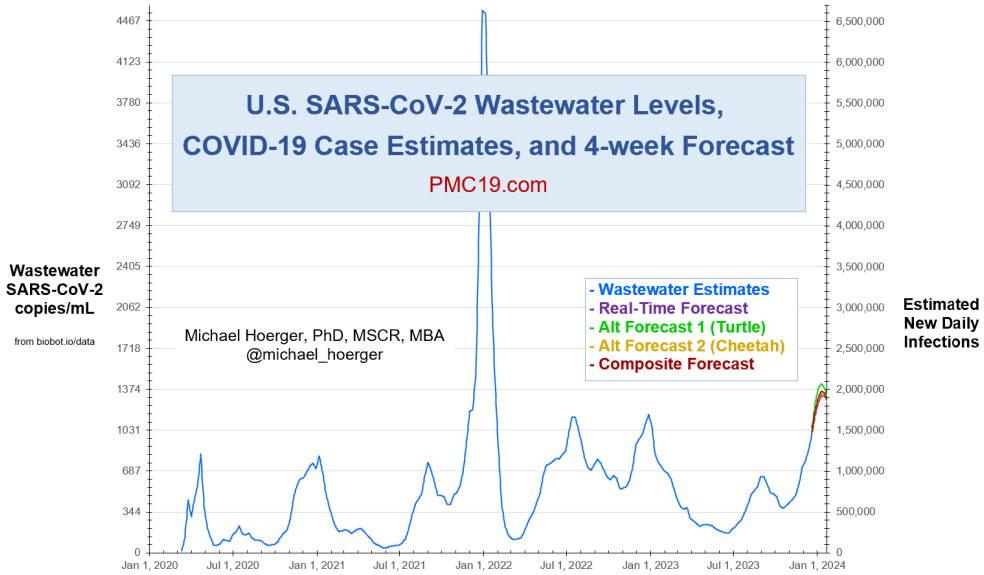
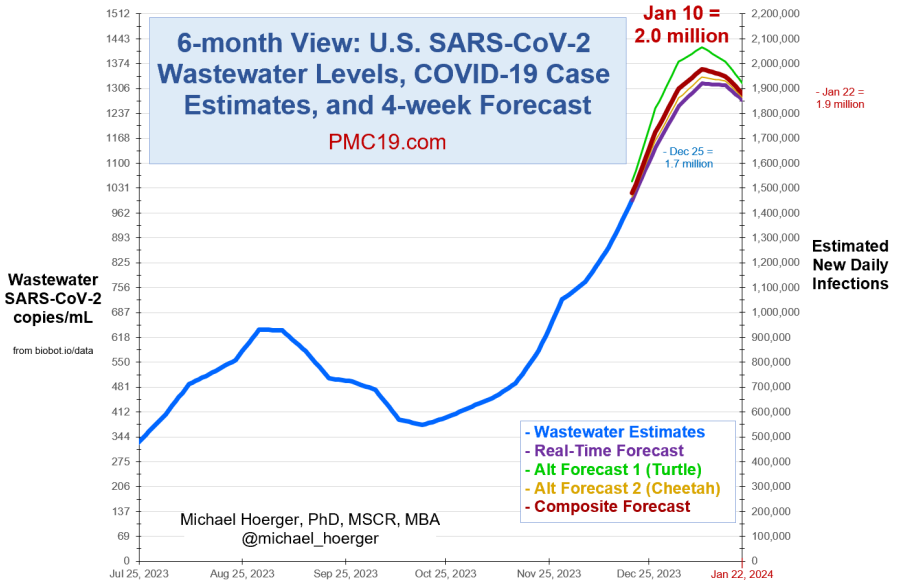
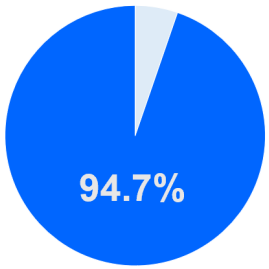
Michael Hoerger, PhD, MSCR, MBA, Pandemic Mitigation Collaborative

U.S. SARS-CoV-2 Wastewater Levels, COVID-19 Case Estimates, and 4-Week Forecast: Report for December 25, 2023, pmc19.com/data







There is more COVID-19 transmission today than during 94.7% of the pandemic.

