

# COVID-19 Counts in Washington State and Select Other Locations

Nathan (*Nat*) Goodman

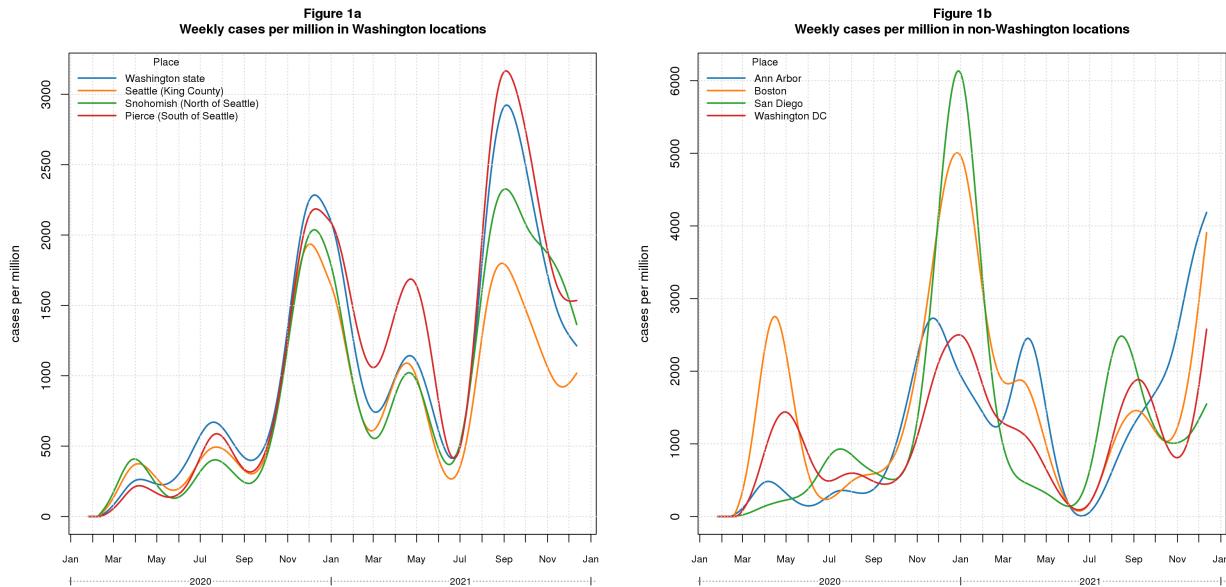
December 22, 2021

*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](#).*

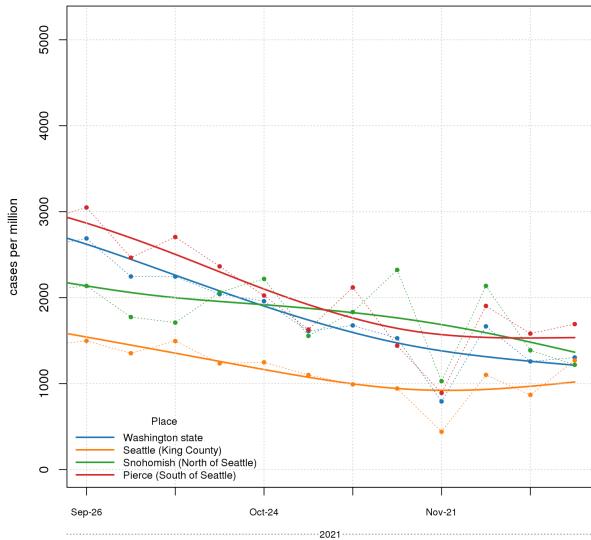
The first several figures (1a-2d) show *case* and *death* counts per million for several Washington and non-Washington locations using data from Johns Hopkins Center for Systems Science and Engineering (JHU), described [below](#). The Washington locations are the entire state, the Seattle area (King County) 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 for the latest 12 weeks to help explain recent trends.

