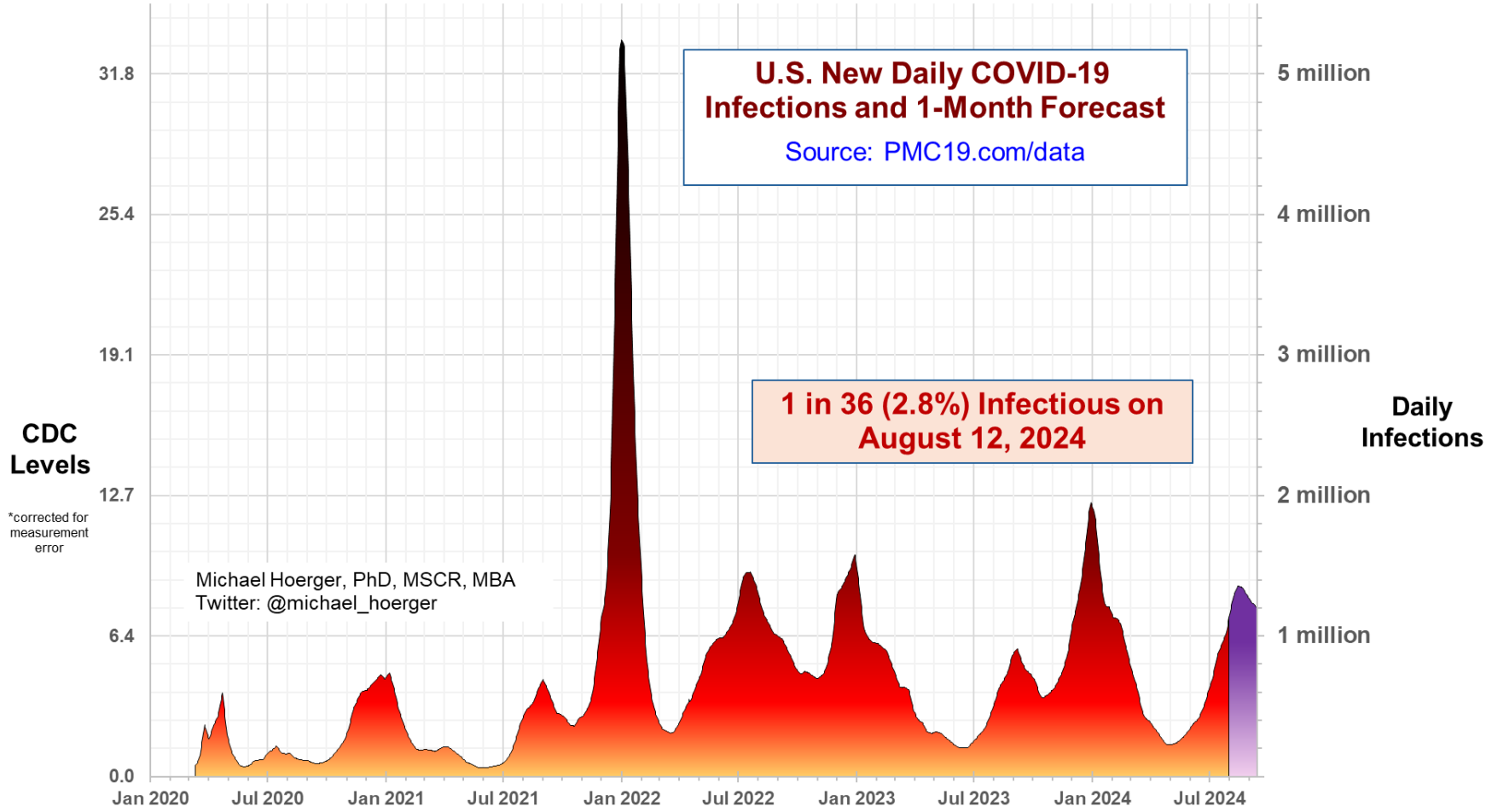
# PMC U.S. COVID-19 Case Estimation and Forecasting Model: Report for August 12, 2024, pmc19.com/data

*Michael Hoerger, PhD, MSCR, MBA, Pandemic Mitigation Collaborative (PMC)*




Cite as: Hoerger, M. (2024, Aug 12). *PMC U.S. COVID-19 Case Estimation and Forecasting Model: Report for August 12, 2024*. Pandemic Mitigation Collaborative. http://www.pmc19.com/data

# Welcome to v2.0 of the PMC Model

## Over 1.3 million people in the U.S. are getting COVID-19 each day during this 9th wave of the ongoing pandemic. Expect >1 million infections per day most of the remainder of the year. After 1 year of public forecasting, we are pleased to offer version 2.0 of the Pandemic Mitigation Collaborative (PMC) case estimation and forecasting model.

Welcome to version 2.0 of the PMC Model. The “C” in PMC is for Collaborative, and the work to improve this model is grounded in feedback from readers like you over the past year. Thank you for your support.

### What’s New?

In short, the new model has substantial data quality improvements by combining multiple data sources for estimating transmission in unique ways that will hopefully increase forecasting accuracy, provide a truer representation of what has happened and is happening during the pandemic, and linkages to some statistics you will find helpful in day-to-day decision making.

Here is a deeper dive into the changes (skip to next section if desired). The new model is designed to provide a “true” picture of what has happened during the pandemic. It integrates three main data sources: the IHME true case estimation model, Biobot SARS-CoV-2 wastewater surveillance data, and the current CDC NWSS SARS-CoV-2 wastewater data. IHME provided a comprehensive case estimation model through April 1, 2023. Biobot was the CDC wastewater subcontractor through last fall and continues to do extensive non-CDC wastewater work. The CDC NWSS data are currently subcontracted with Verily, a subsidiary of Alphabet, which is the parent company of Google. Over the past year, we have seen Biobot scale back their public data and visualizations, and Verily has made steady improvements in their work with the CDC.