CURRENT ESTIMATES FOR December 25, 2023
Wastewater Levels (copies/mL)
1137
New Daily Cases
1,655,000
% of Population Infectious
3.46% (1 in 29 people)
New Daily Long COVID Cases
83,000 to 331,000

WEEKLY ESTIMATES FOR December 25, 2023
New Weekly Cases
11,600,000
New Weekly Long COVID Cases
579,000 to 2,317,000

2023 CUMULATIVE ESTIMATES AS OF December 25, 2023
Total 2023 Cases To Date
241,674,380
Total 2023 Long COVID Cases To Date
12,084,000 to 48,335,000

4-WEEK FORECAST FOR January 22, 2024
Wastewater Levels (copies/mL)
1,293 (14% higher)
New Daily Cases
1,881,000
% of Population Infectious
3.94% (1 in 25 people)
New Daily Long COVID Cases
94,000 to 376,000

How Does Risk Increase with More Social Contacts?			
Number of People	Chances Anyone is Infectious	Number of People	Chances Anyone is Infectious
1	3.5%	25	58.6%
2	6.8%	30	65.3%
3	10.0%	35	70.9%
4	13.1%	40	75.6%
5	16.2%	50	82.8%
6	19.1%	75	92.9%
7	21.9%	100	97.1%
8	24.6%	150	99.5%
9	27.2%	200	>99.9%
10	29.7%	300	>99.9%
15	41.1%	400	>99.9%
20	50.6%	500	>99.9%

Cite as: Hoerger, M. (2023, December 25). *U.S. SARS-CoV-2 wastewater levels, COVID-19 case estimates, and 4-week forecast: Report for December 25, 2023*. Pandemic Mitigation Collaborative. http://www.pmc19.com/data

General Commentary

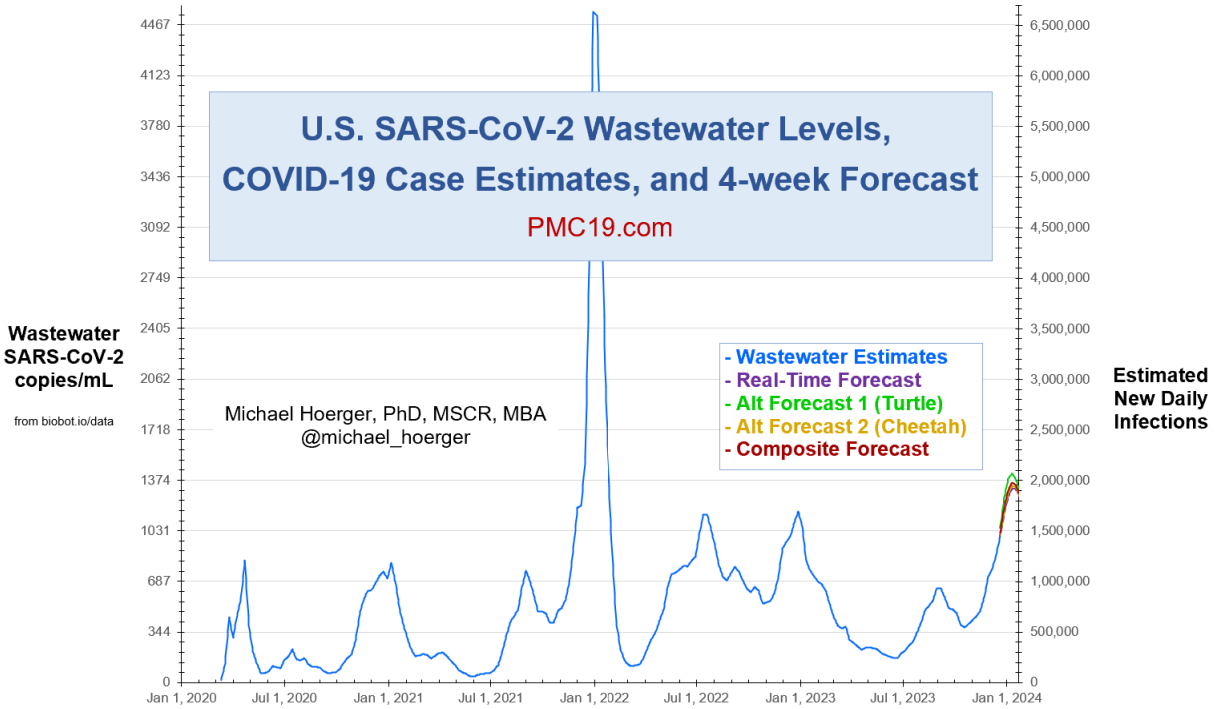
We are in the 8th U.S. COVID wave and 2nd biggest all-time, barring any atypical downward corrections of wastewater levels. We have surpassed the best estimates of the initial COVID wave, the first winter, the Delta wave, and the most recent late-summer wave. Any claims that the “pandemic is over” or “COVID is over” run contrary to stated evidence, will inflict much harm, and should be labeled misinformation.

