Identifying the Effects of COVID-19 in Prisons of Oklahoma

Sandeep Tiwari

1. INTRODUCTION

American prisons have become some of the most dangerous locations with respect to contracting COVID-19. Since the beginning of the pandemic, more than 399075 and 2715 deaths, on average, have been reported inside those facilities each day. The cramped, and often unhygienic conditions of most correctional facilities have become a haven for the coronavirus to breed and transmit from inmate to inmate. Early on in the pandemic, most American prisons did not make testing a priority, and with very little pressure from the American people, policy-makers have been very sluggish to address the issue. The unfortunate fact is that prisons were simply not built to handle the full impact of a pandemic. There is probably not a word environment to prevent its spread—large groups of people living in a crowded indoor space, with indirect interactions to the outside world via correctional staff.

In this project, I considered taking a deeper dive into the impact that COVID-19 had on the United States' prison system, by drilling down into data specifically describing Oklahoma and Oklahoma County. Because of conditions that are not conducive to preventing transmission, coronavirus spreads much faster in prisons than it does in the outside world. So, I hoped to be able to demonstrate the urgency to take proactive measures to address the crisis persisting within American prisons and detention centers. I decided to use data collected by the Marshall Project in order to highlight the discrepancies between rates of transmission of prisons and the general population. The purpose of this project extension is also to understand the ethical dilemma of prioritizing the wellbeing of those who wronged society in some way. I did end up finding out that American prisoners are contracting the coronavirus at a higher rate than the general population, and requires some internal or policy-level remediation to address this crisis.

2. BACKGROUND

A recent New York Times report[1] found that the coronavirus infection rate in American prisons was more than three times that of the general population. But a secondary analysis found that the infection rate in the ICE detentions centers was 20 times more than the infection rate in the general population. Since last March, NYT reporters have tracked every known case of COVID-19 in every prison and detention center across the country, measuring the massive impact that the pandemic had on prisoners by requesting records and interviews with correctional officers and other prison staff, and even the prisoners' families. After one year, they have found that one third of all state prisoners are known to

[1] Burkhalter E. Colòn I, Derr b, Gamio L, Grisbach R, Klein A, Issawi D, Mensah K.B., Norman D, Redl S, Reynolds C, schwing E, Seline L, Sherman R, Turcotte M, Williams T. 2021. Incarcerated and Infected: How the Virus Tore Through the U.S. Prison System.

have had the virus, while almost 40 percent of all federal prisoners have been infected. Clearly these metrics suggest that prisoners are more prone to infection than that of the average American citizen, but this also gives an opportunity for us to examine the ethical dilemma and tradeoffs that are to be made when considering prioritizing a condemned, yet vulnerable subset of the population, to be protected from this sickness, despite their criminal backgrounds.

Here are two juxtaposing stances on the issue[2]:

> "There's no way you're going to get some prisoner a vaccine over a senior citizen"
>
> — Gov. Ron DeSantis

> "People are unwilling to say we should vaccinate people who are incarcerated — who may have done something bad — before us. It's like, 'How dare they?' But that's not the question here. It's not what we value or who we value. This is about who is at risk of disease. That's it."
>
> — Dr. Jaime Meyer, Yale School of Medicine

Before vaccines were made available for COVID, the Center for Disease Control and Prevention (CDC) recommended that all correctional staff and all prisoners be included in the early phases of the initial rollout. Naturally, this was met with mixed responses—those who reprimanded any advocacy for prioritizing prisoners over the majority of the country, and those who employed a more nuanced lens, taking into consideration the objective damage that can be done by the virus. So a panel of domain experts formulated ethical guidelines upon which to base policy decision-making with respect to prioritization of vaccine distribution[3]

1. Maximizing benefits and minimizing harm
2. Mitigating health inequities
3. Promoting justice
4. Promoting transparency

3. METHODOLOGY

3.1 Dataset Description

For this project, I used data collected by the Marshall Project, which is a non-profit investigative newsroom primarily focused on the United States criminal justice system. Since the beginning of the pandemic, they have compiled data on how many incarcerated people, prison workers, correctional officers, and medical staff had tested positive for COVID-19 in federal and state prisons all across the country. This data was reported at a state-level and on a weekly basis. Due to decrease in granularity of this data compared to that used for A4, I made several simplifying assumptions in order to continue my

