near 0.









The shocking devastation of the 80+ age group early in the pandemic jumps off the page. The early death rate for this group in Seattle (King County) hit almost 700 per million reflecting the early outbreak at a long term care facility in the area. Statewide, the death rate in the 80+ group shows three waves with deaths now almost as high as in the first wave. Deaths in Snohomish (north of Seattle) for the 80+ group have climbed to 500 per million, almost as high as the first wave in Seattle; this may reflect an outbreak in a long term care facility in the county. Deaths in Pierce (south of Seattle) for this group is more variable than elsewhere, possibly reflecting delays in reporting deaths to DOH.

Caveats

1. The term *case* means a person with a detected COVID infection. Washington state limits this to "confirmed cases", meaning people with positive molecular COVID tests. Some states include "probable cases", but the data source I use here only includes "confirmed cases" (or so I believe based on the name of the file I download).
2. Detected *cases* undercount actual cases by an unknown amount. As testing volume increases over time, it's reasonable to expect the detected count to get closer to the actual count. Some of the increase in *cases* we see in the data is due to this artifact. Modelers attempt to correct for this. I don't include any such corrections here (but I do extrapolate Washington DOH data as described below).
3. The same issues apply to *deaths* to a lesser extent, except perhaps early in the pandemic.
4. The geographic granularity in the underlyong 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 smooth the graphs using a 3-week rolling mean 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.
7. I extrapolate recent Washington State DOH data to compensate for systematic undercounting in the DOH weekly downloads. See below. The extrapolation probably overshoots by about 25%.

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). DOH updates the weekly data on Sundays.

The weekly downloads lag behind the daily feed causing data for the last few weeks to be incomplete. I correct for this undercount by extrapolating data from the preceeding six weeks using a linear model (R's `lm`). A quick looks suggests the extrapolation is presently overshooting by approximately 25% (below Figures 5a-b).

The weekly DOH download reports data by age group: 20-year ranges starting with 0-19, with a final group for 80+.

The DOH download includes data on hospital admissions (*admits*) in addition to *cases* and *deaths*, although I don't show this data here.

Issues with DOH Extrapolation

Figures 5a-b compare DOH and JHU *cases* and *deaths* for Washington state to illustrate the undercount in the raw DOH data and overcount in the extrapolated data. The discrepancy is most apparent in the *deaths* graph.

Figure 5a
Weekly cases per million: DOH and JHU (Washington state)

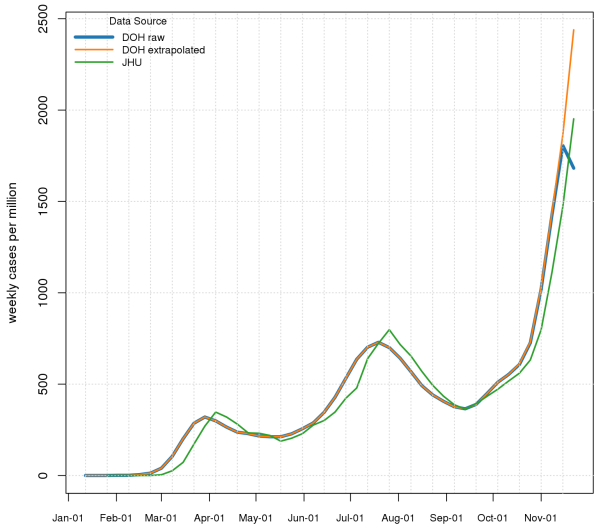
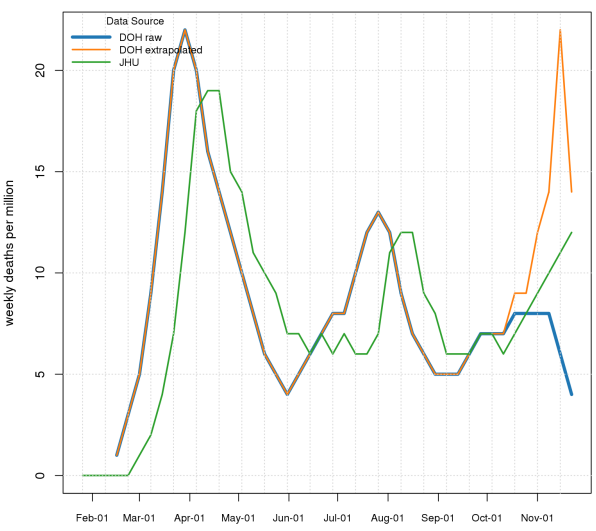


Figure 5b
Weekly deaths per million: DOH and JHU (Washington state)



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 usually download the data on Mondays to align with the DOH weekly data drops.

Other Data Sources

I use two other COVID data sources in my project (although not in this document).

- New York Times COVID Repository (NYT). The file I download is us-counties.csv. Like Washington DOH and JHU, NYT has county-level data. Unlike these, it includes "probable" as well as "confirmed" cases and deaths; I see no way to separate the two categories.
- COVID Tracking Project (TRK). This project reports a wide range of interesting statistics (negative test counts, for example), but I only use the *case* and *death* data. It does not provide county-level data so is not useful for the non-Washington locations I show. The file I download is https://covidtracking.com/data/download/washington-history.csv. I use this only as a check on the state-level Washington data from the other 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.