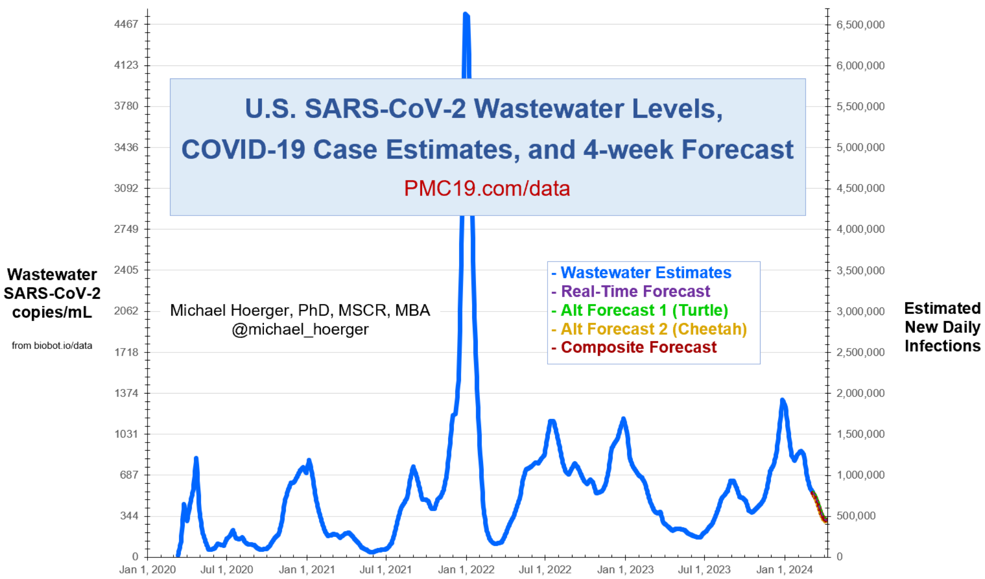
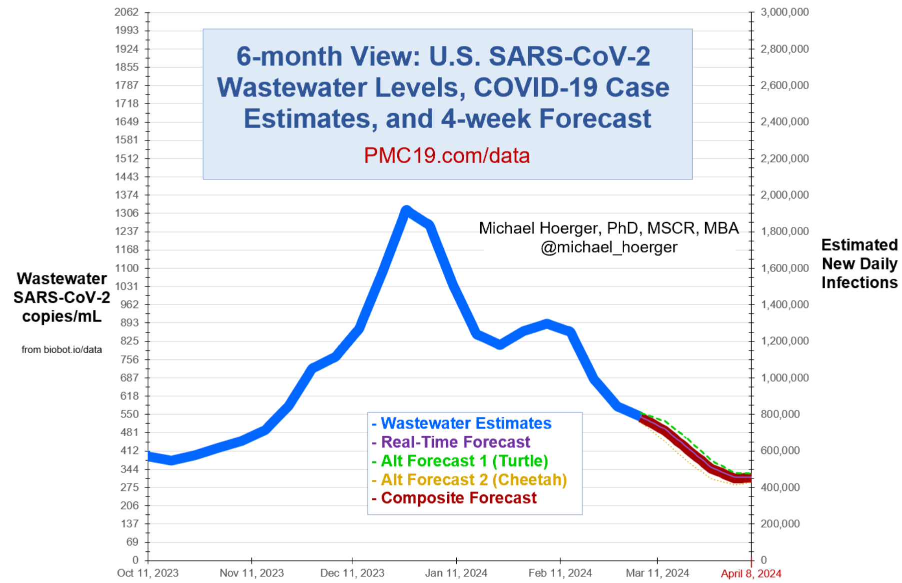
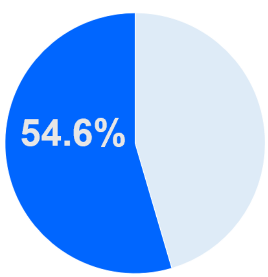
U.S. SARS-CoV-2 Wastewater Levels, COVID-19 Case Estimates, and 4-Week Forecast: Report for March 11, 2024, pmc19.com/data

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







There is more COVID-19 transmission today than during 54.6% of the pandemic.