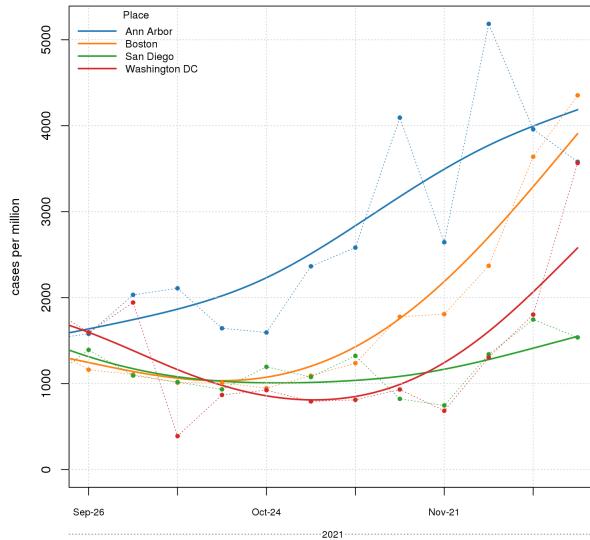
When comparing the smoothed Washington and non-Washington graphs (Figures 1a-b), please note the difference in y-scale; the graphs with raw data (Figures 1c-d) have the same y-scale and may be better for comparing Washington and non-Washington locations. The current raw counts for Washington locations are about 1200-1700 per million; non-Washington raw counts are about 1550-4350 per million.



**Figure 1c**  
Weekly cases per million in Washington locations showing recent raw data



**Figure 1d**  
Weekly cases per million in non-Washington locations showing recent raw data



The smoothed graphs for Washington locations (Figure 1a) show rates decreasing statewide and in Snohomish, flat in Pierce, and starting to head up in Seattle. Trend analysis suggests that the recent trend is flat in all locations, with a hint of increase in Seattle over the most recent 4 weeks. The raw data (Figure 1c) is too variable to be confident in the direction and does not yet support the start of the omicron wave widely predicted by national experts.

Rates in all Washington locations are well below their recent peaks and as well as their previous highs in winter 2020-21 (Figure 1a). Counts remain well above their peaks early in the pandemic (spring and summer 2020).

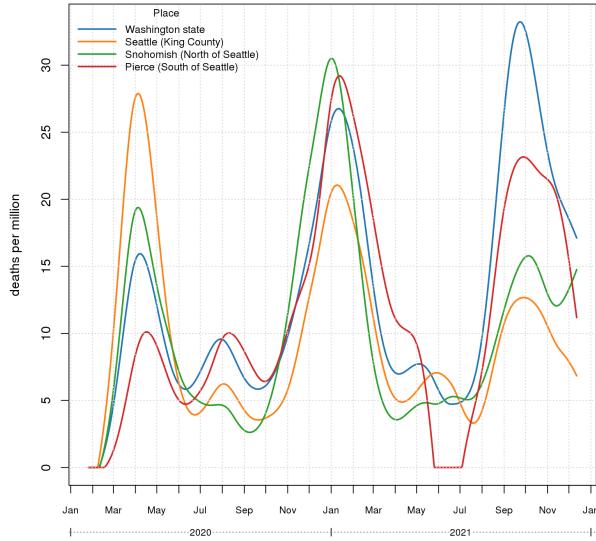
The smoothed graphs for non-Washington locations (Figure 1b) show rates increasing everywhere. The raw data (Figure 1d) and trend analysis are mixed. In Boston and Washington DC, the upward direction is definitive. In Ann Arbor, recent data is too variable to be confident in the direction, although looking back 8 or 10 weeks, the increase is clear. In San Diego, data for the most recent 4 weeks, though variable, seems to be heading up; looking back 6 or 8 weeks, the trend is flat.

Ann Arbor has soared above its previous peak in winter 2020-21; Washington DC has inched above that peak; Boston and San Diego remain well below their winter 2020-21 peaks.

