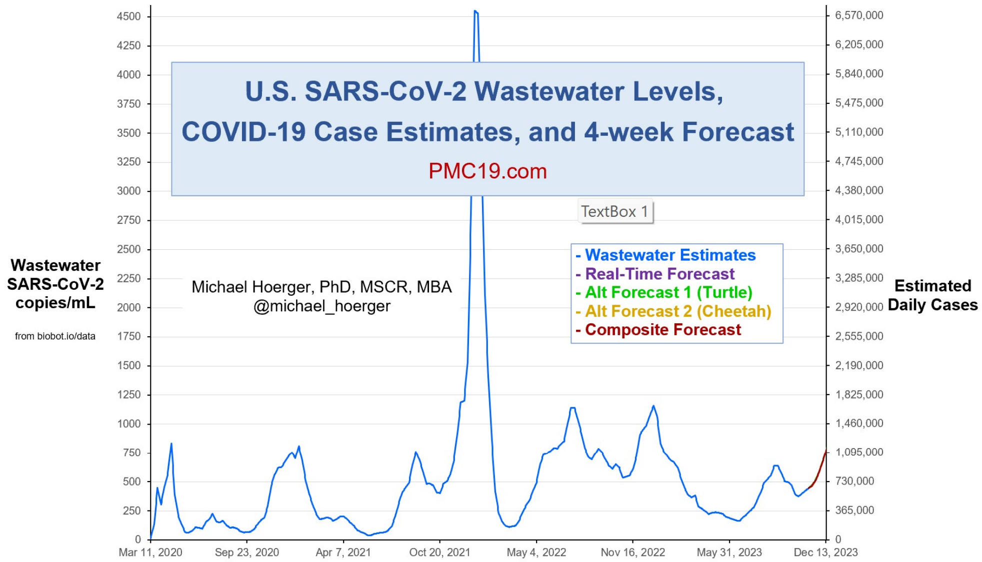
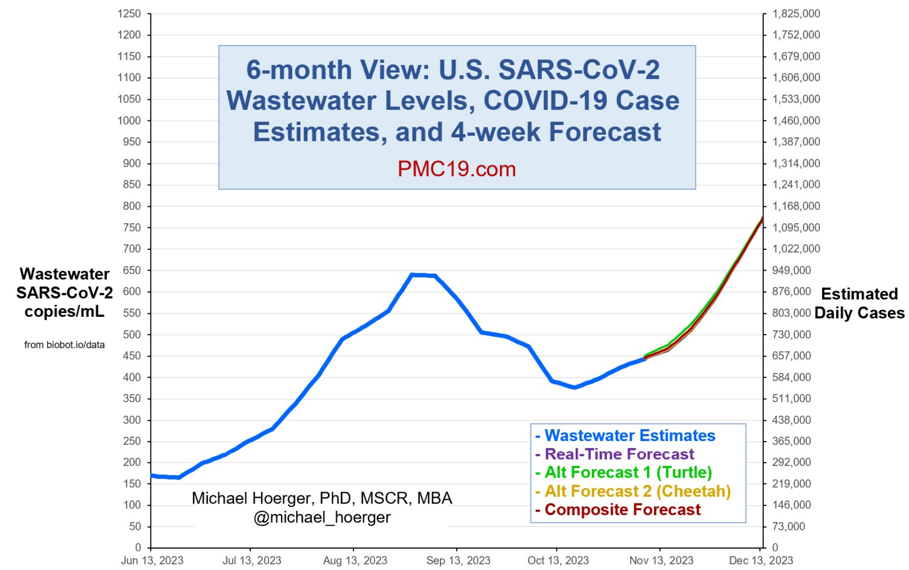
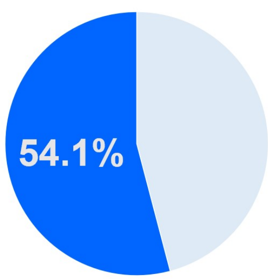
Michael Hoerger, PhD, MSCR, MBA, Pandemic Mitigation Collaborative

U.S. SARS-CoV-2 Wastewater Levels, COVID-19 Case Estimates, and 4-Week Forecast: Report for November 13, 2023, pmc19.com/data







There is more COVID-19 transmission today than during 54.1% of the pandemic.

CURRENT ESTIMATES FOR November 13, 2023
Wastewater Levels (copies/mL)
462
New Daily Cases
672,000
% of Population Infectious
1.41% (1 in 71 people)
New Daily Long COVID Cases
34,000 to 134,000

WEEKLY ESTIMATES FOR November 13, 2023
New Weekly Cases
4,700,000
New Weekly Long COVID Cases
235,000 to 941,000

2023 CUMULATIVE ESTIMATES AS OF November 13, 2023
Total 2023 Cases To Date
194,286,130
Total 2023 Long COVID Cases To Date
9,714,000 to 38,857,000

4-WEEK FORECAST FOR December 11, 2023
Wastewater Levels (copies/mL)
739 (60% higher)
New Daily Cases
1,076,000
% of Population Infectious
2.25% (1 in 44 people)
New Daily Long COVID Cases
54,000 to 215,000

What's the COVID Risk in an Office or in a Classroom?

