

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

Nathan (*Nat*) Goodman

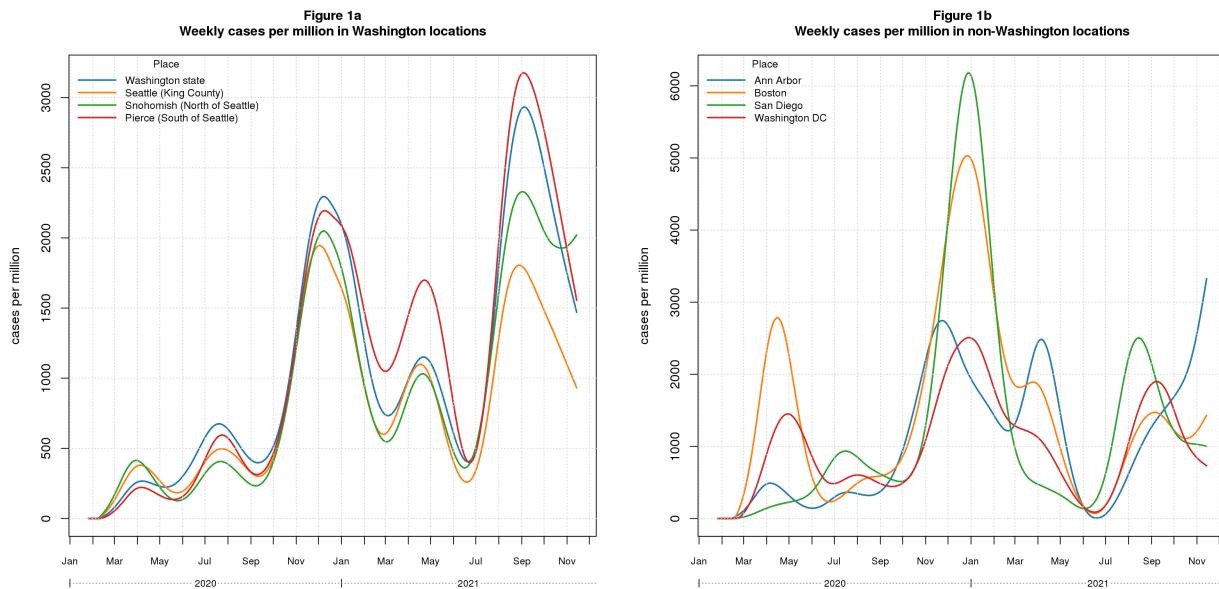
November 24, 2021

*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](#).*

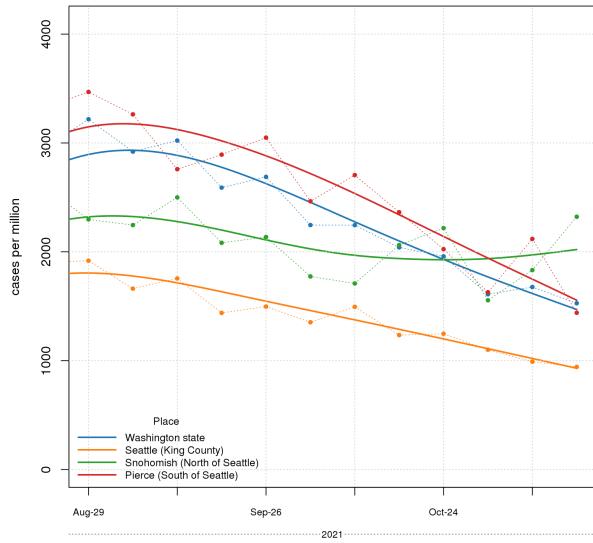
The first several figures (1a-2d) show *case* and *death* counts per million for several Washington and non-Washington locations using data from Johns Hopkins Center for Systems Science and Engineering (JHU), described [below](#). The Washington locations are the entire state, the Seattle area (King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

