

COVID-19 Counts in Washington State and Select Other Locations

Nathan (*Nat*) Goodman

December 29, 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 current raw counts for Washington locations are about 1650-2500 per million; non-Washington raw counts are about 2650-11850 per million. (In versions from August 18 through December 22, 2021, I plotted the raw-data figures with the same y-scale to facilitate comparison of Washington and non-Washington rates; the non-Washington counts are so high now that this obscures the shape of the Washington graph.)

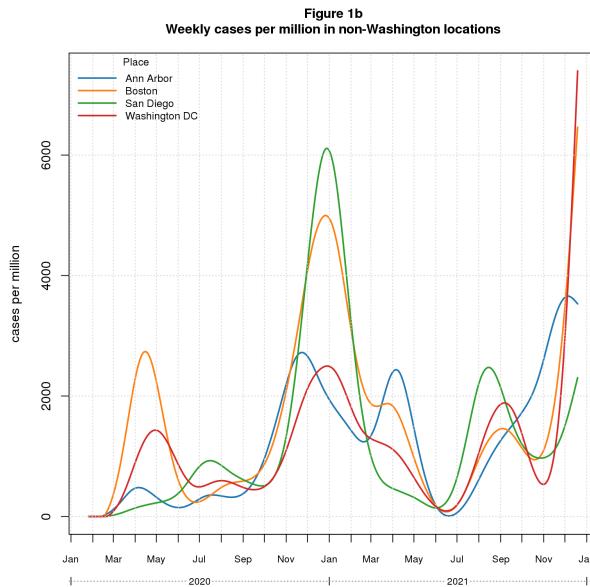
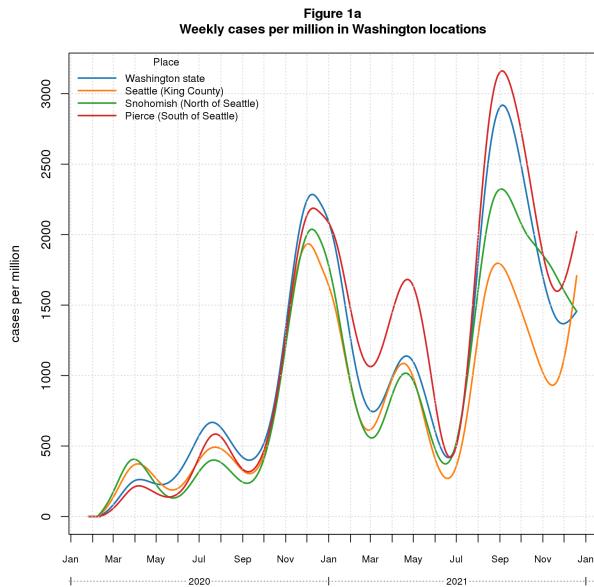


Figure 1c
Weekly cases per million in Washington locations showing recent raw data

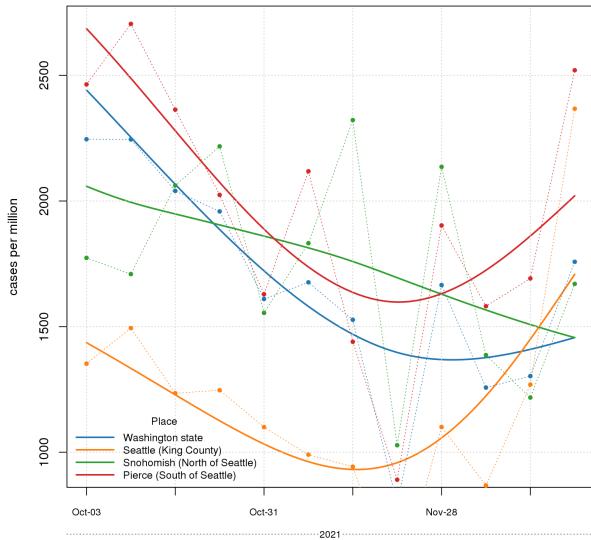
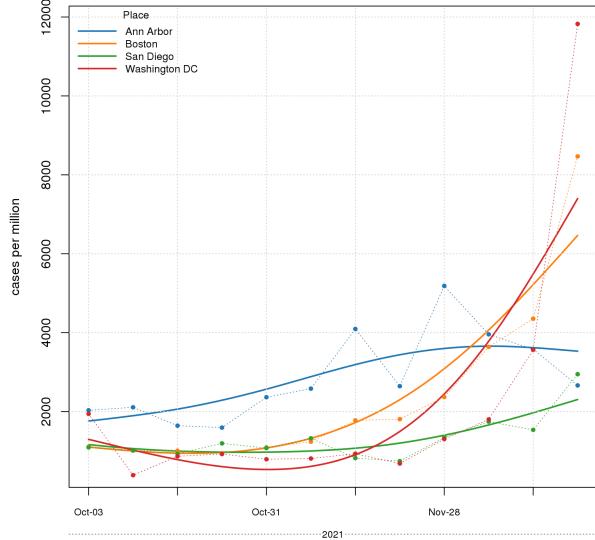