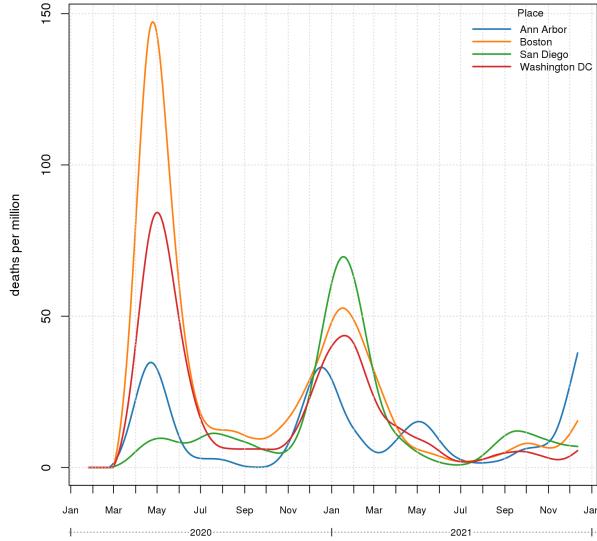
Figures 2a-d show *deaths* per million for the same locations. As with the *cases* graphs, the top row (Figures 2a-b) show smoothed data (see details [below](#)) and the bottom row (Figures 2c-d) overlays raw data onto the smoothed for the latest 12 weeks.

As with the cases data, the graphs with raw data (Figures 2c-d) are probably more useful when comparing the Washington and non-Washington graphs, as they have the same y-scale. The current Washington rates are 4-18 per million and non-Washington rates are 7-41 per million. Two weeks ago, I remarked that the then-high reading in Ann Arbor (38) seemed anomalous and might be an incorrect report; this week's number (41) belies that optimistic interpretation.

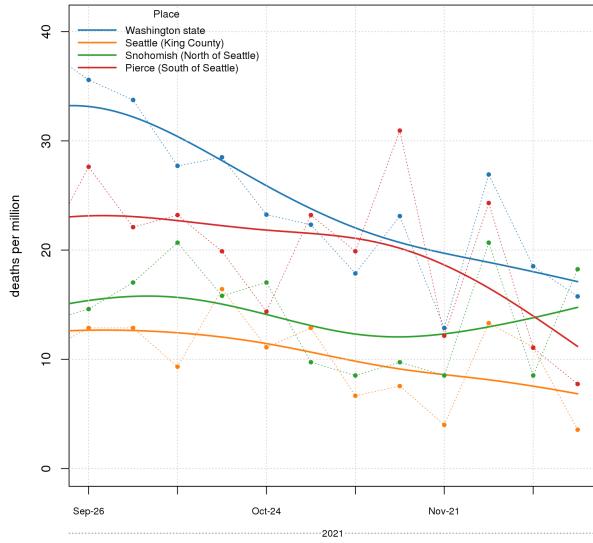
**Figure 2a**  
Weekly deaths per million in Washington locations



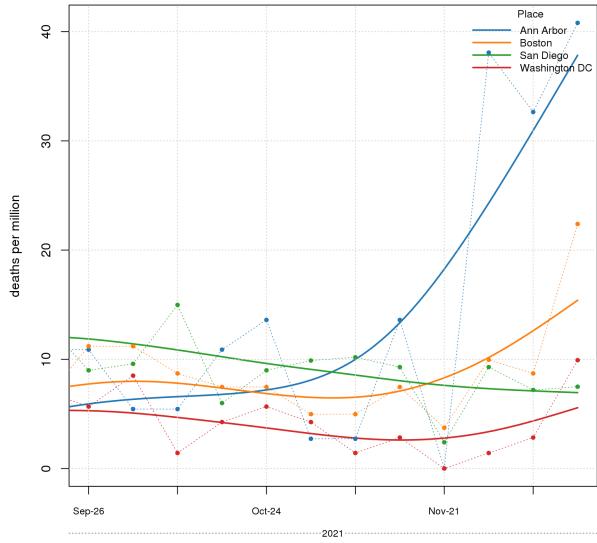
**Figure 2b**  
Weekly deaths per million in non-Washington locations



