UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF NEW YORK

CONEY ISLAND PREP; LESLIE-BERNARD JOSEPH; HOUSING WORKS, INC.; CHARLES KING; MARK LEVINE; and ALEXANDRA GREENBERG,	:	
Plaintiffs,	:	No. 1:20-cv-_____
-against-	:	
UNITED STATES DEPARTMENT OF HEALTH AND HUMAN SERVICES; ALEX. M. AZAR II, *in his official capacity as Secretary of Health and Human Services*; DR. ROBERT KADLEC, *in his official capacity as Assistant Secretary of Health and Human Services*; CENTERS FOR DISEASE CONTROL AND PREVENTION; DR. ROBERT R. REDFIELD, *in his official capacity as Director for the Centers for Disease Control and Prevention*,	:	
Defendants.	:	

DECLARATION OF DR. MICAELA MARTINEZ

I, Dr. Micaela Martinez, under penalty of perjury, state as follows:

1. I am Assistant Professor in Environmental Health Sciences at Columbia University, Mailman School of Public Health and a Visiting Professor at the University of Surrey, United Kingdom. I earned my PhD in Ecology and Evolutionary Biology at the University of Michigan, after which I spent two years as a post-doctoral fellow at Princeton University. I have received numerous grants for my work, including from the National Science Foundation ("NSF") and the National Institutes of Health ("NIH"). I also serve as a Steering Committee Member for the Models of Infectious Disease Agent Study ("MIDAS") at the

University of Pittsburgh, funded by NIH. I am a resident of Manhattan, New York, and live in the neighborhood of Harlem.

2. As an infectious disease ecologist, I apply ecological and evolutionary principles to the study of pathogens: for example, the drivers of seasonality in infectious disease systems and the impact of biological rhythms (i.e., circadian and circannual rhythms) on disease. My specialty has always been the study of vaccines and the prevention of epidemic-prone diseases, including measles, chicken pox, poliomyelitis, zika, and now SARS-CoV-2 (or "COVID-19"). I also conduct research on maternal immunity to better understand the role of immunity passed from mothers to their infants.

3. I utilize cutting-edge statistical inference techniques and mathematical models, coupling disease incidence data with clinical data to gain insight into the transmission dynamics of disease. My research seeks to reveal how demographic, physiological, and environmental factors shape the path of such diseases. Recently, my work addressed how socioeconomic factors play a role in shaping the COVID-19 pandemic. I also currently work on an NIH-funded project called "Hacking Epidemics" that seeks to understand the seasonality of disease outbreaks—for example, why some infectious diseases may be more prevalent in the winter months than in the summer months—as a way to battle vaccine-preventable diseases.

4. I have studied COVID-19 since the beginning of its outbreak in New York City. I began by collating datasets to track the epidemic using data from New York City and New York state. The state released information county by county, and the city was releasing data by borough and NYC zip code tabulation area (ZCTA). Because COVID-19 data from NYC were originally reported in tables within PDF documents, and state data were posted online in tables embedded within web pages, I manually digitized data each night. During the first few months of the pandemic, I was doing this work personally, and now the daily digitization and reconciliation

is manually performed by an undergraduate research assistant in my lab. My COVID-19 research has focused on New York City and has involved the study of environmental and social factors shaping the dynamics of COVID-19.

5. Early NYC data did not include cases reported by race and ethnicity, although data were reported by age and sex. From the earliest days of the pandemic, before the reporting of cases by race and ethnicity, I suspected disparities along race and ethnicity in the City's COVID-19 exposure, infection, and outcomes due to the geographic distribution of cases, hospitalizations, and deaths, with Queens being hit hard early in the pandemic. The city released data by borough, and although Manhattan had the first cases in the city, the number of cases, hospitalizations, and deaths quickly started to stand out as being high in the Bronx and Queens, which have predominantly Black and Latinx populations.

6. By April 5th 2020, the Bronx and Queens had the highest COVID-19 death rates (deaths per 100,000 population) in the city, roughly twice that of Manhattan with the lowest death rate. The Bronx and Queens also had the highest hospitalization rates. As NYC mortality data on race and ethnicity began to be reported the next day (April 6th), it revealed the disparities between boroughs: Black and Latinx New Yorkers were roughly twice as likely to die from COVID-19 as their White counterparts.