Figure 1d
Weekly cases per million in non-Washington locations showing recent raw data



The smoothed graphs for Washington locations (Figure 1a) show rates increasing everywhere except Snohomish. Trend analysis indicates that recent data is too variable to be confident in the direction, but the raw data (Figure 1c) supports the increasing rates seen in the smoothed graphs. The increases are not as great as reported in the news, perhaps because my data processing pipeline aggregates counts for the preceding week ending on Saturday, viz., the last data point is the sum of counts from Sunday December 19 through Saturday December 25; this washes out very recent changes.

Rates in all Washington locations except Seattle are below their recent peaks in fall 2021. All locations remain below their previous highs in winter 2020-21 (Figure 1a).

The smoothed graphs for non-Washington locations (Figure 1b) show rates increasing everywhere except Ann Arbor. The raw data (Figure 1d) and trend analysis concur. Ann Arbor remains above its previous peak in winter 2020-21. Boston and Washington DC have soared to new peaks. San Diego remains below its winter 2020-21 peak and is inching towards its summer 2021 peak.

Figures 2a-d show *deaths* per million for the same locations. As with the *cases* graphs, the top row (Figures 2a-b) shows smoothed data (see details [below](#)) and the bottom row (Figures 2c-d) overlays raw data onto the smoothed for the latest 12 weeks. Current Washington rates are 7-15 per million and non-Washington rates are 1-30 per million.