Number of People	Chances Anyone is Infectious	Number of People	Chances Anyone is Infectious
1	1.4%	25	29.8%
2	2.8%	30	34.6%
3	4.2%	35	39.1%
4	5.5%	40	43.2%
5	6.8%	50	50.7%
6	8.1%	75	65.4%
7	9.4%	100	75.7%
8	10.7%	150	88.0%
9	12.0%	200	94.1%
10	13.2%	300	98.6%
15	19.1%	400	99.7%
20	24.7%	500	>99.9%

Cite as: Hoerger, M. (2023, November 13). *U.S. SARS-CoV-2 wastewater levels, COVID-19 case estimates, and 4-week forecast: Report for November 13, 2023*. Pandemic Mitigation Collaborative. http://www.pmc19.com/data

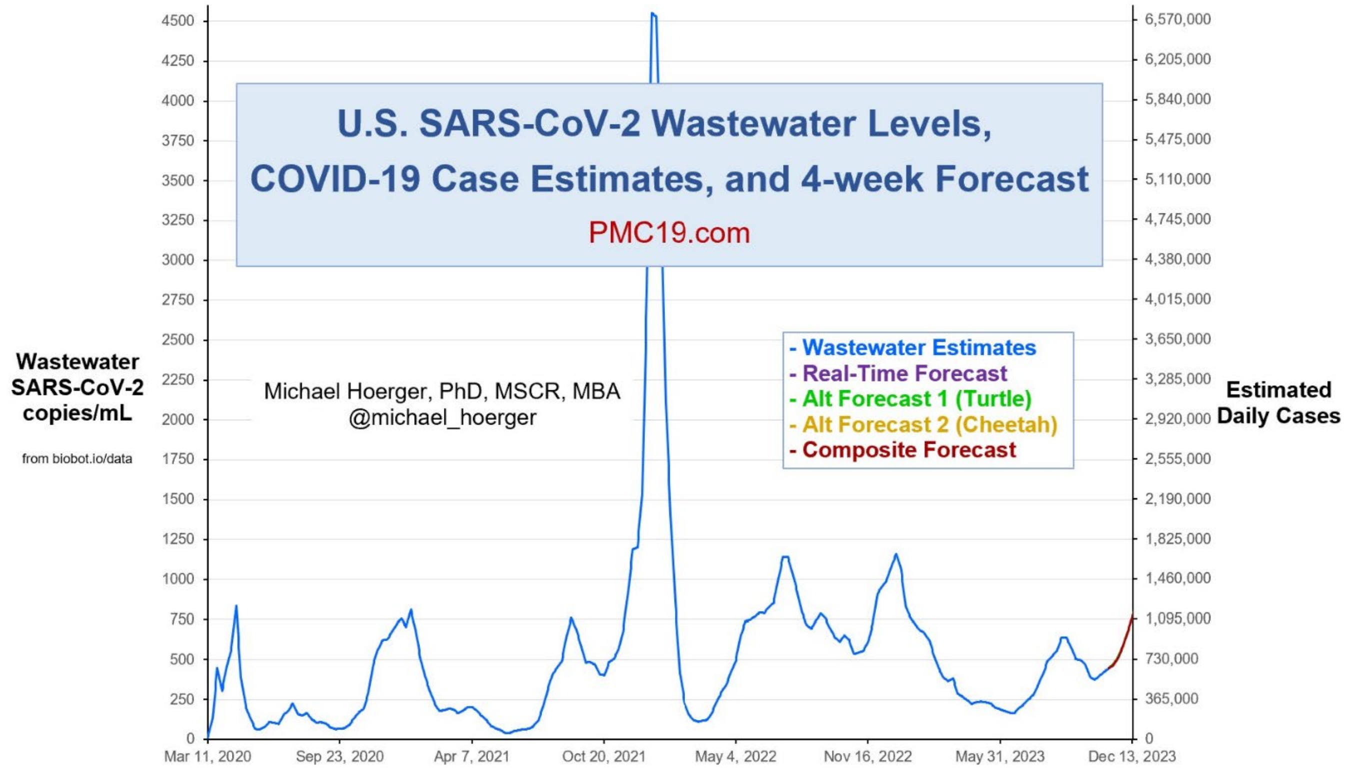
General Commentary

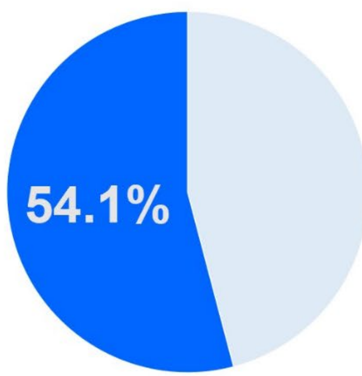
We are in the onset of the 8th U.S. COVID wave. In a few weeks, the wave will match the 7th wave we saw in the late summer. Then, it will get much worse.