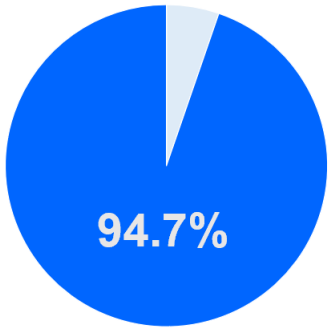
U.S. wastewater levels indicate that COVID transmission is higher than during 94.7% of the days of the pandemic and lower than during 5.3% of the days of the pandemic.

- 3.46% (1 in 29) are infectious
- >1.6 million COVID cases/day
- >80,000 Long Covid cases/day

Weekly new cases stand at 11.6 million. Weekly long COVID cases resulting from these infections stand at >0.5 million.

Nearly 1 in 3 Americans will get infected during the peak 2 months of this winter surge. That’s 105 million infections total. Over half of infections will occur after the peak, which is why it remains important to encourage vaccination among people who are delayed.





There is more COVID-19 transmission today than during 94.7% of the pandemic.

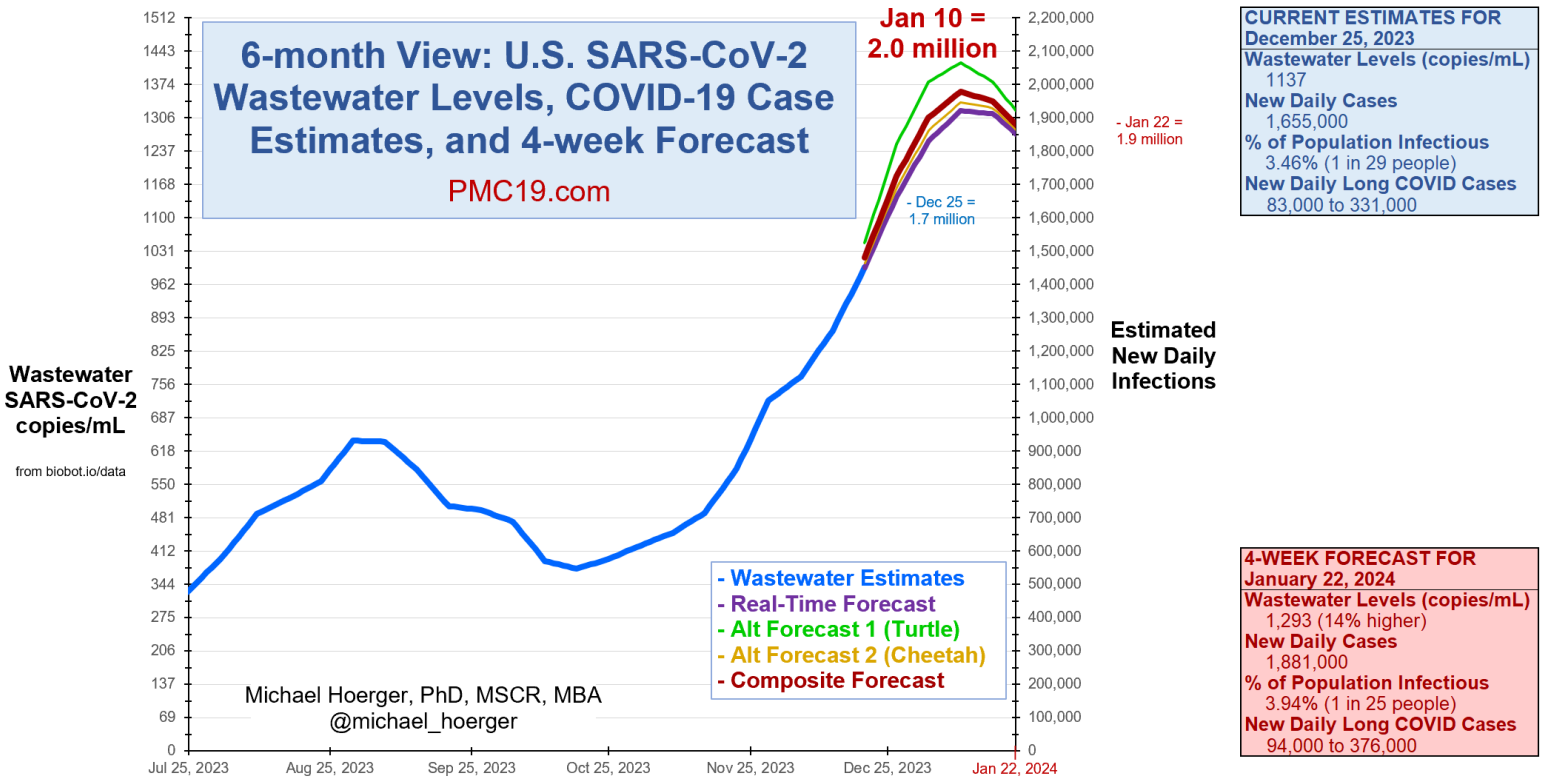
CURRENT ESTIMATES FOR December 25, 2023
Wastewater Levels (copies/mL)
1137
New Daily Cases
1,655,000
% of Population Infectious
3.46% (1 in 29 people)
New Daily Long COVID Cases
83,000 to 331,000

WEEKLY ESTIMATES FOR December 25, 2023
New Weekly Cases
11,600,000
New Weekly Long COVID Cases
579,000 to 2,317,000

2023 CUMULATIVE ESTIMATES AS OF December 25, 2023
Total 2023 Cases To Date
241,674,380
Total 2023 Long COVID Cases To Date
12,084,000 to 48,335,000

Forecast for the Next Month

The forecast indicates that we are continuing toward the 2nd-highest U.S. surge all-time. The different forecasting models show a high level of convergence. At peak, we may see 2 million U.S. infections/day with 4.1% of the population (1 in 24 people) actively infectious around January 10. The specific day of the peak will be unknowable even in hindsight (different data sources, some with weekly versus more frequent reporting), but a reasonable approximation would be around January 10th.



