

COVID-19 Vaccination Rates and their Relationship with COVID-19 Infection Rate: How is it impacted by reporting?

County: Miami-Dade County, Florida

Samuel Perebikovsky

sampereb@uw.edu

Introduction

The COVID-19 pandemic has drastically changed the world since the first cases that occurred in late 2019. With its rapid spread across the world, many scientists became immersed in trying to develop a vaccine for the virus that shut down the global economy and caused the deaths of over 5,000,000 people globally at the time of writing.

When a potential candidate for a vaccine was developed, it was quickly cleared by the FDA and distributed to millions of Americans in an attempt give people protection from COVID and to potentially develop herd immunity so the small percentage of immunocompromised people who are unable to take the vaccine would become safe from the virus by proxy. Although there is a lot of scientific literature on the effectiveness of the vaccines in both lessening the symptoms as well as transmission rate in an experimental setting, I wanted to see how the vaccines impacted COVID case rates in a real-world setting using data provided by Miami-Dade County.

With that in mind I posed the following question:

“What, if any, is the relationship between the COVID vaccination rate in Miami-Dade County and their corresponding COVID case numbers?”

The topic of the effectiveness of COVID vaccines is currently one that is quite divisive and an objective analysis of the question would be interesting to consider. However, throughout the course of the research and analysis, there were some surprising discoveries that forced this investigation into a new direction and forced me to ask a new, updated question:

“How do the reporting practices and lack of transparency of Miami-Dade County’s COVID data impact analysis using that data?”

Instead of starting with my updated question, however, I have framed this analysis to guide the audience along the same journey I took to reach my conclusions, as I feel this method makes the end result more compelling to the reader.

Related Works

There is a litany of work published by multiple sources on the effectiveness of COVID vaccines at preventing their corresponding cases. Most of these are from controlled environment settings, but more and more are starting to appear about the effectiveness

of the vaccines in real-world settings as many of the highly vaccinated places around the countries are reporting lower and lower COVID numbers.

Rosenberg, et al. analyzed the COVID vaccine effectiveness in New York State [1]. They had found that it was effective at preventing hospitalizations, although that effectiveness declined after about 6 months, necessitating a booster for maximum protection from the virus.

Research by Harris *et al.* found that being vaccinated can reduce your transmission rate of the virus by nearly half in many cases [2]. This research used data from people in the general public, so it is quite similar to the question that is trying to be answered in this paper.

Finally, Heath *et al.* also corroborate the previous two works and seems to confirm the efficacy of the vaccines in both preventing COVID cases in general but also lessening the severity of the symptoms in general [3].

All of these works paint a pretty clear picture as to what we should expect to see when looking at the COVID vaccine effectiveness in Miami-Dade County. We continue by describing our methodologies that form the foundations of our analysis.

Methodology

To begin this analysis, an exploratory dive into the data was necessary to start out with. Our analysis began from three data sources, two were from the CDC and had information about COVID cases as well as mask mandates, both at a county level granularity. The final data set was obtained from a survey conducted by the New York Times at a county level granularity about how often people complied with mask mandates and generally how people felt about wearing masks.

The next set of data that needed to be explored was vaccination data. For a general county level view, CDC data was sufficient and provided daily updates across all counties in the US and had the data at a few demographic levels split by age. This information was combined with the COVID case data to be able to see the numbers side by side. However, it quickly became apparent that the county level analysis was not enough to convey results appropriately. Using zip code vaccination level collected by the Miami Herald shed more light onto the vaccination situation and was utilized by creating a choropleth map that colored coded percentage of people vaccinated in every zip code.

Since the data used in this analysis is completely devoid of personally identifiable information (PII) there were not many ethical concerns that would impact people outside of the analysis. This paper is written as a story in the same the original analysis was performed to maximize reader's ability to replicate the results.