U.S. wastewater levels indicate that COVID transmission is higher than during 54% of the days of the pandemic and lower than during 46% of the days of the pandemic:

- 1.41% (1 in 71) are infectious
- >670,000 C0VID cases/day
- >34,000 #LongCovid cases/day

Weekly cases stand at 4.7 million. Weekly long COVID cases resulting from these infections stand at >235,000.





There is more COVID-19 transmission today than during 54.1% of the pandemic.

CURRENT ESTIMATES FOR November 13, 2023

Wastewater Levels (copies/mL)
462
New Daily Cases
672,000
% of Population Infectious
1.41% (1 in 71 people)
New Daily Long COVID Cases
34,000 to 134,000

WEEKLY ESTIMATES FOR November 13, 2023

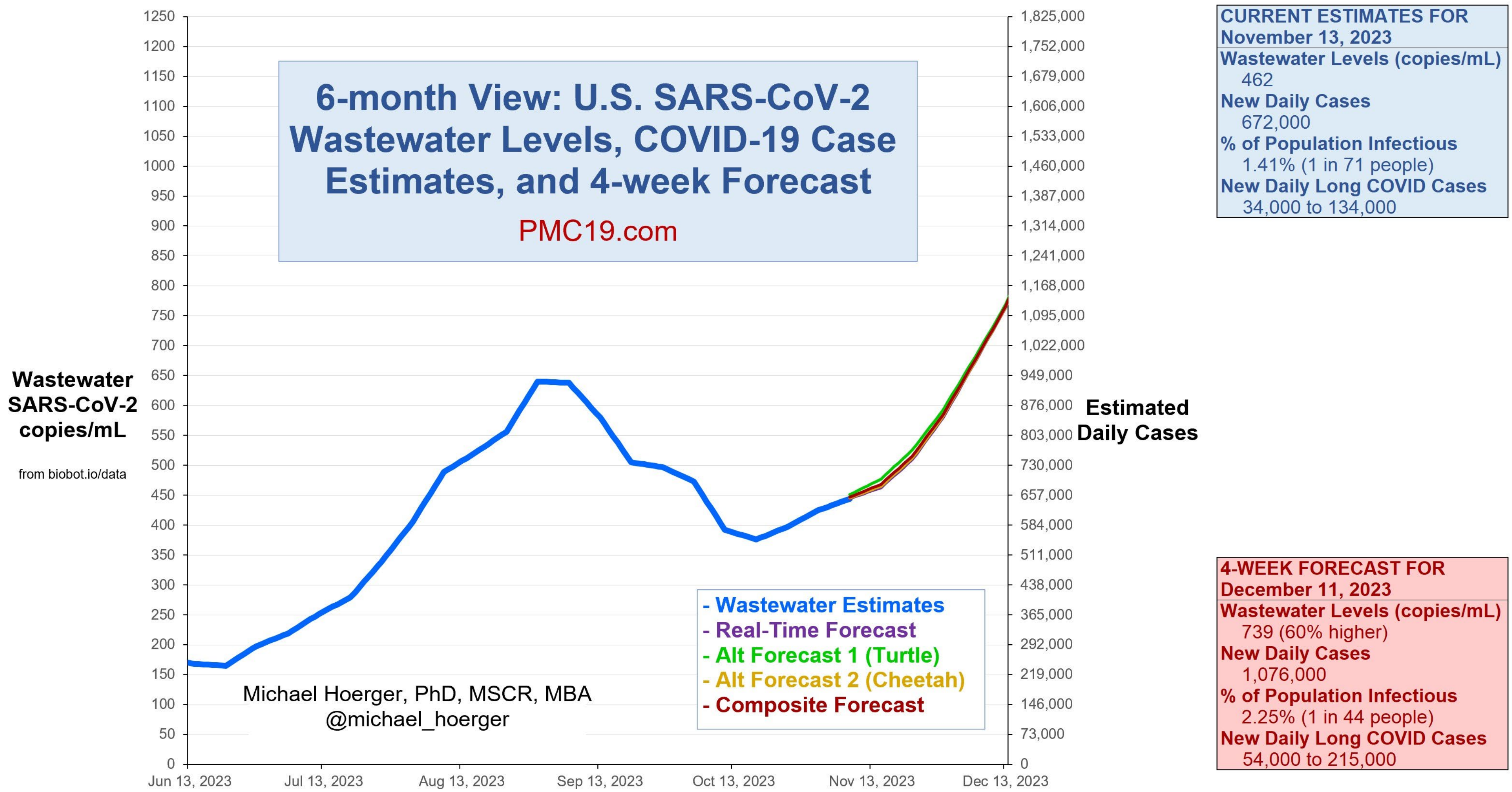
New Weekly Cases
4,700,000
New Weekly Long COVID Cases
235,000 to 941,000