[2] Kleine A, Norman D. 2021. Covid Outbreaks Devastated Prisons, but State Inmates' Access to the Vaccine Varies Widely.
[3] Young-Saver D. 2021. Examining the Math and Ethics of Covid-19 in Prisons.

comparative analysis: (1) I assumed that the the infection and death rates in prisons at a state-level would retain relatively similar values to the infection and death rates at a county level, and (2) I would have to aggregate the data that I used for A4 per 7 days in order to compare weekly values. The final dataset used from the Marshall Project consists of 74 rows for each week that they collected the number of positive test results. This dataset was generated by joining the four datasets, which individually consisted of information about COVID prison cases, COVID prison rates, prison population, and staff population in the state of Oklahoma. The final data keeps track of the number of tests administered to prison staff and incarcerated people in federal and state prisons, as well as the number of cases, deaths, recovery status, all in a time series form.

3.2 Objectives

The main objective of this initiative is to observe the major discrepancies in transmission rates between the prison and general populations in Oklahoma and Oklahoma County, while also gaining some insight into the policy decision-making process for prioritizing vaccine distribution.

3.3 Methodology

3.3.1 Data Preprocessing

There were several preprocessing steps that needed to be taken before that data could be used for analysis. Firstly, because the Marshall Project provided its data in 4 CSV files, I needed to effectively leverage all the necessary information by joining the appropriate files. Since I was not immediately modeling the data, I chose to leave any missing values in the data, as they were typically systematic, and any imputation would have been biased. This is primarily because many prisons across the country underreported the cases within their facilities, which obviously misrepresents the actual count. But Oklahoma, in particular, is known for having inconsistent, and sometimes nonexistent, reporting which eventually led to some gaps in my analysis. In fact, they stopped providing facility-level infection data in late February 2021.

3.3.2 Comparative Analysis

In the original data used to complete assignment A4, I was only provided with county-level time series data of a running count of the number of cases in Oklahoma County, and some information regarding adherence to mask-wearing. Unfortunately, I was not provided with any timeline information regarding mask mandates in Oklahoma County, and so I had no landmark policy changes that would have caused any marked shifts in case trends over time. This only allowed me to observe how the number of cases in Oklahoma County changed over time. So I chose to look at a dataset describing the number of COVID cases in a different population, in order to assess how significant the differences were between the two so that I could highlight the inequities in the healthcare provided to the two populations.

Firstly, I took and deep dive and drilled down into how the number of cases in prisons shifted over time, including the case and death rates compared to that of the American and Oklahoman general populations. My belief was that the most foundational argument to demonstrate the

unsanitary conditions that prisoners live under was to visualize the stark difference in transmission rates, and how quickly the virus was spreading through these prisons. I generated the same visualizations for both populations and placed them side-by-side for utmost clarity.

In order to compute the infection rate, I used the formula as defined on Wikipedia[4] as

$$\text{Rate of infection} = K \times \frac{\text{the number of infections}}{\text{the number of those at risk of infection}}$$

Where *K* represents a value of 100 to represent the rate as a percentage. This is a fairly straightforward equation, however being at risk of infection is open to interpretation and must be clearly defined, accordingly. Using the mask use data, I specified those that were at risk of infection as the percentage of the population that rarely or never wore masks. According to the mask-use dataset, 9.6% of people never wear masks and 6.4% of people rarely wear masks in Oklahoma county. This leads to a grand total of 16% of the uninfected population being at risk of infection. I assumed that those who had been infected could not be infected again, and that these proportions stayed relatively constant as the pandemic progressed, so 16% of those who had not yet been infected would be the total number of those at risk of infection each day. Therefore the daily infection rate was defined as:

$$Infection\ rate\ for\ day\ i\ =\ 100\ \times \frac{the\ number\ of\ infections\ on\ day\ i}{(never\ wears\ mask\ \% + rarely\ wears\ mask\ \%) \cdot (total\ population - total\ infections\ up\ until\ day\ i)}$$

Note that this equation can only be used for the general population as the same standards for mask use as a prevention method cannot be presumed to hold within prison facilities. I just assume that 100% of the remaining uninfected prison population is “at risk” on any given day.