7. In April, my findings were presented to the New York City Council to assist its response to the pandemic.

8. In the recent American experience, most acute infectious disease outbreaks do not present disparities by race and ethnicity, but COVID-19 was an exception. The data suggest that this was the result of environmental and structural racism: the impact of discriminatory policies and practices on the lives of people of color, such as higher levels of pollution; more crowded housing; greater work in "essential" occupations such as food service, retail and construction; a

greater reliance on public transportation; diminished access to health resources; and fewer opportunities to access healthy food and green spaces. These environmental factors contribute to higher rates of pre-existing conditions and comorbidities like asthma, diabetes, hypertension and obesity within communities of color; these medical conditions are high risk factors for adverse outcomes during SARS-CoV-2 infection.

9. The NYC Community Health Profiles clearly demonstrate how environmental and structural racism impacts health. *See* NYC Health, *Community Health Profiles: Mott Haven and Melrose* (2018), available at https://www1.nyc.gov/assets/doh/downloads/pdf/data/2018chp-bx1.pdf. In the South Bronx (i.e., Mott Haven and Melrose) where 73% of the population is Hispanic/Latino and 24% is Black/African American: 29% of people live in poverty (compared to 7% in the Upper East Side of Manhattan); it has elevated levels of air pollution; 76% of rented homes have maintenance defects and 40% have roaches; the incarceration rate is 17 times higher than that of the Upper East Side of Manhattan; child asthma emergency department visit rates are 23 times higher than that of lower Manhattan; and for every supermarket there are 25 bodegas (which are less likely to have healthy food). As a result, 32% of Bronxites are obese, 16% have diabetes and 36% have hypertension and the life expectancy is 77.6 years, 3.6 years shorter than NYC overall. Nearly every resident (97%) of the South Bronx is Black or Latinx.

10. It is no surprise then that neighborhoods like the South Bronx were especially hard hit with COVID-19. Given the virus' high transmissibility and unique pathology—its ability to transmit asymptomatically and damage the body via an overreactive hyperinflammatory immune response—Black and Latinx New Yorkers faced dual risks: greater exposure (due to essential work and crowding conditions) leading to a higher risk of infection, and greater risk of adverse outcomes once infected (due to pre-existing conditions).

11. My colleagues and I used data from the Metropolitan Transit Authority ("MTA") to analyze subway use in 126 neighborhoods (i.e., ZCTAs) in NYC in the first months of the pandemic. We also used COVID-19 case data and demographic data from the U.S. Census Bureau. A true and correct copy of a manuscript reporting my team's findings, which is available at https://www.medrxiv.org/content/10.1101/2020.05.28.20115949v1.full.pdf, is attached hereto as Exhibit A. This analysis produced a number of critical insights. Subway use declined earlier and more rapidly in wealthier neighborhoods; by contrast, poorer neighborhoods, communities of color, and those with a higher percentage of essential workers had more subway use during the height of the pandemic. The latter neighborhoods had a greater percentage of confirmed cases of COVID-19.

12. Our data revealed a phenomenon we called "social distancing inequity"—various factors, such as your race, where you live, your income, your subway use, and whether you perform "essential" work impact your potential to contract COVID-19. For example: lower income neighborhoods in NYC have carried a higher burden of COVID-19; these are also communities of color, and the communities carrying the burden of essential work for the city; these communities continue to use the subway (presumably out of need, not by choice). The privilege of social distancing is not a luxury afforded to everyone. In order to keep their families afloat, many essential workers *must* leave their homes and ride public transit with others who are similarly situated (compounding everyone's risk of infection). As a result, these are communities carrying layered exposures and vulnerabilities to COVID -19.

13. Our MTA study focused on how mobility (i.e., subway use) impacts transmission within neighborhoods and the association with neighborhood demographics (e.g., race and ethnicity). Our study was a necessary substitute for understanding how race and socioeconomics impacted COVID-19 cases, at a time when race and ethnicity data were not being reported for

COVID-19 cases; it was only being reported for deaths and the focus of the disparity was pre-existing conditions. Even in New York, where nightly case data were available by age, sex, and borough in March, information with respect to race and ethnicity was only released with respect to deaths on April 6th, and for all cases on May 18th (where they continue to be posted on the NYC Github site). *See NYC Coronavirus Disease 2019 (COVID-19) Data*, available at https://github.com/nychealth/coronavirus-data. This means that for the crucial first months of the pandemic we were missing data that could have helped us understand how to mitigate racial and ethnic disparities. My study (and others like it) provided ways to puzzle piece together useful insights from the imperfect data that were available, but there should have been a plan in place to ensure that data stratified by race and ethnicity were made readily available for cases, hospitalizations, deaths, and testing.

