

Covid-19 Testing Equity in New York City

Dan Rosenfeld MSc
Sean Brennan MPH
Sharzhad Divsalar MPH
Sarah Joseph Kurien MPA
Chris Keeley MUP
Andrew Wallach MD
Ted Long MD

March 23, 2023

Introduction

New York City emerged in March 2020 as one of the global epicenters of Covid-19 infections and deaths, roughly two months after the first detected case in the United States and three months following the first case in Wuhan, China. (Thompson et al, 2020, p.1). A distinct lack of testing characterized the beginning of the pandemic in the United States and especially in New York City, with no positive tests recorded in the city until the beginning of March 2020 (Lieberman-Cribbin et al, 2020, p.1). Over two years into the pandemic, over 43,000 deaths and over 3 million infections have been documented in New York City, though mortality and infection rates have varied widely based on a number of spatial and socioeconomic factors (Dong et al, 2022). Individuals with chronic conditions such as obesity and diabetes have faced higher mortality rates than the general population as a whole, while socioeconomic factors such as poverty, housing overcrowding, de facto racial segregation, and an inability to do some jobs remotely have intersected with aforementioned chronic health issues to create a perfect storm of Covid-19 mortality in some neighborhoods (Bilal et al, 2021, p. 1). Following a dire lack of testing at the onset of the pandemic in New York City, various levels of the government sought

to implement a mass testing strategy that had to deal with a number of fundamental barriers to access, the largest likely being that the United States lacks national health insurance and has a substantial number of uninsured or minimally insured individuals who often have traditionally avoided seeking care for all but the most serious of emergencies, out of a fear of incurring substantial adverse financial consequences. The New York City Health + Hospitals Corporation (HHC), a public benefit corporation founded in 1969 that operates public hospitals and clinics, became one of the largest free testing providers in the city. In this paper we intend to analyze the relative success of the HHC directly administered Covid-19 testing program in addressing racial and income equity issues seen in testing rates in New York City and elsewhere in the country. An analysis of the six million directly administered HHC tests shows an inverse relationship between testing rates and income, something not seen when examining citywide testing rates instead of HHC tests. Findings were more mixed when analyzed against other demographic variables related to Covid-19 mortality risk. While the very neighborhoods that HHC tested heavily did often suffer from extremely high Covid-19 mortality, we find that HHC testing likely did address some citywide testing access issues that led to overall higher testing rates in more socioeconomically affluent and privileged populations.

Discussion: An analysis of testing rates from all providers citywide reveals a substantial disparity in overall testing rates, with wealthy, predominantly white areas having the highest Covid-19 testing rates, even after free testing has been widely available for much of the last two years (nychealth, 2022). Testing at the beginning of the pandemic was extremely limited, and even though severe citywide testing shortages had been largely eliminated by the summer of 2020, surges in Covid-19 infection and mortality could still cause testing shortages even as late as the Omicron variant surge that struck the city in December 2021. Consistent and easily

accessible mass testing has proven to be essential in devising successful strategies to avoid mass mortality. One factor that led to the extremely high mortality rates seen in New York City during the height of the first wave in April 2020 was that the first case in New York City was not detected until March 2020, one month earlier, likely due to a lack of testing, as the virus had been confirmed in the United States as early as January 2020. (Thompson et al, 2020, p.1). At this time, the City officials had minimal data with which to make decisions about when and whether to institute a lockdown. Due to a number of factors, mortality in New York City has not since equaled or surpassed the extremely high mortality rates seen in April 2020, one reason among them likely being the proliferation of mass testing, even though mortality rates did spike again during the winters of 2020-2021 and 2021-2022 (nychealth, 2022). We do want to sound a note of caution in interpreting disparities in testing rates across space and time as related to Covid-19 mortality, due to a number of factors. At the onset of the pandemic, testing was extremely limited due to supply constraints, in contrast to now in 2022 when at-home rapid testing has become very common, with the federal government even supplying free tests of this sort in the mail. Results from these tests are likely rarely reported to health authorities, with negative at-home test results very unlikely to be reported. At the same time, as we compare testing rates and results to geographic and demographic variables, we want to be mindful of potential ecological fallacies that may arise when equating jurisdiction level data with individuals who live in those jurisdictions. New York City is an extremely diverse city with substantial income inequality, meaning that even the smallest zip codes or ZCTAs (ZIP Code Tabulation Areas) analyzed here may be expected to contain substantial variation along socioeconomic and demographic lines. For instance, homeless rates may be higher in wealthier cities with high costs of living, though high income individuals who pay large amounts in

housing costs themselves are likely not necessarily at a higher risk of homelessness. Equating this population with risk of homelessness would be an example of the type of ecological fallacy that we would like to avoid in this analysis. Still, we have found an inverse relationship between HHC testing rates and median income, at odds with the citywide testing rate's relationship with median income, and we intend to illuminate and hopefully explain the foundational underpinnings and consequences of this disparity.