CURRENT ESTIMATES FOR December 25, 2023
Wastewater Levels (copies/mL)
1137
New Daily Cases
1,655,000
% of Population Infectious
3.46% (1 in 29 people)
New Daily Long COVID Cases
83,000 to 331,000

4-WEEK FORECAST FOR January 22, 2024
Wastewater Levels (copies/mL)
1,293 (14% higher)
New Daily Cases
1,881,000
% of Population Infectious
3.94% (1 in 25 people)
New Daily Long COVID Cases
94,000 to 376,000

U.S. Winter 2023-24 COVID Surge

	Best Estimate	Range	
Rank among COVID waves	2nd	2nd	4th
Date of peak	Jan 10	Jan 3	Jan 17
Daily infections at peak	2.0 million/day	1.7 million/day	2.1 million/day
Percentage of population infectious at peak	4.1% (1 in 24)	3.7% (1 in 27)	4.4% (1 in 23)

The differences across the four forecasting models are marginal. The real-time forecast (purple) anticipates a peak at 1.9 million daily infections. The cheetah model (orange) corrects the real-time model for over/under-reporting in last week's real-time Biobot data, so a slightly higher estimate. The turtle model (green) ignores the most recent week's real-time data as a potential aberration and is a bit more liberal than the other models based on preceding weeks' transmission acceleration. The composite model (red) is the average of all three. These differences are not particularly important.

More importantly, consider optimistic and pessimistic scenarios not captured by these models. A rosy scenario would be that the peak occurs a week earlier at a slightly lower level (1.7 infections/day like last winter or the preceding summer, which would require some downward corrections). The level of acceleration in transmission argues against that, in favor of a higher peak, but Biobot is reporting some unusual regional variation (delayed acceleration in transmission in the U.S. South and West). Moreover, historical patterns of how transmission should or should not accelerate cannot account for existing variation on population-level immunity due to variation in prior exposure history, recency of vaccination, and how well the current vaccine matches disseminating subvariants relative to prior vaccines. Finally, Biobot wastewater sites could be overreporting, and levels could get corrected downward. Each of these factors is highly plausible, but the "rosy" scenario remains quite bleak and suggests the pandemic remains far from "over."

Also, consider more pessimistic scenarios. Current vaccination rates remain extremely low, and several other countries are reporting atypically high acceleration via wastewater data. Placing plausible hypothetical values in the model, it is difficult to imagine a scenario where the U.S. reaches 2.4 million infections/day. Sometimes, people draw graphs showing a continued acceleration like BA.1, but such models seem to reflect imagination rather than data. The data do not suggest any evidence for a BA.1-level surge.

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 remain abysmally low. Use the PMC data and local vaccination location information to help people get vaccinated.

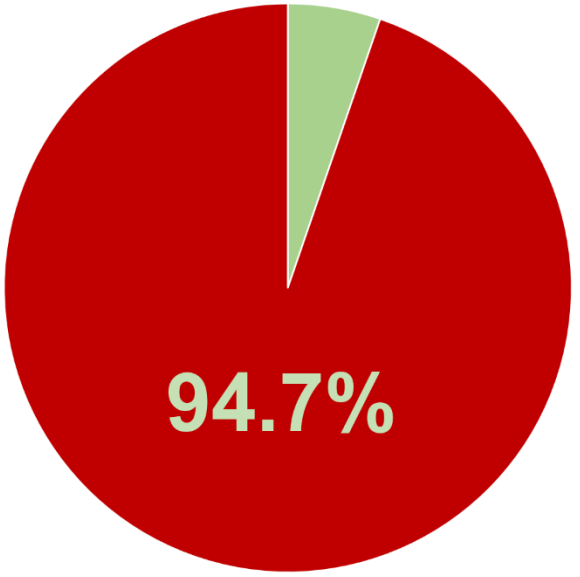
Current Risk Based on Number of Social Contacts – Christmas Day