2023 CUMULATIVE ESTIMATES AS OF November 13, 2023

Total 2023 Cases To Date
194,286,130
Total 2023 Long COVID Cases To Date
9,714,000 to 38,857,000

Forecast for the Next Month

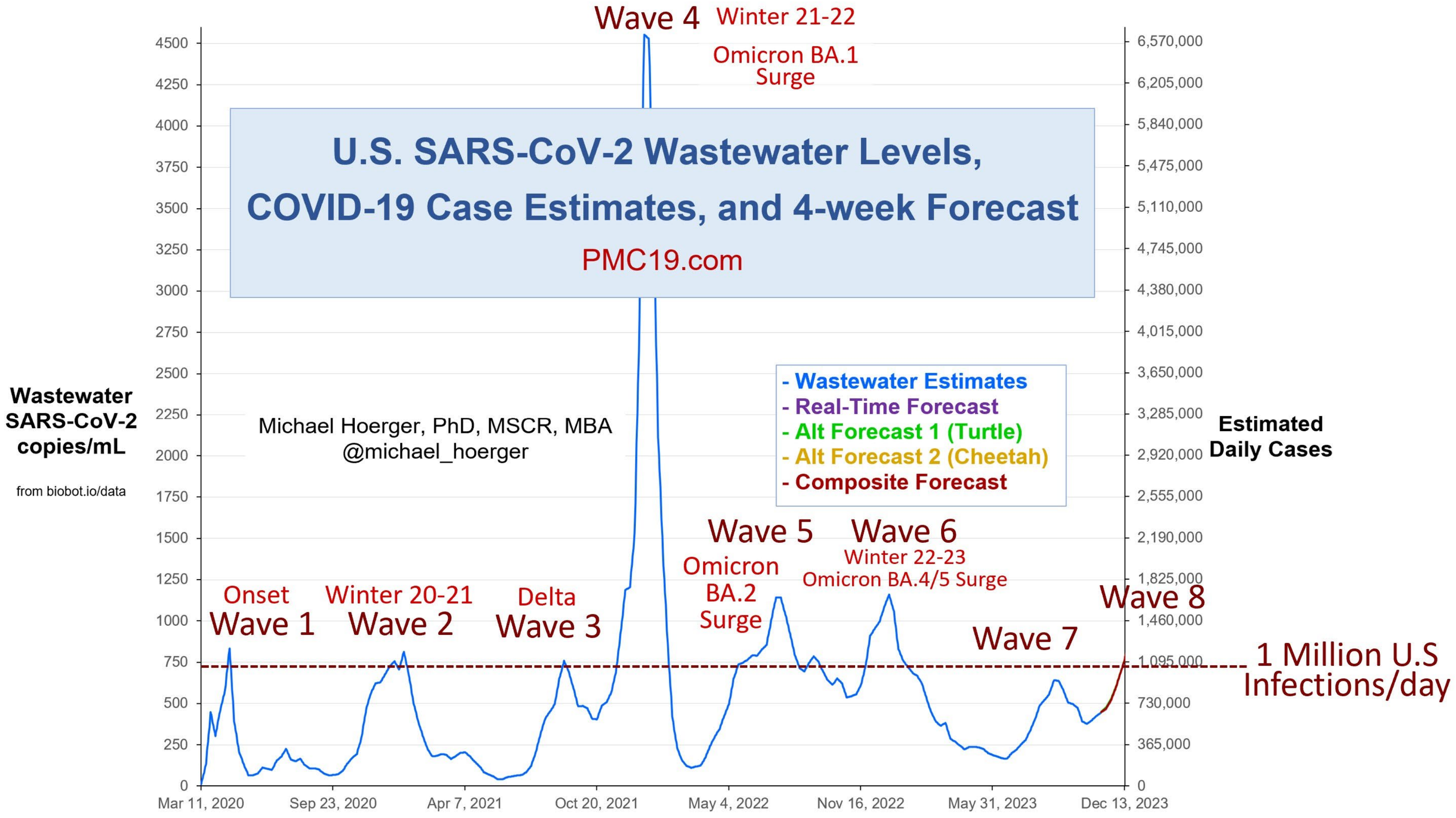
The different forecasting models are in near perfect agreement that a month from now we will see about 1.1 million U.S. cases per day, with 2.25% of the U.S. population or 1 in 44 people actively infectious.



*Note, some of the color coding has been updated this week, so review forecast lines thoughtfully.

Let us briefly review each forecast. The technical notes describing each forecast are summarized near the end of the report. The real-time (purple line) forecast uses the Biobot wastewater data. The Alt 2 (cheetah, yellow line) model corrects the real-time forecast base on the level of over/under-reporting in the prior Biobot update. The recent real-time estimates have been accurate, so the cheetah model matches the real-time model very closely. The Alt 1 (turtle, green line) model ignores the most recently reported wastewater level as unreliably inaccurate. You will see that in ignoring the most recent data point, the turtle model is still very similar to the other models, which means we are following historical trends closely, and the most recent wastewater levels offer no surprise. The composite model (red line) is the average of the individual models and is used for estimates of future cases. Note that all models agree on what we are headed toward: very bad transmission.