CURRENT ESTIMATES FOR March 11, 2024
Wastewater Levels (copies/mL)
504
New Daily Cases
734,000
% of Population Infectious
1.54% (1 in 65 people)
New Daily Long COVID Cases
37,000 to 147,000

WEEKLY ESTIMATES FOR March 11, 2024
New Weekly Cases
5,100,000
New Weekly Long COVID Cases
257,000 to 1,028,000

2024 CUMULATIVE ESTIMATES AS OF March 11, 2024
Total 2024 Cases To Date
84,968,250
Total 2024 Long COVID Cases To Date
4,248,000 to 16,994,000

4-WEEK FORECAST FOR April 8, 2024
Wastewater Levels (copies/mL)
312 (-38%, lower)
New Daily Cases
453,000
% of Population Infectious
0.95% (1 in 106 people)
New Daily Long COVID Cases
23,000 to 91,000

How Does Risk Increase with More Social Contacts?

Number of People	Chances Anyone is Infectious	Number of People	Chances Anyone is Infectious
1	1.5%	25	32.1%
2	3.0%	30	37.1%
3	4.5%	35	41.8%
4	6.0%	40	46.2%
5	7.4%	50	53.9%
6	8.9%	75	68.7%
7	10.3%	100	78.7%
8	11.6%	150	90.2%
9	13.0%	200	95.5%
10	14.3%	300	99.0%
15	20.7%	400	99.8%
20	26.6%	500	>99.9%

Cite as: Hoerger, M. (2024, March 11). *U.S. SARS-CoV-2 wastewater levels, COVID-19 case estimates, and 4-week forecast: Report for March 11, 2024*. Pandemic Mitigation Collaborative. http://www.pmc19.com/data

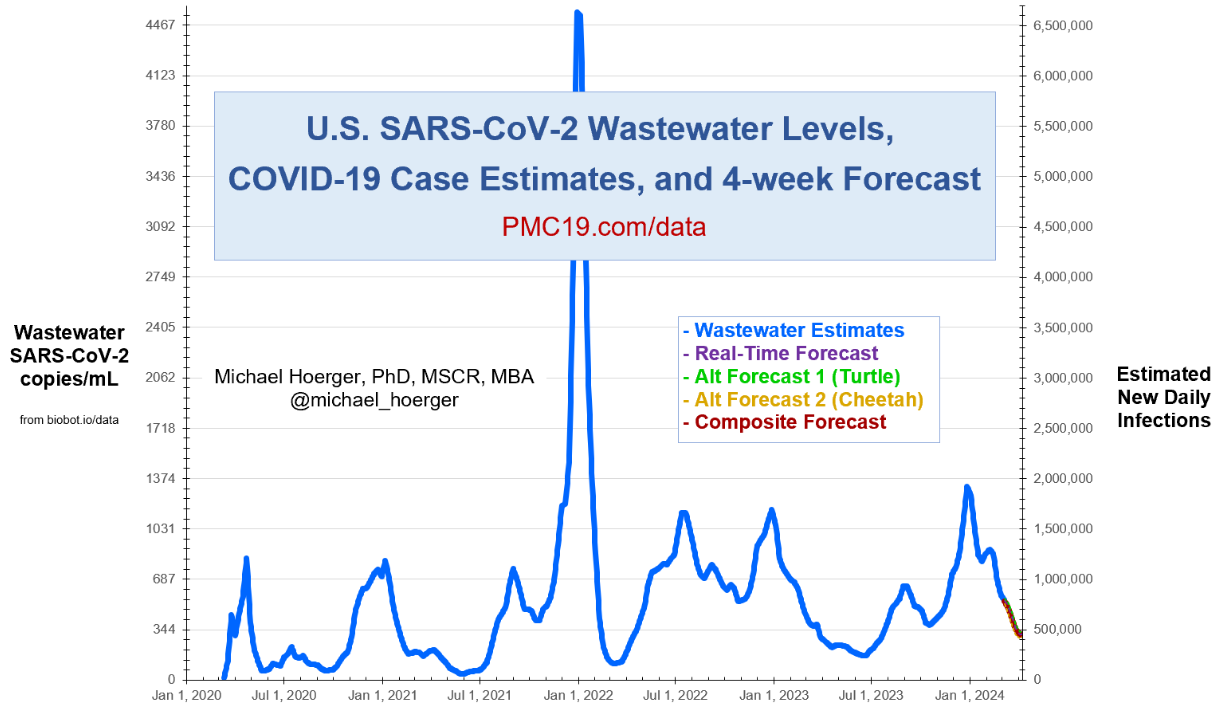
Overview of the Current State of the U.S. COVID-19 Pandemic

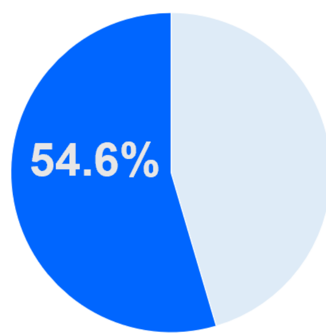
We are approaching the tail end of the 8th U.S. COVID wave and 2nd biggest all-time.

