The Influence of Mobility Trends on COVID-19 transmission

Muqing Fan, Yuchen (Eva) Liu, Xiangyu (Alaric) Wang, Juanshu Wu, Hyojeong (Jenny) Yoo

I. Introduction

In light of record-breaking COVID-19 cases in the U.S, the control of spreading, accurately estimating human mobility and measuring its relationship with virus transmission is crucial. Various studies investigated the influence of mobility on COVID-19 transmission, which suggest that mobility has a strong positive correlation with the number of COVID-19 cases in most U.S. counties. [3,6]

However, there are some limitations of previous U.S.-specific research on the relationship between mobility and COVID-19 transmission. Most of these studies focus on the early stages of COVID-19 breakout. However, the more recent phase of COVID-19 may suggest a different relationship of covid transmission and mobility. Moreover, the mobility data that previous research uses does not differentiate between mobility types, especially low-risk and high-risk mobility types (e.g. a trip to park might be lower risk than a trip to the grocery store). [9] Therefore, it may be more informative to categorize different types of mobility and investigate whether different types of mobility influence COVID-19 transmission differently. Lastly, previous studies do not address how sociodemographic factors may have influenced the relationship between COVID-19 transmission and mobility. It is worth investigating how sociodemographic factors such as low-income rate or elderly rate contribute to the relationship between mobility and COVID-19 transmission.

Overall, our aim for 201A thus builds on the knowledge gap of previous covid-mobility studies. Particularly, we plan to investigate the relationship between mobility and COVID-19 transmission after the early breakout phase. Moreover, we aim to involve sociodemographic factors in our study to see how it influences the relationship between mobility and COVID-19 transmission. We also aim to include mobility data with specific categories to distinguish which category could influence COVID-19 transmission the most.

II. Data collection

We collected data from 3 sources: the Social Vulnerability Index of 2018 (SVI) provided by Centers for Disease Control and Prevention (CDC) [2], Community Mobility Reports of 2020 by Google [7], and the COVID-19 Data Repository from the Johns Hopkins University, Center for Systems Science and Engineering (CSSE) [5].

As aforementioned, much of the current research has several limitations in that they do not take into account the uniqueness of each state in terms of its social vulnerability. Social vulnerability, as defined by the CDC, is the impact factor of natural, man-made disasters, or disease outbreaks on a particular region or community, determined by 15 social factors such as poverty, minority status, and per capita income [1]. The SVI ranks communities based on these social factors to discern which regions are in most need of assistance [1]. Pertinent to our research, the SVI dataset provides crucial information on which states may receive the highest impact due to COVID-19, which may also affect mobility trends.

The Community Mobility Reports contain information on mobility trends in both US county and state level on 6 specific categories: retail and recreation, grocery and pharmacy, parks, transit stations, workplaces, and stay-at-home rates [7]. A particular number of a given day in the dataset is calculated as the percent change from a baseline day, which is “the median value from the 5-week period” between January 3rd and February 6th, 2020 [7]. Since each day of the week (e.g.Monday vs Saturday) has different baselines (7 baselines in total), our data analysis did not include a day-to-day difference in mobility, but rather a biweekly average percent change.

The COVID-19 Data Repository by JHU CSSE includes comprehensive, daily updated time-series data on statistics such as hospitalized rates, and recovered rates, active cases, deaths, confirmed cases, positive test rates, and case fatality ratio on a worldwide level. Among these statistics, only deaths and confirmed cases have historical data from Jan 22, 2020.

III. Data clean up

Within the SVI dataset, we extracted 7 relevant factors that may have a larger impact on mobility and abilities of prevention measures of COVID-19: population density, poverty, uninsured, minority, disabled, elderly, education level (no high school diploma at age 25 or older), and per capita income. Among these factors, PCI and population density spanned the largest magnitudes. Therefore, we log-transformed PCI and population density to ensure the numerical results of the linear regressions are within a more manageable range. For all other variables, the percent rate for all values within each variable was calculated by taking into account the population of each 51 states (New Mexico was excluded because there were many missing values). This SVI dataset is used as categorical data to later categorize the states in regards to social vulnerability.

In the mobility dataset, we took out the non-responsive data points and selected only mobility percentage changes on the state level. Biweekly average mobility percent change in each category was also computed. Note that our main analysis was based on average percent rates for COVID-19 active cases/mortality and average percent rate change for mobility. To be consistent with the time frame of our COVID-19 dataset, we only included mobility data from April 13th to October 24th.