14. Then and now, research on the disparate impact of COVID-19 is subject to the quality and scope of testing and reporting. It has been widely reported that, in the early months of the pandemic, testing materials were in short supply in New York, and that results were delayed significantly. I have personally experienced a more than ten-day wait for test results from an official COVID-19 testing center in Harlem, NYC. Add to that the public health guidance that encouraged people to abstain from seeking treatment until they were confident their symptoms needed care. Many cases went untested and unreported altogether. According to the CDC Seroprevalence Survey Data, by May 6th, 23.2% of New Yorkers had been infected with SARS-CoV-2, which was at least 10 times higher than the number of reported cases. *See* CDC COVID Data Tracker, *Commercial Laboratory Seroprevalence Survey Data*, available at https://covid.cdc.gov/covid-data-tracker/?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fcases-updates%2Fcommercial-labs-interactive-serology-dashboard.html#serology-

surveillance. This discrepancy between the number of infections and reported cases may be exacerbated in communities with fewer health resources and where seeking medical care can be difficult or discouraged. Accordingly, colleagues and I have recommended that there be increased testing and access to medical care, especially in the hardest hit and most vulnerable neighborhoods. This means paying special attention to assisting NYCHA residents and neighborhoods with the highest risk of exposure (i.e., due to essential work) and pre-existing conditions.

15. My colleagues and I have suggested a number of other policy interventions would be effective: 1) identifying at-risk households and coordinating efforts to provide services or run errands on their behalf during lockdowns; 2) strategically setting up field hospitals to provide accessible facilities for hard-hit neighborhoods and relieve the burden on local hospitals; 3) providing supplies directly to housing complexes so that food security and other basic needs are met; 4) subsidizing and safeguarding WiFi and other communications in these areas to ensure access to health information and services; and 5) creating multilingual services information to be posted in high-traffic public areas. Such efforts would help residents bridge the inequalities that COVID-19 has revealed and help insulate the most vulnerable from exposure, illness, and death.

16. At the federal level, I am very familiar with notifiable disease data reporting by CDC. I, and other infectious disease modelers, require robust, granular and reliable data for our work, and, in my experience, the CDC is the federal agency that compiles the highest quality data on infectious disease outbreaks. My work modeling epidemic-prone diseases, for example, extensively used the CDC's datasets. The public health community, especially those of us using large volumes of data, rely on the CDC in order to have the information we need to make progress in our research.

17. Because of my involvement with MIDAS, I participated in regular teleconferences that included CDC officials, medical experts, and public health researchers. In the early stages of the pandemic, those meetings occurred two-to-three times per week and became less frequent in the months since. Attendees would use these meetings as an opportunity to present modelling results, and the CDC would inform us of their new data websites intended for public access (such as the Seroprevalence Survey website mentioned above).

18. Generally speaking, during the pandemic, the CDC's data has not been as up-to-date or detailed as the NYC or NY State data. Data often lagged by days or weeks, limiting its usefulness for public health experts and researchers like me modelling the outbreak and evaluating potential countermeasures. The CDC Wonder database, which reports mortality by underlying cause of death (including infectious diseases) by state, county, and census region, has lagged behind by nearly 2 years, with data only available through 2018. The National Notifiable Disease Surveillance System (NNDS) weekly tables include a column for SARS-CoV "*Severe acute respiratory syndrome-associated coronavirus disease*", but do not appear to report SARS-CoV-2/COVID-19. Until June, the CDC's real-time data was significantly behind the Johns Hopkins tracker. CDC serology data (regarding COVID-19 antibodies) was also not typically up to date and I would have to get updated numbers from the NY Governor's COVID-19 press briefs, as opposed to CDC's website. In response to criticism of these delays during MIDAS meetings, CDC representatives have responded that they must wait to get approval before they could make any new tools or data available to the public.

