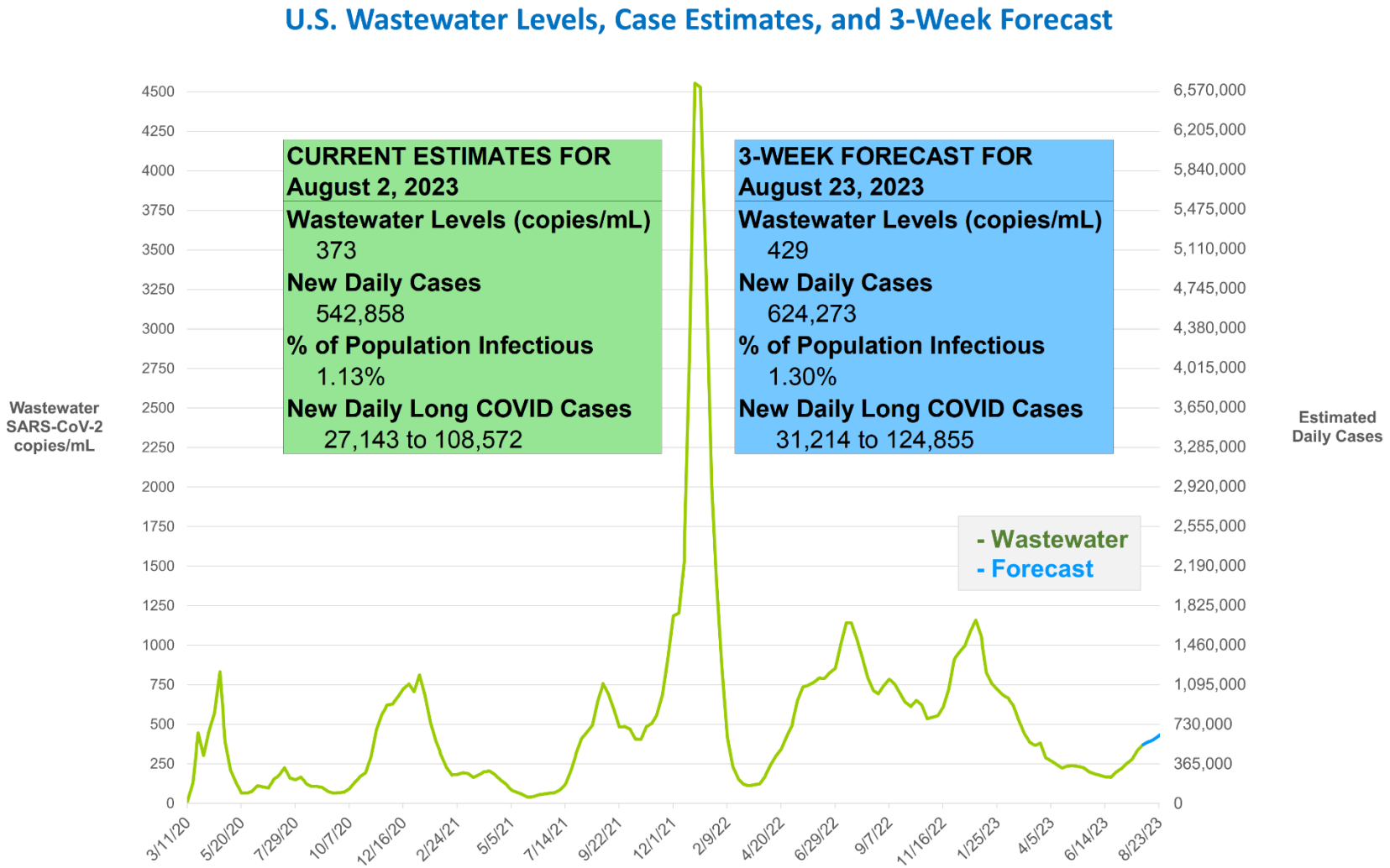
Michael Hoerger, PhD, MSCR, MBA, Pandemic Mitigation Collaborative
U.S. SARS-CoV-2 Wastewater Levels, COVID-19 Cases, and 3-Week Forecast: Report for August 2, 2023, pmc19.com/data



Cite as: Hoerger, M. (2023, August 2). *U.S. SARS-CoV-2 wastewater levels, COVID-19 cases, and 3-week forecast: Report for August 2, 2023*. Pandemic Mitigation Collaborative. http://www.pmc19.com/data

Informal Commentary:

U.S. Pandemic:

- Wastewater: 373 copies/mL
- Estimated new daily cases: 543,000
- % of population infectious: 1.1%
- New daily Long C0VID cases: 27,000-109,000
- 3-week forecast: 15% worse than today

How bad is it?
Wastewater levels are now higher than during 48.9% of the pandemic, lower than during 51.1% of the pandemic. 51% is an F grade, in my view. It's a typical day in the pandemic, except many are behaving as though transmission is lower than 95-99% of the pandemic (A or A+ grade), that the pandemic is "over." Instead, we're in the thick of it.