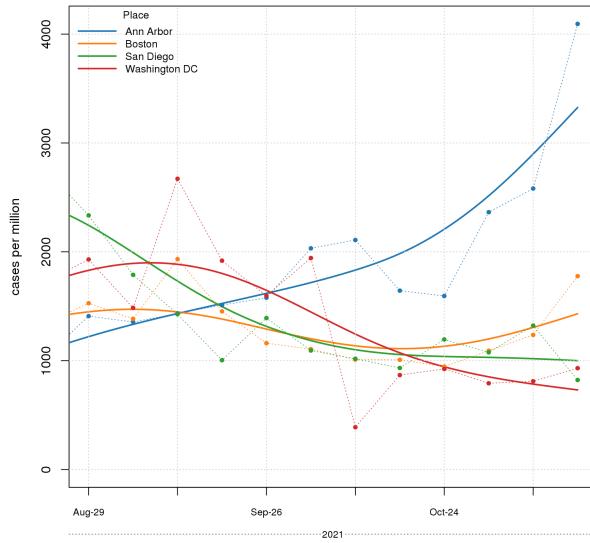
When comparing the smoothed Washington and non-Washington graphs (Figures 1a-b), please note the difference in y-scale; the graphs with raw data (Figures 1c-d) have the same y-scale and may be better for comparing Washington and non-Washington locations. The current raw counts for Washington locations are about 950-2350 per million; non-Washington raw counts are about 800-4100 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



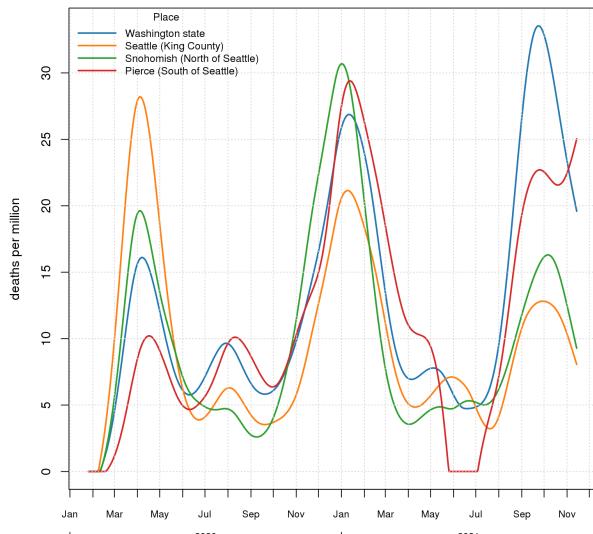
The smoothed graphs for Washington locations (Figure 1a) show rates decreasing everywhere except Snohomish. The raw data (Figure 1c) and trend analysis confirm the declining rates; the Snohomish data is quite variable but the trend looks flat. Rates in all Washington locations are well below their recent peaks and as well as their previous highs in winter 2020-21 (Figure 1a), although Snohomish is inching back up toward its winter 2020-21 high. Counts remain well above the peaks early in the pandemic (spring and summer 2020).

The smoothed graphs for non-Washington locations (Figure 1b) show rates increasing strongly in Ann Arbor, starting to head up in Boston, and decreasing in San Diego and Washington DC. The raw data (Figure 1d) and trend analysis concur with the increases in Ann Arbor and Boston. The data for San Diego and Washington DC is too variable to be confident in the direction; the real trend is probably flat. Ann Arbor has soared above its previous peak in winter 2020-21 (Figure 1b), but all other non-Washington locations remain well below their winter 2020-21 peaks.

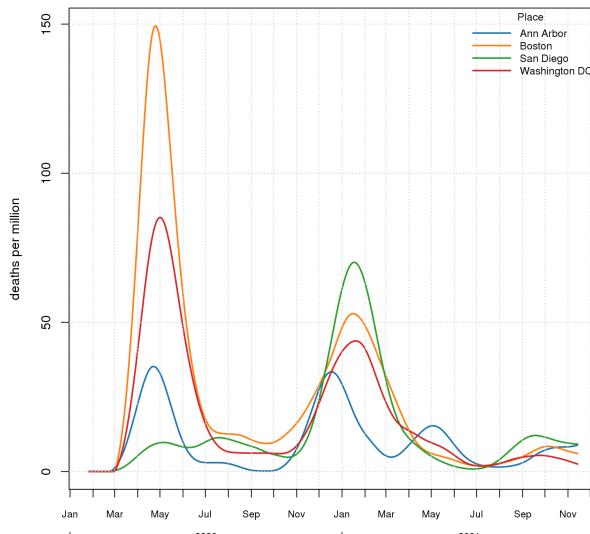
Figures 2a-d show *deaths* per million for the same locations. As with the *cases* graphs, the top row (Figures 2a-b) show smoothed data (see details [below](#)) and the bottom row (Figures 2c-d) overlays raw data onto the smoothed for the latest 12 weeks.