19. In my experience, CDC did not raise the issue of racial and ethnic disparities in COVID-19 morbidity and mortality until recently. Data stratified by race/ethnicity was by and large unavailable during the early surges of the pandemic and continues to be unevenly reported and disclosed. Importantly, the CDC's Human Infection with 2019 Novel Coronavirus Case

Report Form has a section for reporting race and ethnicity, but such data have not been available until recent months. *See* CDC, *Human Infection with 2019 Novel Coronavirus Case Report Form*, available at: https://www.cdc.gov/coronavirus/2019-ncov/downloads/pui-form.pdf. The CDC COVID Tracker now supplies the national totals for cases and deaths by race/ethnicity after public reporting revealed such data collection efforts had fallen severely short. *See* CDC COVID Tracker, available at: https://covid.cdc.gov/covid-data-tracker/#demographics.

20. The data available, however, on the CDC's website is presented in a misleading way. Racial and ethnic information is presented in raw numbers and as a percentage of total cases nationwide, rather as a percentage of that racial or ethnic group. This has the effect of downplaying the severity of minority health disparities and obscuring the impact of COVID-19 on communities of color. The CDC has now made available the "COVID-19 Case Surveillance Public Use Data" (available at: https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data/vbim-akqf) which is patient-level data that includes sex, age, and race/ethnicity variables. As of Oct 26th, this dataset had not been updated in over three weeks (since Sept 30th). More importantly, of the 4,481,062 cases reported in the dataset 1,895,793 had race/ethnicity reported as "Unknown"; yet, only 5,188 had age listed as "Unknown". This clearly shows that CDC, and those reporting cases, have prioritized some demographic information (i.e., age) over others (i.e., race/ethnicity). I have not seen any indication that CDC intends to backfill the missing race and ethnicity data—leaving a crucial gap in the data I would need for my research.

21. Another problem is that CDC largely relies on states to report case data, and state reporting of racial and ethnic information with respect to COVID-19 has not been complete. Some states may not report these data at all to the CDC, since nearly 2 million of the cases in the CDC's COVID-19 Case Surveillance Public Use Data are missing that information entirely.

22. *The Atlantic*'s COVID Tracking project provides a useful comparison, not only because the project updates twice weekly with data, but also because the CDC website directs visitors to *The Atlantic*'s online database. *The Atlantic* requests race and ethnicity data for COVID-19 cases and deaths from each U.S. state and territory. *See COVID Tracking project*, THE ATLANTIC, available at https://covidtracking.com/race/dashboard. *The Atlantic* requests these directly from state public health officials and governors (not the CDC). According to *The Atlantic*'s data, some states have reported race and ethnicity in high fidelity.

23. For instance, Washington DC has reported race and ethnicity for 99% of cases and 99% of COVID-19 deaths, while New York has not reported race and ethnicity for any of the cases, but it has reported race and ethnicity for 91% of deaths. New York State's COVID-19 dashboard also only reports race and ethnicity for deaths and not cases. *See* N.Y. Dep't of Health, *COVID-19 Tracker*, available at https://covid19tracker.health.ny.gov/views/NYS-COVID19-Tracker/NYSDOHCOVID-19Tracker-Map?%3Aembed=yes&%3Atoolbar=no&%3Atabs=n.

24. It is not possible to tell which states are reporting race/ethnicity to the CDC since the CDC's COVID-19 Case Surveillance Public Use Data does not provide geographic information such as state, county, or ZIP code. However, given the gaps in the CDC's COVID-19 data with respect to race and ethnicity and the degree to which states are inconsistently reporting to the *Atlantic*'s project, I assume that the same states are likely reporting inconsistently to the CDC.

25. Different states' health agencies use different racial and ethnic criteria, making it difficult to harmonize data from different locations. For example, some report "Hispanic/Latino" as a race like Black/African American or White, although "Hispanic/Latino" is an ethnicity and Hispanic/Latino people can also be of any race. This can make it difficult not only to harmonize

data across agencies, but also to adjust cases by population size for each racial/ethnic group using data from the US Census.