We previously relied solely on Biobot for forecasting and a Biobot-IHME data linkage for case estimation. It was a Biobot-heavy model. The current model is not tied strictly to any data set, but rather the PMC’s best estimate of the truth, a true-case model that uses multiple data sources in the spirit of IHME’s original work in this area. Essentially, we link all three data sources, which have been active over different points of the pandemic to derive a composite

“PMC” indicator of true levels of transmission. The indicator is weighted based on which data sources were available and their perceived quality at each point in time. We scale this composite PMC indicator to the metric the CDC uses when helpful for comparisons with their website, and scale it with the true case estimates of the IHME otherwise, as true cases are more relevant than arbitrary wastewater metrics.

A great feature of the model is that it continues to integrate real-time data from Biobot and the CDC. From the perspective of Classical Test Theory, this is a huge advantage, as it provides a much more reliable indicator of what is currently happening with transmission. Both sources often make retroactive corrections for the most recent week’s data, sometimes sizable, and pitting the two indicators against one another reduces measurement error on average, which offers vital improvements in forecasting.

**What are the Biggest Improvements in the Model?**

- ***Accuracy in Real-Time Data*** – In integrating two active surveillance data sources, the real-time data will be more accurate. The biggest predictor of next week’s transmission levels, and the shape of how transmission is increasing or decreasing, accelerating or decelerating, is the current week’s real-time data. If the real-time data are off by 5% or 10%, the big-picture take on the forecast will still be reasonable, but a more precise estimate allows for greater accuracy in estimating the height and timing of waves.
- ***Regional Statistics*** – We are already integrating some regional data. Like you, we miss the vast and high-quality regional data and visualizations Biobot provided. We are hoping to take back some of those advantages through the new model and will improve them over time.
- ***Credibility*** – Although Biobot and CDC have unique strengths and limitations, a clear strength of adding the current CDC data set is that many people prefer to defer to the credibility of the CDC. The PMC model can be characterized fairly as a “CDC-derived case estimation and forecasting model,” which should lend more credence with those who are not deep enough in the weeds to evaluate the data as critically and prefer appeals to authority. We also provide some statistics that will allow you to draw more useful inferences from the CDC website.

## What's the Same in the Current Model?

The analytic assumptions underlying the forecasting model remain the same. It uses regression-based techniques common across all industries, using a combination of historic data (median levels of transmission for each day of the year) and emerging data from the past four weeks to characterize how transmission is growing or shrinking. Holidays and routine patterns of behavior that map on well to a calendar are "baked in" to the historic data. "New variants" and atypical patterns of behavior are baked into the data on recent patterns of transmission. It's a top-down big picture model.

## What are the Biggest Drawbacks of the New Model?

- ***Disruptions in Longitudinal Comparisons*** – You will notice some inconsistencies between the current and prior model that use additional data to form more accurate estimates, which is sometimes frustrating. A few examples. In the early pandemic, we estimated cases linking Biobot to IHME case estimates. Biobot transmission estimates were a bit "hotter" than others during that time period, the IHME estimates "cooler." Our composite model depicts each of the first 4 waves somewhat smaller, which we believe provides a better picture of the "truth" as we can estimate it, but it is annoying psychologically to re-envision what has happened. This also throws off some of the big-picture statistics; for example, as of August 12, 2024, we estimate that Americans have had about 3.3 infections on average. A few months ago, we estimated nearly 3.5, so this is consistent with "cooler" picture of early-pandemic transmission. Presently, the CDC transmission estimates are running much hotter than those of Biobot, leading to estimates of a larger and earlier peak in the present wave. We would have preferred the CDC re-up with Biobot at the potential contract renewal to promote continuity in the data, but these sorts of changes in model estimation are the expected consequences of such a transition.
- ***Constantly-updating Historical Data*** – The CDC updates all of their historical estimates of transmission frequently, any time a new site comes on board, and twice annually to standardize the data longitudinally. This can sometimes create weird issues, where transmission is going up, but real-time values are lower than what was reported in real time the prior week because recent data were corrected downward. It will also throw off some of the helpful statistics we provide. These are minor nuisances, but be aware of them in case you spot something that seems strange.