Methodology

NYC Health + Hospitals directly-administered testing data was collected from the in-house HHC Microsoft SQL Server Management Studio 18 Clarity database, representing 6,347,533 total tests and 449,721 positive tests. These tests have been sourced from the in-house HHC instance of Epic Systems. Please note that this does not include tests conducted by external testing vendors contracted by HHC, or the schools testing program that mandated surveillance testing in New York City public schools from 2020-2022. Demographic data was taken from the most recently available American Community Survey data from the United States Census Bureau, which in this case was from 2020. Citywide testing data came from the New York City coronavirus data repository. We initially sought to run a Poisson regression analysis comparing counts of tests and positive tests to demographic variables while using population as an offset, though we found the data to be over dispersed and chose to use negative binomial regressions. In terms of variables to analyze, we were influenced by two studies from King County, Washington and New York City, respectively, that among other things analyzed testing rates against demographic variables. The Kings County, Washington study collapsed non-white racial data into a variable called People of Color (POC) due to collinearity among non-white groups (Seto et al, 2020). This is something we have chosen to do in this study. The New York City study takes a number of different

demographic variables like income, education, and housing overcrowding to calculate a single socioeconomic index score (Lieberman-Cribbin et al, 2020). While this is one way to handle collinearity between socioeconomic variables like income and poverty, we chose to use median income as our main predictor of socioeconomic status. As variables related to Covid-19 mortality and testing rates, we included age and a sex breakdown, as being elderly and male is positively correlated with higher mortality. As a measure of presumed difficulty in accessing healthcare, the percentage of English speakers was included. Regressions were run in a Jupyter Notebook using the pandas, numpy, statsmodels, matplotlib, scipy.stats, and sklearn.linear_model packages in Python. Data visualizations were created using Tableau.

Results

The overall pseudo r-squared value when comparing HHC tests by zcta against the selected variables was 0.1101. The number of HHC tests significantly increased as median income fell (exponentiated $\beta=0.99998$, SE= 1.00000155, $p=0.00000$). The number of tests increased at a slightly significant rate as the percentage of English-only speakers rose (exponentiated $\beta=1.59824$, SE= 0.245, $p = 0.056$). No other variables correlated at a significant level with the number of tests. When considering positive tests by zcta, positive tests also significantly increased as median income fell (exponentiated $\beta=0.99998$, SE= 1.57e-06, $p=0.000$), and as the percentage of female residents fell (exponentiated $\beta=0.04014$, SE= 0.957, $p=0.001$). Positive tests by zcta rose at a less significant level alongside the percentage of English-only speakers ($\beta= 1.81630$, SE= 0.271, $p= 0.027$). No other predictor variables correlated with positive tests by zcta.

Conclusion

Despite overall testing patterns that displayed testing clustered in relatively affluent areas of the city, the New York City Health + Hospitals corporation was able to provide free testing that was weighted towards less wealthy parts of the city. Despite the extremely negative consequences of Covid-19 to New York City, this testing regime could be seen as a bright spot that might be able to be deployed again in the event of a future pandemic. Mortality was dramatically higher in the first wave in the Spring of 2020 when little testing was available, clearly speed of deployment in the event of a quickly spreading emergency is just as important as an equitable distribution of testing resources. However, there is evidence that Covid-19 caused higher mortality in areas with high income inequality, making equitable distribution of testing a key component of successful pandemic mitigation strategies in the future (Liao et al, 2021).