**Figure 2c**  
Weekly deaths per million in Washington locations showing recent raw data



**Figure 2d**  
Weekly deaths per million in non-Washington locations showing recent raw data



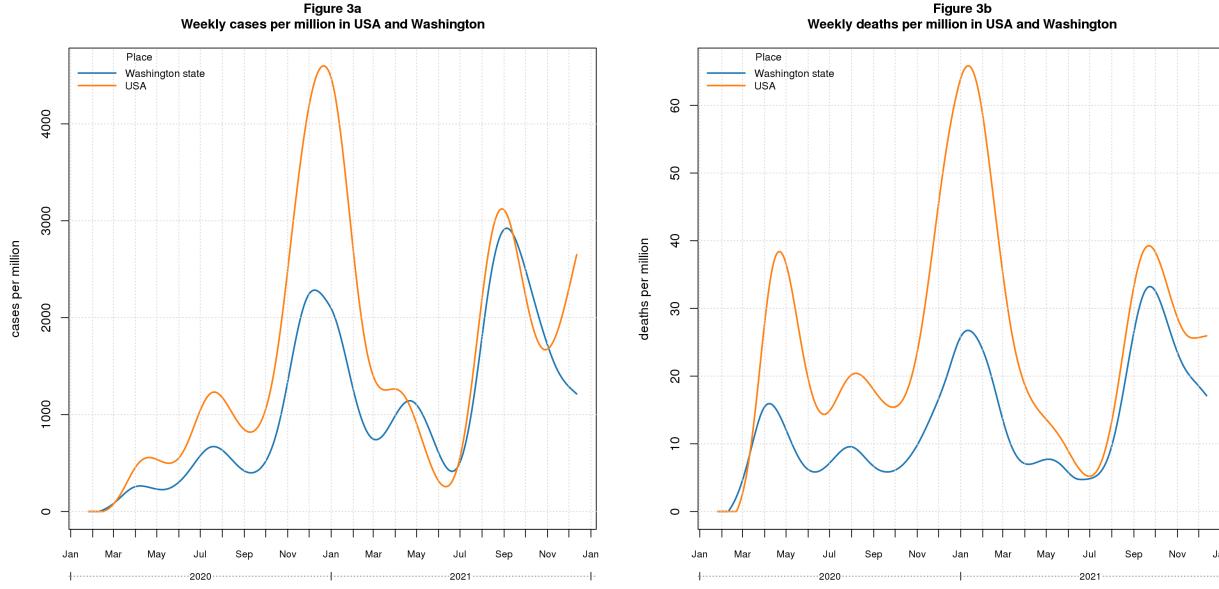
The smoothed graphs for Washington locations (Figure 2a) show rates decreasing everywhere except Snohomish. Trend analysis and raw data (Figure 2c) suggest the trend is flat everywhere. For the state, looking back 12 weeks, the raw data shows a pronounced decline. Rates are well below their earlier peaks in winter 2020-21 (Figure 2a). The picture is mixed compared to peaks early in the pandemic: Seattle and Snohomish are well below the first wave (spring 2020), while the state and Pierce are above. All locations are above the second wave (summer 2020).

(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 increasing in Ann Arbor, Boston, and Washington DC, but flat in San Diego. Trend analysis and raw data (Figure 2d) concur, with the caveat that recent data in Ann Arbor is too variable to be confident in the direction; looking back 6 or 8 weeks, the

increase in Ann Arbor is clear. Ann Arbor has climbed above all previous peaks. All other locations remain below their peaks from the start of the pandemic through spring 2021.

Figures 3a-b show counts for the USA as a whole, using Washington state for comparison. Figure 3a are *cases*; Figure 3b are *deaths*.