Transmission is declining, though slower than in historical trends. Transmission is now higher than during 54.6% of the pandemic (only marginally better than last week). Transmission is lower than during 45.4% of the pandemic.

- 1.54% (1 in 65) are infectious
- >730,000 COVID cases/day
- >37,000 Long Covid cases/day

These are troubling figures as we experience the 4th anniversary of the WHO pandemic declaration.





There is more COVID-19 transmission today than during 54.6% of the pandemic.

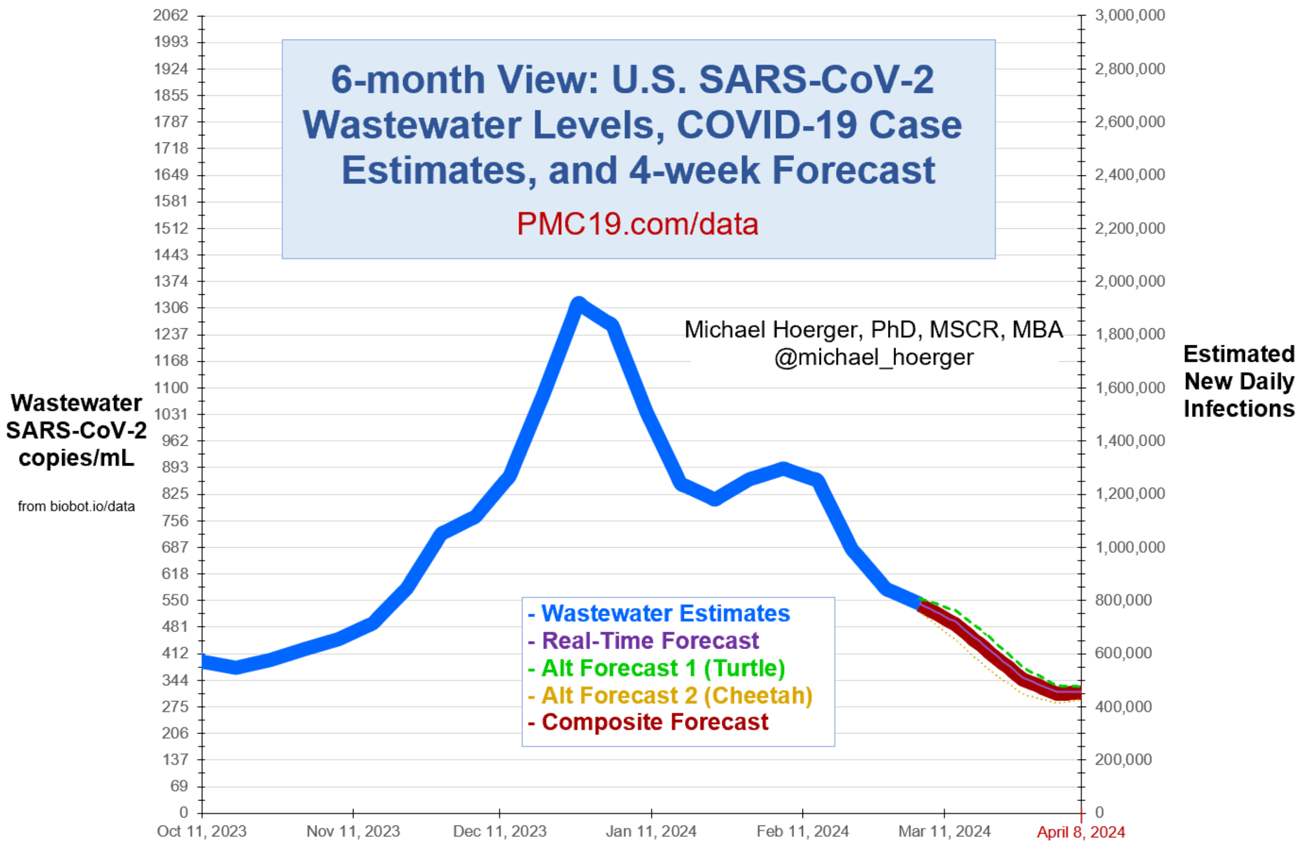
CURRENT ESTIMATES FOR March 11, 2024
Wastewater Levels (copies/mL)
504
New Daily Cases
734,000
% of Population Infectious
1.54% (1 in 65 people)
New Daily Long COVID Cases
37,000 to 147,000