Figure 2a
Weekly deaths per million in Washington locations

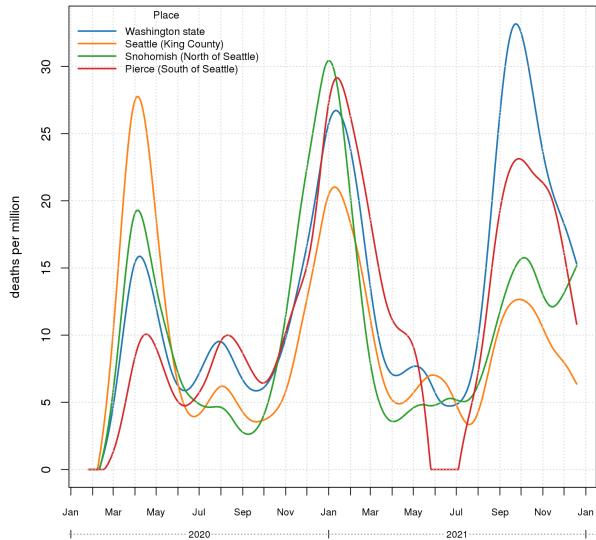


Figure 2b
Weekly deaths per million in non-Washington locations

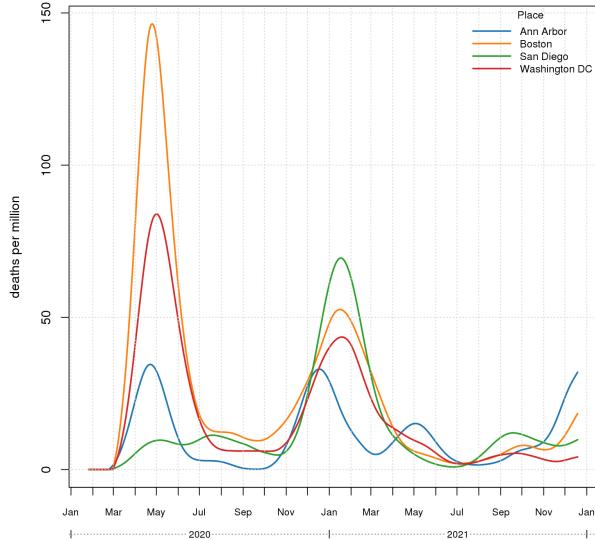


Figure 2c
Weekly deaths per million in Washington locations showing recent raw data

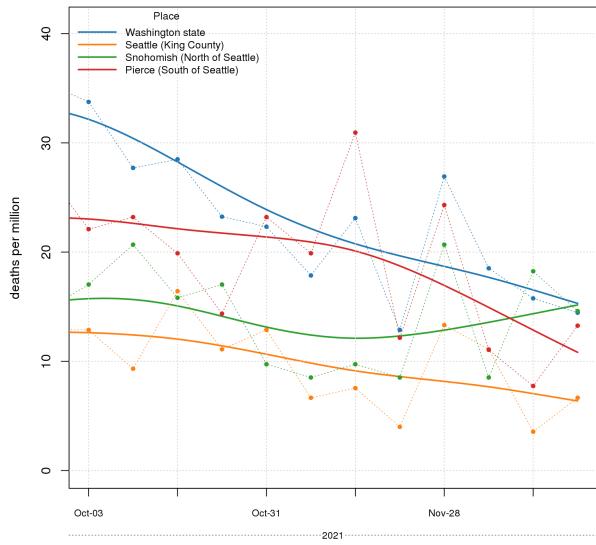
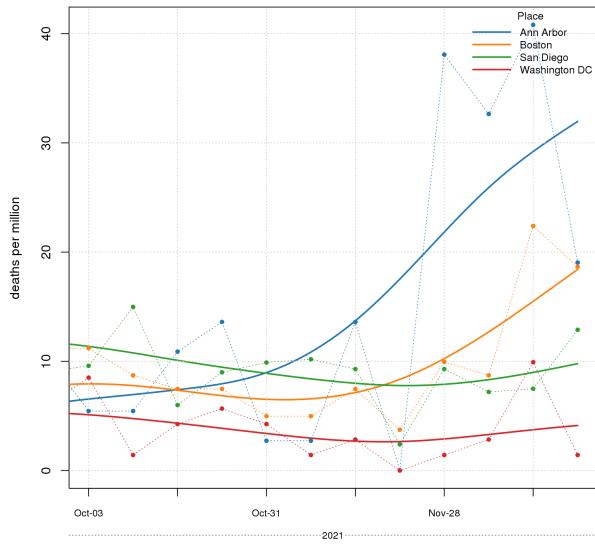


Figure 2d
Weekly deaths per million in non-Washington locations showing recent raw data