26. Even in locations that have provided data by race/ethnicity, COVID-19 cases in Native Americans/Indigenous people have not been reported. For instance, the NY State COVID-19 tracker lists deaths by race as Hispanic, Black, White, Asian, or Other. This pushes Native Americans and Indigenous peoples into the "Other" category, which masks disparities that are unique to individual racial groups. "American Indian or Alaska Native" and "Native Hawaiian or Other Pacific Islander" should be reported as individual categories. This problem was highlighted in a recent report in the journal *Science* which states:

> "Today, American Indians and Alaska Natives have higher rates of obesity, diabetes, asthma, and heart disease than white Americans, as well as higher rates of suicide. The system of oppression in the United States, Echo-Hawk says, '*has built a perfect environment to kill us in a pandemic*.'"

See Lizzie Wade, *COVID-19 data on Native Americans is 'a national disgrace.' This scientist is fighting to be counted*, SCIENCE, Sep. 24, 2020, *available at* *https://www.sciencemag.org/news/2020/09/covid-19-data-native-americans-national-disgrace-scientist-fighting-be-counted*. The embedded quote is from Abigail Echo-Hawk, the director of the Urban Indian Health Institute (UIHI) and the chief research officer for the Seattle Indian Health Board. The full impact of COVID-19 on Indigenous communities remains unknown due to racial misclassification and the exclusion of indigenous communities from data sets and analysis, such as has occurred in my own work in New York, where "American Indian or Alaska Native" and "Native Hawaiian or Other Pacific Islander" are not included in the COVID-19 data, even though these are official race categories in the US and these categories are included on the official CDC COVID-19 case report form.

27. I also study the impact of vaccines on epidemic diseases and am particularly concerned about the effect that unreliable or missing data will have on the distribution and

effectiveness of a vaccine in the United States. We need to better understand the disparate impacts of the virus, especially within communities vulnerable to exposure and death. In order to direct resources to those who need it most, we need a better understanding of the distribution of cases among communities through time. Black, Latinx, and Indigenous people have been disproportionately impacted by this pandemic, and have also faced both historical and contemporary abuses by the U.S. medical establishment and the CDC (with the notorious Tuskegee Experiments being an example). Without reliable data, I am concerned that vaccine distribution plans will not sufficiently weigh the needs of communities of color, sensitivities surrounding stigmatization, and vaccine hesitancy that may arise due to historical and contemporary abuses, resulting in further inequities during this health crisis.

28. The CDC's databases are uniquely valuable to public health researchers and cannot easily substituted with state and local data. Many state health agencies, like New York, do not digitize their data in downloadable excel or csv files; instead, they are posted on COVID-19 online dashboards (as cumulative totals, rather than time series and/or the new number of cases daily). For example, were I to attempt to study the racial and ethnic disparities nationally that I studied in New York, even if all states/counties were reporting cases by race/ethnicity on their websites, it would be impossible for my lab to scrape the data from all state and local jurisdictions daily, and backtrack all the prior data using internet webpage archivers to get the case counts going back to March. There is not enough time in the day. If the CDC does not fill in "Unknown" race/ethnicity within the COVID-19 Case Surveillance Public Use Data and also release geographic information for cases (i.e., at least the state where the case was reported), these data will remain hidden and there is no adequate substitute that can be cobbled together.

29. I have also reviewed the new HHS Protect database from which the CDC's National Healthcare Safety Network migrated its collection from hospitals with respect to

COVID-19 cases, treatment and deaths. I have designed multiple websites over the years, and in my opinion the layout of HHS Protect resists the user's easy access to the underlying data. In my assessment, HHS Protect is a problematic data portal, and it is not designed to make line-by-line case data readily available for public access let alone public health research. The prominent pages are focused mainly on representing the fidelity of reporting (e.g., what percentage of hospitals are reporting data to HHS and the number of data fields each hospital entered). Similarly, for COVID-19 testing data, some time series are provided regarding testing. However, these data do not provide any information beyond the state and date on which testing results were reported. This ignores a number of criteria worthy of collection, as I have emphasized in detail above—not just by race and ethnicity, but localized geographic information, such as zip code.

30. As an infectious disease ecologist who collaborates extensively with federal health agencies already, I would be eager to attend and participate in any and all opportunities that the government provides where I could help establish standards and guidelines related to data collection and reporting.

31. The lack of reliable and granular nationwide data has hampered my research into COVID-19, intended to help health officials and the public to understand and better mitigate the effects of the pandemic.

I swear under penalty of perjury pursuant to 28 U.S.C. § 1746 that the foregoing is true and correct.

Dated: November 2, 2020

DR. MICAELA MARTINEZ