WEEKLY ESTIMATES FOR March 11, 2024
New Weekly Cases
5,100,000
New Weekly Long COVID Cases
257,000 to 1,028,000

2024 CUMULATIVE ESTIMATES AS OF March 11, 2024
Total 2024 Cases To Date
84,968,250
Total 2024 Long COVID Cases To Date
4,248,000 to 16,994,000

Forecast for the Next Month

Over the next month, we should see transmission fall from 740,000 infections/day toward more like a range of 400,000-500,000 infections per day, depending on post-hoc corrections by Biobot and better or worse scenarios across the different models, though the models largely agree on the next month's course (see Technical Notes to understand the different models). In a month, we will likely see <1% of the population actively infectious on any given day for the first time since approximately July 24, 2023. In the spring, the forecast suggests more and more that we will hover around 400,000-600,000 daily infections, with the possibility of a small hill in May. The low point of the summer will likely be worse than the low point in prior summers of the pandemic.



CURRENT ESTIMATES FOR March 11, 2024

Wastewater Levels (copies/mL)
504

New Daily Cases
734,000

% of Population Infectious
1.54% (1 in 65 people)

New Daily Long COVID Cases
37,000 to 147,000

4-WEEK FORECAST FOR April 8, 2024

Wastewater Levels (copies/mL)
312 (-38%, lower)

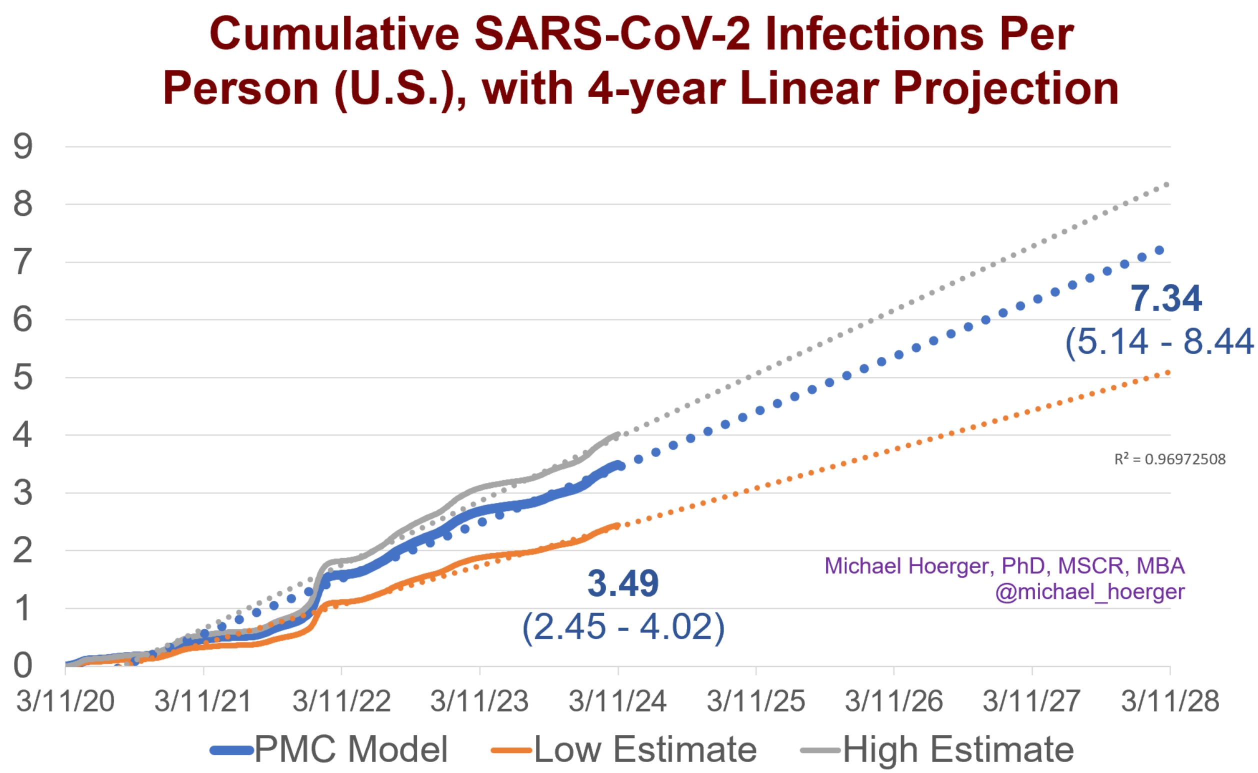
New Daily Cases
453,000

% of Population Infectious
0.95% (1 in 106 people)