References

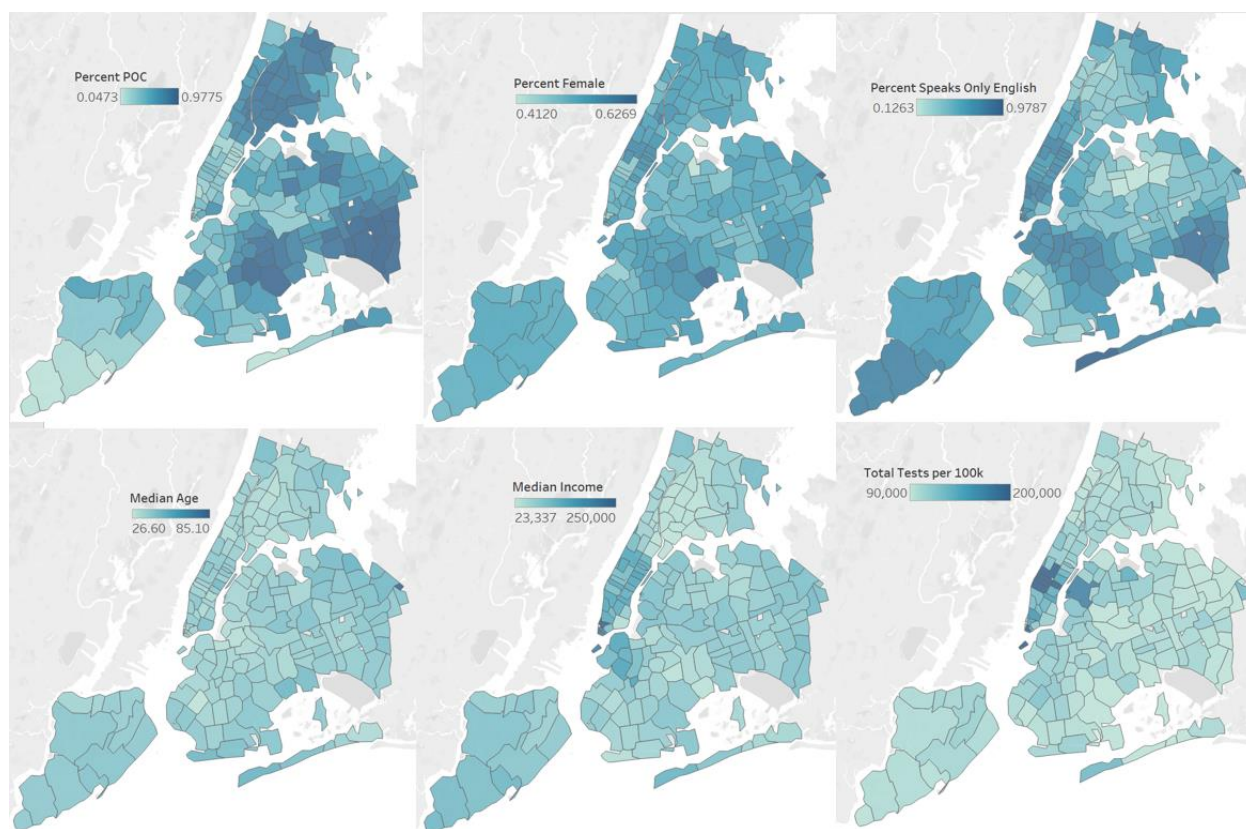
- Bilal, U., Tabb, L. P., Barber, S., & Diez Roux, A. V. (2021). Spatial inequities in COVID-19 testing, positivity, confirmed cases, and mortality in 3 US cities: an ecological study. *Annals of internal medicine*, 174(7), 936-944.
- Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Inf Dis*. 20(5):533-534. doi: 10.1016/S1473-3099(20)30120-1
- Liao, T. F., & De Maio, F. (2021). Association of social and economic inequality with coronavirus disease 2019 incidence and mortality across US counties. *JAMA network open*, 4(1), e2034578-e2034578.
- Lieberman-Cribbin, W., Tuminello, S., Flores, R. M., & Taioli, E. (2020). Disparities in COVID-19 testing and positivity in New York City. *American journal of preventive medicine*, 59(3), 326-332.
- nychealth. (n.d.). *coronavirus-data/data-by-modzcta.csv at master · nychealth/coronavirus-data*. GitHub. Retrieved September 29, 2022, from <https://github.com/nychealth/coronavirus-data/blob/master/totals/data-by-modzcta.csv>

Seto, E., Min, E., Ingram, C., Cummings, B., & Farquhar, S. A. (2020). Community-level factors associated with COVID-19 cases and testing equity in King County, Washington. *International journal of environmental research and public health*, 17(24), 9516.