3.3.3 Linear Regression

Due to several crucial intervals of unreported cases in Oklahoma, I chose to model the data using a linear regression. Taking into consideration how many variables go into accurately predicting how case rates would change over time, as well as the fact that this is time series data, this approach was simply to visualize the overall trend over time. This was to get a grasp of the behavior of prison cases as the pandemic progresses. However, whatever result was derived from this experiment would have to be taken with caution simply due to the inconsistent reporting of cases, as well as the blatant gaps in the data.

3.3.4 ANOVA

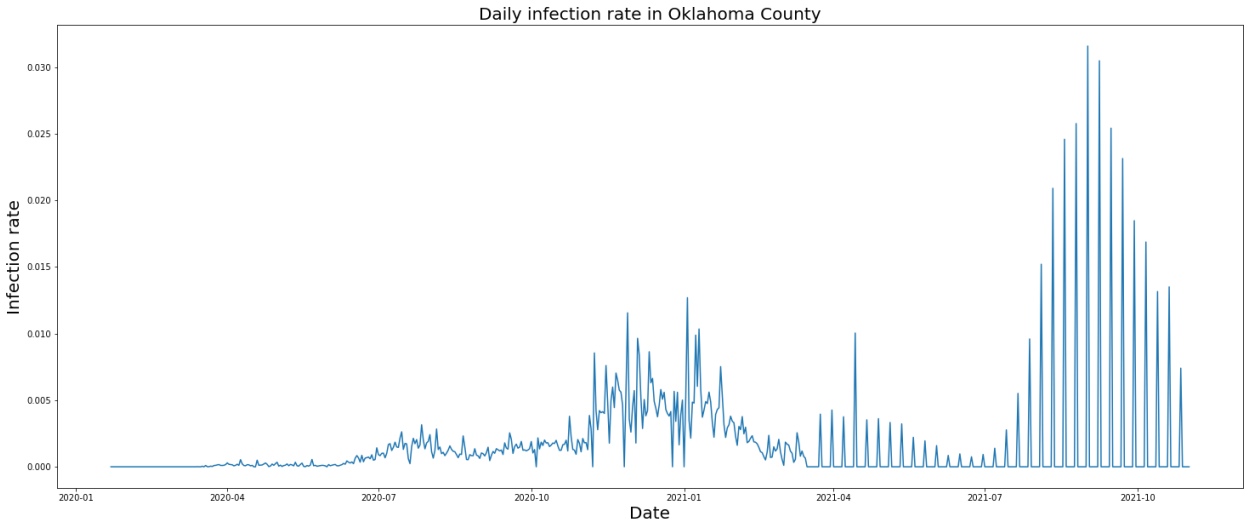
So to circumvent these results that were guaranteed to be time-dependent, I decided to perform an ANOVA analysis. ANOVA (ANalysis Of VAriance) is a test that can be used to compare the

[4] “Infection Rate.” *Wikipedia*, 28 July 2021, en.wikipedia.org/wiki/Infection_rate.

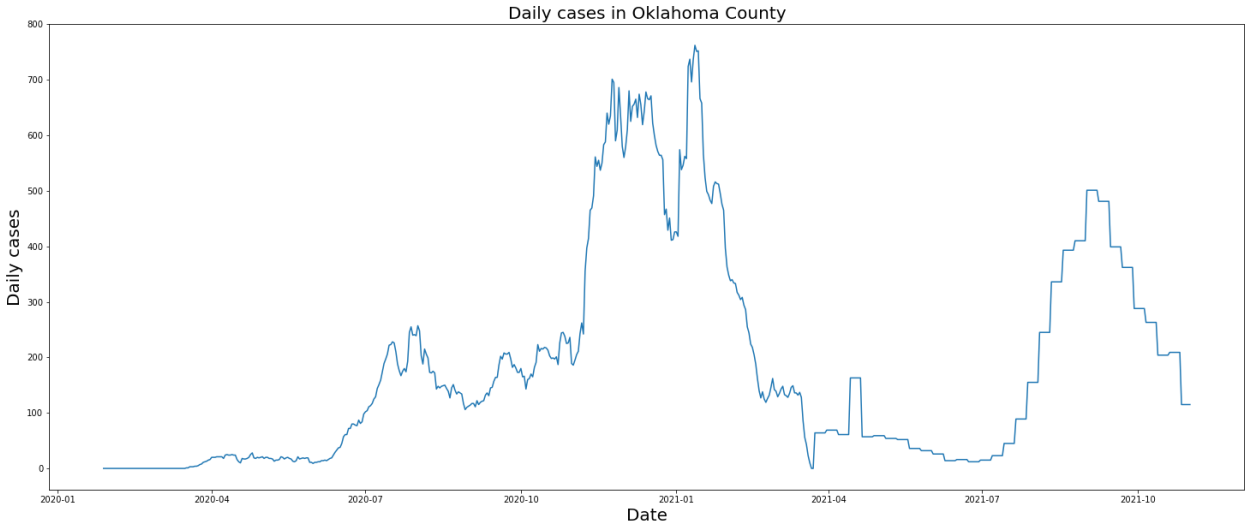
means of more than two groups. It uses a variance-based *F* test to check the group mean equality, testing a non-specific null hypothesis that all group means are equal. In the context of this problem, I compared the mean infection rates of the prison and general populations of Oklahoma since March of 2020. Looking at this problem from a more aggregated approach would zoom out and compare the two populations by taking a more encompassing view of the week-to-week infection rates, and be less prone to skew from the underreporting of cases within prisons.