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 8-31 per million and non-Washington rates are 3-14 per million.

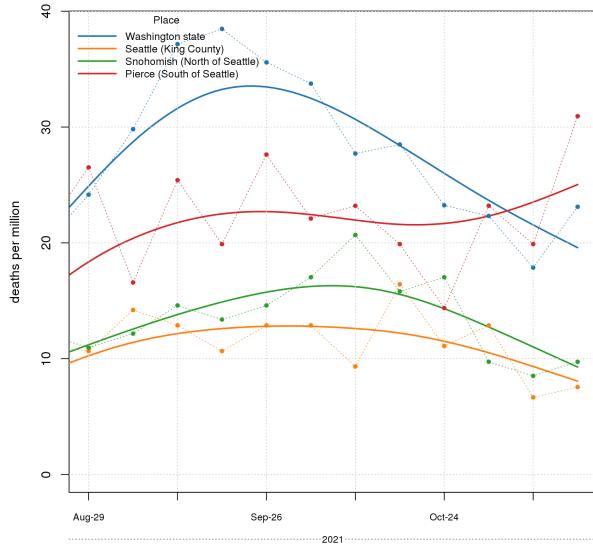
**Figure 2a**  
Weekly deaths per million in Washington locations



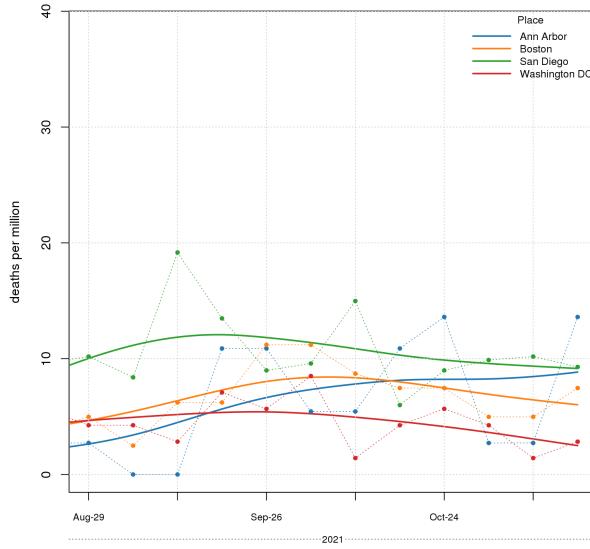
**Figure 2b**  
Weekly deaths per million in non-Washington locations



**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 Pierce. Trend analysis and raw data (Figure 2c) agree with the statewide decline. In King and Snohomish, the long term trend is down, but recent data is too variable to be confident in the direction. In Pierce, the recent trend is marginally up, but the long term trend is flat.

Rates are below their earlier peaks in winter 2020-21 (Figure 2a); the picture is mixed compared to the peaks early in the pandemic (spring and summer 2020): Seattle and Snohomish are well below, while the state and Pierce are above. 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.