Are we in a surge?
We are currently in a very bad place, and I forecast it will be worse in 3 weeks (blue line). I like this definition of a surge: to rise suddenly to an excessive or abnormal value, as with a new highly immune-evasive variant. Contrary to "surge" enthusiasts, I see no evidence in the data of a surge, simply a steady, predictable, school-induced increase on the way, likely for August and September, followed by a dip (not shown), then high levels from November-January. There is every reason to increase universal precautions (mask utilization, mask quality and fit, indoor air cleaning, remote meetings, testing, boosters, contact reduction). I encourage accurate, balanced framing so people are more cautious and can tolerate avoiding denial.

How does this compare with other models?
On Twitter, @JPWeiland also estimates cases. I estimated cases from wastewater in several ways. My model estimates somewhat higher cases than his. However, about 1 in 7 of my models produced estimates similar to JPW. You'll note that our estimates of the % of the population who is infectious (slightly >1%) are eerily similar. I think he estimates the infectious window at 10 days, whereas I prefer 7 days, so we alternate our high and low estimates. All of this is in the ballpark of what I would anticipate from independent estimates.

How do you estimate cases?
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. I compared wastewater with their case estimates on the 1st of each month. The correlation was r=.94. 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 also reasonable. 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). Elegant is good.

How does the forecast work?
It's a reasonably simple forecast. It incorporates data based on the month and each of the prior 4 weeks. Believe it or not, "month" remains a pretty good predictor, as I have posted previously. On average (or median), cases are higher from August through January, and lower in the other months,

with a hill in August-September and a mountain from November through January. Note, predictable is different from “seasonal,” unless you believe in 6-month seasons. The model also incorporates data from 7, 14, 21, and 28 days ago. The 7 day estimate is very good, and the other recent estimates get at how steeply transmission is rising or falling. This is good. The model will update each week and quickly incorporate emerging data, so if there’s a potential surge starting in the U.S., we will have a little more of a warning signal. I’m forecasting through December 2025 already (not a typo), but only showing a very limited range. I have said repeatedly throughout the pandemic that there is too much uncertainty to believe anyone who says they know what the state of the world will be in 2 months.

What are the limitations of these estimates?
1) Nobody knows the true number of daily cases, but these estimates are useful in that cases are associated with LC and other long-term health complications. They document the significance of the ongoing pandemic. 2) It’s very difficult to predict the onset of a genuine surge a few weeks in advance using these data. More granular data, such as sub-variant level estimates, might better inform these models. Behavioral data could inform models. Whereas we used news headlines (e.g., BA.1 surge in India) to predict the state of the U.S. weeks later (e.g., Project Bandura), international variation in exposure histories is too varied for me to see this as highly useful anymore. 3) I do not model hospitalizations. One, most U.S. hospital CEOs have dropped masking and testing requirements, so we are in a state where the people reporting hospital statistics are likely underestimating and undercounting C19. Two, in my view, after rapidly killing the most acutely vulnerable, C19 is slowly killing many more through a chronic disease model, similar to organ failure models of death trajectory (long roller coaster where any dip in functioning could = death), so hospitalizations would be a red flag, but mostly aren’t relevant to 6-month or 15-year death trajectories. 4) I do not model excess deaths, at least yet. As I have mentioned many times, modeling excess deaths is extremely complicated, and most models vastly underestimate excess deaths due to poor underlying assumptions (survivor bias that doesn’t adjust for age or comorbidities, using 2020-2022 data to casually predict expected deaths in a simplistic and biased fashion, etc.). Seeing flat mortality rates should deeply disturb everyone after C19 has killed off many of the most vulnerable people.

What are the main strengths of this model?
1) It puts everything in context. Everyone should have these data to have a general sense of how good or bad things are, how today compares to the Delta wave, or when to escalate precautions. 2) Panic button. This model is not perfect but it will give you solid evidence with 1-2 weeks notice if we are headed into a wave or surge. That matters. It could mean cancelling a trip or surgery. 3) Beating the market. If you are C0VID cautious, you want to engage in mandatory riskier activities (e.g., dental visit, throat exam, colonoscopy) when safest. As a psychologist, the key is to do so at a low-transmission time period (good) immediately after a high-transmission time period (when people have not yet dropped precautions or dropped tolerance of precautions). Think late February, but the model will guide us.

Are you qualified to make this model?
Yes. I wish we had 100 experts smarter than me modeling and making national and worldwide dashboards. We don’t have that. I use these data to provide guidance to people I care about, cite in articles, and keep patients safe in clinical research studies. Each of my degrees had an analytics emphasis, and my MBA focused on financial analytics, which included modeling C0VID mortality data and very similar models with arbitrarily different outcomes of interest (e.g., thrillers like predictors of energy usage). Someone more experienced in modeling could do 3% better on wastewater level prediction. They probably are privately, for $$$$$, not here. I did some modeling of positivity ratios earlier in the pandemic, but the data were too volatile and low quality to provide helpful guidance. Wastewater data are very strong. “Summer colds” lie. Poop doesn’t.

What else would help?
I’d like to know what information would be most helpful to you. Do you care about the total number of cases? % of the population infectious? Something else? Do you have a compelling article or database suggesting something I assumed that could be more accurate? Are there adjectives, terms, or colors that would help signal to you to do something different? Is there a useful statistic I should add?