4. FINDINGS

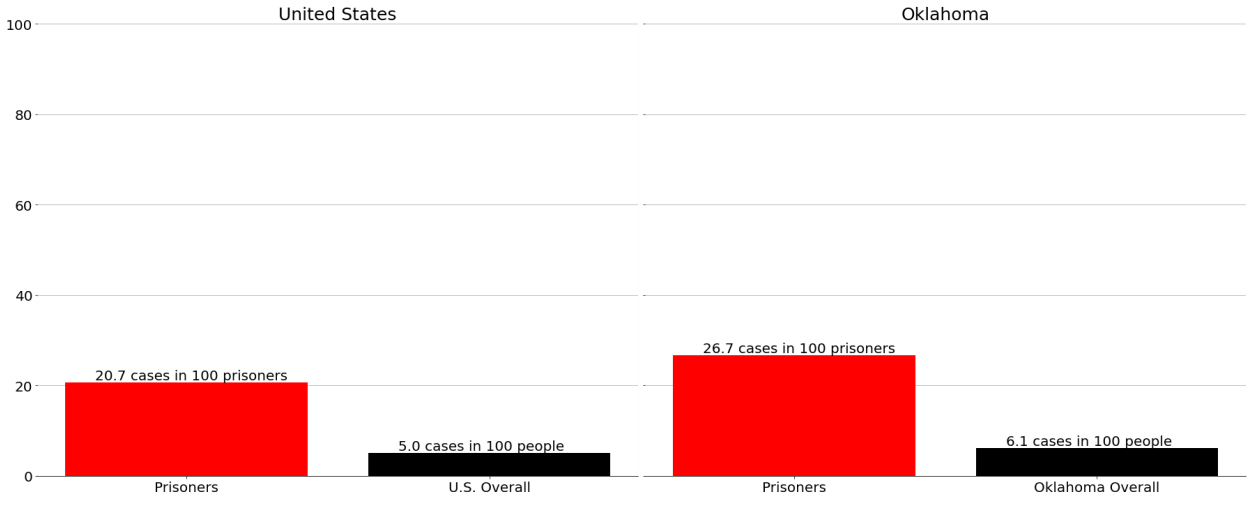
In order to understand how much faster coronavirus was spreading inside of the prisons compared to outside of the prisons, I generated several visualizations that would help me get a more direct insight into the crisis. Using the original data, I found several interesting trends:



We see that the infection rate remains fairly stable through March 2021. However, this occurs at a very low rate (<1%). It's difficult to understand why this is the case, as it could be due to a low volume of testing, or people simply not getting infected. But if we also look at this accompanying (below) that displays the daily cases, we see that the number of cases spikes during the Fall and Winter of 2020.