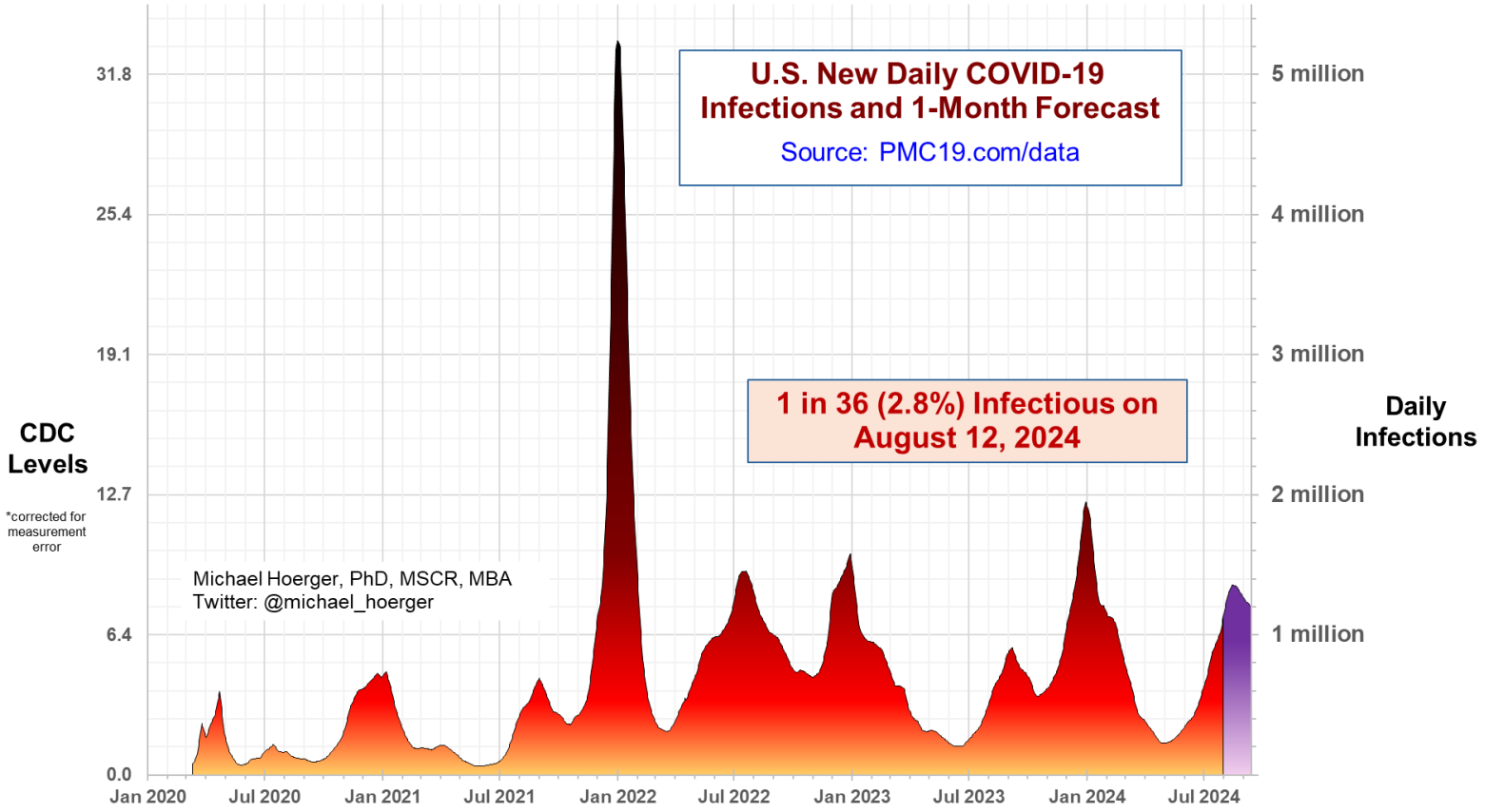
- ***Documentation of Accuracy*** – We have excellent data on the accuracy of the prior model and will submit a report for publication shortly. All prior reports are publicly available. Many report quick facts on longitudinal accuracy, international comparisons, use in news articles, and references to use in peer-reviewed scientific journal articles. We cannot document the real-time accuracy of the new model yet, but know that when using historical data, the model accounts for 98% of the variability in wastewater transmission 1-week into the future, which is 2% higher than our prior model. The vast majority of forecasting errors have been and will continue to be based on inaccuracies in the real-time data wastewater surveillance companies report, and the model changes reduce those issues. We hope you will trust our history and that the methodologic changes represent improvements.

## What Improvements Should We Expect in the Future?

There are many improvements we hope to roll out in the future. These include changes based on your feedback, the addition of confidence intervals in some of the graphs, and regional forecasting models. We may incorporate additional data sets if they can improve real-time estimates of current transmission.