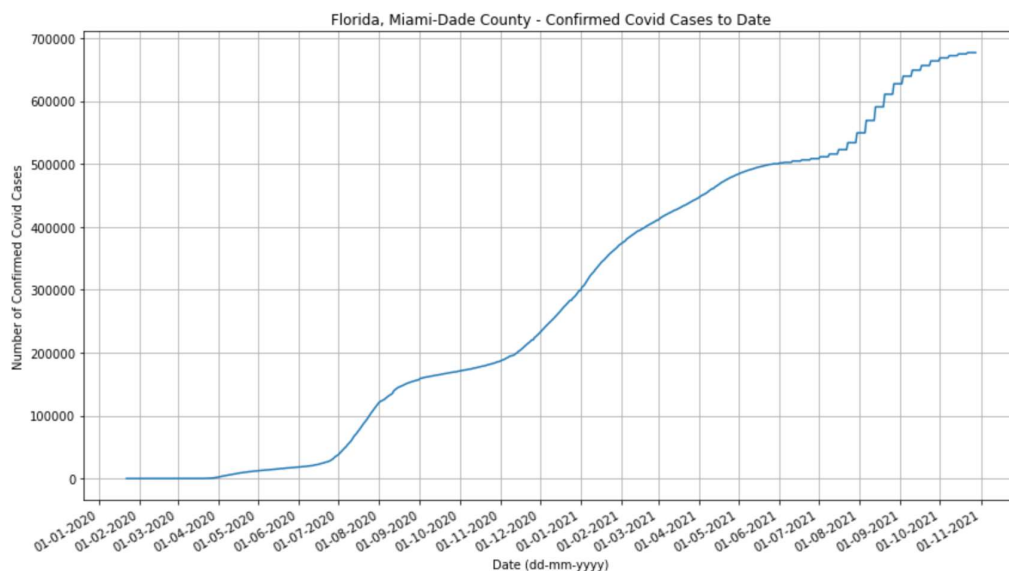
There were multiple visualizations created that were created to help readers quickly understand the data through a visual medium. To answer the original question, the initial analysis began by mostly focusing on quantitative methods. Correlation and

regression were all used to obtain results and look at relationship between our two main variables, vaccination rates and case rates. However, the methods of analysis performed in this paper shifted to include both quantitative and qualitative flavors to try to capture the story as a whole, especially after the original question became obsolete for reasons that will become clear in the 'Findings' section. The shift to more qualitative methods through the use of quotes and images from news articles and other sources ended up being necessary to paint the whole picture and display just how widespread the problems are in Miami-Dade County.

Findings

The analysis began with looking at the COVID case data from a few different perspectives. These would form the foundation of our analysis and is the first step to guide readers to reach the conclusion.

Looking at total COVID cases confirmed to date, we see a function behaving as we would expect it to - which is to monotonically increase. But there is a strange behavior at the end of the end of the function starting around June in 2021. The step-function behavior implies that Miami-Dade County had multiple days in a row with no increase in COVID cases.

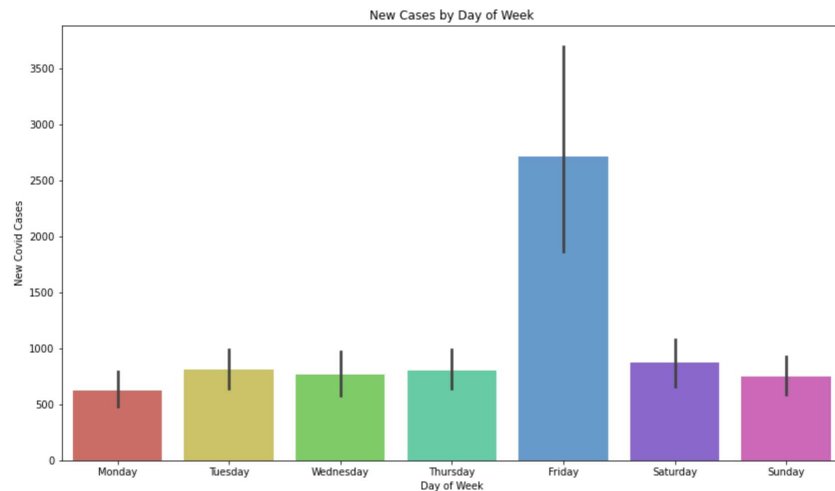


If we just look at the numbers, we can reach one of two conclusions:

1. Either the county has somehow been able to prevent COVID for 6 days of the week
2. Or, the reporting of the data shifted from once a day to once a week.

