Factors Influencing Human Mobility During The COVID-19 Pandemic in Selected Countries of Europe and North America

Maryam Sadat Hosseini, Alice Masterangelo Gittler
Systems Science and Industrial Engineering Department
Binghamton University
Binghamton, NY, USA
{mhossei1, agittle2}@binghamton.edu

2020 IEEE International Conference on Big Data (Big Data) | 978-1-7281-6251-5/20/$31.00 ©2020 IEEE | DOI: 10.1109/BigData50022.2020.9377932

Abstract— **A novel coronavirus (COVID-19) was first reported in Wuhan, China in December 2109 and was declared a global pandemic by the World Health Organization on 11 March 2020. Identification of the critical factors that predict reduced mobility and human interaction is critical to developing successful transmission mitigation efforts globally. Governments and localities around the world have responded with wide-ranging policies related to containment and closure such as travel restrictions and stay-at-home-orders, as well as economic benefits, expanded testing, and public health education, among others. Anonymized GPS-enabled smartphone data is a novel tool to track human mobility and is becoming widely available. This study explores the relationships between containment and closure policies, disease trends, and human mobility patterns in 40 countries in Western, Eastern, Northern, and Southern Europe and North America. The principal component analysis was applied to reduce variable dimensions, followed by multivariate multiple regression. The model parameter estimations indicate that total cases, canceling activities (school, work, events), mask policies and the pandemic declaration all were significant predictors (p<.001) of change in workplace mobility from baseline.**

***Index Terms* - COVID-19, human mobility, multivariate regression, smartphones pandemic, principal component analysis.**

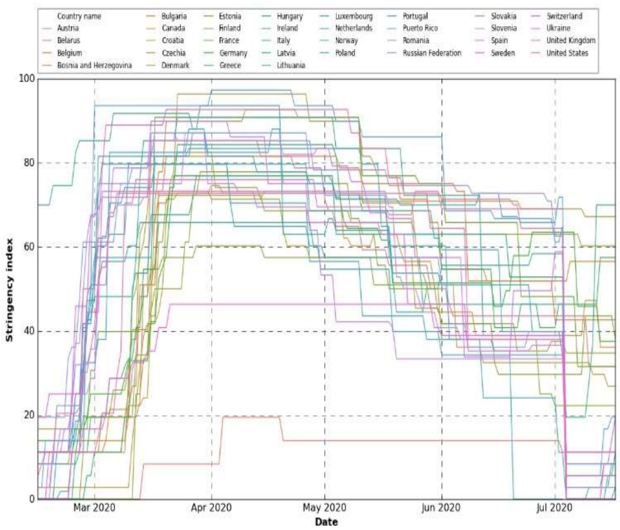
I. Introduction

A novel coronavirus (COVID-19) was first reported in Wuhan, China in December 2019. The World Health Organization (WHO) issued a statement declaring COVID-19 a Public Health Emergency of International Concern on 30 January 2020. On 11 March 2020, it was escalated to the status of a global pandemic [1]. Officially known by the name of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), it is a member of a large family of viruses [2], [3] which include MERS-CoV [4] and SARS-CoV [5]. While the primary means of transmission and associated risk of acquiring COVID-19 are not yet fully understood, it is currently thought that close person-to-person interactions pose the highest transmission risk, allowing disease spread through respiratory droplets [6]. Air pollution-to-human transmission has also been identified as a mechanism for disease spread [7]. As of July 30, 2020, 17,057,700 people have contracted the disease (confirmed cases) and 676,744 deaths have been attributed to COVID-19 worldwide [8]. Between 15 February and 17 July 2020, the period within the scope of this study, the number of confirmed cases grew by 20450% (from 67,104 to 13.79 million) globally [9]. COVID-19 is more infectious than previous coronaviruses [10]. Due to its novelty, humans have no natural immunity. It has been estimated that the person-to-person transmission rate is between 1.4 and 6.7, meaning that each infectious person may transmit COVID-19 to an average of 1.4 to 6.7 people [11]. As of 29 July 2020, the highest proportion of confirmed COVID cases have been in Asia (32%), North America (30%) and South America (27%); with Europe and Africa comprising 6%, and Oceana less than 2% [9]. However, It is believed that these figures represent an underestimation of the true infection pattern since the total number of infections have reached almost 3 to 20 times higher than what reported as confirmed cases mainly due to incomplete testing and low accuracy of several testing techniques[12].