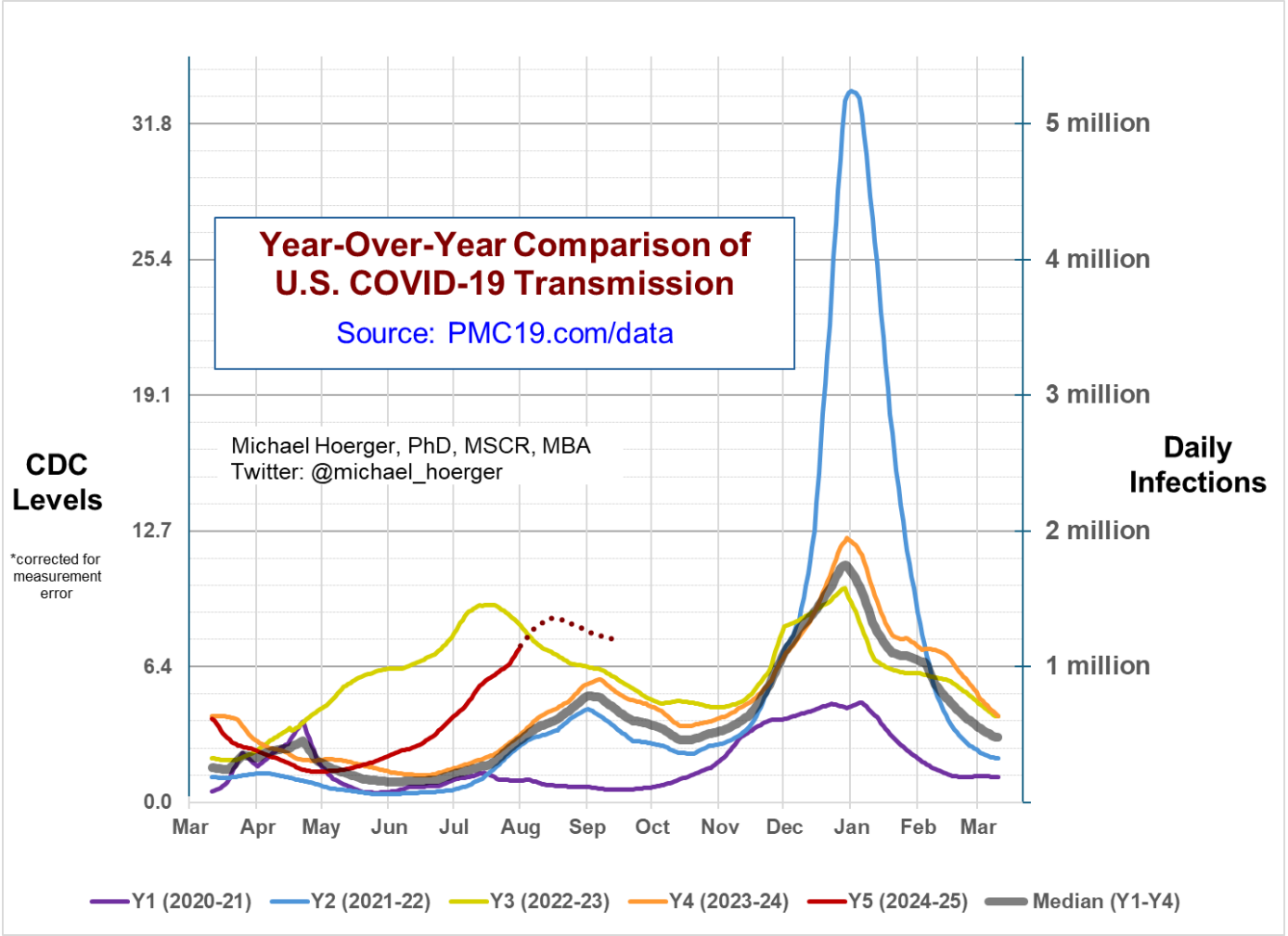
# The Big-Picture View of the Pandemic

We are well into the 9th wave of the pandemic, with approximately 1.3 million infections per day. Based on the statistical criteria we use to differentiate a wave from a larger "surge," we are possibly in surge territory. Note that the CDC data are running much hotter than the Biobot data. Our Biobot-based analyses still suggest a peak in early September. The CDC data show much higher transmission and suggest a peak nationally late this week. Relative to our prior model, note that the earlier waves of the pandemic appear smaller when incorporating multiple data sources. For August 12, we estimate that 1 in 36 Americans (2.8%) are actively infectious with COVID-19. Wastewater data do have a reporting lag, so today's estimate is actually a slight forecast into the data's future. The purple hill represents the current forecast, and imagine that we're walking up the purple hill and are already a few steps from the peak. This assumes reasonable accuracy in real-time wastewater reports, so the possibility of a higher or later peak remains, and different locations will peak at different times.



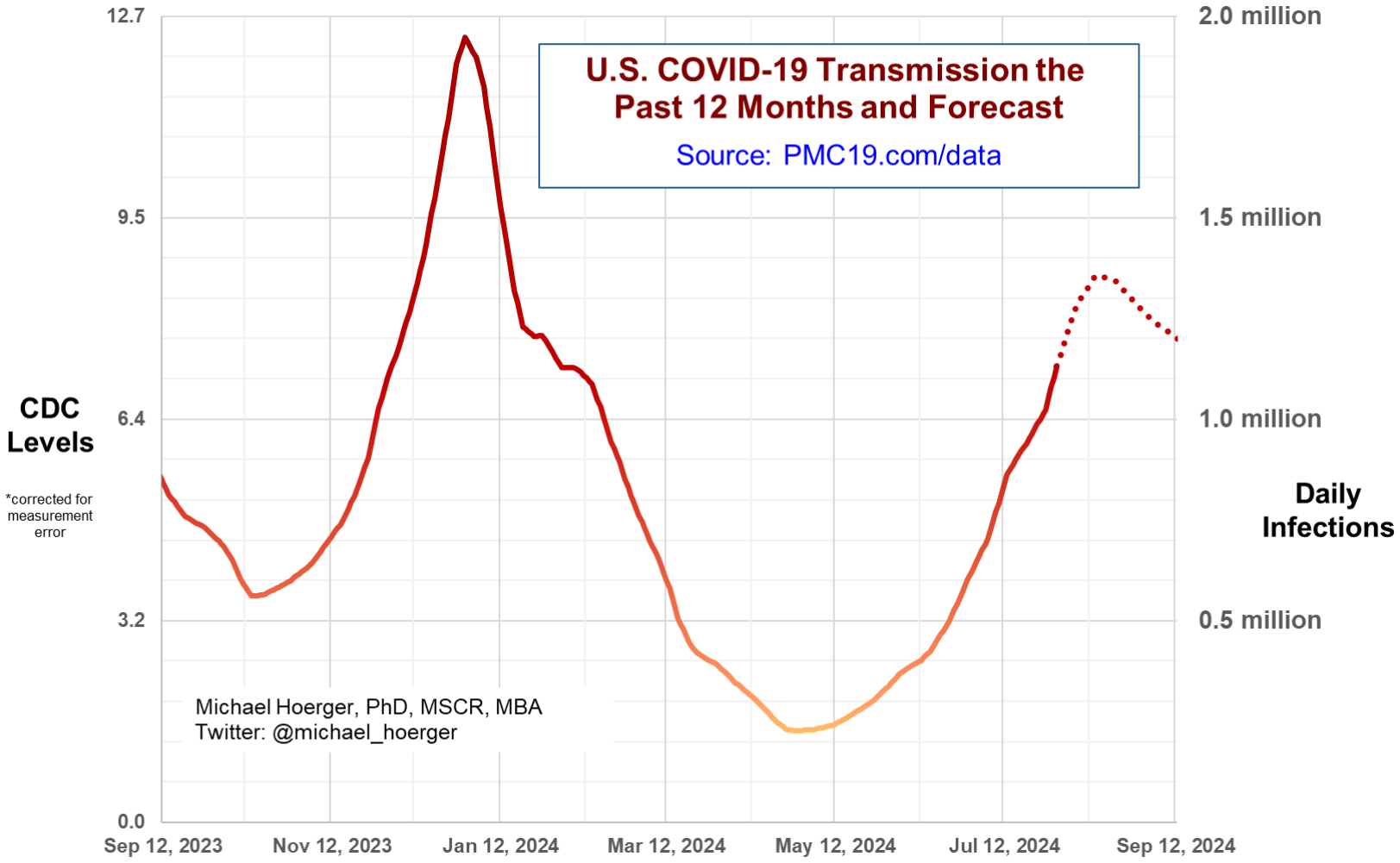
# Year-Over-Year Comparisons

We missed the elegant year-over-year graph that Biobot used to supply, and with it gone, we have provided a PMC version. This is very helpful for comparing the same time point of different years without having to get out a ruler. Note, we are near-equal to our worst summer wave (2022). Our model suggests a slightly lower peak, but we will not know that for sure until the peak is in hindsight. This is the highest level of transmission at the time of schools starting, so expect K-12 schools and universities to be hotbeds for COVID outbreaks unless they are using multilayered mitigation like indoor air quality that meets ASHRAE Standard 241 (if they have never heard of this or cannot explain how they are meeting the standard, they likely are not meeting the standard), surveillance testing, free on-demand testing, and universal masking. This is uncharted territory in terms of such low mitigation coupled with high transmission with school starting.