How Does the Oncoming 8th Wave Compare with Prior Waves?



I suspect we're headed somewhere between the magnitude of Wave 2 (winter 2020-21) and Wave 6 (last winter), and if so, I would characterize us as heading into a winter surge. I do not use the word "surge" lightly. In comparing waves, I see a qualitative distinction at about 1 million daily infections (wastewater levels of 750 copies/mL). Waves 1, 2, 3, and 7 all hovered around that mark. I refer to these as waves, not surges. Wave 4, 5, and 6 all had sustained time periods of weeks with over 1 million infections per day. The "area under the curve" is substantially higher, and I refer to these as "surges." If one ventures to use this distinction, will Wave 8 be a wave like Waves 1, 2, 3, and 7, or more of a surge with sustained daily infections of >1 million per day? Remember, the model uses a combination of historical data plus what is happening during the preceding 4 weeks. This means we will know best when a couple weeks out. Beyond 4 weeks, the model is relying heavily on historical data, and the three prior winters (Waves 2, 4, and 6) are very different. Wave 4 is an outlier among the waves, obviously. Wave 2 was more typical of a non-winter wave. Wave 6 (surge) piggybacked on Wave 5 (surge) and may have been unique in that regard due to any fleeting population-level immunity. My model suggests we're headed toward a surge slightly bigger than last year, and I could see that, because any fleeting immunity from Wave 7 is likely less than that derived from Wave 5, and people are not up-to-date on vaccinations. Other behavioral precautions are similar for this versus last winter: mostly non-existent. However, the model does not capture any of that. It's predicting worse (more infections) than last year because it's influenced by the BA.1 wave. Acknowledging that, I suspect we will peak slightly lower than Wave 6, perhaps around the 1.5 million infections/day rate. Anything between Wave 2 and a 20% bigger version of Wave 6, I would consider a reasonable estimate at this point.

What's the Current Risk in an Office or in a Classroom?

The office and classroom risks remain quite bad. In a group of 10 people (daycare, team meeting, etc.), there's a 13% chance someone will have infectious COVID. In a group of 20 people (e.g., K-12 classroom, department meeting, busy hospital waiting room, etc.), there's 25% chance someone would have infectious COVID. In a university classroom of 50 people, it should be assumed someone has infectious COVID. This is quite troubling for instructors or students who mix time with multiple groups of classmates each week.

What's the COVID Risk in an Office or in a Classroom?

Number of People	Chances Anyone is Infectious	Number of People	Chances Anyone is Infectious
1	1.4%	25	29.8%
2	2.8%	30	34.6%
3	4.2%	35	39.1%
4	5.5%	40	43.2%
5	6.8%	50	50.7%
6	8.1%	75	65.4%
7	9.4%	100	75.7%
8	10.7%	150	88.0%
9	12.0%	200	94.1%
10	13.2%	300	98.6%
15	19.1%	400	99.7%
20	24.7%	500	>99.9%

Not all classrooms and meetings are the same. Virtual meetings reduce risk close to zero. Outdoor meetings are often safer than indoors. Testing reduces risk, as do policies that encourage people to stay home when symptomatic. High-quality, well-fitting masks greatly reduce risk. Air quality monitoring and improved air cleaning reduce risk. Recent boosters reduce risk. It remains troubling that elected leaders and public health officials choose to model poor mitigation when ongoing risk is so high.

The CDC has recently approved an updated booster, available to anyone in the U.S. older than 6 months. It is becoming widely available for adults, and now increasingly available for children. Utilization rates are only 7% for adults and 2% among children. These 'abysmal' figures indicate poor access and awareness. Use the PMC data and local vaccination location information to help people get vaccinated.

Forecast for Thanksgiving

Based on forecasted transmission on Thanksgiving day, this figure shows the chance anyone would be infectious in a group based on group size. If seeing 10 people for Thanksgiving, there's a 15% chance anyone would be infectious. In a group of 20, there's a 28% chance someone would be infectious. These figures are virtually unchanged from last week's estimates, so nothing newly surprising here.

In the U.S., What's the COVID Risk for Thanksgiving?

Number of People	Chances Anyone is Infectious	Number of People	Chances Anyone is Infectious
1	1.6%	25	33.2%
2	3.2%	30	38.3%
3	4.7%	35	43.1%
4	6.2%	40	47.5%
5	7.7%	50	55.3%
6	9.2%	75	70.1%
7	10.7%	100	80.0%
8	12.1%	150	91.1%
9	13.5%	200	96.0%
10	14.9%	300	99.2%
15	21.5%	400	99.8%
20	27.6%	500	>99.9%

@michael_hoerger Estimated November 13, 2023

Forecast for Christmas

Based on forecasted transmission on Christmas day, this figure shows the chance anyone would be infectious in a group based on group size. Transmission will peak around New Year's Day, so Christmas will be very bad for transmission. If seeing 10 people for Christmas, there's a 27% chance anyone would be infectious. In a group of 20, it should be assumed that someone may be infectious. This estimates are also virtually unchanged from last week, but expect better precision within 2-4 weeks of the holiday.