Since the disease was first identified as a major public health concern, response strategies have varied from highly coordinated federal-level responses in countries such as South Korea to a mosaic of state-issued mandates in countries such as the United States. While guidance from national and international sources did not offer coordinated policies with regard to transmission mitigation efforts, the best available evidence supports social or physical distancing, use of a face mask, hand washing, and reducing interactions with others as the most effective means to limit transmission [13]–[16]. Eye protection has also been found to be associated with lower infection transmission [14]. Various restrictions on travel and protective measures such as masks have been implemented at different points since January of 2020. Stay-at-home orders (self-isolation) and restrictions on travel have been at the center of most efforts to mitigate disease transmission through physical distancing to "flatten the curve". One measure of governmental response to COVID-19 is the government response stringency index or GSRI developed by Oxford University to track pandemic mitigation interventions globally [17]–[19]. GSRI is a composite measure of nine response strategies and includes containment and closure strategies as well as economic or financial support, and public education (see Fig.1).

The volume of publicly available COVID-19 data, from epidemiological models of disease progression to dashboards providing real-time daily updates on leading indicators has grown as the pandemic has progressed. Access to COVID-19-related information is almost universally free of cost. Public data, academic models and dashboards, social media platforms, search queries, and purchased products offer researchers a rich source of data for analysis.

Fig. 1. Stringency index by countries during timeframe Feb. 15 to Jul. 17



Location data gathered through GPS-enabled smartphones has been identified as an important potential data source for understanding and predicting human movement [17], [18]. Geographic information systems (GIS) have been in use since the 1990s [20] to track disease transmission and to develop and refine responses and preparedness. Responses to SARS-CoV and influenza are two such examples [10], [16], [21]. The utility of aggregated population mobility data in assessing the impact of social distancing interventions and refining them in real-time has been specifically identified with respect to COVID-19 [10], [22]. Google has made mobility change data publicly available, specifically to be used to assist in the COVID-19 response [23]. This data provides daily, aggregated percent change from baseline in mobility patterns covering six location types, including parks, workplaces, residences, grocery/pharmacies, retail/recreation, and transit. Apple has also made mobility data publicly available that is based on requests for directions to locations compared to a baseline value.

In this study, we explore the factors that influenced human mobility during the COVID-19 pandemic in 34 countries and seek to shed light on which public health-related strategies might be most effective in reducing mobility and disease transmission rate. Using data from multiple sources, we explore governmental response factors such as stay-at-home orders and school closings; disease prevalence factors such as new cases, deaths, hospitalizations, and whether they predict changes in mobility.

II. Literature Review

Understanding human mobility patterns and the resulting high rate of face-to-face interactions and shared spaces or surfaces that contribute to disease transmissions has been long recognized as fundamental to a better understanding of the spread of disease as well effects of armed conflict and environmental disasters on migration and displacement of communities [24]. The widespread use of GPS-enabled smartphones has suddenly made available aggregated electronic mobility data not previously widely recorded or shared. This has accelerated with the COVID-19 pandemic where data of all types are being made available to the public at no cost. Google Location History (GLH) was identified as a novel information source in a 2018 study that noted the underutilization of smartphone location tracking for infectious disease response [18]. In April of 2020, Google made its anonymized Community Mobility Reports freely available [23]. That paper also recognized the limitations of these tools, noting a lack of universal access as well as user understanding of opt-in and opt-out settings.

Many risk factors for COVID-19 transmission and COVID-19-related deaths have been identified including socioeconomic factors, underlying health conditions (Fang et al., 2020), mobility factors [25]–[27], social distancing rules, demographic [28] and environmental variables [27]. Finding these risk factors help public health administrators to identify populations at greater risk and develop different health intervention plans to reduce and mitigate disease transmission.

Badr et al. (2020) utilized aggregated cell-phone mobility data representing unique daily trips to developing a mobility ratio as a proxy for social distancing. A generalized linear model was developed for counties in the US to evaluate the extent to which social distancing reduced COVID-19 case rates. A COVID-19 growth rate ratio was computed based on publicly available epidemiological data. The study found that individual behavior, rather than official mandates, drove social distancing practices as measured by mobility. It also suggested that behavior change often preceded official response policies.