# Close-up on the Current Forecast

This is our most granular graph of the current wave. The light gray vertical line near the peak of the wave shows August 12. It indicates that we are likely near the peak of the wave, unless the unprecedented context of back-to-school with no emphasis on mitigation pushes transmission higher in ways the model cannot predict statistically. We anticipate that transmission will be at over 1 million infections per day in the U.S. for most of the rest of the year.



# Supplemental Statistics

These supplemental statistics may prove useful in conversations about transmission and mitigation. To highlight a few, 1 in 36 people are actively infectious, with >1.3 million infections/day. Over the next month, expect a similar average for both statistics, as we venture a little higher, and then hold steady only marginally lower. In a school classroom of 25-30 people, it should be assumed that someone (>50% chance) has infectious COVID. Transmission is higher than during 91% of the pandemic, lower than just 9% of pandemic days. The impact on potential Long COVID cases the next month will be staggering, and the rest of 2024 will be worse.

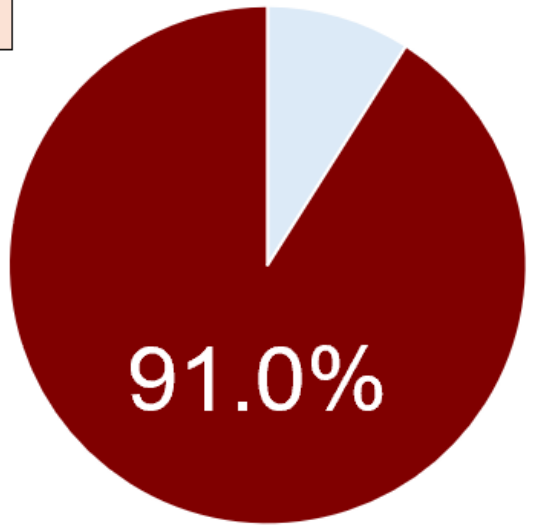
| Current Levels for Aug 12, 2024 |
|---|
| % of the Population Infectious |
| 2.8% (1 in 36) |
| New Daily Infections |
| 1,336,000 |
| New Weekly Infections |
| 9,352,000 |
| Resulting Weekly Long COVID Cases |
| 468,000 to 1,870,000 |

| Monthly Forecast |
|---|
| Average % of the Population Infectious |
| 2.7% (1 in 37) |
| Average New Daily Infections |
| 1,295,900 |
| New Infections During the Next Month |
| 38,877,000 |
| Resulting Monthly Long COVID Cases |
| 1,944,000 to 7,775,000 |

| Running Totals |
|---|
| Infections Nationwide in 2024 |
| 164,431,000 |
| Average Number of Infections Per Person All-Time, U.S. |
| 3.27 |

**How Does Risk Increase with More Social Contacts?**

| Number of People | Chances Anyone Is Infectious | Number of People | Chances Anyone Is Infectious |
|---|---|---|---|
| 1 | 2.8% | 15 | 34.6% |
| 2 | 5.5% | 20 | 43.3% |
| 3 | 8.2% | 25 | 50.8% |
| 4 | 10.7% | 30 | 57.3% |
| 5 | 13.2% | 35 | 62.9% |
| 6 | 15.6% | 40 | 67.8% |
| 7 | 18.0% | 50 | 75.8% |
| 8 | 20.3% | 75 | 88.1% |
| 9 | 22.5% | 100 | 94.1% |
| 10 | 24.7% | 300 | 99.9% |