Both of these would explain this problem. But we cannot be sure without further analysis. This was the first step into the rabbit hole that would ultimately shift the goal of this project and completely change the nature of the question considered here.

Considering the distribution of new cases on a ‘day of the week’ level we can see that Friday is by far, the worst day to “catch COVID”.

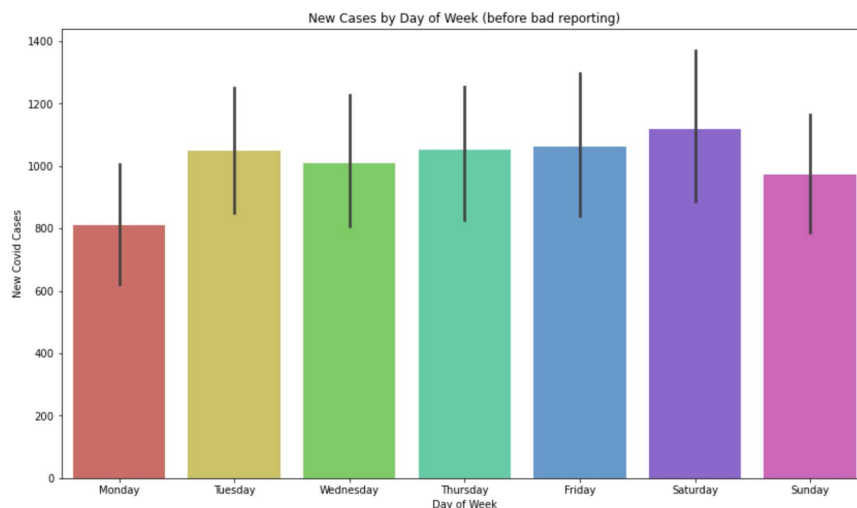


This behavior is certainly not what we would expect to see from a virus with presumably no understanding of time or what day of the week it is. But it is difficult to say for certain why something like this would occur until we look at some qualitative sources. There were a few articles that revealed what appeared to be a much more viable explanation for these numbers:

“The state stopped posting daily reports, which included local level data, and took down its COVID-19 dashboard on June 3.” [4]

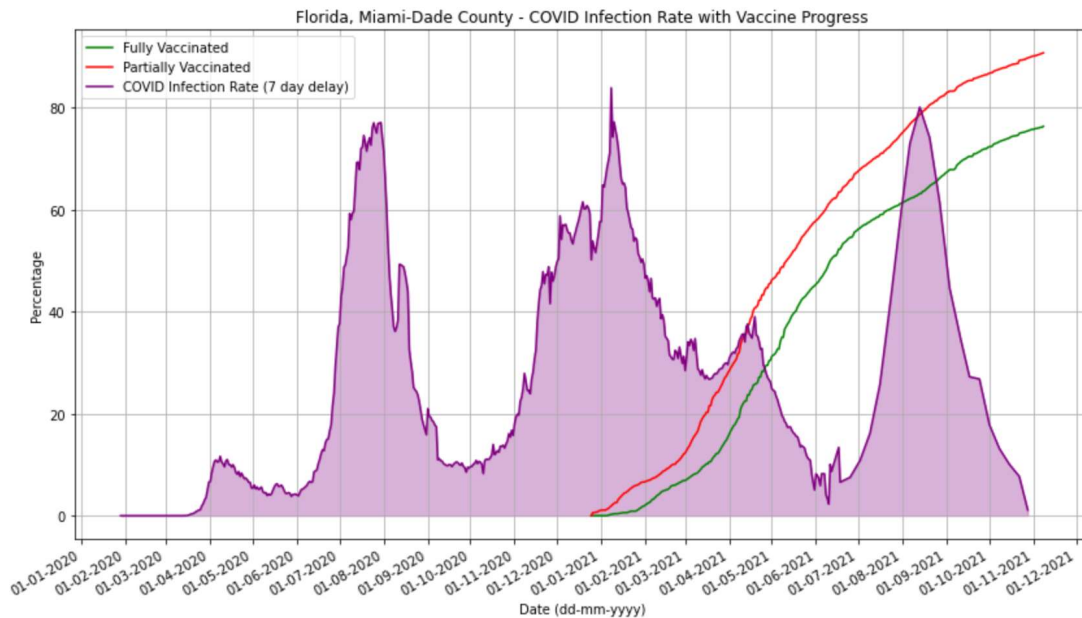
“When Florida stopped publishing its daily COVID reports in June, the COVID picture was improving.” [5]

Both pointing to the fact that sometime in June, the entire state of Florida, not just Miami-Dade County, stopped reporting COVID cases on a daily basis. Which is the cause of the irregularity of the bar chart above. By plotting only data from before June 2021 tells us a more accurate story of how cases were distributed.