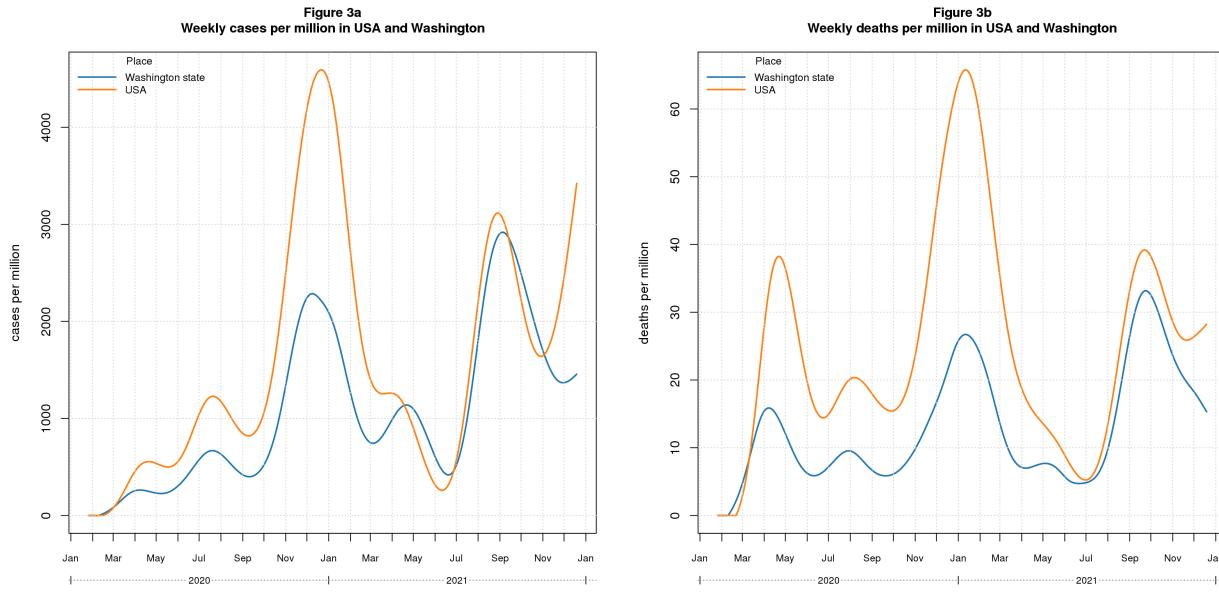
The smoothed graphs for Washington locations (Figure 2a) show rates decreasing everywhere except Snohomish. Trend analysis and raw data (Figure 2c) are mixed. For the state and Seattle, the decline is clear looking back 10 or 12 weeks. For Snohomish, the short and long term trend is flat. For Pierce, the data is quite variable but is generally down when looking 6-12 weeks back. Rates are well below their earlier peaks in winter 2020-21 (Figure 2a) and below or near the first wave (spring 2020). Seattle has dropped to the same level as the second wave (summer 2020) while all other locations 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 increasing everywhere. Trend analysis and raw data (Figure 2d) are mixed. In Ann Arbor, recent data is bouncing around, though looking back 8-12 weeks, the direction is clearly up. In Boston, looking back 6-12 weeks the direction is strongly up. In

San Diego and Washington DC, the trend seems flat. All other locations except Ann Arbor remain below their peaks from the start of the pandemic through spring 2021. Ann Arbor is below its peaks early in the pandemic but above its peaks in spring and summer 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 (159,000 vs. 107,000 per million) and *deaths* about 1.9x (2,500 vs. 1,300 per million). *Cases* are increasing sharply in the USA and starting up Washington state (Figure 3a). *Deaths* are climbing in the USA but continue to decline in 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. The undercount for *cases* and *admits* is 10-20% for most places and ages, while for *deaths* it's larger and variable.

Figures 4a-d are *cases*.

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

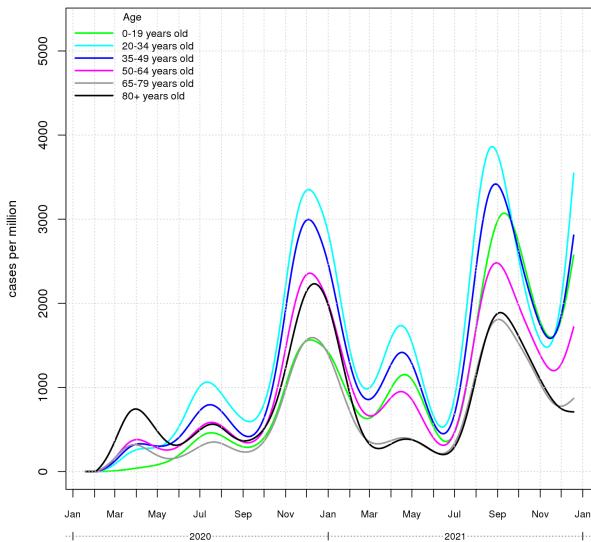


Figure 4b
Weekly cases per million by age in Seattle (King County)