Abramov and Junior (2020) studied the relationship between lockdown period and coverage in controlling the growth rate of a pandemic comparable to COVID-19 by developing a multivariate prediction network model. They concluded that a 65% to 75% increase in lockdown duration will result in reduction of the spread of the virus by 40%. They concluded that lockdown and social distancing should be relatively long, as early release of quarantine will cause a speedy spike in the number of incidences [29]. Couto et al. (2020) used a susceptible-exposed-infected-recovered (SIER) model along with a logistic regression model to predict the transmission dynamics of COVID-19 and investigate the impact of each Google Map mobility restriction components on the spread of the virus. They found that, out of all six Google Map mobility variables, when people stayed at home more (residential mobility variable) it was associated with greater impact in controlling the outbreak in the country. One drawback of their model is that they overlooked the interrelationships among parameters by making separated logistic models for each mobility component. Roda et al. (2020) investigated the impact of strict quarantine measures undertaken in the city on the time course of the epidemic and used the SEIR and SIR model to assess the potential return of a second outbreak after return to a work situation [30].

With respect to other factors influencing disease spread, a study done by Li et al. (2020) developed linear and logistic regression to explore the effect of different factors on death rates by using a publicly available COVID-19 data collected from 661 U.S. counties. They found that factors such as GDP per capita, social distancing, age, percentage of different races, income level, presence of heart disease, and access to primary care have the most significance on the death rate. In addition, they found that independent of other factors, a higher proportion of Black people were dying from COVID-19, and that lower daily temperatures are associated with higher death rate and caseload, respectively [11]. One of the limitations in the study was imbalanced data as they excluded counties having lower than 50 cases or lower than 10 deaths. Another limitation of this model is that it offers a snapshot of the specific timeframe, rather than real-time data, and does not considering all counties and their related cases in the analysis. In fact, as the model uses outdated data, it also fails to capture accurate changes in respect to the updated number of cases, the number of deaths, and changes in the temperature. Furthermore, additional latent variables were not considered in the model.

III. METHODOLOGY

A. Data sources and description

We used openly-licensed country-level data from the DELVE Global COVID-19 dataset which provides aggregated data updated daily from multiple sources including Google, Our World in Data, Oxford Blavatnik School of Government's Coronavirus Government Response Tracker (OxCGRT), and the European Center for Disease Prevention and Control (ECDC). A total of 23 variables were extracted from the DELVE dataset including Google Human Mobility Data (6 variables), COVID-19 Cases and Deaths (ECDC) (7 variables), mask policies (1 variable) (ACAPS Government Measures Dataset, Masks4All), as well as testing information and government containment and closure policies (7 variables) [31]. Data points were extracted from the period beginning 15 February 2020 and ending 17 July 2020. For the purposes of this study, each country was coded by geographic region as defined by the UN (UNSTATS). A single binary variable denoting whether a day was a national or major religious holiday was added. An additional binary variable is added to the dataset indicates the pre/post-pandemic period declared by WHO. A description of variables and sources can be found in Table I.

A convenience sample of 34 countries was chosen representing four geographic regions: North America, and Western, Eastern, Southern, and Northern Europe. Within this sample, countries represent a wide range of infection and response dynamics and fall within the top and bottom range of total cases worldwide.

B. Data Pre-Processing and Dimension Reduction

Due to a large number of observations/magnitude of the dataset and a high number of highly correlated variables, we applied user-controlled variable selection, dimension reduction, and clustering of variables in the preprocessing phase to manage our variables in the research and create real value out of the COVID-19 data. Data analyses were carried out using R 4.0.2. First, data was preprocessed to address missing data points and assess for outliers.