For our research purposes, we modified the COVID-19 dataset to only include US cases on a state level. Both the COVID-19 and mobility dataset contain values starting from January 2020, but there are many values missing (e.g. active cases) within the COVID-19 data until April 12th. Thus, we omitted the data before April 12th in further data analysis. Since our investigation centered on the influence of mobility on covid transmission trend, the cumulative confirmed cases might not be a good indicator of covid severity as the cumulative confirmed cases are always increasing regardless of the number of people that are currently contracting COVID-19. Similarly, the confirmed cases each day indicate how many people are newly contracting COVID-19 but fail to capture the previous confirmed cases that are still active and have the danger of transmission. Thus, what may be more informative of current COVID-19 transmission trend is the current active case, which equals to the cumulative confirmed cases excluding the cumulative death cases and cumulative recovered cases. In order to capture the active rate of the state, we divided the active cases of each day by the total population of the state to form the rate of active cases per day of the state. In order to take into account the various delayed onset of symptoms and the lag effect of COVID (median incubation period of 5.1, whereas the 95th percentile was 11.7 days) [8], we grouped active cases to coarser bins by computing the biweekly percent averages of active case rate. We also included mortality rate as another main variable for COVID-19 since mortality rate reflects how vulnerable people are when contracting COVID-19. We chose the mortality rate from the last date October 24th in our COVID-19 dataset, which equals the total number of cumulative deaths divided by total number of cumulative confirmed cases by the last date in our COVID-19 dataset.

IV. Data analysis

IV.a. Relationship between Social Vulnerability Index and COVID-19 mortality rates

To incorporate sociodemographic factors in our main data analysis between mobility and COVID-19 transmission, we first conducted a preliminary analysis to categorize high and low risk states. In order to figure out which SVI has the strongest influence on COVID-19, we performed a linear regression of COVID-19 mortality rates on selected SVIs and included only the SVI category with the largest R-squared value. As SVIs are categorical variables, we incorporated a non-time series COVID-19 data (i.e. mortality rates) to select the SVI factor that may have the highest impact. We assumed that SVIs have a similar relationship to mortality rates and active rates. In this case, a highly populated state potentially has similar trends for active rates and mortality rates. An individual in highly-populated states might have less access to hospitalization resources and higher chances of contracting COVID-19. Here, R-squared was used as our indicator of which SVI category had the most predictive power. Our linear regressions indicated that log population density had the largest R-squared value (p = 1.36e-10, r^2 = 0.580), suggesting that it had the most predictive power of COVID-19 mortality rates. We then selected the top and bottom 10 states based on log population density with the top 10 states being at high risk and the bottom 10 states being at low risk.

IV.b. Relationship between Mobility Data and COVID-19 Active Case Rates

Our main data analysis was to determine which one of the 6 categories in the mobility dataset is most predictive of the COVID-19 daily active case rates, and how the predictive powers differ between the high-risk vs. low-risk states. We did a partial regression analysis to investigate the effect of specific mobility category on covid active rates after factoring out the average mobility. To do this, we examined the proportion of active rates that cannot be explained by average mobility rates, but can be explained by the individual mobility categories, which is referred to as the *disproportionate effect*. This allowed us to investigate the unique contribution of each mobility category to active rates, separated from the average mobility trend.

Rank	High-risk states (R^2)	Low-risk states (R^2)
1	Residential 0.9056 ***	Residential 0.658 ***
2	Retail and recreation 0.891 ***	Workplace 0.6329 ***
3	Workplace 0.8148 ***	Transit 0.6077 **
4	Transit 0.7711 ***	Parks 0.5391 **
5	Parks 0.7102 ***	Retail and recreation 0.4398 ***
6	Grocery 0.04013 not significant	Grocery 0.1935, not significant

Table 1: linear model examining disproportionate effect (lm(residuals(lm((Mobility Type ~ Average Mobility) ~residuals(lm(Active Rate ~ Average Mobility)) and reported R^2 values for high population density states (high risk) and low population density states (low risk).