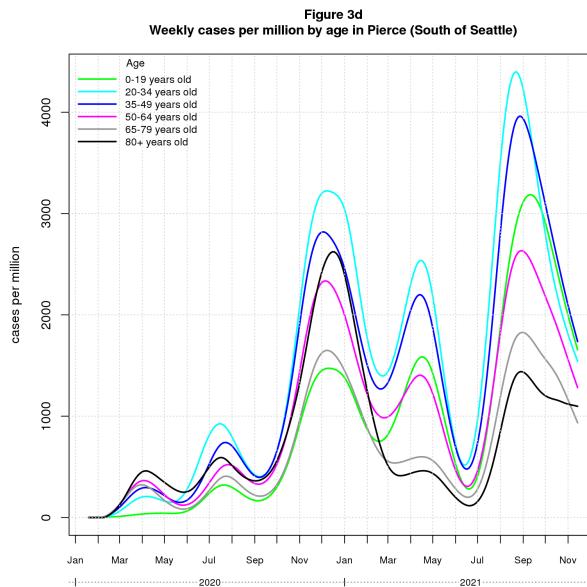
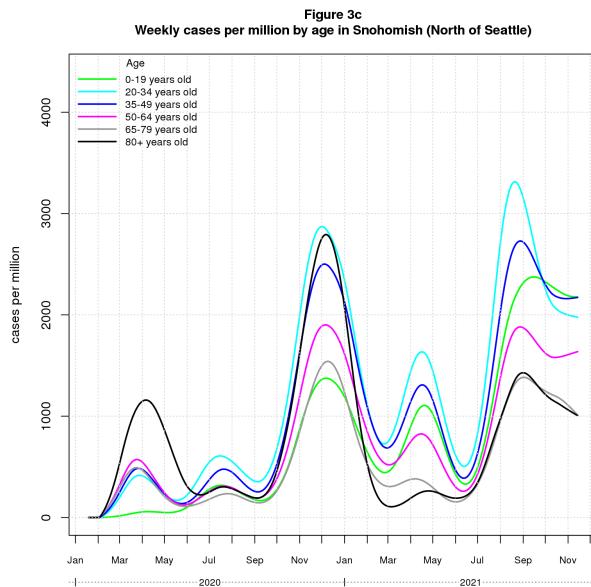
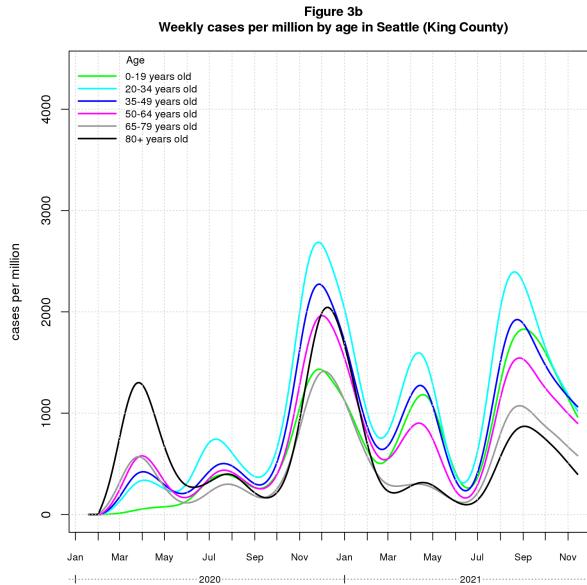
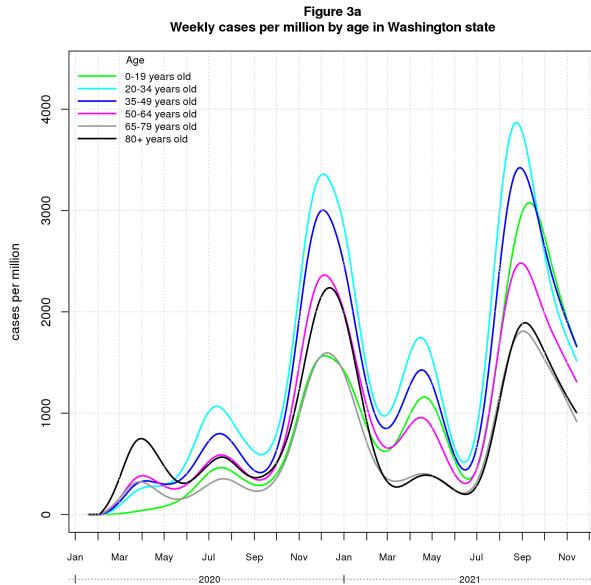
The smoothed graphs for non-Washington locations (Figure 2b) show rates decreasing or flat everywhere. Trend analysis and raw data (Figure 2d) concur.

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, but now seems to be erratic, making data near the end unreliable. I'm working to improve this.

In most versions between March 17 and September 1, 2021, I showed graphs for the entire state and not individual locations, since all locations were similar. In the current wave, geographic differences are substantial with Seattle doing much better than the rest of the state. As the wave declines, the differences are becoming smaller.

Figures 3a-d are *cases*.