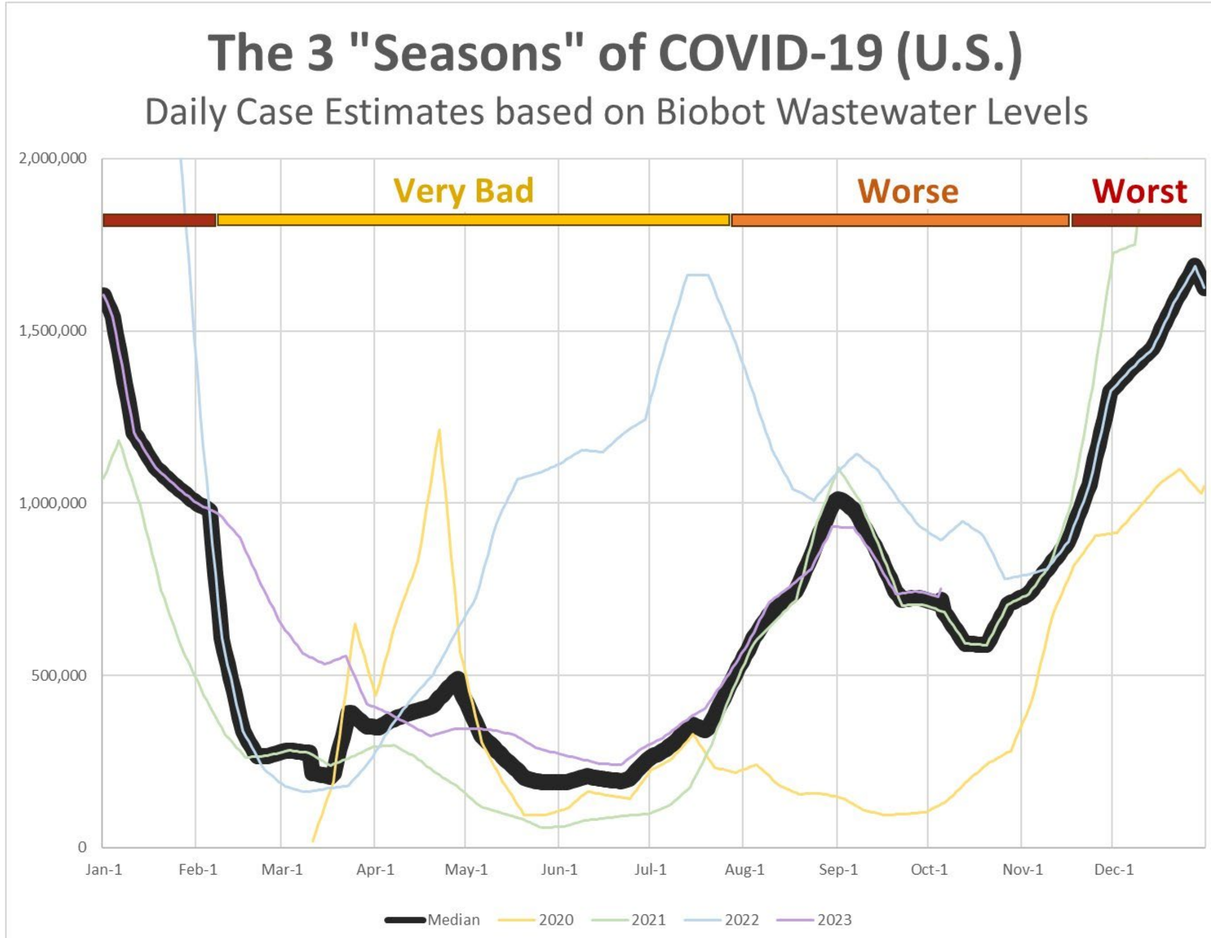
In the U.S., What's the COVID Risk for Christmas Day?

Number of People	Chances Anyone is Infectious	Number of People	Chances Anyone is Infectious
1	3.1%	25	54.8%
2	6.1%	30	61.4%
3	9.1%	35	67.1%
4	11.9%	40	71.9%
5	14.7%	50	79.5%
6	17.3%	75	90.7%
7	19.9%	100	95.8%
8	22.4%	150	99.1%
9	24.8%	200	99.8%
10	27.2%	300	>99.9%
15	37.9%	400	>99.9%
20	47.0%	500	>99.9%

@michael_hoerger Estimated November 13, 2023

Forecast for the Longer-Term – Annual Trends

*This section is being reshared from a prior report, as some people may need to make decisions about travel, surgery, and other important events several months out.