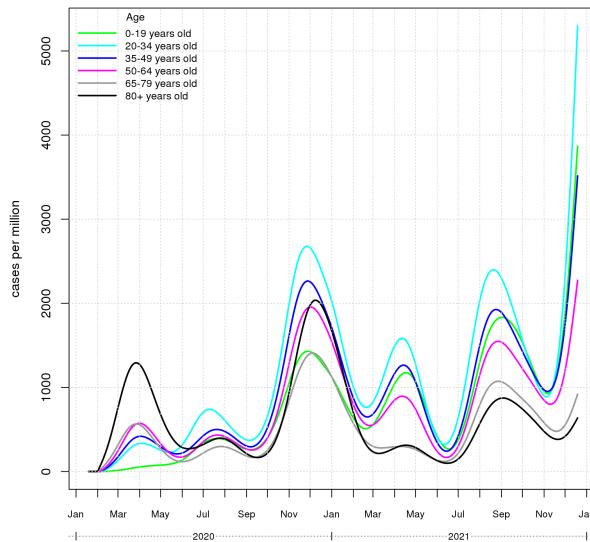


Figure 4c
Weekly cases per million by age in Snohomish (North of Seattle)

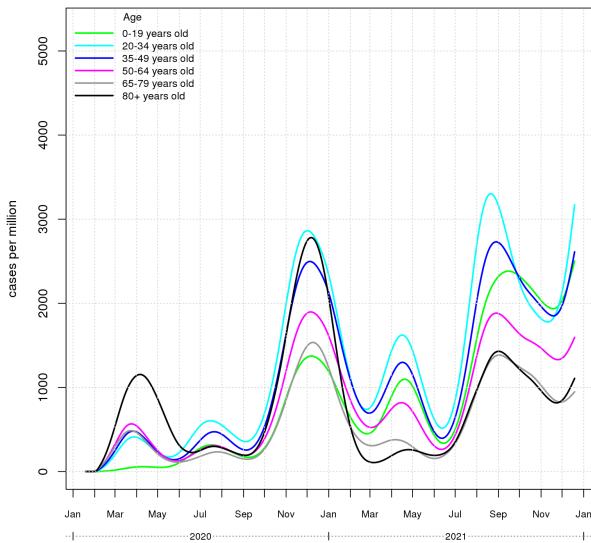
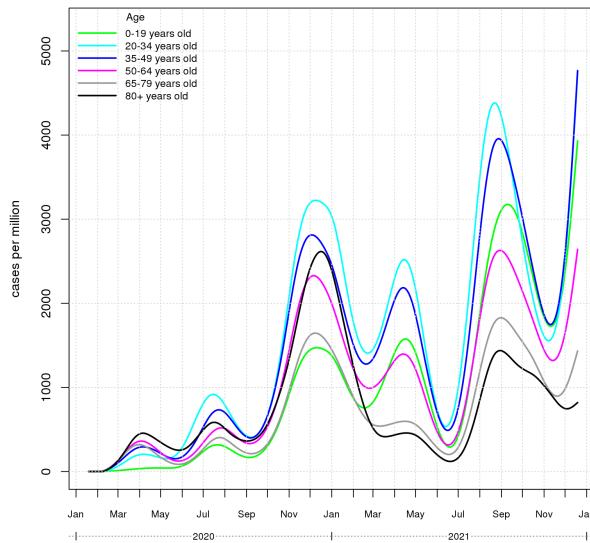


Figure 4d
Weekly cases per million by age in Pierce (South of Seattle)



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 climbing again in all but the oldest group (80+ 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. In the past four weeks, young adults (20-34 years) have regained the top spot. 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).