Thompson, C. N., Baumgartner, J., Pichardo, C., Toro, B., Li, L., Arciuolo, R., . . . Fine, A. (2020). COVID-19 Outbreak - New York City, February 29-June 1, 2020. *MMWR Morb Mortal Wkly Rep*, 69(46), 1725-1729. doi:10.15585/mmwr.mm6946a2

Images

Figure 1, Selected demographic, testing, and mortality variables by Zip Code Tabulation Areas (ZCTAs).



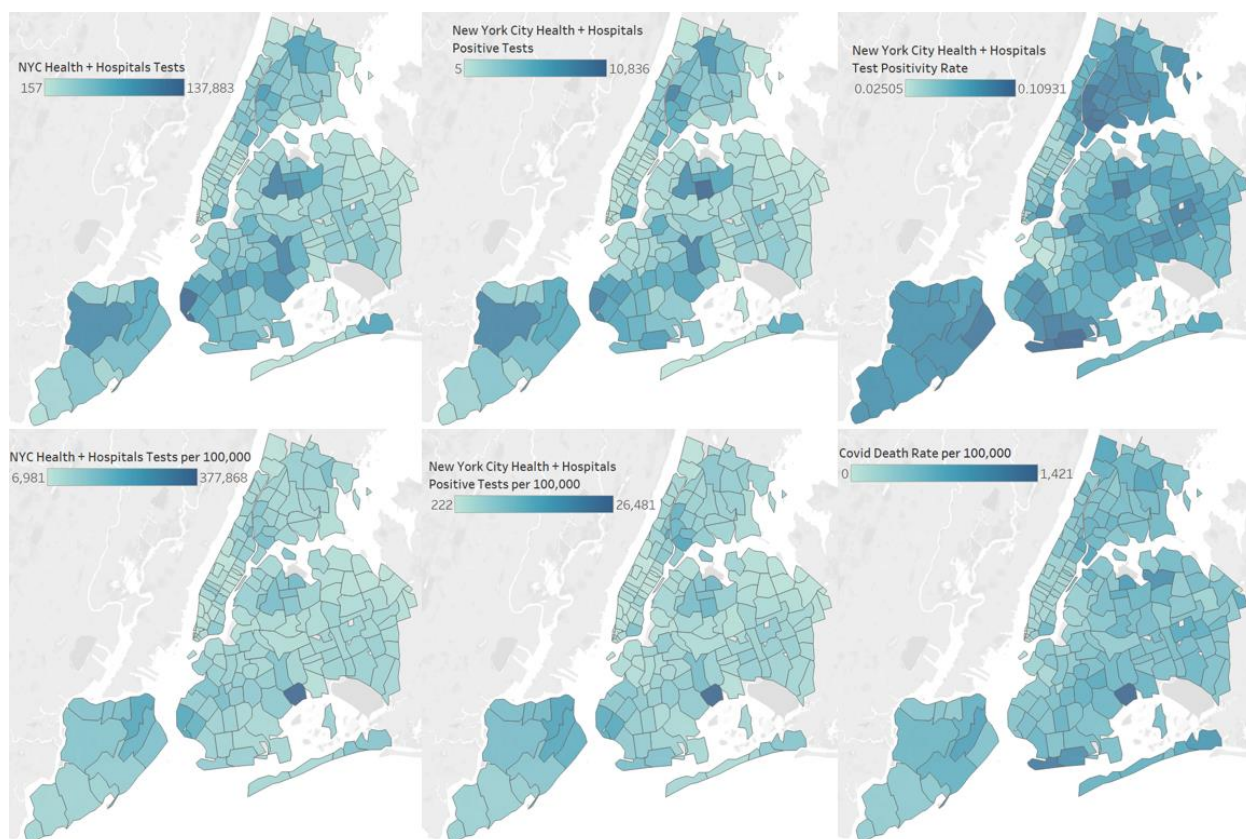


Figure 2

Selected Demographics compared to New York City Health + Hospitals Covid-19 tests per 100,000 by ZCTA