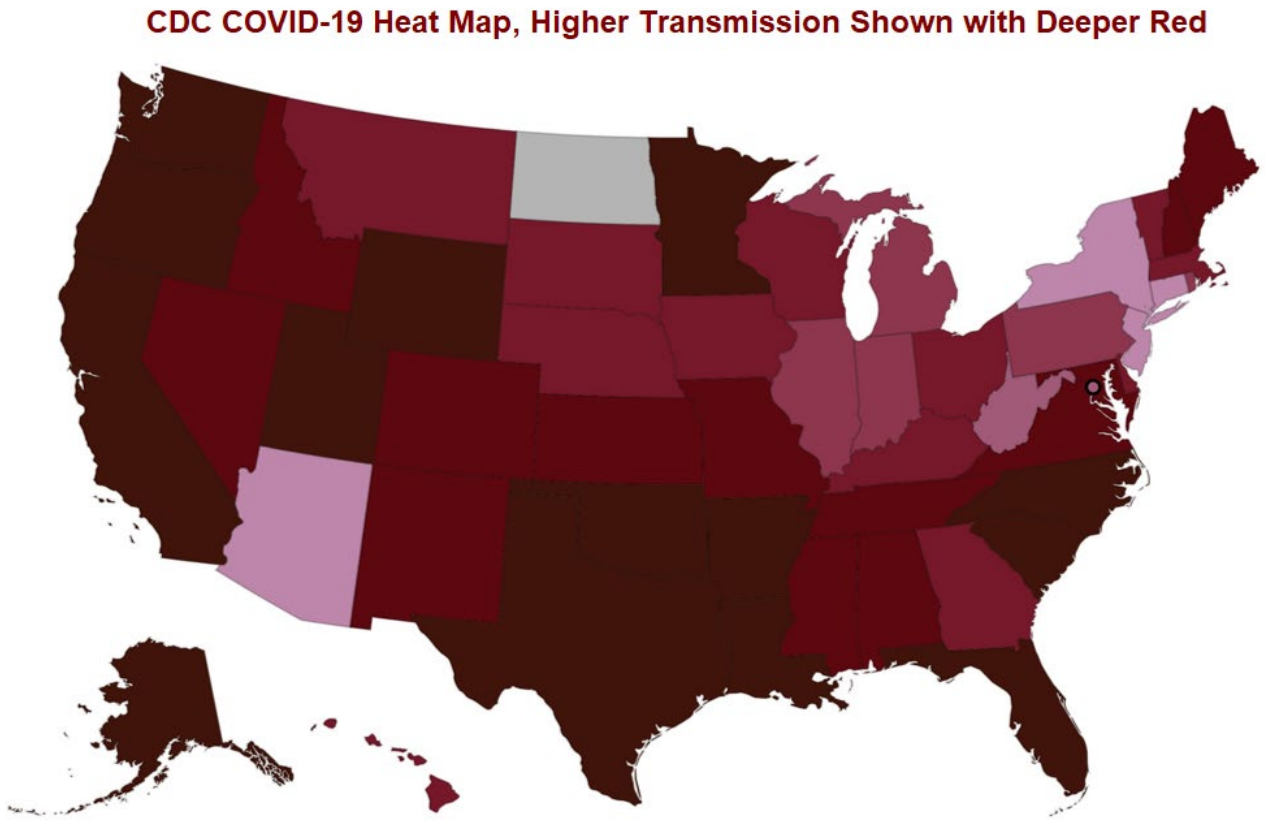

**There is more COVID-19 transmission today than during 91% of the pandemic.**

pmc19.com/data

Michael Hoerger, PhD, MSCR, MBA
Twitter: @michael_hoerger

# CDC COVID-19 Heat Map

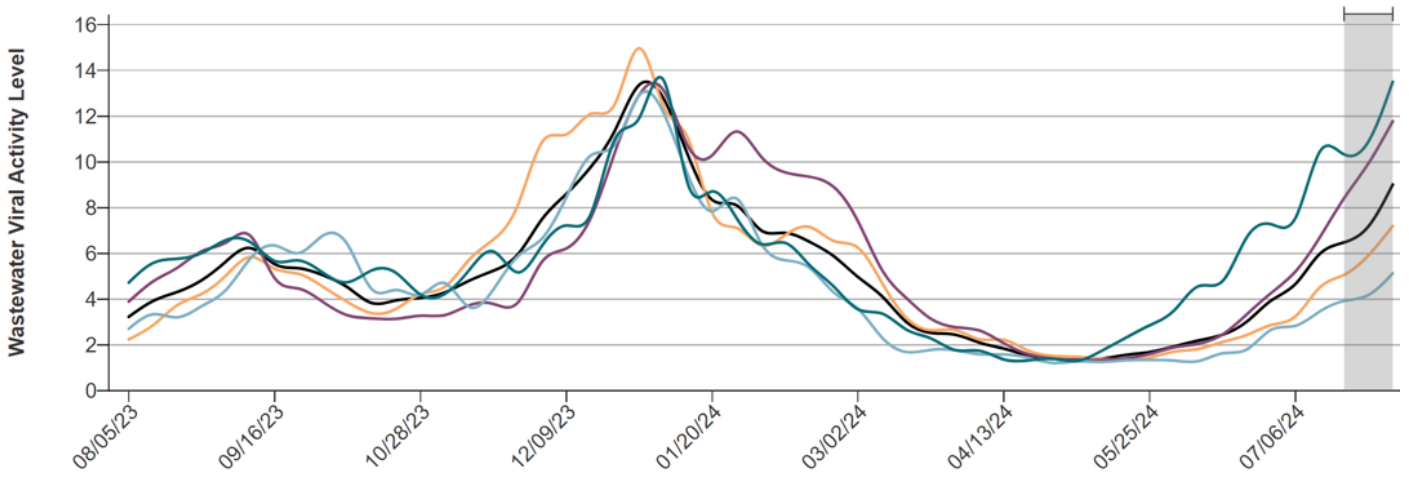
This is the CDC COVID-19 Heat Map of COVID-19 transmission with the colors scheme switched from blue to red. You can recreate it yourself by using Photoshop to shift the hue to the right by 150. There were no manipulative changes, such as altering the contrast or levels. In courses in Geographic Information Systems (GIS), the consensus is to use colors like red to indicate something is “hotter” or greater cause for concern, and blue for cooler or better outcomes. We encourage the CDC to switch to a color scheme more like this to improve messaging.



# Regional Case Estimation

This graph from the CDC shows regional variation in transmission. You can use the “PMC Regional Multiplier” to estimate transmission levels by region. Just take the transmission level reported on the CDC website and multiply it by .329. For example, if the regional level says “9,” just take 9 x .329 = 2.961. That means 2.96% of people there would be estimated to be infectious with COVID. You will note that this is very close to just taking the CDC level and dividing it by 3 – a quick shortcut! We have provided regional estimates in the table. Note that the values in the CDC website have a lag and they do not account for Biobot data suggesting levels are slightly lower than CDC, so we provide both the raw estimates (using the simple Multiplier) and our hopefully marginally more precise PMC Model estimates. These are crude indicators. Estimates around the proportion of a population get less precise when focusing on smaller geographic entities, so take them with a grain of salt. We had previously provided such conversion calculations for Biobot data, so this process is familiar to many readers. Note, the Multiplier may vary marginally from week to week as the CDC updates historical data or Biobot and CDC disagree on transmission levels, but even if using a slightly “stale” few-week-old Multiplier, it should still be in the ballpark.