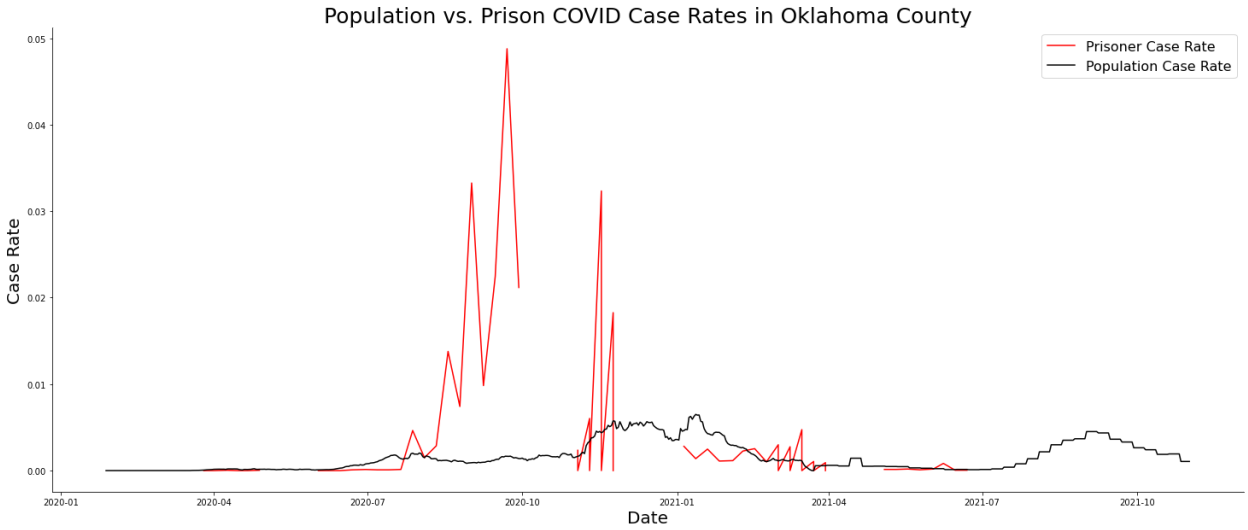


For full transparency, this daily cases plot has been smoothed out over 7-day windows to account for the delay between contraction and testing positive for COVID-19. One thing to note is that the infection rate spikes at seemingly constant intervals after April 2021. This could potentially be due to the behavior and when they tend to get tested, resulting in this seasonal trend in these spikes of daily infection rates. But using the complimentary graph of daily cases, the stepwise nature of the plot indicates that the cases are reported, perhaps, once a week in which we lose the daily nature of our data. This is why I chose to begin with a more aggregated approach, viewing the data in a coarser time scale so that I can more effectively compare the two populations.

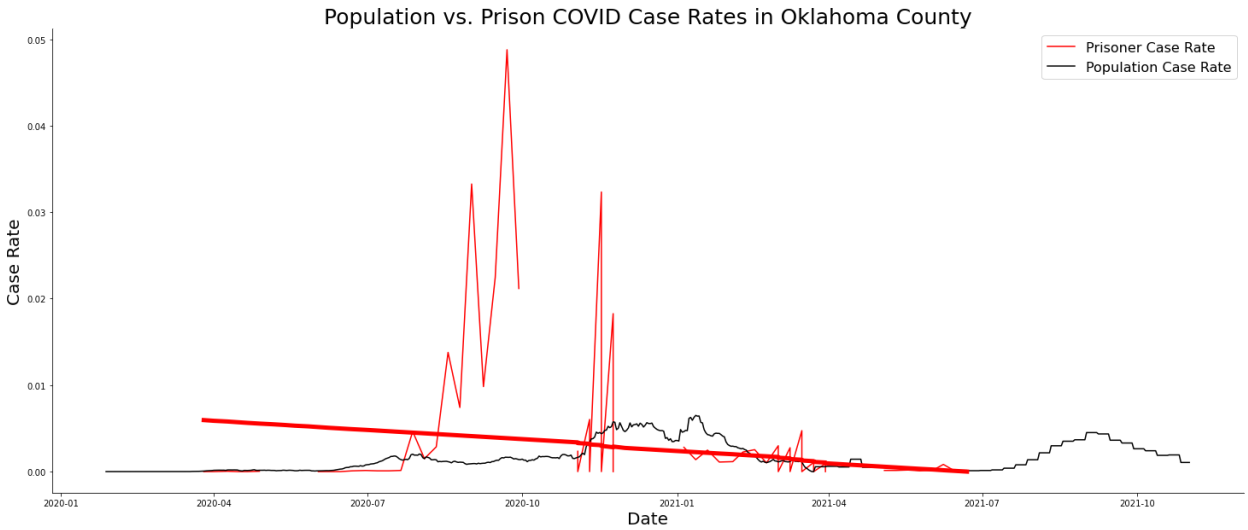


Over the course of the pandemic, we can see that about one in five state and federal prisoners in the United States have tested positive for coronavirus, a rate that is four times as high as that of the general population. I also found that in some states, such as Michigan, Arkansas, Kentucky, South Dakota, and Kansas, more than half of the prisoners had tested positive for coronavirus. In fact, Oklahoma had almost 30% of its prisoners test positive. But we can see that prisoners, particularly in Oklahoma, had an

infection rate more than 5 times the national rate and over 4.5 times the infection rate of the general population of Oklahoma. While this aggregated perspective is quite telling, I also observed just how quickly prison rates were skyrocketing.