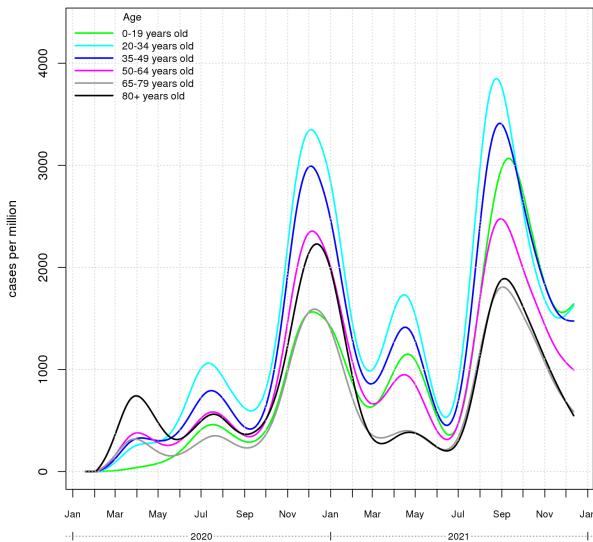


The two locations are traveling the same trajectory, indicating that the pandemic waves have played out similarly in the USA and Washington state. For most of the pandemic, USA *cases* and *deaths* are higher than Washington state. Overall, USA *cases* per capita are about 1.5x the Washington counts (156,000 vs. 106,000 per million) and *deaths* almost 2x (2000 vs. 1000 per million). *Cases* are starting to increase again in the USA (Figure 3a) but not yet Washington state.

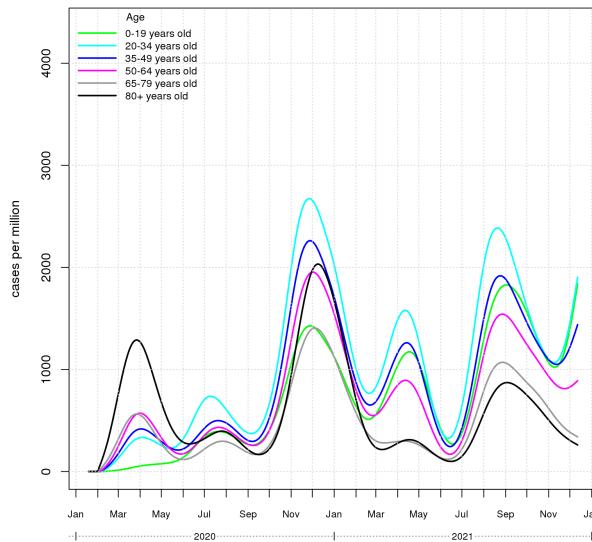
The next graphs show the Washington results broken down by age. This data is from Washington State Department of Health (DOH) weekly downloads, described [below](#). An important caveat is that the DOH download systematically undercounts events in recent weeks due to manual curation. I extrapolate data for late time points to correct the undercounting as discussed [below](#). The extrapolation worked well in past; presently, the extrapolation seems reasonable for *cases* and hospital admissions (*admits*) but is doing poorly for *deaths*.

Figures 4a-d are *cases*.

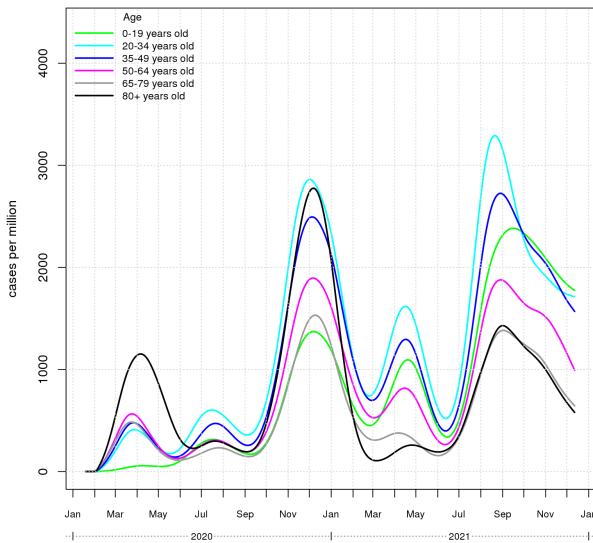
**Figure 4a**  
Weekly cases per million by age in Washington state