**CDC Regional Levels with PMC Estimates of the Percentage Actively Infectious**




| Estimated Percentage Actively Infectious* | PMC Model | Raw CDC Data |
|---|---|---|
| National | 2.8% (1 in 36) | 3.0% (1 in 34) |
| Northeast | 1.6% (1 in 63) | 1.7% (1 in 59) |
| Midwest | 2.2% (1 in 45) | 2.4% (1 in 42) |
| South | 3.7% (1 in 27) | 3.9% (1 in 26) |
| West | 4.2% (1 in 24) | 4.4% (1 in 23) |

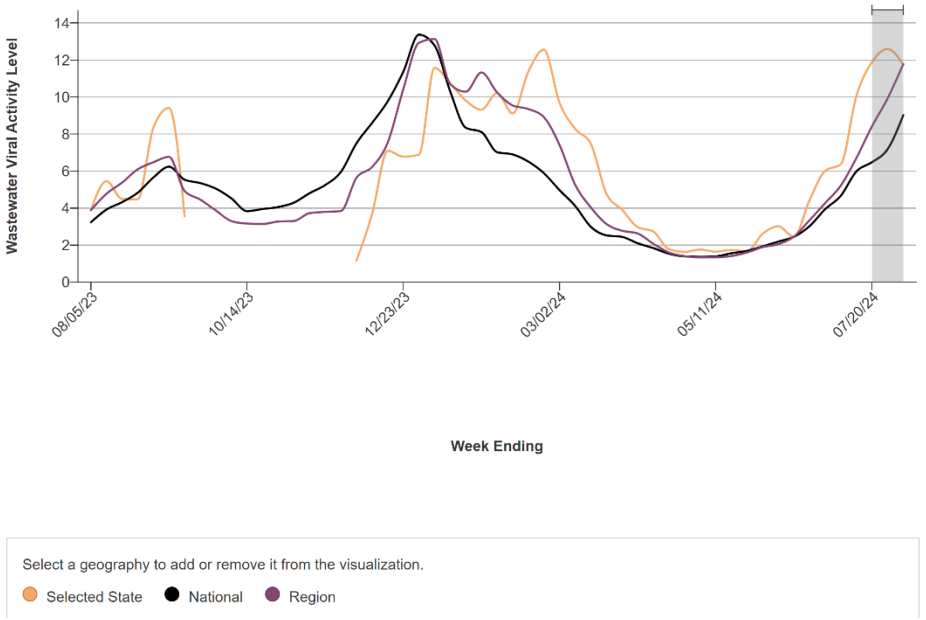
| PMC Regional Multiplier* |
|---|
| 0.329 |

* CDC level multiplied by the PMC Regional Multiplier provides an approximate estimate of the percentage actively infectious.

* The "Raw CDC" values are simply the value in the CDC chart multiplied by the PMC Regional Multiplier. The "PMC Model" estimates adjust thosedata by accounting for reporting time lag and to a marginal degree whether Biobot data suggest higher or lower transmission levels.

Example for Louisiana:

Try playing around with the PMC Regional Multiplier, including with smaller geographic entities, and please give us your feedback. For example, in Louisiana, the CDC website reports a current level of “11.75.” Using the Multiplier (11.75 x .329), we estimate that 3.89% of people in Louisiana are actively infectious. Just take 100 and divide it by that number (100 / 3.89) to get the “1 in ____” statistic. For example, 100/3.89 = 25.87. Thus, we would venture to infer that approximately 1 in 26 people in Louisiana are actively infectious with COVID-19. The main challenge with going down to a smaller geographic level is that the data quality gets lower because different sites vary in how effectively they estimate transmission. The CDC does their best to provide meaningful metrics, but the metrics are always better when averaged across many locations. We suggest testing out the Multiplier because many people are already trying to come up with meaningful local/regional data anyway. In our view, it is much better to have a process for doing so than to offer no process at all and leave people to guess what an arbitrary wastewater metric might mean.



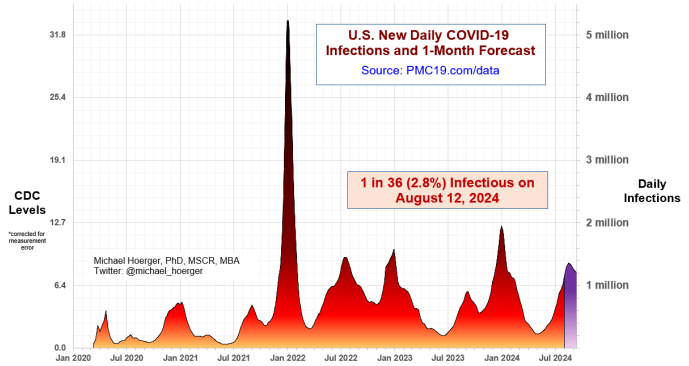

The PMC Regional Multiplier only works with the CDC levels. There is no way to convert from WastewaterSCAN or various local wastewater surveillance companies because they use different methodologies.

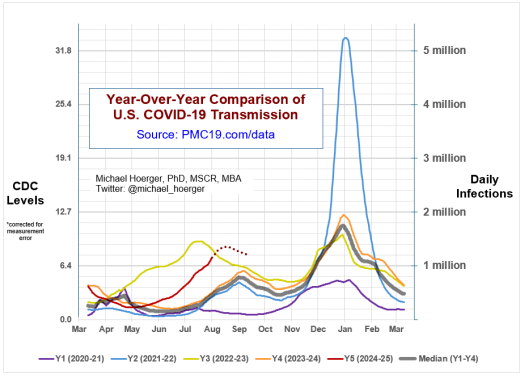
Regional graph:
https://www.cdc.gov/nwss/rv/COVID19-nationaltrend.html

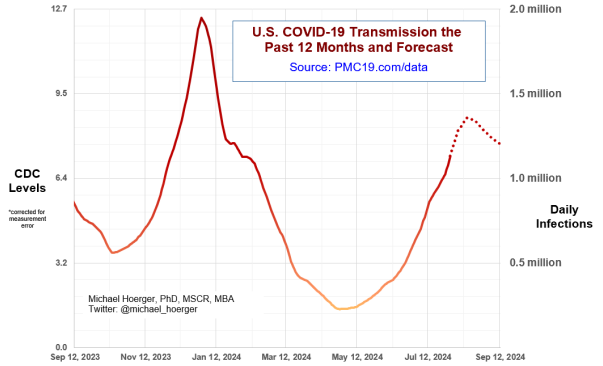
State selector:
https://www.cdc.gov/nwss/rv/COVID19-statetrend.html

# Complete Dashboard

Here is the complete dashboard. Please feel free to use any of the images, update or improve them, and share across other websites and social media. No permission is required. Try to include the website, a name, or a Twitter account in what you share to add credibility that people can find and evaluate the primary source.