TABLE I. DESCRIPTION AND SOURCES FOR VARIABLES

Variables: Description and Source		
Original Variables	**Description**	**Source; *Original Source* URL**
Predictor Variables		
Containment/Closure Policies		
npi_school_closing (schools and universities) npi_workplace_closing npi_cancel_public_events npi_gatherings_restrictions npi_close_public_transport npi_stay_at_home	Governmental Containment and closure policies issued in response to the pandemic. Included six of eight total available. Ordinal scale.	DELVE Global COVID-19 Dataset; Oxford COVID-19 Government Response Tracker (OxCGRT) https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker Codebook at https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/codebook.md
npi_stringency_index	Summary of governmental response (school, workplace, public transport closing, restrictions on gatherings, cancellation of public events, stay at home orders, internal and interational travel/movement restructions, and public information. Ordinal scale.	DELVE Global COVID-19 Dataset; Oxford COVID-19 Government Response Tracker (OxCGRT) https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker Codebook at https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/codebook.md
masks	Mask policy in place. Ordinal variable.	DELVE Global COVID-19 Dataset; ACAPS + Masks4all + Manual https://github.com/rs-delve/covid19_datasets/blob/master/docs/codebook.md
Disease Prevalence		
cases_total cases_new deaths_total deaths_new	COVID-19 prevalence and deaths	DELVE Global COVID-19 Dataset; Our World in Data Coronavirus Pandemic (COVID-19) https://ourworldindata.org/coronavirus
cases_days_since_first deaths_days_since_first		Derived from Human Mortality Database / The Economist excess mortality tracker / EuroStat
days since the first recorded case		Derived from Human Mortality Database+ The Economist excess mortality tracker +EuroStat
Descriptive		
National or Major Religious Holiday	Binary variable 1=holiday; 0=no holiday	Holidays and Observances Around the World https://www.timeanddate.com/holidays/
Region Code. Western, Eastern, Northern, South Europe, North America	Geographic region assigned. Categorical variable.	UNSTATS https://unstats.un.org/unsd/methodology/m49/
Pre/post pandemic period	Before or after 11 March 2020. Binary variable.	WHO declaration
country_name	English name according to ISO. Categorical variable.	DELVE Global COVID-19 Dataset; ACAPS + Masks4all + Manual https://github.com/rs-delve/covid19_datasets/blob/master/docs/codebook.md
Response Variables		
Human Mobility		
mobility_retail_recreation mobility_grocery_pharmacy mobility_parks mobility_transit_stations mobility_workplaces mobility_residential	Daily change in mobility to locations measured by percent change over baseline days (median value from 5-week period 3 Jan - 6 Feb 2020). Aggregated, anonymized cellphone data recorded by Google Maps.	DELVE Global COVID-19 Dataset; Google COVID-19 Community Mobility Reports Dataset https://www.google.com/covid19/mobility/

In the dataset, invalid inputs such as too big values, mismatch between related columns, and outliers were checked and cleaned. To address multicollinearity and high correlations among variable set, Principal Component Analysis (PCA) using the covariance technique was applied to three sets of variables to create new, uncorrelated variables and reduce dimensions of a set of variables while preserving the total amount of original variance among variables. No standardization was required as the PCA equalizes the correlation matrix with the covariance matrix. In this study, we assume that all variables included in the model have some measurement error and thus add noise to the results.

Due to the fact that the DELVE dataset was aggregated from multiple sources acquired in different time intervals, and reporting was not even across all countries for all indices, rows with all missing values were eliminated. Outliers were detected through boxplot and quantile analysis and extreme outliers that were inconsistent with other data were removed. To account for differences in the ordinal scales for the government containment and closure variables, these were unified to a three-level scale.

As noted, significant correlations are present among the variables. For the purposes of the analysis, the variables were grouped into three sets. Set 1 included government responses - containment and closure policies, Set 2 included disease prevalence/testing variables; and Set 3 included the six mobility variables. Sets 1 and 2 represent predictor variables and Set 3 represents response or dependent variables.

We aim to identify the most important variables in these 3 sets. We cannot examine all possible models, as the size of the model will grow exponentially, as we add one new variable to the model. A correlation analysis establishes that some variables in each set are highly correlated to each other, therefore we choose to implement PCA on each set to reduce the size of each set into small number of informative variables. A description of the PCA methodology and resulting new variables are discussed next.

C. Set 1 Government Response Containment and Closure PCA

Variables describing government response containment and closure policies were significantly correlated (see Fig. 2). For Set 1, three principal components were retained based on an assessment of the Scree Plot (Fig. 3). The first three principal components (PC) account for 81% of the original total variance (Table II). Additional PCs do not account for a significant portion of variance and were disregarded. The first retained principal component (PC1) can be understood as representing policies activity cancelling orders (Eigenvectors = 0.432, 0.447, 0.426, and 0.421). The second retained principal component (PC2) can be understood as representing mask wearing directives (Eigenvector = 0.943). The third retained principal component (PC3) is positive and can be understood as representing closures in public transit (Eigenvector = 0.683).