This figure shows the chance anyone would be infectious in a group based on group size. If seeing 10 people for Christmas, there's a 30% chance anyone would be infectious. In a group of 20 (or two family gatherings totaling 20), it's a coin toss whether someone will be infectious. If going to a packed restaurant or traveling on a plane with 100 other people, assume 97% chance someone there may be infectious. These estimates are similar to those suggested over the past several weeks, so hopefully they have been helpful in planning. Note that these estimates do not account for "dependency" in the data, that sub-sets of people gathering may be more likely to have nobody infectious (one cautious family) or more likely to have multiples infectious (one incautious symptomatic family). This skews the odds toward either safety or superspreaders. The table can be thought of as averages that account for both scenarios. Cognitive biased lead people to consider the "safer" option rather than the superspreader option. Moreover, people likely overestimate dependency, when in fact, many people within U.S. families have diverse recent exposure histories (youngest in daycare, tween playing indoor basketball, college student who flew home after two nights of parties, parents in unmitigated work environments). These are quite independent risk scenarios. The utility of the table as an average that checks against cognitive biases prevails as the best estimate.

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	3.5%	25	58.6%
2	6.8%	30	65.3%
3	10.0%	35	70.9%
4	13.2%	40	75.6%
5	16.2%	50	82.8%
6	19.1%	75	92.9%
7	21.9%	100	97.1%
8	24.6%	150	99.5%
9	27.2%	200	99.9%
10	29.7%	300	>99.9%
15	41.1%	400	>99.9%
20	50.6%	500	>99.9%

@michael_hoerger Estimated December 25, 2023



There is more COVID-19 transmission today than during 94.7% of the pandemic.

Christmas and family travel and gatherings will seed transmission, which will seed more transmission around New Year's Eve, and drive the peak around the 10th. There is a tendency to focus on new sub-variants ("scariants"), or variation in air travel, which will show in media headlines that ignore the fact that nothing surprising is happening. The surge is the result of quite predictable behavior under a lack of public health leadership.

The PMC forecast focuses on the next 4 weeks and gains precision within 2 weeks, but for planning purposes it can be helpful to have a longer-range forecast at times, as long as people understand that it's more speculative. On October 30, we posted our first "Christmas Risk" table (right), and on the whole, you can see it performed pretty well. People who changed their holiday plans based on that speculative table would likely be pleased today, particularly given that the current state is marginally worse than anticipated.

Posted on October 30

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	6.2%	30	61.5%
3	9.1%	35	67.1%
4	11.9%	40	72.0%
5	14.7%	50	79.6%
6	17.4%	75	90.8%
7	19.9%	100	95.8%
8	22.5%	150	99.1%
9	24.9%	200	99.8%
10	27.2%	300	>99.9%
15	37.9%	400	>99.9%
20	47.0%	500	>99.9%

@michael_hoerger Estimated October 30, 2023

Forecast for New Year's Day

This figure shows the chance anyone would be infectious in a group based on group size. Transmission will hopefully peak around New Year's Day if we are lucky, though likely 1-2 weeks later. Expect transmission to be exceedingly bad. If seeing 10 people on New Year's Eve or for a New Year's Day brunch, there's likely about a 33% chance anyone would be infectious. In a group of 15-20, it's a coin toss whether someone will be infectious. If going to a packed restaurant or traveling on a plane with 100 other people, assume there's a 98% chance someone there may be infectious. None of these statements have changes since last week's estimates.