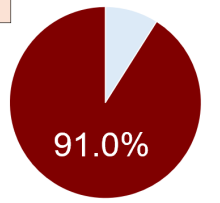

| Current Levels for Aug 12, 2024 |
|---|
| % of the Population Infectious |
| 2.8% (1 in 36) |
| New Daily Infections |
| 1,336,000 |
| New Weekly Infections |
| 9,352,000 |
| Resulting Weekly Long COVID Cases |
| 468,000 to 1,870,000 |

| Monthly Forecast |
|---|
| Average % of the Population Infectious |
| 2.7% (1 in 37) |
| Average New Daily Infections |
| 1,295,900 |
| New Infections During the Next Month |
| 38,877,000 |
| Resulting Monthly Long COVID Cases |
| 1,944,000 to 7,775,000 |

| Running Totals |
|---|
| Infections Nationwide in 2024 |
| 164,431,000 |
| Average Number of Infections Per Person All-Time, U.S. |
| 3.27 |

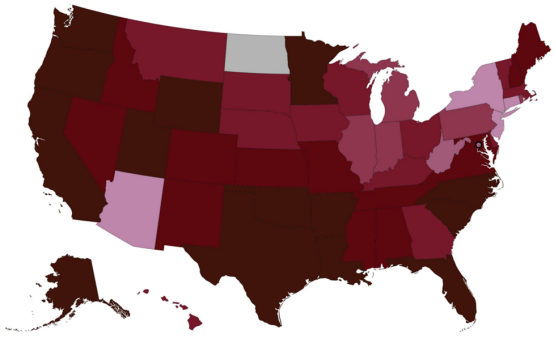
| How Does Risk Increase with More Social Contacts? | | | |
|---|---|---|---|
| Number of People | Chances Anyone Is Infectious | Number of People | Chances Anyone Is Infectious |
| 1 | 2.8% | 15 | 34.6% |
| 2 | 5.5% | 20 | 43.3% |
| 3 | 8.2% | 25 | 50.8% |
| 4 | 10.7% | 30 | 57.3% |
| 5 | 13.2% | 35 | 62.9% |
| 6 | 15.6% | 40 | 67.8% |
| 7 | 18.0% | 50 | 75.8% |
| 8 | 20.3% | 75 | 88.1% |
| 9 | 22.5% | 100 | 94.1% |
| 10 | 24.7% | 300 | 99.9% |

pmc19.com/data Michael Hoerger, PhD, MSCR, MBA Twitter: @michael_hoerger

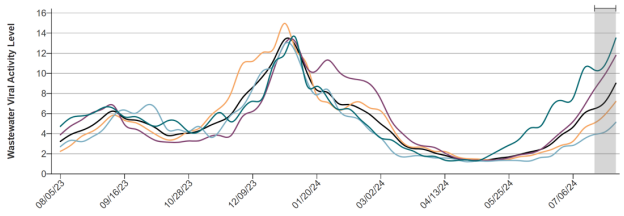



There is more COVID-19 transmission today than during 91% of the pandemic.

CDC COVID-19 Heat Map, Higher Transmission Shown with Deeper Red



CDC Regional Levels with PMC Estimates of the Percentage Actively Infectious




| Estimated Percentage Actively Infectious* | PMC Model | Raw CDC Data |
|---|---|---|
| National | 2.8% (1 in 36) | 3.0% (1 in 34) |
| Northeast | 1.6% (1 in 63) | 1.7% (1 in 59) |
| Midwest | 2.2% (1 in 45) | 2.4% (1 in 42) |
| South | 3.7% (1 in 27) | 3.9% (1 in 26) |
| West | 4.2% (1 in 24) | 4.4% (1 in 23) |

* The "Raw CDC" values are simply the value in the CDC chart multiplied by the PMC Regional Multiplier. The "PMC Model" estimates adjust thosedata by accounting for reporting time lag and to a marginal degree whether Biobot data suggest higher or lower transmission levels.

| PMC Regional Multiplier* |
|---|
| 0.329 |

* CDC level multiplied by the PMC Regional Multiplier provides an approximate estimate of the percentage actively infectious.

# Announcements

**July 11**
Recent COVID chat on Twitter had >2,000 listeners:
https://x.com/AnciraBecky/status/1808429122831401145

**July 24**
TODAY covers the PMC Forecast for the summer wave:
https://www.today.com/health/coronavirus/states-with-highest-covid-rates-2024-rcna163403

**Aug 1**
Check out our new empirical article in JAMA-NO framing masking in healthcare as a healthcare quality indicator.
Article: https://jamanetwork.com/journals/jamanetworkopen/article-abstract/2821699
Summary: https://www.msn.com/en-gb/health/other/masking-policies-prevalent-in-top-cancer-centers-amid-winter-covid-wave/ar-BB1qZWnr
Twitter Spaces Conversation: https://x.com/i/spaces/1OdKrXllryAJX
*If new to Twitter, it is not terribly challenging to create an account. Do so, and check in once a month or so. You may find it more useful than realized. I did.
PPT for the Space: https://pmc19.com/jama.pdf

**Sep 17**
Pencil in a Data Discussion between Drs. Hoerger and Moriarty who run the top public COVID case estimation models in the U.S. and Canada, respectively. The tentative time is 8-10 PM ET (7-9 PM CT). More details to follow.

**These reports typically conclude with extensive, repetitious technical information. In future reports, we will have a streamlined report and then a methodologic primer with more scientific details. Expect fewer words next time, but we wanted to ease the transition to the new model.**