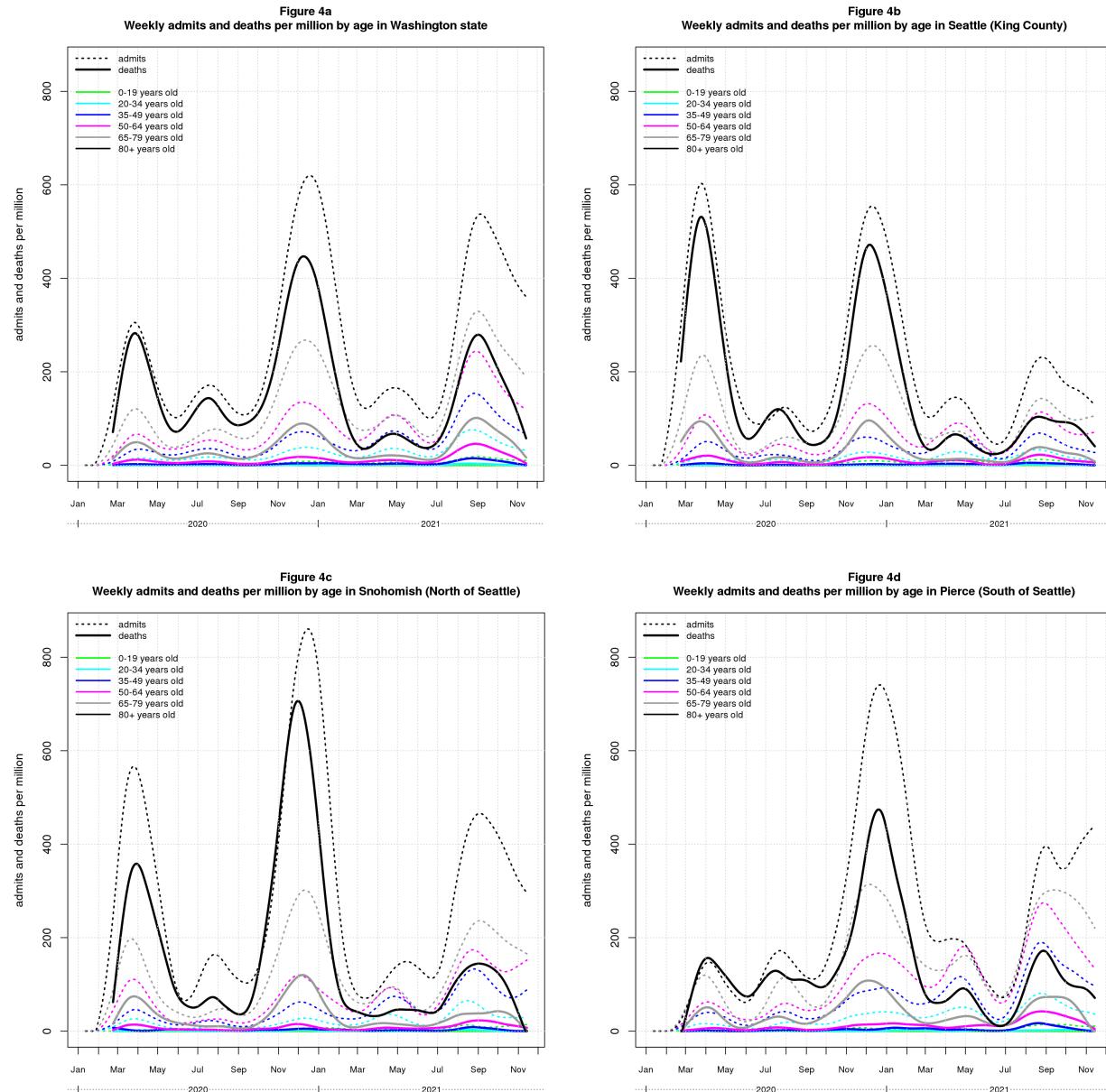


Early on, the pandemic struck older age groups most heavily, but *cases* quickly spread into all age groups, even the young. Rates greatly increased in late summer and early fall 2021 but are now declining in all age groups except for small bumps in Snohomish (may be artifacts of DOH undercounting and my [extrapolation](#)

correcting this). 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, but has dropped to a tie with the 35-49 year group. 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).