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

Figure 5a
Weekly admits and deaths per million by age in Washington state

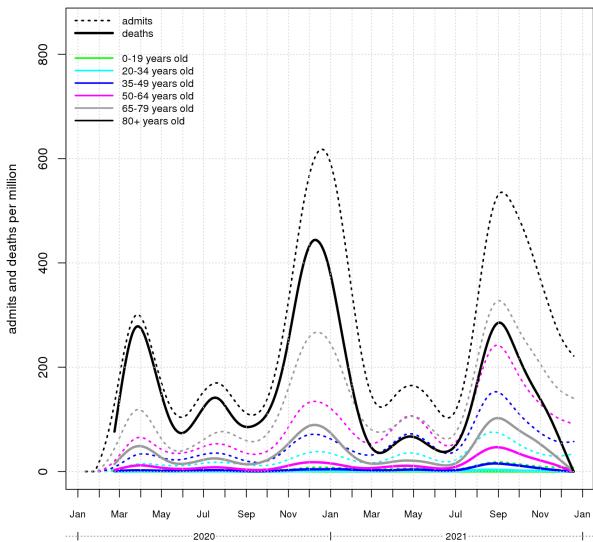


Figure 5b
Weekly admits and deaths per million by age in Seattle (King County)

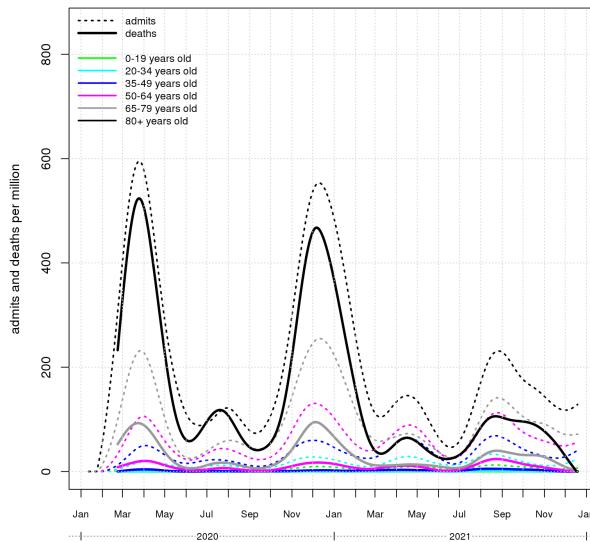


Figure 5c
Weekly admits and deaths per million by age in Snohomish (North of Seattle)

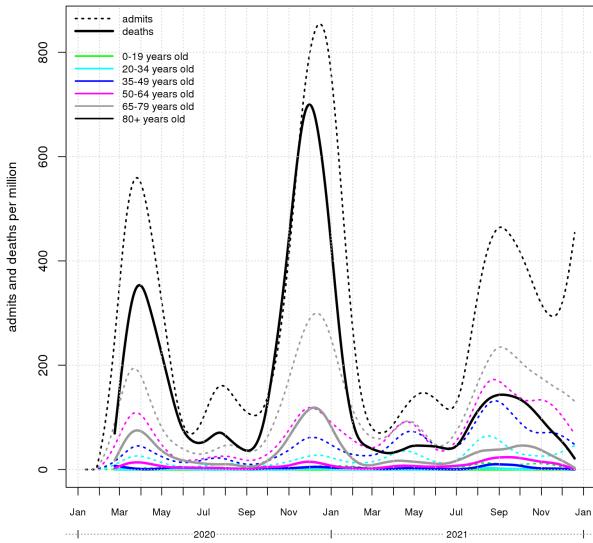
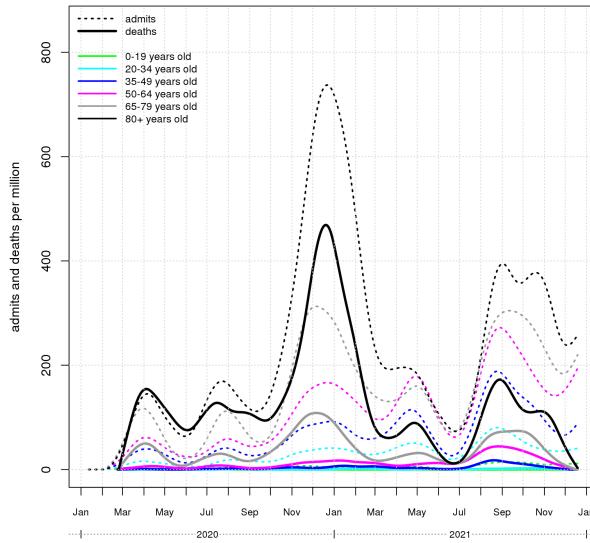