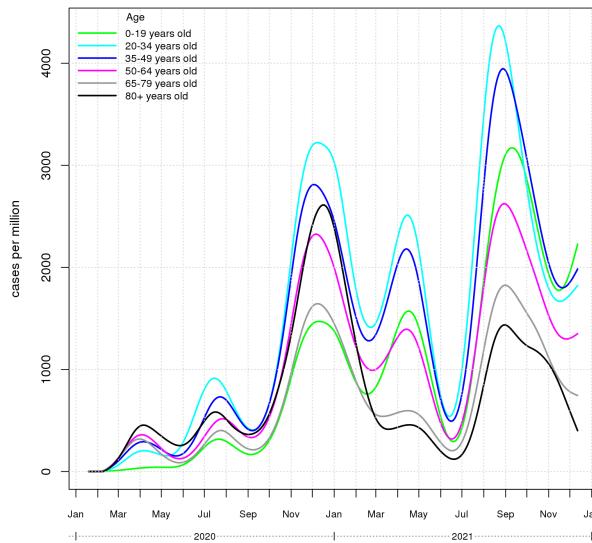
**Figure 4b**  
Weekly cases per million by age in Seattle (King County)



**Figure 4c**  
Weekly cases per million by age in Snohomish (North of Seattle)



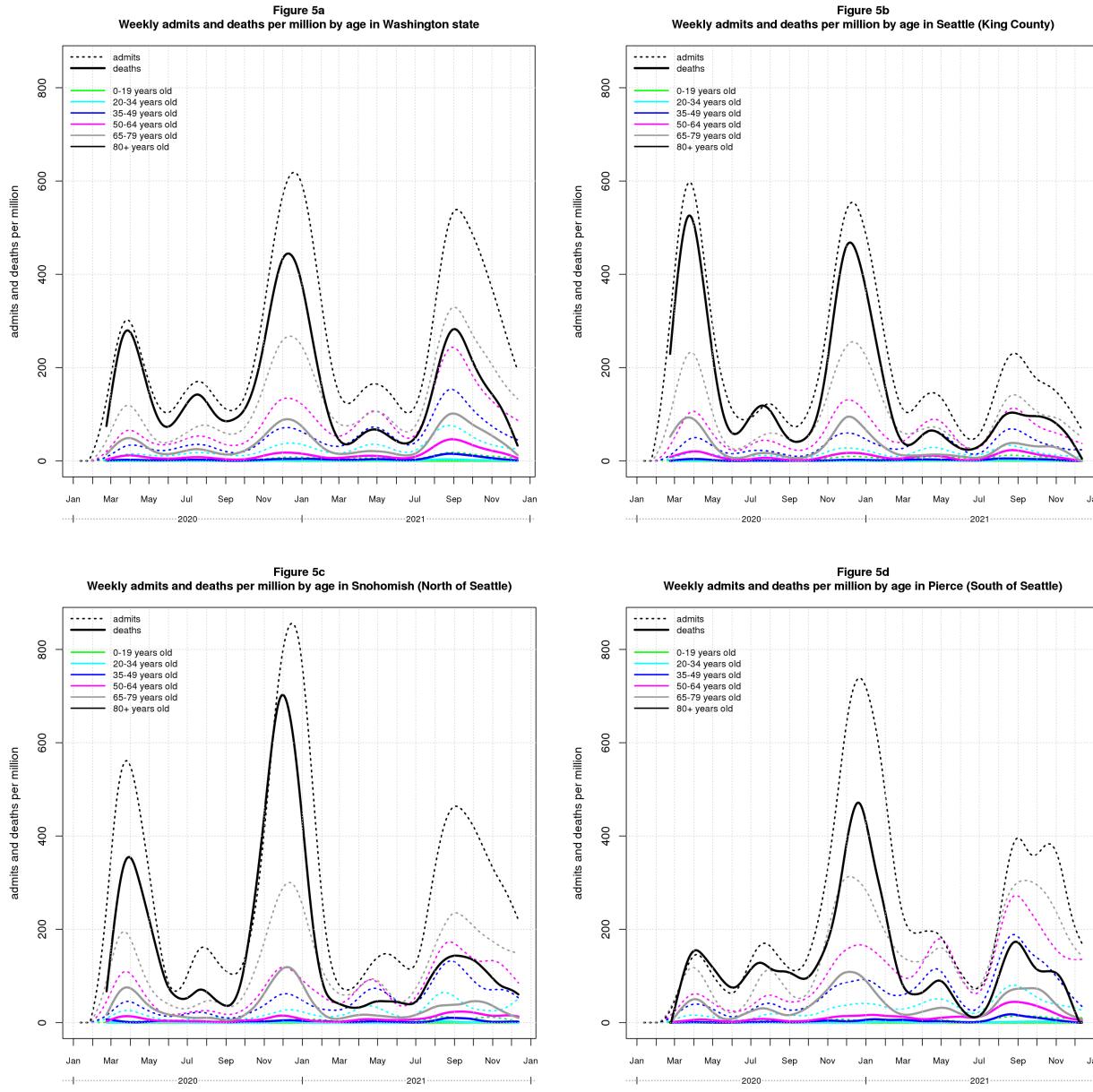
**Figure 4d**  
Weekly cases per million by age in Pierce (South of Seattle)