The ratio between worst and best in the statewide data (Figure 3a) is about 1.9x (about 1650 for 0-19 vs. 900 for 64-79); between youngest and oldest, the ratio is about 1.6x (about 1650 for 0-19 vs. 1000 for 80+). The general pattern is similar in Seattle (Figure 3b) but the counts are a lot lower: 1.7x lower for 0-19 year olds (1650 vs. 950) to 2.5x lower in the 80+ group (1000 vs. 400). The Snohomish rates (Figure 3c) are slightly worse than the state for most age groups; Pierce is about the same as the state (Figure 3d).

Figures 4a-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 except for small bumps in Seattle and Snohomish (may be artifacts of DOH undercounting and my [extrapolation](#) correcting this) and a pronounced increase in Pierce for the 80+ group which is probably real. As with *cases*, *admits* are much higher statewide than in Seattle (Figure 4a vs. 4b): more than 2x in all groups except 40-64 and 65-79 where it's 1.7x and 1.8x resp. Snohomish rates (Figures 4c) are about the same or a little worse than statewide. Pierce rates are worse than the state in all age groups. *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, but increased above their spring 2021 levels in all locations during the current wave, and are now falling in all locations. *Deaths* for the oldest groups are generally higher statewide (Figure 4a) than any of the individual locations (Figures 4b-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. It's implausible for *deaths* in the 80+ group to be higher in Seattle than Snohomish as the graphs suggest. In the raw data (not shown), *deaths* for the 80+ group are 0 in Snohomish and Pierce compared to 42-106 four weeks ago. I take this to indicate that reporting of *deaths* to DOH is lagging.

## 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.

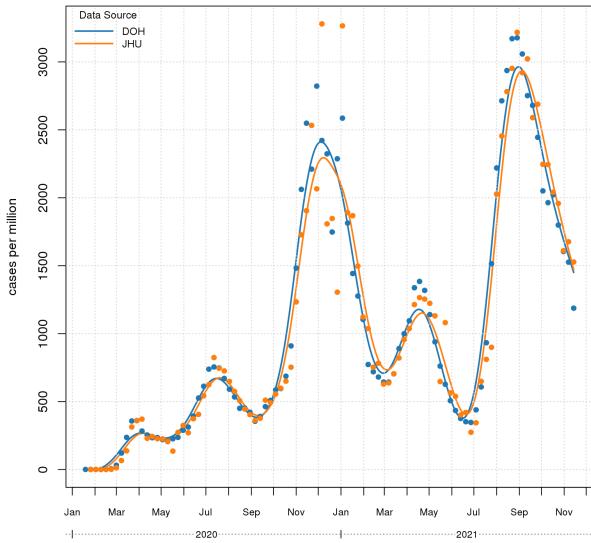
1. [Washington Disease Reporting System \(WDRS\)](#) provides daily “hot off the presses” results for use by public health officials, health care providers, and qualified researchers. It is not available to the general public, including yours truly.
2. [COVID-19 Data Dashboard](#) provides a web graphical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

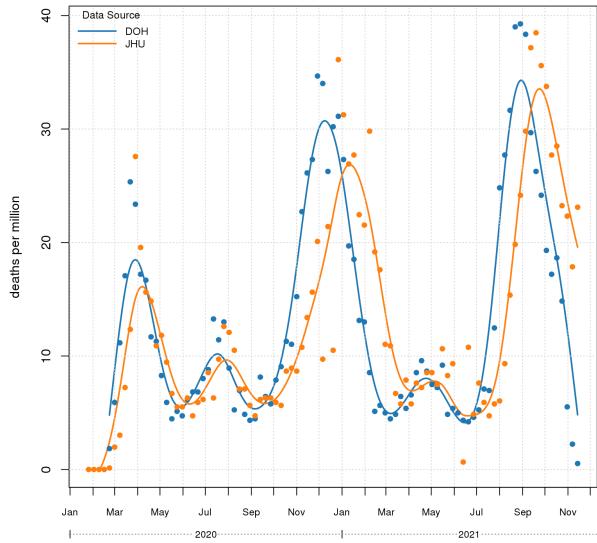
### Issues with DOH Undercounting

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

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



**Figure 5b**  
Weekly fitted and raw 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 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 [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](#).