Figure 5d
Weekly admits and deaths per million by age in Pierce (South of Seattle)



Admits, like *cases*, went up a lot in early fall 2021, then declined, and are now flat or increasing in all locations and age groups. *Admits* are much higher statewide than in Seattle (Figure 5a vs. 5b). *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. In the raw data (not shown), *deaths* for the 80+ group are 0 in Seattle and Pierce for the latest timepoint compared to 71 and 84 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. In previous versions, I attempt to correct for this through a linear extrapolation function (using R’s `lm`). I’ve tweaked the extrapolation repeatedly, and have now decided to turn it off.

Analysis of recent data indicates that for *cases*, the undercount is 10-20% throughout; for *admits*, the undercount is more variable, esp. for places and ages with low numbers; for *deaths*, the undercount is much greater, approaching 100% in some cases. I found that my extrapolation scheme didn’t reduce the error much but made interpretation more difficult, because the extrapolation would undershoot or overshoot unpredictably. The analysis compared the raw data to extrapolation that created models for each Washington location and age group (as used in the distant past) and to extrapolation that creates a single model for the state as a whole and all age groups summed together then blithely applies that model to all locations and ages (as used in the recent past).

3. The trend analysis computes a linear regression (using R’s `lm`) over the most recent 4-12 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.

Johns Hopkins Center for Systems Science and Engineering (JHU)

JHU CSSE operates 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: files [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.

New York Times COVID Repository (NYT)

The [New York Times COVID Repository](#) provides data similar to JHU. The relevant file is [us-counties.csv](#). I used both JHU and NYT early in the project and found they had nearly identical data.

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.

Figure 6a
Weekly fitted and raw cases per million: DOH and JHU (Washington state)

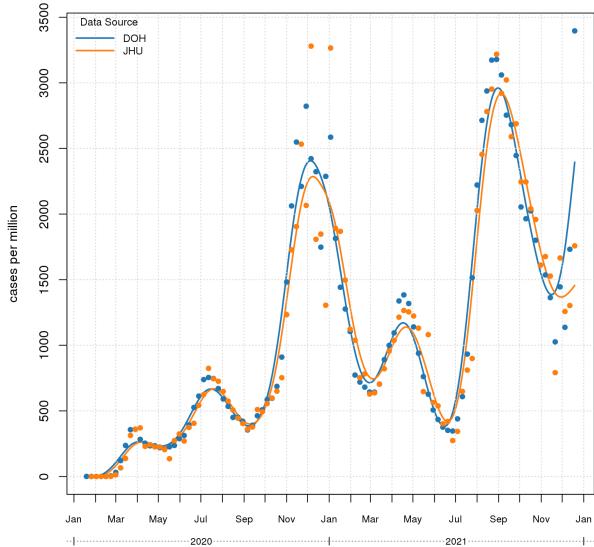
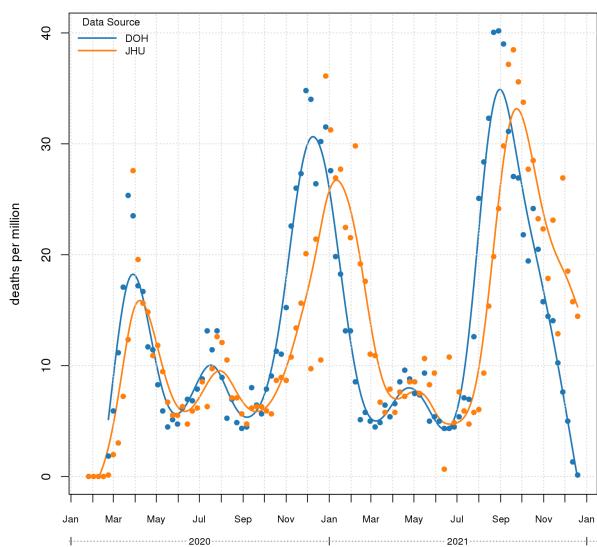


Figure 6b
Weekly fitted and raw deaths per million: DOH and JHU (Washington state)



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](#).