Early on, the pandemic struck older age groups most heavily, but *cases* quickly spread into all age groups, even the young. Rates greatly increased in late summer and early fall 2021, then declined in all age groups, and now starting back up in the youngest groups (0-19 and 20-34 years old). Statewide, the youngest group (0-19 years old) surpassed the 20-34 and 35-49 year olds in late September 2021 to take the dubious honor of being the highest group. The two youngest groups are now essentially tied while the next group (35-49 years) is slightly below. In all locations, the youngest three groups are at the top although the ranking varies. Pre-seniors (50-64 years) are below the younger groups, followed by the two oldest age groups (65-79 and 80+ years).

Counts for the youngest two groups (0-19 and 20-34 years) are higher in the individual locations than in the state. Counts for the next group (35-49) are the same in Seattle and the state, but higher in Snohomish and Pierce. The oldest three groups (50-64, 65-79, and 80+ years) are doing better in Seattle than statewide, and are mixed in Snohomish and Pierce.

Figures 5a-d show hospital admissions (*admits*) and *deaths*.



*Admits*, like *cases*, have gone up a lot in recent weeks but are now declining in all locations for most age groups. As with *cases*, *admits* are much higher statewide than in Seattle (Figure 5a vs. 5b). Snohomish and Pierce (Figures 5c-d) are worse than the state in some age groups and slightly better in others. *Admit* rates vary with age groups exactly as one would expect: lowest in the young and increasing with age.

Throughout the pandemic, the *death* rate for the 80+ group was much higher than any other group. *Deaths* even in this group dropped to near zero in early summer 2021, then increased above their spring 2021 levels in all locations during the fall 2021 wave, and are now falling in all locations. *Deaths* for the oldest groups are generally higher statewide (Figure 5a) than any of the individual locations (Figures 5b-d), reflecting the much higher death rate in the rural eastern part of the state (data not shown).