Is C0VID-19 "seasonal"? Not in any meaningful sense of the word. The following graph uses historical Biobot #wastewater data to estimate daily case rates using the PMC model. Thin lines show 2020 (yellow), 2021 (green), 2022 (blue) and 2023 (purple). The black line is the median. It is not really a forecast, but merely a summary of historical data. To the extent the median provides a reasonable approximation of the future, it is a useful starting point for a gift-level forecast.

Season 1: "Very Bad Transmission." Focusing on the median, you'll see that case rates tend to be lower (but still in the 250-500K/day range) from mid-Feb through the end of July. These are valuable data. If I needed to schedule a non-urgent surgery, when would I do it? Late February, when transmission has often dropped, but before the general public not monitoring wastewater has realized so, perhaps meaning some people are still using airborne precautions. You're basically hopefully beating the transmission "market." I'd also be prepared to cancel an appointment or push back 6 weeks if needed.

Season 2: "Worse Transmission." Again, focusing on the median, you see a late summer wave from August through mid-October. This is the clearest indication that C19 is not "seasonal," if people are using that term to mean an annual event. If we were doing two boosters a year, it seems like booster 1 would roll out in July. Why do we have this wave? Schools have very little mitigation (poor air quality, little/no testing, little/no masking, low vax rates). Also, the fleeting immunity from winter boosters and infections has waned. If I had an urgent maskless medical/dental visit, I'd schedule mid-October through early November and cross my fingers (around my HEPA). It's still high transmission but about to get worse. This is also a good time to stock up on N95s, rapid tests, and HEPA filters before the prices may increase, scarcity may become a problem, or one has an infection in the home. Travel insurance is wise.

Season 3: "Worst Transmission." From mid-November to mid-Feb, transmission is extremely problematic, according to the median line. Everybody should be wearing high-quality masks, testing as frequently as possible, improving indoor air quality, and moving activities outdoors and remote.

A Couple Caveats. *Seasonality*. Some people use the word "seasonal" to mean predictable, rather than merely a discrete 2-3 month season of transmission. In some ways, transmission is predictable. You'll see the 2023 purple line has followed the median very closely. However, we're talking about a very small sample size of years, so one would expect one of the years to mostly follow the median. Also, there are clear discrepant cases. BA.1 goes off the chart (winter 2021 to early 2022). The 2022 summer wave was also sizable. My approach is to make longer-term plans based on the median line and then be prepared to shift plans toward more remote activities if a large wave picks up. Hopefully, transmission becomes more predictable as years go by, but I'm not betting on it yet.

Case estimates. If you have followed the PMC dashboard, you'll know these are estimated by linking Biobot wastewater levels to IHME true case estimates. I would find case estimates 15% higher or 30% lower also reasonable and discuss these estimates with many modeling experts. There are also some more sophisticated models, where I believe an argument can be made that waves are actually marginally more leptokurtic (spikier mountains and deeper valleys than shown here).

General Technical Notes, Not Specific to the Current Week's Report

Status of Biobot wastewater reporting. The estimates and forecast described here use wastewater data reported by Biobot. Biobot is now updating their data on Fridays or Mondays, and the CDC has awarded several prior Biobot sites to a company called Verily. The transitionary phase at Biobot seems mostly through, though Biobot is contesting the contract reassignment in court. As long as national wastewater data are being reported, the PMC reports will continue.

Case estimates. Case estimates were used by evaluating various potential multipliers to go from wastewater levels to cases. To identify true cases, not merely just reported cases, I used the IHME's case estimates for January 1, 2021 through April 1, 2023 (https://covid19.healthdata.org/united-states-of-america?view=cumulative-deaths&tab=trend). I compared wastewater with their case estimates on the 1st of each month. The correlation was r=.94. The maximum possible correlation is 1.00, so that is freakishly high, higher than just about any of the 10,000 or so correlations I've ever run. I was hoping for a correlation of r=.70 or higher, which still would have been great. Basically, wastewater is a supreme indicator of case rates. Next, I examined multipliers. Are cases 10x the arbitrary wastewater metric? 10,000x? Something else? Take cases and divide by wastewater at each data point, then find a summary metric (mean, median, trimmed mean, etc.). The metric I found most defensible was to use a +/-10% trimmed mean (average that excludes extreme data points, where case estimates are more error-prone), where each unit of wastewater translated into 1455 cases. I would find multipliers of 1000 to 1700 (31% lower to 17% higher) also reasonable. Arguably, case rates are magnitudes (10-100 times) higher than many people expect, so these details have minimal practical significance for everyday decision making. There are also more sophisticated strategies, such as regression models, but I found those results to be counter-intuitive (e.g., positive intercept, where I would have expected zero or negative). One can set the intercept to zero, use various heteroscedasticity-related techniques, and correct for the lack of imperfect reliability, but most of that is over the heads of people using this model and would accomplish little more than the trimmed multiplier method. The multiplier method has also led to techniques (only posted on Twitter thus far) for making regional estimates using very simple multipliers. Elegant is good.