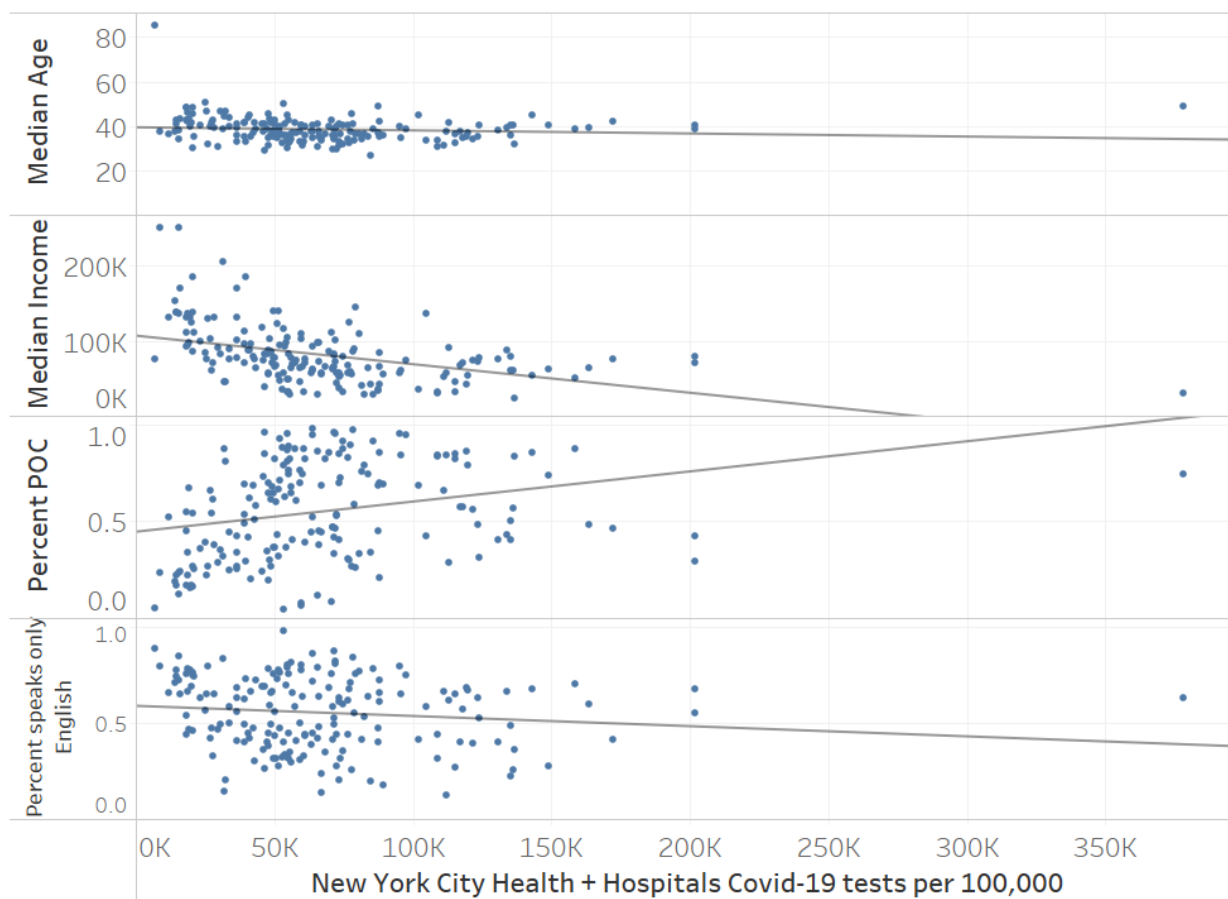


Figure 3

Selected Demographics compared to New York City Health + Hospitals Covid-19 test positivity rate by ZCTA