** $p < 0.001$, *** $p = 0$.

We first constructed a general linear model of average mobility activities on COVID-19 active rates (e.g. lm(Active Rate ~ Average Mobility)). Then we performed a set of linear regression of each category of mobilities on the average mobility activities (e.g. lm((Mobility Type ~ Average Mobility Percent Change)). We extracted the residuals from the models above, then performed a final linear regression of the residuals of the individual linear model on the residuals of the general linear model, respectively for all individual mobility categories (e.g. lm(residuals(lm(Mobility Type ~ Average Mobility Percent Change)) ~ residuals(lm(Active Rate ~ Average Mobility Percent Change)))). Here, r^2 represents the proportion of variance that can be explained by the specific mobility type after excluding the variance that is explained by average mobility. Except for the grocery mobility category, we found that the disproportionate effects of each category were all significant ($p < 0.001$). Mobility categories in high-risk states appear to have higher r^2 than in low-risk states. Our findings suggested that grocery mobilities do not have a unique contribution to covid active rate separating from the overall mobility. Subtracting the influence of the overall mobility, each mobility type also appears to have a higher contribution to covid active rate in high risk states than low risk states.

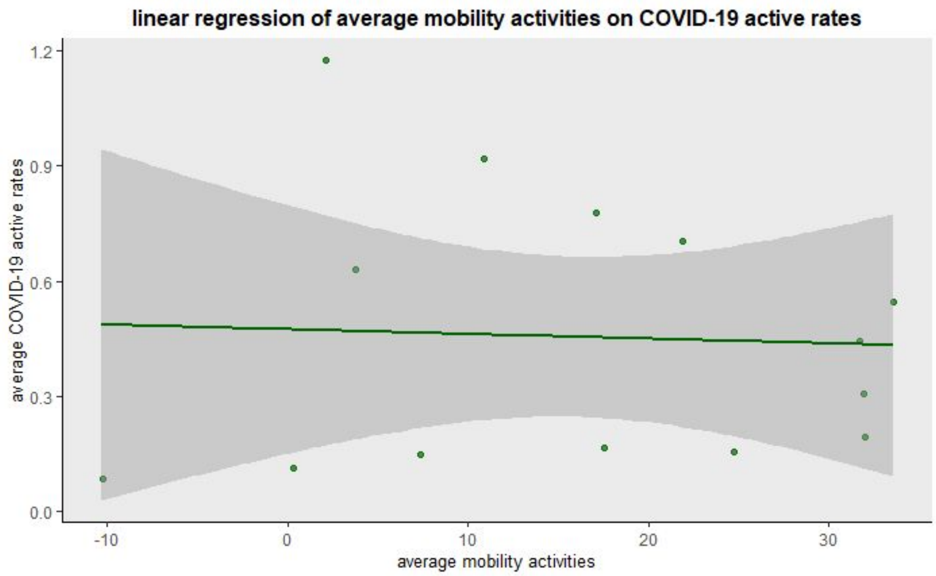

Figure.1 The relationship between average mobility activities and Covid-19 active rates.

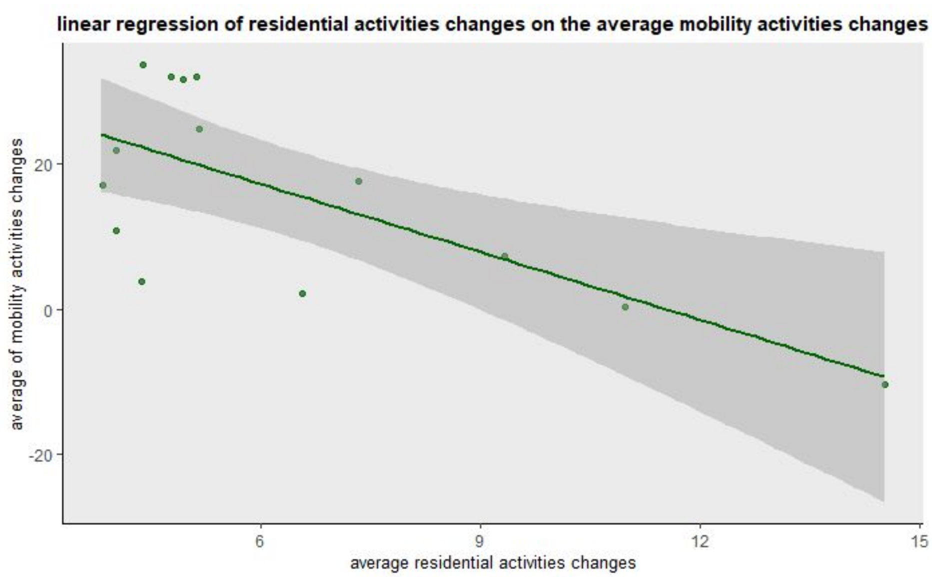

Figure.2 The relationship between average residential activities change and mobility activities change.