Now we see that Saturday was the day with the most COVID cases reported. But more importantly is that the distribution across the days is much more even with no clear majority occurring on a particular day--much closer to what we would expect to see. This was a small setback, but not a major one.

Continuing with the original question, a visual understanding of the vaccination rates compared with case infection rate would be useful for a high-level comparison before diving deeper into the numbers. Using CDC's County vaccination and case data generated the following graphs.



Notice that this is graphing the COVID infection rate averaged over the previous seven days as a percentage to make it easier to compare with vaccination rate. The results here are not encouraging in showing that vaccines prevent COVID. Despite being at an 80% partially vaccinated and 65% fully vaccinated rate in September of 2021, the infection rate is just as bad as before the vaccine rollout. But this is the picture that the data is painting. We continue the dive into the vaccination data by looking at how it splits across age demographics and obtain the following result from the CDC's data.

Vaccinations in Miami-Dade County, Florida

People Vaccinated	At Least One Dose	Fully Vaccinated
Total	2,542,938	2,114,869
% of Total Population	93.6%	77.8%
Population ≥ 12 Years of Age	2,520,551	2,114,092
% of Population ≥ 12 Years of Age	99.9%	90%
Population ≥ 18 Years of Age	2,356,285	1,989,991
% of Population ≥ 18 Years of Age	99.9%	91.8%
Population ≥ 65 Years of Age	504,174	425,701
% of Population ≥ 65 Years of Age	99.9%	94.1%

This one dashboard about COVID vaccination data from November 30th 2021 was the tipping point and should cause some concern for most readers. There is not a single place with a 99.9% partial vaccination rate for any county in the country. In fact, 99.9% appearing almost anywhere should immediately be cause for concern about the validity of the data. How can there be a 90% vaccination rate in Miami-Dade County when what are generally regarded as the most vaccinated places such as San Francisco or New York don't have anywhere near those numbers? These results prompted further investigation.

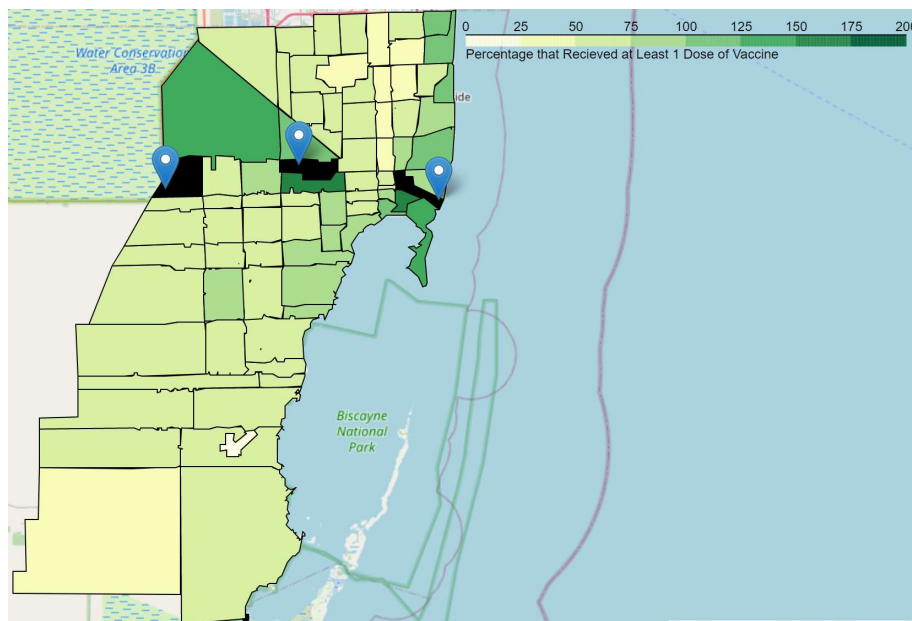
Here we begin our dive into the reporting of vaccination rates in Miami-Dade County and Florida in general. Our qualitative analysis becomes much more relevant in this section as it uncovers some shocking truths. According to the Miami Herald:

“As of Friday, 24 Miami-Dade ZIP codes logged a mathematically impossible vaccination rate of greater than 100% of eligible residents (those over 12) who have received at least one dose ... A smaller number of ZIP codes – 17 in Miami-Dade – have a fully vaccinated rate that is higher than 100% of the population, according to the health department’s data” [6]

Another article explained the most likely reason for this occurrence:

“The true percent vaccinated is nowhere near 90%,” said Dr. Aileen Marty, infectious disease expert at Florida International University. “It’s really giving people a false sense of security. “It’s just not true,” Marty added. “It’s based on a lot of people who gave Miami-Dade addresses who do not live here.’ [7]

In reality, it seems that Miami-Dade County is reporting the data for vaccination rates in a way that is somewhat disingenuous and is painting a picture of deceit. When looking at the zip code data for Miami-Dade County, we can see the following choropleth map showing us exactly where the problem zip codes are.



If we observe the sale of the graph on the top right of the image. We see that it ranges from 0% to 200% and indicates how much of the population has at least 1 vaccine dose. The 3 zip codes with the darkest green color indicate a vaccination rate of over 200%, one of which, contains the international airport. But we still cannot ascertain the reason for these inflated numbers and how they are being counted. The Miami Herald was able to find a likely source of the inflated numbers.

“one of Miami-Dade’s biggest vaccine providers said it could not validate the local addresses reported by people who used a passport as ID ... That provider, NOMI Health, said vaccination sites at Miami International Airport and Dolphin Mall in Sweetwater saw a jump in people using passports as ID in June, after Florida opened vaccine eligibility to anyone who had a reason to be in the state.” [6]

And we see similar conclusions reached from other sources as well.

“People age 18 and over in the 33122 area code had more than a 2,700% vaccination rate, according to the data.

Why is that zip code significant?

“That’s the airport,” Gelber said

Miami International Airport has been a pop-up vaccination site open to anyone — not just county residents.” [7]

The quotes from these articles help explain exactly why and how these numbers are being inflated. Anyone who gets vaccinated in Miami-Dade gets counted towards the vaccination rate of the county and state, regardless of actual residence. Effectively, when they create the fraction to calculate the percentage, they are fixing the denominator as the number of vaccine eligible residents in that zip code while having no restriction on the numerator. Which leads to the results we see on the map above.

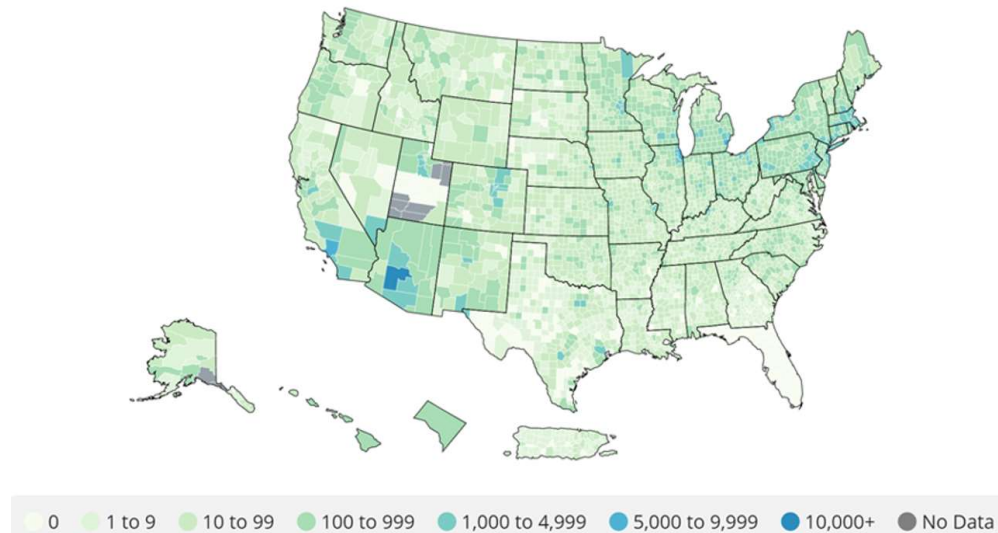
At this point, we realize that the original question is no longer possible to answer accurately. The saying in Data Science is “garbage in, garbage out” refers to the fact that if the data that is being used in an analysis is poor, the conclusions and results of the analysis will be just as poor. These numbers are being used to justify how amazingly Florida is handling the COVID situation. Governor of Florida, Ron DeSantis is quoted with saying “Without mandates or lockdowns, COVID-19 cases in Florida have decreased 90% since August,” which is a disingenuous statement at best. We will see why there are similar problems with how Florida reports its COVID cases later. Given this information, however, the question pivoted from analyzing vaccination and case rates to analyzing reporting practices and data transparency.