Fig. 2. Correlation Matrix for Government Response PCA (Set 1)

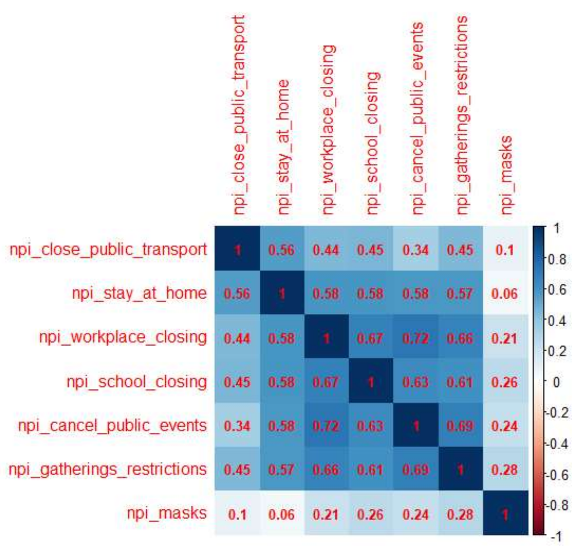


Fig. 3. Scree Plot for Government Response PCA (Set 1)

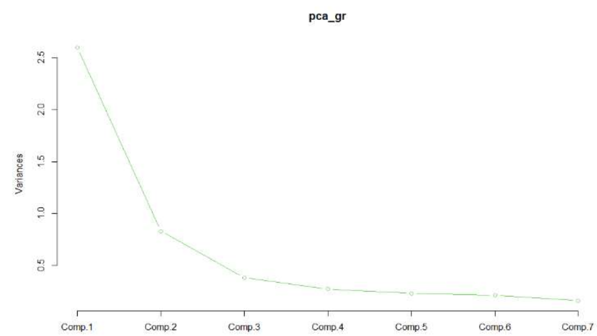


TABLE II. EIGENVECTORS AND EXPLAINED VARIANCE FOR GOVERNMENT RESPONSE PCA (SET 1)

Government Response Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7
school closing	**0.432**	-0.018	0.002	-0.807	0.019	0.400	0.039
workplace closing	**0.447**	-0.077	-0.282	-0.109	0.218	-0.671	**-0.453**
cancel public events	**0.426**	-0.032	-0.467	0.207	-0.235	-0.100	0.701
gatherings restrictions	**0.421**	0.014	-0.122	**0.500**	0.415	**0.561**	-0.265
close public transport	0.245	-0.146	**0.683**	0.051	0.476	-0.236	**0.409**
stay at home	0.388	-0.286	0.427	0.199	**-0.693**	0.025	-0.252
masks policy	0.209	**0.943**	0.198	0.044	-0.133	-0.088	-0.021
Proportion of Explained Variance	0.550	0.180	0.080	0.060	0.050	0.050	0.030
Cumulative of Explained Variance	0.550	0.730	0.810	0.870	0.920	0.970	1.000

D. Set 2: Disease Prevalence Variables PCA

Variables describing the prevalence, spread, and death rates for COVID-19 during the period were likewise significantly correlated (see Fig. 4).

Fig. 4. Correlation Matrix for COVID-19 Prevalence PCA (Set 2)

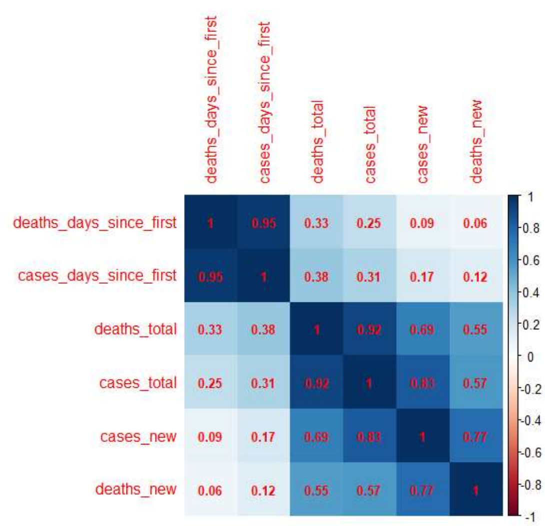


Fig. 5. Scree Plot for COVID-19 Prevalence PCA (Set 2)