Percentage infectious. After estimating the current number of new infections, it is relatively straightforward to estimate the percentage of the U.S. population actively infectious with COVID-19, but there are several caveats worth noting. One, the U.S. population is assumed to be 334,565,848. This was the CDC-estimated U.S. population on the final day of the IHME case estimation model. The number of new daily cases divided by the population tells one the percentage of the population newly infected today, often small at around 0.3% or less. Two, consider the infectious window. The percentage of the population infectious depends on the percentage of new people infected but also the duration people stay infectious. The model assumes people stay infectious for 7 days. Low estimates are that people are infectious for an average of 5 days (this defies the preponderance of the evidence, in my view), and high estimates are more like 10 days (too high in my view, based on a preference for round numbers). Other compelling estimates are more like 8-8.5 days. This duration may change over time, based on new variants, new vaccines, vaccine utilization rates, and treatments. If assuming the infectiousness duration is 10% longer, multiply by 1.10. If assuming 20% shorter, multiply by 0.80. New cases divided by the population equals new daily infections. Note also, these are merely averages and do not reflect individual variation, as some get infected and are not contagious, whereas others get infected and remain infectious likely for months (extremely rare). New daily infections multiplied by the number of days infectious indicates the percentage of the population actively infectious.

Long COVID. Long COVID case estimation. The lower and upper bounds for Long COVID case estimates assume that 5-20% of people infected with SARS-CoV-2 will develop Long COVID as a result of that infection. Some published reports and analysts have suggested lower (1%) or higher (40%) values. A useful framework for thinking about these estimates is that the low value is more indicative of people experiencing serious, enduring, known harms, whereas the upper estimates are closer to the number experiencing disruptive symptoms for at least several months, perhaps with full or partial recovery. These estimates do not indicate unknown long-term harms. For example, if infections increase the risk of cancer or cardiovascular disease substantially and with increasing risk over 10-30 years, that is not captured well by these metrics. The metrics also do not encompass the 1.2 to 1.8 million Americans who have died of COVID-19. Future models may incorporate estimates of mortality. Finally, the estimates project the number who will ultimately experience Long COVID from a new infection, but that is several months down the line. The estimates reflect future implications. For simplicity of interpretation, they are not modeling the number of new Long COVID cases today that resulted from infections three months ago.

General forecasting model specification. The forecasting models are elegant, meaning simple and effective. In regression analyses using historical pandemic wastewater data, the model explains 96% of the variance in the following week's forecast. The model is simple. It includes the year (2020, 2021, 2022, or 2023). It includes the historical average for the current half month; imagine the year sliced into 26 pieces, and it incorporates data on the historical average for that half month (e.g., second half of September). The model also incorporates four lagged variables, the wastewater levels 1, 2, 3, and 4 weeks ago. Overall, you can think of the model as having two main processes. One incorporates what we know historically. The other incorporates what has been happening the past several weeks. The historical data are useful because transmission mostly, but not always, follows a particular monthly pattern. It is not seasonal in that there are not just three bad months a year, but there is month-to-month variation, and sometimes even useful differences between the first versus second half of the month. The use of recent wastewater estimates helps in several ways. It lets the model know if something about the current point in time differs dramatically from the historical data, and it quickly adapts the model to changes, such as if a wave is starting or ending,

Real-time model (purple line). This model assumes that real-time data reports of wastewater levels are accurate. However, real-time data often get corrected. Some sites may be slow reporting, and if there is a bias built in, such as places with high transmission being late to report, that would be a problem. Often, the real-time reports are quite accurate, but occasionally they have been corrected substantially a week later. The general model places a lot of weight on the most recent data, so any errors here can lead the model to assume a wave is picking up that really is not (false alarm) or that things are improving better than expected (false hope).

Alt model #1, turtle (green line). The turtle model moves slow and steady. It completely ignores the most recent week's worth of data from Biobot, treating it as unreliable. It will ignore false fluctuations inferred from inaccurate real-time reporting. However, it will be slower to respond to real changes, such as the onset in a new wave or the decline in a wave that has peaked.

Alt model #2, cheetah (orange line). The cheetah model moves fast. It aims to correct for biases in real-time data reports. If last week's real-time report overestimated levels by 10% upon correction, it assumes this week's real-time report suffers the same bias. If last week's real-time report underestimated true levels, it assumes the same for this week. If last week's real-time report was accurate, it will look similar to the real-time model. This model is very good if there is a bias, such as if areas with high transmission experience delays in reporting. However, it can also be overreactive. If there was some error in a real-time report that was just "random" rather than biased in a particular correction, it will tend to overcorrect the next week's model.

Composite Model (red line). This is the arithmetic average of the three models. It's what's used for deriving all of the statistics reported. When all of the individual models are very close to the average, that suggests high confidence. When the models make vastly different predictions, that suggests more uncertainty in the data, largely based on perceptions of the accuracy of real-time wastewater reporting.