New Question: “How do the reporting practices and lack of transparency of Miami-Dade County’s COVID data impact analysis using that data?”

The scope of the reporting problem does not just stay within the vaccination rates of Miami-Dade County, it also spread to how Florida as a state reports its COVID case

numbers. If we look at the CDC's COVID weekly COVID case dashboard on the week of Thanksgiving in 2021 we observe the following.

Reported cases of All Counties in US



Notice that the entire state of Florida was reported as having zero COVID cases for the week of Thanksgiving in 2021. This is unquestionably the state with the least COVID cases for that entire week. Given that Thanksgiving is a time we would expect to see a rise in COVID cases as people and families gather to celebrate, this result is unexpected, to say the least. Once again, our qualitative analysis is able to provide us with the reasons behind these numbers and paint a completely different picture.

“The Florida Department of Health announced Wednesday night that its dashboard will not be updated on Thanksgiving.” [8]

And here we see the truth. The state has decided not to update numbers for thanksgiving and essentially take a break for the holiday. While this is not a huge problem by itself, notice that there is an option for ‘No Data’ on the dashboard. So, if Florida wanted to be truthful about the numbers that were being reported for that week, instead of saying they had ‘o’ cases they should have selected the option for ‘No Data’ which would be much closer to the truth. Thanksgiving week was not the only time Florida’s reporting strategies have concealed the truth from the public either. There are many instances of people saying the way Florida counts its COVID case numbers follows a different pattern from most other states and skews the results to favor them. When interviewed, Dr. David Cutler said

“Because Florida only reports those positive cases among people who are full-time residents of Florida. So, if your full-time residence is in Ohio, and you have a condo in Miami Beach and you go down there and you get COVID, you’re not a case in Florida. You’re not even a case in Ohio. You’re nothing, and that’s why I don’t put a lot of impact on that Florida case rate.” [9]

The selection bias presents itself clearly here. When it is convenient to have a higher number, such as vaccinations, a person doesn't need to live there to get counted towards the results. However, if it is more beneficial to have lower numbers, such as for COVID cases, then even if a person got sick in Florida, if they don't live there, that person will not be counted towards the results. This clearly impacts how people will view the progression of COVID throughout the state and can even influence policymakers who are operating under the assumption that the data reporting has no problems.

Discussion/Implications

The reporting practices of Miami-Dade and even the State of Florida clearly have the potential to cause problems that range from an individual scale all the way up to country level implications.

Individuals are likely to look at this data without doing a deep dive and see the result that seemed to say the infection rate of COVID is just as bad even though a large majority of the population was vaccinated against it. This would lead many to the conclusion that the vaccines don't work at preventing the spread of COVID. In reality, while vaccinated people can get sick, there is a litany of work that shows it slows the spread of the disease as well as makes the symptoms less severe.

State and Country policymakers might look at the data coming out of Florida, similar to how Ron DeSantis did and think that a lack of mask mandates and lockdowns do not lead to an increase in COVID cases as previously thought. Then they could begin to base their policies on data that is not at all representative of real life, causing problems for hundreds of thousands and potentially millions of people who would be directly or indirectly impacted by these policies.

Essentially anybody that thinks they can reach meaningful conclusions while looking at the data from Miami-Dade County and Florida will be doomed to reach conclusions that are not indicative of real-life and will likely be mis-informed until they are convinced otherwise.