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

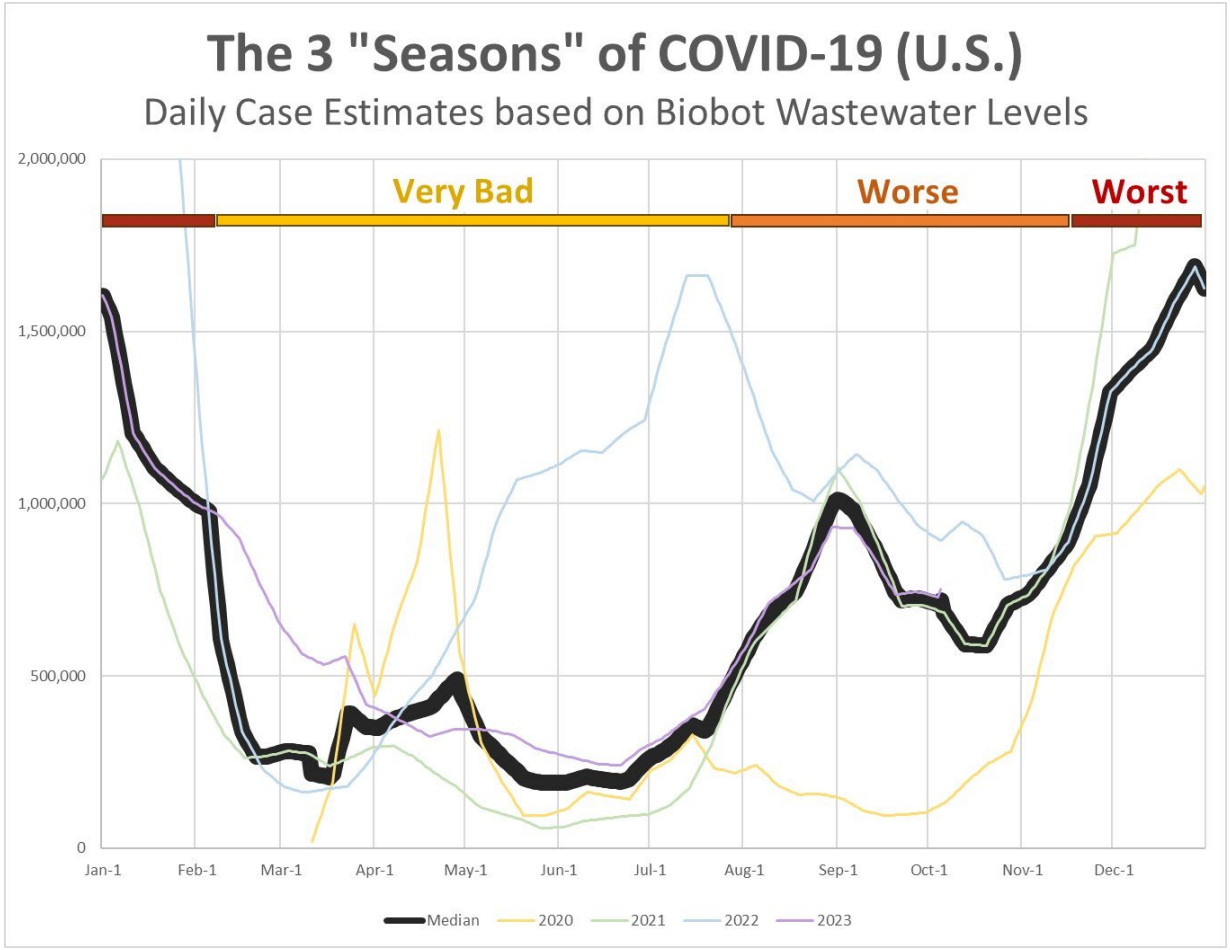
Number of People	Chances Anyone is Infectious
1	3.9%
2	7.6%
3	11.2%
4	14.6%
5	17.9%
6	21.1%
7	24.1%
8	27.1%
9	29.9%
10	32.6%
15	44.7%
20	54.6%

Number of People	Chances Anyone is Infectious
25	62.7%
30	69.4%
35	74.9%
40	79.4%
50	86.1%
75	94.8%
100	98.1%
150	99.7%
200	>99.9%
300	>99.9%
400	>99.9%
500	>99.9%

@michael_hoerger Estimated December 25, 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 median (switched from average on 12/11/23) for the current half month; imagine the year sliced into 26 pieces, and it incorporates data on the historical median 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.