New Daily Long COVID Cases
23,000 to 91,000

Four Years Since the WHO Pandemic Declaration: Looking Back and Looking Forward

At the 4-year anniversary of the WHO Pandemic Declaration, the U.S. has seen an average of 3.5 cumulative infections per person (bold blue line), with a credible interval of 2.5 to 4.0 cumulative infections per person on average (other bold lines). A linear projection suggests that IF infections continue at the current rate, in 4 years the U.S. will reach an average of 7.34 cumulative infections per person, with a credible interval of 5.1 to 8.4 cumulative infections (dotted lines).



Averages do not characterize individual variation in number of infections, which is likely substantial. These are averages, meaning that nearly half of people have had less than 3, and a little over half have likely had more than 3 infections. For each person that has had 2 infections, about as many people have had 5 infections (these are broad strokes). Moreover, individuals using multilayered mitigation are often on a much more gradual slope than individuals who are less cautious. Many people likely underestimate their number of infections due to a reduced emphasis on testing, false negatives, and asymptomatic cases. In publications on reinfections, serious long-term outcomes often start to escalate around the 3rd infection.

Note that from 2020-21, transmission was much slower (bold lines) than the overall trend (dotted lines). It took a long time to reach an average of 1 cumulative infection per person. Then, mitigation dwindled, with reduced emphasis on masking, testing, and isolation. Since the original omicron BA.1. surge, cumulative infections have increased at a relatively steady clip as public health has downplayed mitigation.

It would be a mistake to take this graph for a fatalistic interpretation that we can do nothing on COVID. It shows that transmission was low when mitigation was in place (gradual slope). Now, transmission is much higher (steeper slope). The vertical access is cumulative infections, which can only increase, not decrease. Indeed, time accounts very well (R^2=.97) for the cumulative number of infections since the cumulative total only increases. That means that 97% of the variability in cumulative infections in our world as it exists can be explained by time passing. However, that statistic could just as easily be the case under a low-transmission graph (gradual slope) as in this high transmission graph.

The PMC has long cautioned never to take projections beyond two months as having a high degree of certainty. New subvariants make forecasting challenging beyond that time frame. Over the long-term, the potential for new vaccines and treatments creates substantial uncertainty. However, a “conditional forecast,” can be used to model what the world would be like, if one assumes X, Y, and Z to be true. This is much less useful as a forecasting tool. It’s a bit like a sports pundit saying, “The Warriors will win if Steph Curry scores 50 points.” What people want is the unconditional prediction, the forecast for the game outcome without qualifying it based on obvious stats. We cannot make unconditional and accurate long-term forecasts. The conditional forecast merely tells you what we would expect IF transmission proceeded according to the status quo.

The next 4 years could be much different from this linear projection. On the one hand, the CDC recently reduced the isolation period to 1 day, when most people remain highly infectious. Many people are not keeping up with vaccines. This could lead to a steeper slope and worse situation 4 years from now. However, improvements in vaccines could cause a more gradual slope or even a non-linear increase in cases such that cumulative infections increase at a slower and slower rate. As the long-term outcomes of repeat infections become more known, there could also be a reimplementation of other types of mitigation.

Some might wonder how to reconcile this graph with those of daily transmission showing 8 pandemic waves. How is it so linear? Consider a large bowl of ice cream with 7 scoops. If the 8^{th} scoop is big or small, it’s not affecting the cumulative total as much. One might not even discern the relative change. Same for cumulative reinfections. Since BA.1, infections have accumulated at a steady relatively linear fit. You can see waves and surges at the areas where the bold lines are higher than the dotted lines and times of lower transmission the reverse. The nature of a cumulative total, however, is that it obscures a lot of the month-to-month variation.

Finally, this graph suggests that there is not credibly evidence of enduring immunity against COVID infections. If transmission were slowing, we would see a non-linear trend where the cumulative total increases more slowly and slowly. That does not seem to be the case. It could be that marginal gains in enduring immunity are being obscured by the reduction in public health mitigation, but such a hypothesis is not obviously supported in these data.