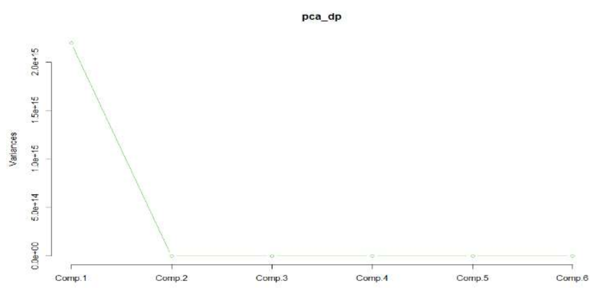


TABLE I. EIGENVECTORS AND EXPLAINED VARIANCE FOR DISEASE PREVALENCE PCA (SET 2)

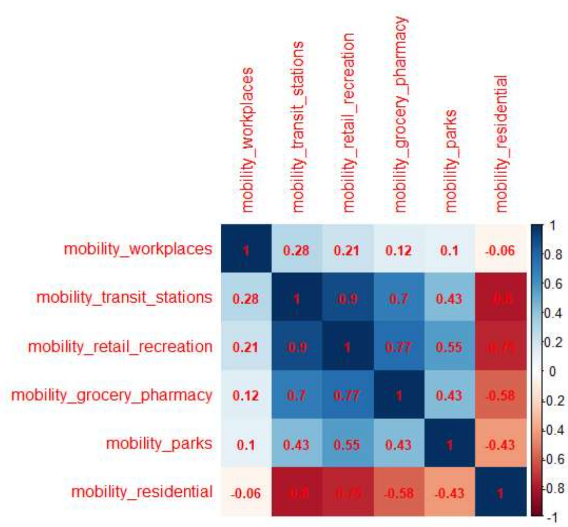
Disease Prevalence Variables	PC1	PC2	PC3	PC4	PC5	PC6
total cases	**1.0E+00**	-5.4E-02	2.4E-02	-1.1E-03	3.6E-05	-4.2E-06
new cases	1.5E-02	-1.6E-01	**-9.9E-01**	6.6E-02	1.4E-03	-4.4E-04
total deaths	5.7E-02	**9.9E-01**	-1.5E-01	1.2E-02	-2.6E-03	2.0E-05
new deaths	6.2E-04	2.0E-03	-6.7E-02	**-1.0E+00**	2.5E-02	-8.7E-04
days since first case	6.4E-05	1.9E-03	1.5E-03	1.7E-02	**7.2E-01**	**7.0E-01**
days since first death	4.8E-05	2.0E-03	2.2E-03	1.8E-02	**7.0E-01**	**-7.2E-01**
Proportion of Explained Variance	**1**	0	0	0	0	0
Cumulative of Explained Variance	**1**	0	0	0	0	1

For the disease prevalence variables (Set 2), the first principal component is retained based on an assessment of the Scree Plot (Fig. 4). The first principal component (PC) accounts for 100% of the original total variance (Table III). The first retained principal component (PC1) can be understood as representing total new cases of COVID-19 (Eigenvector = 0.996).

E. Set 3: Mobility Variables PCA

The correlation matrix for the six mobility variables are described in Fig. 6.

Fig. 6. Correlation Matrix for Mobility Variables (Set 3)



For the mobility variables, three principal components were retained based on an assessment of the Scree Plot (Fig. 7). The first two principal components (PC) account for 98% of the original total variance (Table IV). Additional PCs do not account for a significant portion of variance and were disregarded. The first retained principal component (PC1) can be understood as representing changes in workplace mobility (Eigenvector=0.998). The second retained principal component (PC2) can be understood as representing changes in visits to parks (Eigenvector = 0.915).

Fig. 7. Scree plot mobility pca (set 3)