*Deaths* data near the end is especially unreliable due to DOH undercounting and my [extrapolation](#) correcting this. In the raw data (not shown), *deaths* for the 80+ group are 5 in Seattle and 0 in Pierce for the latest timepoint compared to 63 and 81 resp. four weeks ago. I believe this reflects delayed reporting of *deaths* to DOH.

## 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. In past, DOH only reported confirmed cases but as of August 29, 2021 they seem to be including probable ones, too. This doesn’t seem to have affected the totals much.
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.
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 recency 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.

## 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).

- 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 the March 14, 2021 data release (corresponding to document version March 17), they changed the groups to 0-19, followed by several 15 year ranges (20-34, 35-49, 50-64, 65-79), with a final group for 80+. They changed age groups again in the August 29, 2021 data release (corresponding to document version September 1) and again a week later in the September 5, 2021 release (document version September 8), but I chose to map the new age groups to the previous ones to avoid difficulties in aligning DOH age groups with population data. The new groups are 0-11, 12-19, 20-34, 35-49, 50-64, 65-79, 80+; I reconstitute 0-19 by summing 0-11 and 12-19.

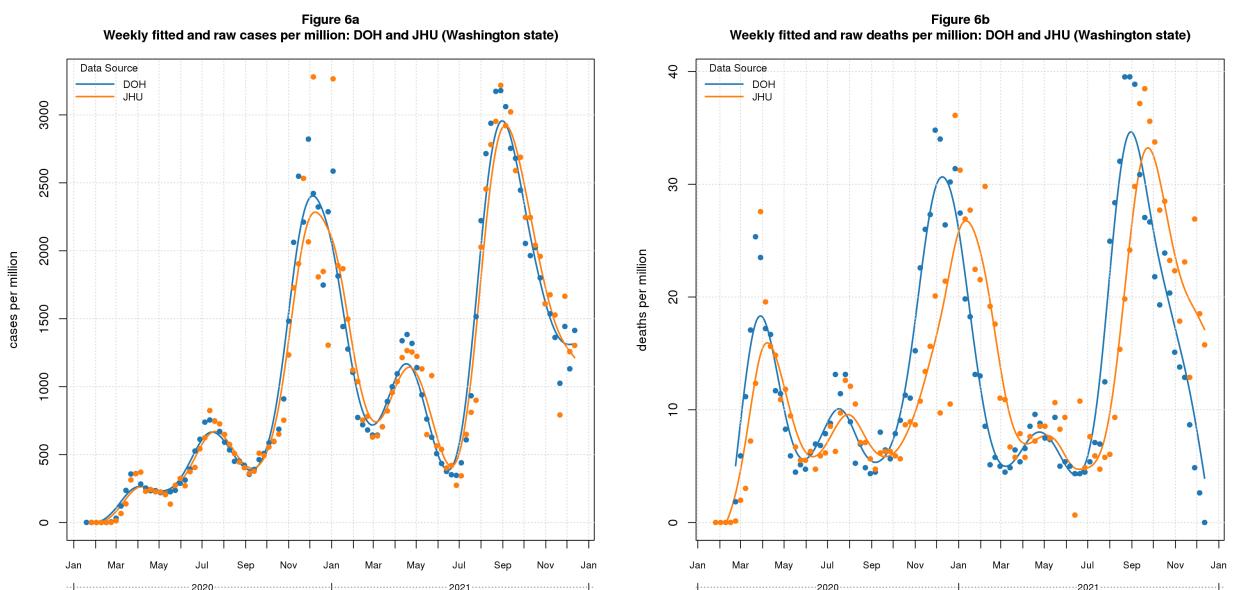
### Johns Hopkins Center for Systems Science and Engineering (JHU)

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.

### Issues with DOH Undercounting

Figures 6a-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 weeks in winter 2020 and the most recent few weeks. The *deaths* data matches less well and DOH is presently much lower than JHU. I believe the discrepancy in the *deaths* data reflects the consistent undercount of recent DOH data.



## 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 [LinkedIn](#) or [Twitter](#).

## 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](#).