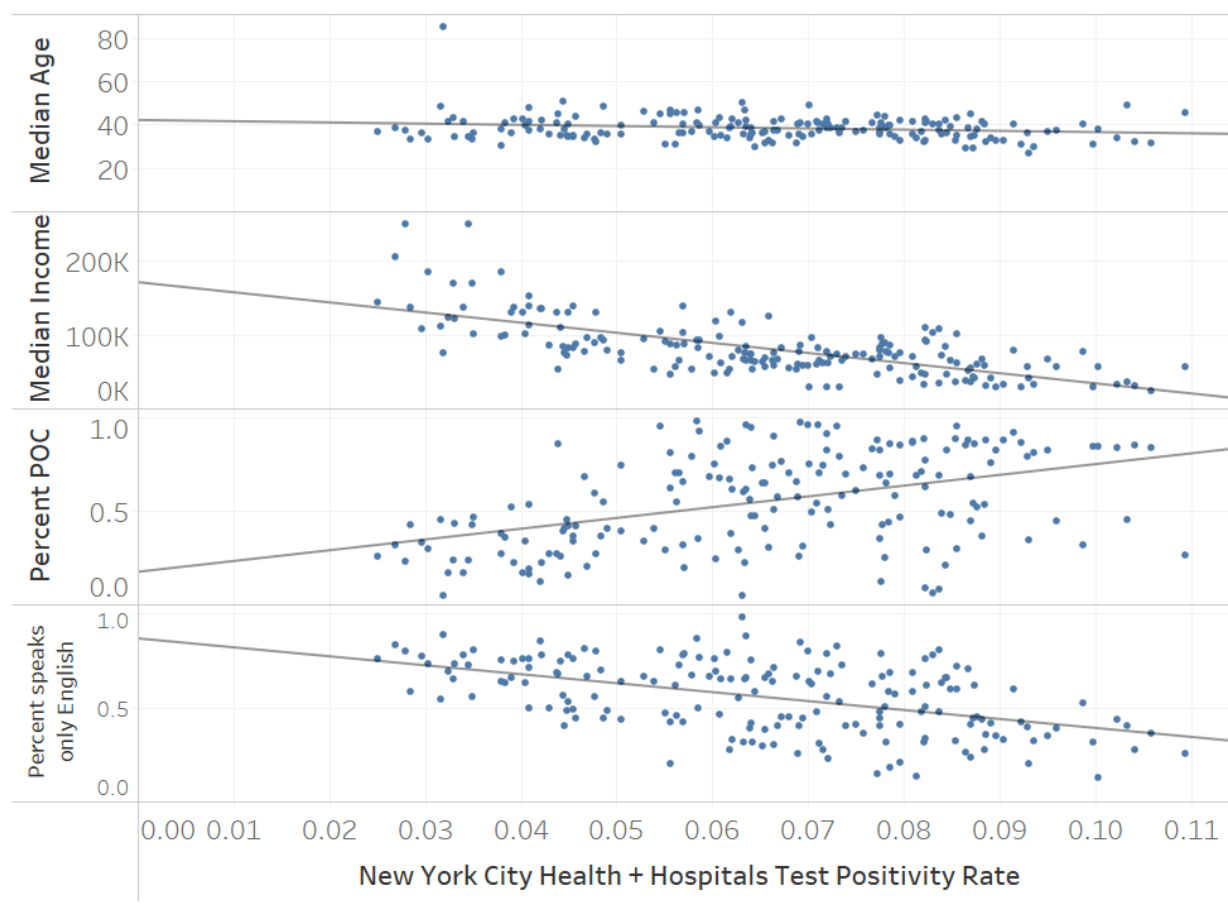


Figure 4

Selected test and death rates compared to New York City Health + Hospitals Covid-19 tests per 100,000 by ZCTA