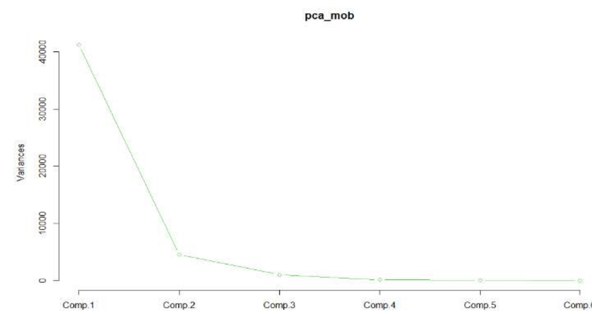
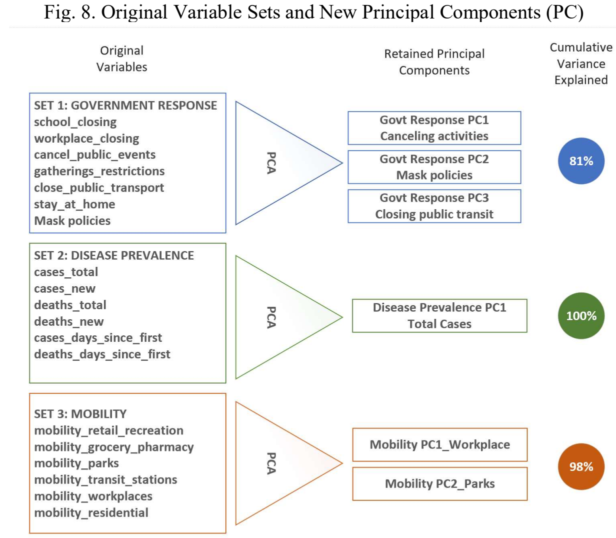


TABLE II. EIGENVECTORS AND EXPLAINED VARIANCE FOR MOBILITY VARIABLES PCA (SET 3)

Mobility Variables	PC1	PC2	PC3	PC4	PC5	PC6
retail recreation mobility	-0.032	0.290	**-0.592**	-0.145	-0.731	0.098
grocery pharmacy mobility	-0.014	0.166	-0.388	**0.883**	0.206	0.003
parks mobility	-0.037	**0.915**	0.399	-0.007	0.040	-0.019
transit stations mobility	-0.036	0.203	-0.553	-0.409	**0.554**	-0.422
workplaces mobility	**-0.998**	-0.053	0.030	0.008	-0.002	0.010
residential mobility	0.003	-0.083	0.185	0.179	-0.339	**-0.901**
Proportion of Explained Variance	0.88	0.1	0.02	0	0	0
Cumulative of Explained Variance	0.88	0.98	1	1	1	1

A summary of the original and new principal components for the three sets of variables can be found below.

Fig. 8. Original Variable Sets and New Principal Components (PC)



The new principal component scores for each set of new, uncorrelated variables were computed and used in the multiple regression analysis to assess which factors were the most significant predictors of mobility.

IV. Multivariate Multiple Regression Analysis

Multivariate regression analysis was conducted to assess which of the predictor variables – government response PC 1, PC 2 and PC 3, disease prevalence PC 1, holiday or not, or pre/post-pandemic period or not – significantly predicted the two dependent variables: workplace mobility (PC1), park mobility (PC2). We implemented the analysis using SAS 9.4. β_0 is the intercept and β_i, $i = 1, 2, \ldots, 8$ show the regression coefficients and the symbol ε represents the error of the model. Therefore, our first model is:

(1)

$$\text{Workplace mobility} = \beta 0 + \beta 1 \times \text{pc1}_{\text{totalcases}} + \beta 2 \times pc1_{activitycancelling} + \beta 3 \times pc2_{maskspolicy} + \beta 4 \times pc3_{publictransport} + \beta 5 \times pre_{pandemic} + \beta 6 \times pandemic + \beta 7 \times not\ holiday + \beta 8 \times holiday + \varepsilon$$

And so, the second model becomes:

(2)

$$\text{Parks mobility} = \beta 0 + \beta 1 \times \text{pc1}_{\text{totalcases}} + \beta 2 \times pc1_{activitycancelling} + \beta 3 \times pc2_{maskspolicy} + \beta 4 \times pc3_{publictransport} + \beta 5 \times pre_{pandemic} + \beta 6 \times pandemic + \beta 7 \times not\ holiday + \beta 8 \times holiday + \varepsilon$$

V. Results

Model (1) assessed the effects of disease prevalence, government responses, period (pre- or peri-pandemic), and holidays on the workplace mobility PC1. The global F Test ($p<.001$) indicates that the model is significant for predicting workplace mobility based on the independent variables. The model gives a R-square value of 0.3261 for workplace mobility, or that 32.61% of the variance in workplace mobility is explained by that model. If we assume that all predictors are uncorrelated, we are dealing with a balanced design and the estimates of coefficients as they minimize the sum of squared of residuals are shown in Table V. The model parameter estimations indicate that total cases, canceling activities (school, work, events), mask policies and the pandemic declaration all were significant predictors ($p<.001$) of change in workplace mobility from baseline.