One majorly concerning result we're seeing here is that there is a single week in 2020 where 5% of all Oklahoma prisoners tested positive for coronavirus when less than 0.5% of the Oklahoma population had tested positive. Even with the possibility of underreporting of cases, this is quite a disturbing result that clearly demonstrates that prisoners were more than 5 times as likely to contract the virus than the general population at a certain point. There are also several massive gaps in the data as there were intervals where cases were not being reported, throwing into question the consistency and reliability of the reported cases. Because some doubt has been cast upon this analysis due to the data, I decided to try modelling the trend of the prison cases, to get a general sense of which direction it was going over time by using a simple linear regression modelling the cases using only time.



We can observe a distinct downward trend in the data, suggesting that the cases decreased significantly by the start of 2021. However, due to the questionable nature of reporting COVID cases in prisons, I'm apprehensive to accept this result just yet. It simply does not seem feasible that prisons in Oklahoma suddenly reined in the number of cases when 1 in 20 prisoners were testing positive just a few months earlier. So I performed an ANOVA analysis to observe the significance of any difference in case rates between the incarcerated and the free populations. For my ANOVA analysis, I used the hypotheses:

> H_0: The mean rate of infection of prisoners is equal to the mean rate of infection for the general population
> H_A: The mean rate of infection of prisoners is different than the mean rate of infection for the general population

	df	sum_sq	mean_sq	F	PR(>F)
C(population)	1.0	0.000423	0.000423	23.167162	0.000002
Residual	706.0	0.012896	0.000018	NaN	NaN

Already looking at these results, we find the *p*-value to be significant (< 0.05), and therefore, we conclude that there are significant differences among the populations with respect to infection rates. It should be noted that the *F* value is inversely related to the *p*-value and a higher *F* value indicates a significant result, suggesting that prisoners are more vulnerable to the virus than the remaining population who can socially distance, wear masks, and properly sanitize.

5. DISCUSSIONS/IMPLICATIONS

There are several takeaways we can collect from this analysis. Firstly, prisoners are contracting COVID-19 at a significantly higher rate than is the general population. It is difficult to pinpoint what the exact reason is for these significantly higher rates, but it is clear that some measure must be taken within the prison walls before it's too late. Each step of my analysis points to this same conclusion, and so there are a few interpretations for what the appropriate response should be, given all these results. Stemming from the underreporting of cases, and large infection rates within prisons, there is no doubt that prisoners are probably being put in the backseat with respect to pandemic response priority. Despite their criminal backgrounds and societal infractions, perhaps, they need to be prioritized a little bit more before it becomes too late.

While we have the hard data and statistics to back it up, we begin to enter murky waters as this starts to become an ethical dilemma. Should we be giving first access to resources for fighting the pandemic to those who have previously wronged society and are still serving their sentence? Many people would not, while many others might take a more nuanced stance, by taking into consideration who is most at risk, and how much damage this virus can do. Ultimately, this analysis was an attempt to garner some insight into the ethical guidelines highlighted by external advisors to the CDC for policy decision-making.

Naturally it is difficult to promote justice through raw numbers, but I aimed to make these results extremely transparent, to demonstrate inequities in public health, and highlight the urgency with which proactive measures need to be taken to improve the conditions for prisoners during this pandemic, but was a bit hindered by the quality of the dataset and the information it provided. I believe if underreporting was not such an issue within prisons, nationwide, I would be able to present a more reliable analysis founded on high-quality data. But even with some missing data from these prisons, we still find concerning patterns that presumably would only get worse, given more data.