Current Risk Based on Number of Social Contacts

This figure shows the chance anyone would be infectious in a group based on group size. For example, in a group of 40-50, there's a 50% chance at least one person is infectious. In a group of 150, there's a >90% chance at least one person is infectious. These numbers are virtually identical to last week, owing to the only slight reduction in transmission.

How Does Risk Increase with More Social Contacts?			
Number of People	Chances Anyone is Infectious	Number of People	Chances Anyone is Infectious
1	1.5%	25	32.1%
2	3.0%	30	37.1%
3	4.5%	35	41.8%
4	6.0%	40	46.2%
5	7.4%	50	53.9%
6	8.9%	75	68.7%
7	10.3%	100	78.7%
8	11.6%	150	90.2%
9	13.0%	200	95.5%
10	14.3%	300	99.0%
15	20.7%	400	99.8%
20	26.6%	500	>99.9%

Trustworthy Health Evidence

Transparency

The pandemic has reminded us that trust is essential in public health, and we have made every effort to provide transparency. Every PMC report remains posted publicly. A methodology video describes the gist of the analytic approaches used, with enough detail that others can construct similar models. Dr. Hoerger's biography and publication record are publicly available. Many of those making similar dashboards and forecasts work in private industry or wish to avoid scrutiny so do not post information publicly or do so using anonymous accounts or pseudonyms.

Lack of Conflicts of Interest

These reports are written by Dr. Hoerger who has no financial conflicts of interest. This means no ownership of equity stakes or individual stocks. Dr. Hoerger actively avoids funding from the pharmaceutical industry and other private corporations. He has no investment in mask, HEPA, or other mitigation companies. As a scientist, he is required to complete financial forms as a part of being a university professor as well as when publishing at top medical journals, like JAMA Oncology. An underappreciated aspect of financial conflicts of interest is that clinicians and scientists sometimes gain substantial revenue from social media advertisements (e.g., YouTube, Twitter) and newsletter subscriptions (e.g., Substack, Patreon); this can lead to significance "audience capture" in which the scientist gains increasing financial remuneration for telling an audience what they want to hear, and this is a huge problem among those posting disinformation online. Dr. Hoerger intentionally gains no revenue through this website, social media, nor other platforms, as doing so would undermine his scientific credibility. The closest PMC collaborators are the co-authors and co-investigators noted in the various projects and publications on the PMC homepage, and there are no noteworthy conflicts of interest there either.

Documented Evidence of Accuracy

The PMC Forecasting model can be conceptualized as "participatory action research," meaning that it is public health activism first, iterates and improves with feedback, and is research second. An alternative approach would be to publish a model first and then implement, but that would be a horribly unsuccessful strategy. Using historical data, our forecasting model explains 96% of the variance in 1-week forecasts, which is exceptional. However, implementing forecasts in real-time adds new challenges. Wastewater levels are reported with inconsistent cadence and often corrected retroactively, so forecasting in real-time is more challenging, meaning that real-world estimates of accuracy are more important than those derived from a historical database. Moreover, it has been a learning experience understanding what metrics matter most to the public, and these metrics vary across different points of the pandemic. Accordingly, all reports are public so that users can evaluate accuracy themselves in real-time. However, a scientific report will ultimately document accuracy. Accuracy levels will guide a 2.0 model of the forecast likely to launch in spring or summer of 2024, which may include the same or different alternative forecasts, and add confidence intervals based on historic levels of accuracy.

A useful way to evaluate accuracy is to take one of our public reports and then compare it with what unfolds four weeks later. We believe you will be impressed with the results and that any discrepancies are minor and not of public health significance, given that much of the public is magnitudes off in their subjective estimates of transmission. As an example, in our December 25 report, we noted that our first-posted Christmas Day risk table (posted October 30, 2023, with that report cited in JAMA Oncology) was quite accurate in characterizing the eventual risk on Christmas Day (noted in the December 25, 2023 report, though Biobot may retroactively adjust wastewater levels marginally).

Posted on December 25

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

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

Negligible differences between the long-range forecast and today's table. In a group of 20, for example, the forecasted risk was 47%, and today we estimate it at 50.6%. Hopefully this helped with planning.