Limitations

There are a few notable limitations throughout this analysis that essentially limit how much we can say that the conclusions reached in this analysis are the absolute truth. One of the biggest limitations is the nature of the second question we are trying to answer is very subjective. The biggest limitation that transpired here was the realization that the original question could not be answered in a way that would be satisfactory. Given the findings about how the data is reported and the lack of transparency of it, there should be hesitation about any conclusion reached from using Florida's data. It is also interesting to note that it seems Florida is trying to hide many of the specifics about COVID data and usually only provides the minimum required information. This is exemplified by the following quote:

‘Making matters more difficult, said Rep. Carlos Smith of Orlando, is that the Department of Health is withholding materials from the public by citing exemptions from the state’s Sunshine Law.

Smith said the department denied his office’s public records request for daily Orange County pediatric COVID-19 hospitalizations and cases, telling him the data he requested was “confidential and exempt from public disclosure.” [10]

This kind of behavior, while understandable, is cause for suspicion when it is considered in context with the rest of the findings in this paper.

Another limitation is the use of qualitative sources which may be biased in a significant way to influence the analysis. Another limitation is our data of vaccinations at the zip code level, which was collected from the Miami Herald, and although not used in this analysis they did estimate the true vaccination rate using a variety of techniques to compare with the vaccination rate reported by the government. Of course, one of the main points of this analysis is to show that there are times when quantitative data can also be limited in similar ways to qualitative data. But despite our extensive use of both quantitative and qualitative data, it does not necessarily translate into truth. The explanations given by the qualitative data explain the problems we see throughout our analysis and fills in the gaps of our story nicely.

Conclusion

Despite our unsatisfactory findings and inability to answer the original question, there are still some interesting points that can be taken away from this analysis. It should be clear just how much impact the method of reporting data can impact analysis by misleading anyone trying to use that data. And with the lack of transparency, most people will not be able to find out just how bad the reporting is of Miami-Dade and Florida when it comes to COVID data.

It is practically impossible to draw any meaningful conclusions just from using the quantitative data that Florida provides and has implications ranging from the individual to the country level. Essentially, anything that anyone says about Florida, COVID, and the data surrounding it should be taken with a grain of salt unless they acknowledge that the data is not reflective of the real-world and have attempted to adjust for that accordingly.

References

- [1] <https://www.nejm.org/doi/full/10.1056/NEJMoa2116063>
- [2] <https://khub.net/documents/135939561/390853656/Impact+of+vaccination+on+household+transmission+of+SARS-COV-2+in+England.pdf/35bf4bb1-6ade-d3eb-a39e-9c9b25a8122a?t=1619551571214>
- [3] <https://www.nejm.org/doi/full/10.1056/NEJMoa2107659>
- [4] <https://www.tallahassee.com/story/news/local/state/2021/08/13/floridas-covid-numbers-cases-data-difficult-but-not-impossible-find-heres-how/8099274002/>
- [5] <https://www.wfla.com/community/health/coronavirus/lawsuit-filed-for-return-of-daily-covid-19-reports-in-florida/>
- [6] <https://www.miamiherald.com/news/coronavirus/article254801602.html>
- [7] <https://www.local10.com/news/local/2021/09/22/90-of-miami-dade-vaccinated-its-just-not-true-experts-say/>
- [8] <https://www.clickorlando.com/news/local/2020/11/26/florida-wont-release-covid-19-data-on-thanksgiving/>
- [9] <https://www.healthline.com/health-news/why-florida-now-has-the-lowest-covid-19-case-rate>
- [10] <https://floridapolitics.com/archives/449831-florida-health-reports-impossible-vaccination-rates-withholds-confidential-covid-19-data-dems-say/>

Data Sources

- [1] [https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university?select=RAW us confirmed cases.csv](https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university?select=RAW+us+confirmed+cases.csv)
- [2] <https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i>
- [3] <https://github.com/nytimes/covid-19-data/tree/master/mask-use>
- [4] <https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh>
- [5] <https://www.miamiherald.com/news/coronavirus/article254801602.html> *

* Inside this article you will find a link to take you to the data (can no longer get past the paywall as my free trial ended)