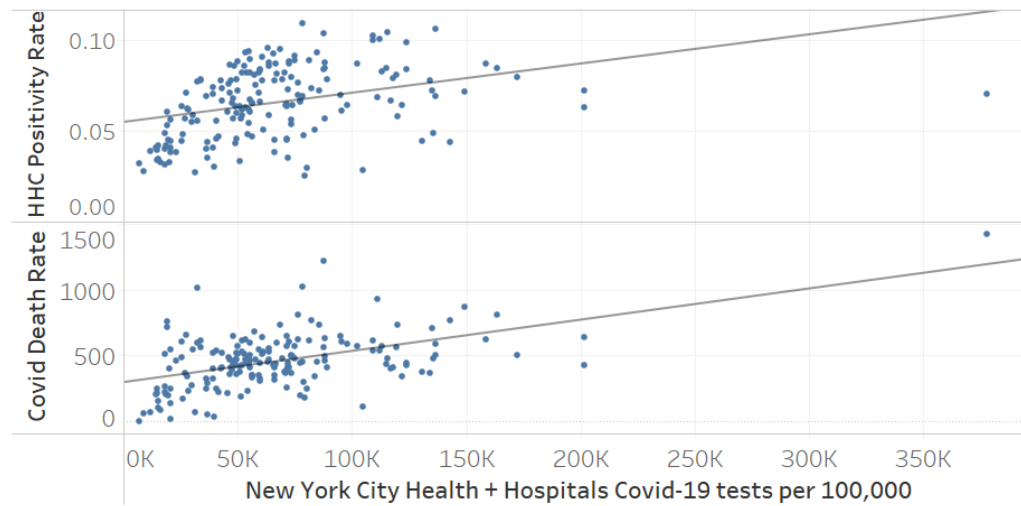
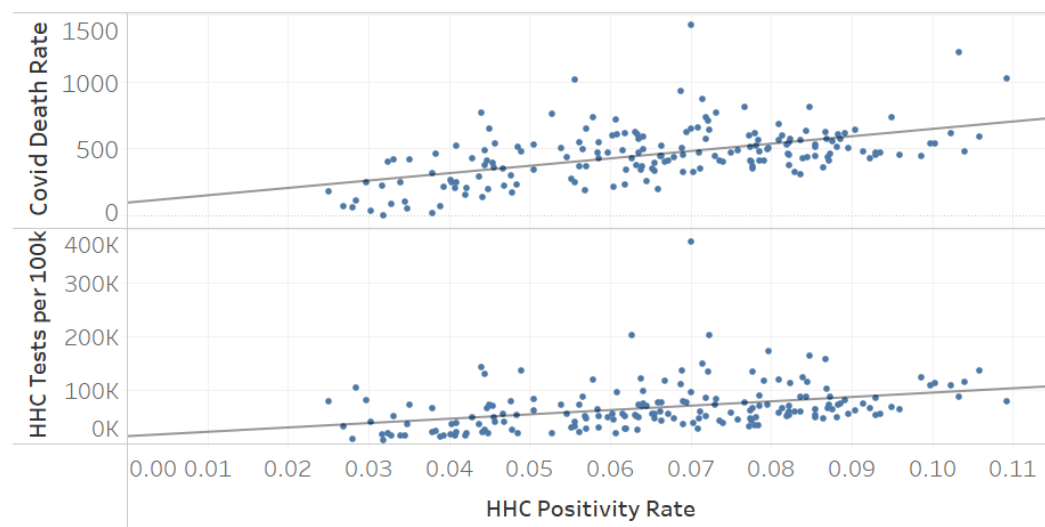


Figure 5

Selected test and death rates compared to New York City Health + Hospitals Covid-19 positivity rate by ZCTA



Appendix

All underlying data by ZCTA can be found on GitHub at

https://github.com/rosenfed/testingequitydata/blob/main/2022-10-12_acsimport.xlsx

Python code: <https://github.com/rosenfed/testingequitydata/blob/main/2022-12-02%20updated%20testing%20equity%20python%20code.ipynb>

```

import pandas as pd
import numpy as np
import statsmodels.api as sm
from statsmodels.genmod.families import Poisson
import matplotlib.pyplot as plt
import scipy.stats as stats
from sklearn.linear_model import PoissonRegressor

plt.rcParams['figure.figsize'] = (20.0, 10.0)

acsimport = pd.read_csv (r"ACSimport4.csv")

#negative binomial model, total tests
famnegbin = sm.families.NegativeBinomial()
offset = np.log(acsimport['total_pop'])
exog = acsimport[['Median Income', 'Median Age', 'Percent Female', 'Percent POC', 'Speaks only
English %']]
endog = acsimport[['updatedhhctests']]
glm_negbin = sm.GLM(endog,
exog,M=sm.robust.norms.HuberT(),family=famnegbin,link='log',offset=offset)
negbin_results=glm_negbin.fit(cov_type="hc1")
print(negbin_results.summary())

# exponentiating results for total tests, repeat for positive tests
print('Median Income')
print(np.exp(-1.007e-05))
#print(np.exp(-0.000),np.exp(0.000))

print('Median Age')
print(np.exp(-0.0115))
#print(np.exp(-0.060),np.exp(-0.048))

print('Percent Female')
print(np.exp(1.0481))
#print(np.exp(-0.000),np.exp(0.000))

print('Percent POC')
print(np.exp(-0.0228 ))
#print(np.exp(-0.0000966),np.exp(-0.000))

print('Speaks only English %')
print(np.exp( 0.4689))
#print(np.exp(-0.000),np.exp(0.000))
#negative binomial model positive casesfamnegbin = sm.families.NegativeBinomial()

```

```
offset = np.log(acsimport['total_pop'])
exog = acsimport[['Median Income', 'Median Age', 'Percent Female', 'Percent POC', 'Speaks only
English %']]
endog = acsimport[['updatedhhcpositivetests']]
glm_negbin = sm.GLM(endog,
exog,M=sm.robust.norms.HuberT(),family=famnegbin,link='log',offset=offset)
negbin_results=glm_negbin.fit(cov_type="hcl")
print(negbin_results.summary())
```