Documented Evidence of Validity Relative to Other Data Sources

The public has genuine reasons to ask questions about wastewater-derived case estimates and forecasts. These concerns fall into two main themes. One, much of the general public has observed how important metrics (reported case counts, positivity ratio) have gone from valuable to fallible as testing has declined and largely switched from reported PCR tests to unreported home rapid tests. Two, with each new subvariant, some also wonder whether wastewater data could over or underestimate transmission. With regard to the fallibility of testing-based data, a tremendous benefit of wastewater is that it does not rely on self-report. With regard to subvariants, it is important to know that the process of converting raw wastewater into useful, normalized metrics of viral levels is oversee by wastewater scientists with specialized expertise in addressing our most obvious concerns (e.g., new subvariants, regional differences in demographics, changes in rain water). It is a good example of the Dunning-Kruger Effect for anyone to assume they have more expertise than the scientific experts who translate sewage into viral level estimates, but our public health infrastructure fosters much skepticism. Moreover, by outsourcing wastewater measurement to third parties who must compete with one another, methodology surrounding wastewater quantification remains proprietary to protect competitive interests, and this reduces transparency and trust. Fortunately, there are two easy checks on wastewater validity.

Validity Check #1: Independent Cross-Cultural Comparisons. If the PMC model provides a useful estimate of cases, it should converge with other independent estimates. Unfortunately, other U.S. based models are private or anonymous, so not appropriate comparisons. Nonetheless, there are high-quality estimates outside the U.S. For example, Dr. Moriarty's lab uses wastewater data in Canada to make similar estimates, and the U.K. runs a strong testing surveillance program to assess case rates. Despite being from independent groups, using slightly to considerably different methodology, and in different countries, all of the estimates are within the same ballpark and have followed a similar pattern.

International COVID Statistics from High-Quality Data Sources

Compiled by the Pandemic Mitigation Collaborative (PMC19.com), March 11, 2024

	U.S.A.	Canada	U.K.
% of Population Actively Infectious	**1.5% (1 in 65)**	**2.1% (1 in 47)**	**0.9% (1 in 111)**
Chances Someone is Infectious…			
In a Group of 10	**14.3%**	**19.3%**	**8.6%**
In a Group of 30	**37.1%**	**47.5%**	**23.8%**
In a Group of 50	**53.9%**	**65.9%**	**36.4%**
Primary Data Source	Wastewater	Wastewater	Surveillance Testing
Reference	Michael Hoerger, PhD, MSCR, MBA @michael_hoerger pmc19.com/data	Tara Moriarty, PhD @MoriartyLab covid19resources.ca	Alex Glaser, PhD UK Health Service Agency tinyurl.com/pmcukhsa
Last Updated	March 11, 2024	February 17, 2024	February 21, 2024

Validity Check #2: Examining Consistency Relative to Other Metrics. For the past 18+ months, the relationship between self-reported cases and estimates of true cases has remained remarkably consistent across various indicators of each variable. A simple heuristic is that there are 25 actual cases for every 1 self-reported case tracked in federal statistics. This fluctuates a bit depending on whether in a lull, rising surge, or falling surge, often in the 15-30x range. Simply review IHME self-reported versus actual case data, available through April 1, 2023 (https://covid19.healthdata.org/united-states-of-america?view=infections-testing&tab=trend&test=infections). The same is true for our wastewater-derived estimates of true cases, relative to self-reports of actual cases provided by states and aggregated nationally by BNO News (https://twitter.com/BNONews) and others. If this relationship were to decouple suddenly, such that estimates of true cases were 100x or 1,000x that of self-reported cases, that would suggest that a particular subvariant may be disrupting wastewater estimates in ways that were outsmarting wastewater scientists, but there is no evidence for such a claim at present.

PMC Trusted in Peer-Reviewed Scientific Journals, News Media, and Funded Grant Applications

Examples:

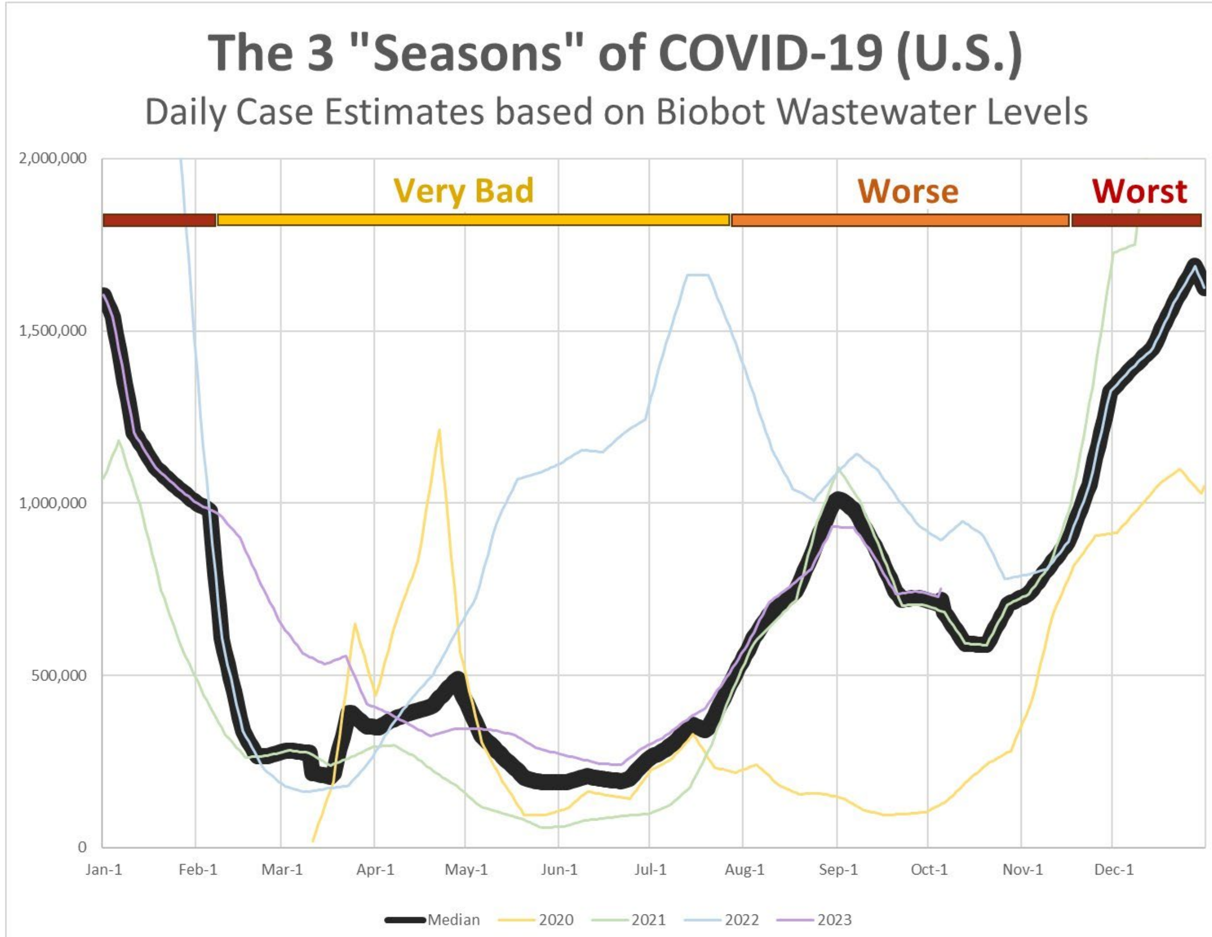
- JAMA Oncology: https://jamanetwork.com/journals/jamaoncology/fullarticle/2813585
- BMC Public Health: https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-023-16787-1
- TODAY: https://www.today.com/health/news/covid-wave-2024-rcna132529
- Forbes: https://www.forbes.com/sites/judystone/2023/12/01/cdc-improves-their-covid-19-reporting-with-a-new-wastewater-dashboard
- Salon: https://www.salon.com/2023/10/19/a-lapse-in-wastewater-detection-is-worrying-scientists-about-distorted-data/
- The New Republic: https://newrepublic.com/article/177849/biden-democrats-covid-pandemic-2024
- Yahoo! News: https://news.yahoo.com/us-sees-largest-covid-wave-165811252.html
- Washington Post: https://www.washingtonpost.com/health/2024/01/12/covid-surge-january-2024/
- Time: https://time.com/6554340/covid-19-surge-2024/
- Stateline: https://stateline.org/2024/01/23/wastewater-tests-show-covid-infections-surging-but-pandemic-fatigue-limits-precautions/
- PRISM: https://prismreports.org/2024/01/29/covid-surges-senate-hearing-california/
- SELF Magazine: https://www.self.com/story/cdc-new-covid-19-isolation-guidelines
- Grants funded intramurally by Tulane University and externally by the American Cancer Society, and an additional grant submission with PCORI under review

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 COVID-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.