6. LIMITATIONS

The primary limitations that I had to endure during this analysis stemmed from missing components of my data. Firstly, the mask mandate data from A4 simply did not exist for Oklahoma County, and with further research, I found that this was because Oklahoma never implemented a mask mandate. This meant that I could not use any sort of implementation of pandemic-response policy as a landmark for how the cases accordingly responded. To mitigate this problem, I tried looking at surrounding states, i.e. Texas, Kansas, Arkansas, Missouri, and Louisiana. However, none of these states seemed to have much in the name of mask mandates either, and so I was left with no choice but to proceed without this information, prompting me to compare this data with a different dataset rather than drill down into the behavior of cases in Oklahoma County. Secondly, infection data for nearly all detention centers were almost certain to be undercounted due to a lack of testing. I have already discussed how testing within prison facilities was simply not a priority, and so it is possible that the infection rates are biased. Due to the aggregated nature of this data, taken on a week-to-week basis, I chose to make the simplifying assumption that any underreporting was negligible and would not drastically affect the outcome.

7. CONCLUSION

The pandemic response in the nation's prisons and detention centers has improved ever so slightly over this past year, by performing more testing, especially when intaking new inmates, and enforcing internal mask mandates. The fact is that prisons were built with security as their main priority, not to act as hospitals. And given the old age and poor health of many inmates forced into cramped, unsanitary conditions, incarcerated people have become an especially vulnerable population during the pandemic. Despite this elevated risk of infection and death, most states were not prioritizing them to get immunized. However, by April 2021, most states had finally decided to make a plan to vaccinate its prisoners. But as reported by many state prison facilities, many prisoners and prison staff still remain reluctant to get vaccinated, all but ensuring future outbreaks. But I do believe that progressively more availability to testing inside and outside prisons will significantly mitigate death rates, particularly correctional officers that have direct access to the outside world and could carry the virus into prisons.

At the end of the day, it is difficult to pinpoint what the exact reason is for these substantially higher rates, but it is clear that measures must be taken and that there is much more to be understood about the impact of the pandemic in prisons. In particular, I hope that the nationwide reporting of prisoner

infections at a facility-level and state-level can become more consistent and reliable. The underreporting of cases was a bit of an obstacle for this project, and I hope to collect more information so that a more comprehensive time series analysis can be done on the data. In the coming weeks, I presume the Marshall Project will continue to track testing data on coronavirus and analyze its effects in the American prison systems.

8. REFERENCES

Bedre, R. (2021, October 24). *ANOVA using python (with examples)*. Data science blog. Retrieved December 14, 2021, from https://www.reneshbedre.com/blog/anova.html

Burkhalter, E., Colón, I., Derr, B., Gamio, L., Griesbach, R., Hinga Klein, A., Issawi, D., Mensah, K. B., Norman, D. M., Redl, S., Reynolds, C., Schwing, E., Seline, L., Sherman, R., Turcotte, M., & Williams, T. (2021, April 10). *Incarcerated and Infected: How the Virus Tore Through the U.S. Prison System*. Retrieved December 14, 2021, from https://www.nytimes.com/interactive/2021/04/10/us/covid-prison-outbreak.html

Hinga Klein, A., & Norman, D. M. (2021, March 17). *Covid Outbreaks Devastated Prisons, but State Inmates' Access to the Vaccine Varies Widely*. The New York Times. Retrieved December 14, 2021, from https://www.nytimes.com/2021/03/17/us/covid-prisoners-vaccine.html

The Marshall Project. (2021, June 24). *A state-by-state look at 15 months of coronavirus in prisons*. The Marshall Project. Retrieved December 14, 2021, from https://www.themarshallproject.org/2020/05/01/a-state-by-state-look-at-coronavirus-in-prisons

Park, K., Meagher, T., & Li, W. (2020, April 24). *Tracking the spread of coronavirus in prisons*. The Marshall Project. Retrieved December 14, 2021, from https://www.themarshallproject.org/2020/04/24/tracking-the-spread-of-coronavirus-in-prisons

Young-Saver, D. (2021, May 4). *Examining the Math and Ethics of Covid-19 in Prisons*. The New York Times. Retrieved December 14, 2021, from https://www.nytimes.com/2021/05/04/learning/examining-the-math-and-ethics-of-covid-19-in-prisons.html

9. DATA SOURCES

- Raw Confirmed Cases in the US
    - https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university?select=RAW_us_confirmed_cases.csv
- Mask Mandates by County

    - https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i
- Mask Compliance Survey
    - https://github.com/nytimes/covid-19-data/tree/master/mask-use
- The Marshall Project: COVID Cases in Prisons
    - https://data.world/associatedpress/marshall-project-covid-cases-in-prisons