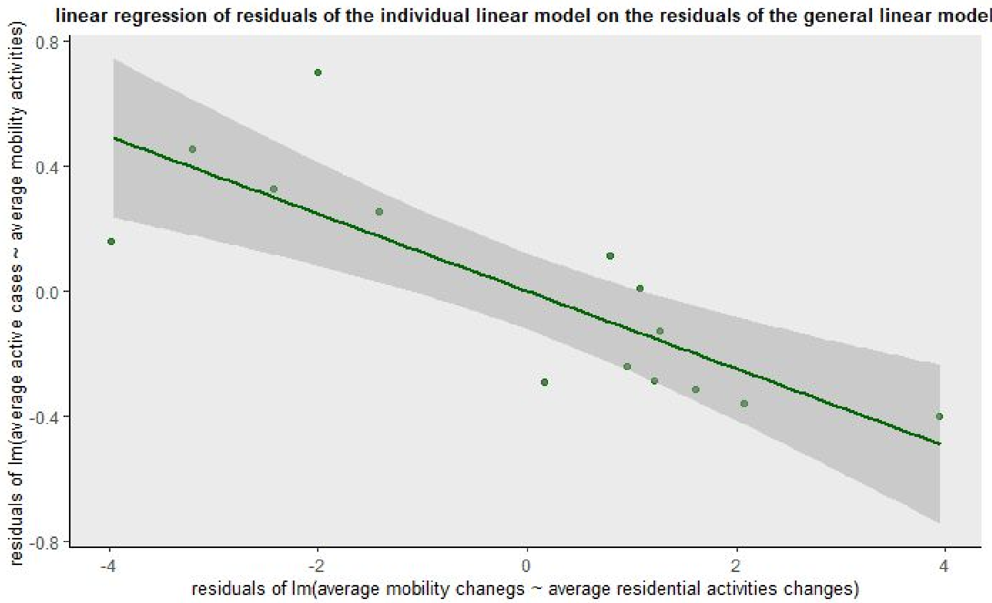

Figure.3 A partial regression model showed a linear relationship between the residuals from the two models, suggesting a disproportionate effect of residential activities changes.

V. Discussion

Through a partial regression analysis, we were able to find differences in disproportionate mobility for more socially vulnerable states vs. less vulnerable states, as defined by the state's population density. One limitation in our study is that when determining which SVI factor to use in categorizing the states, we ran a linear regression with each of the SVI factors to mortality rates instead of active cases, which is the main variable of interest. This is largely due to the fact that while SVI factors are not a time series, active cases in the COVID-19 dataset is a time series and mortality rates are not. Thus, we assumed that the SVI factors have the same relationship to mortality rates as to active rates. We hope to address this limitation in 201b and implement a better method to categorize states as high and low social vulnerability, and possibly conduct multidimensional modeling that incorporates these demographic data more than binary categorization into high vs. low risk states.

References

[1] Agency for Toxic Substances and Disease Registry (ATSDR). (2018). *CDC SVI Data and Documentation*. https://www.atsdr.cdc.gov/placeandhealth/svi/data_documentation_download.html

[2] Agency for Toxic Substances and Disease Registry (ATSDR). (2020, September 15). *CDC's Social Vulnerability Index (SVI)*. https://www.atsdr.cdc.gov/placeandhealth/svi/index.html

[3] Badr, H. S., Du, H., Marshall, M., Dong, E., Squire, M. M., & Gardner, L. M. (2020). Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases*, *20*(11), 1247-1254.

[4] 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.

[5] Du, H. D., Goyea, T. G., Katz, A. K., Dong, E. D., Zhang, A. Z., & Hang, Y. H. (2020, October 24). *CSSEGISandData/COVID-19*. GitHub. https://github.com/CSSEGISandData/COVID-19

[6] Gatalo, O., Tseng, K., Hamilton, A., Lin, G., & Klein, E. (2020). Associations between phone mobility data and COVID-19 cases. *The Lancet. Infectious Diseases*.

[7] Google. (2020, October 23). *COVID-19 Community Mobility Report*. Overview - Community Mobility Reports Help. https://support.google.com/covid19-mobility/answer/9824897?hl=en&ref_topic=9822927

[8] McAloon, C., Collins, Á., Hunt, K., Barber, A., Byrne, A. W., Butler, F., ... & Wall, P. (2020). Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. *BMJ open*, *10*(8), e039652.

[9] Vanselow, M. V. (2020, July 8). *Which activities put you at the greatest risk of contracting COVID-19? Texas doctors rank them*. KWTX. https://www.kwtx.com/2020/07/08/activities-ranked-by-covid-19-risk-level