If we assume that all predictors are uncorrelated, we are dealing with a balanced design. In a balanced design, a one-unit change in total COVID cases predicts a mean change in workplace mobility of 24.74, while holding all other predictors fixed . Activity cancellations (workplace closure, school closure, event cancellations) is a more effective way to reduce mobility; having a negative correlation with change in mobility (reduction in workplace mobility). Due to correlations amongst predictors, uncertainty is introduced to the interpretation and model predictors will tend to change together. In this sense, we avoid the claim of causality, which implies that one single predictor causes the mobility change.

TABLE V. Multiple Regression Model – DV= Mobility PC1_Workplace

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	R-Square
Model	6	3617541.64	602923.61	204.5	<.0001	0.326151
Error	2535	7474076.63	2948.35			
Corrected Total	2541	11091618.28				

TABLE VI. Model Parameter Estimation – DV= Mobility PC1_Workplace

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	10.11	B	3.25	3.12	0.00
pc1_total cases	24.74		0.97	25.62	<.0001
pc1_activity cancelling	-20.91		1.21	-17.23	<.0001
pc2_masks policy	13.65		1.78	7.68	<.0001
pc3_public transport	0.01		0.01	1.38	0.17
pre pandemic	-68.32	B	4.54	-15.05	<.0001
pandemic	0.00	B	.	.	.
not a holiday	-0.85	B	3.38	-0.25	0.80
holiday	0.00	B	.	.	.

Model (2) assessed the effects of the same independent variables on the parks mobility PC 2 variable. Table VII indicates that this model is also significant ($p< .001$). The R-square value for park mobility is equal to 0.3565 which means that approximately 35.65% of the variance in this park mobility is explained by that model.

The estimates of parameters represent the mean change in the response variable (y) for every unit increase in the corresponding (xi) when all other xi are held fixed. For Model 2, the estimates of correlation coefficients of each independent variables are small and indicate that these variables are not powerful predictors of parks mobility PC2 (Table VIII).

TABLE VII
Multiple regression model – DV=Mobility PC2_PARKS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F	R-Square
Model	6.00	108.37	18.06	234.07	<.0001	0.3565
Error	2535.00	195.61	0.08			
Corrected Total	2541.00	303.98				

TABLE VIII
model parameter estimation – DV=Mobility PC2_PARKS

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	0.91	B	0.02	55.06	<.0001
pc1_total cases	-0.06		0.00	-11.95	<.0001
pc1_activity cancelling	0.03		0.01	4.75	<.0001
pc2_masks policy	0.03		0.01	3.84	0.00
pc3_public transport	0.00		0.00	-0.87	0.39
pre pandemic	-0.36	B	0.02	-15.48	<.0001
pandemic	0.00	B	.	.	.
not a holiday	0.00	B	0.02	-0.15	0.88
holiday	0.00	B	.	.	.

VI. Discussion and Conclusion

In this study, we examined the extent to which multiple factors related to government responses and disease prevalence were significant predictor of changes in human mobility during the COVID-19 pandemic period of February – July 2020. We implemented principal component analysis prior to the regression in order to decrease the correlation among independent variables. This also allowed for dimension reduction. The PCA results showed that the independent variables were highly correlated. For future studies of mobility, changes in visits to workplace and parks capture the majority of variance in mobility changes (as measured by Google data). For disease prevalence, the PCA indicates that total cases are an important factor. Reducing multicollinearity among the variables decreases the unique variance explained by that variable which leads to an increase of the shared prediction percentage. Therefore, the use of principal components eliminates multicollinearity and may maximize the prediction power. In both models, due to large number of observations, the normality assumption of the data is maintained and the minimum R- square that can be found statistically significant is reduced. The sign, magnitude, and statistical significance of each independent variables shown in Table VI and Table VIII is representing the individual contribution of each in predicting mobility variables. These estimated can be assessed more with the relative importance of each independent variables.

Additional factors, variables need to be examined to ensure the validity and to maximize the overall predictive power of mobility variables by the independent variables as represented in two models.

REFERENCES

[1] "Coronavirus." https://www.who.int/health-topics/coronavirus (accessed Aug. 08, 2020).

[2] M. U. G. Kraemer *et al.*, "The effect of human mobility and control measures on the COVID-19 epidemic in China," *Science*, vol. 368, no. 6490, pp. 493–497, May 2020, doi: 10.1126/science.abb4218.

[3] "Coronavirus." https://www.who.int/westernpacific/health-topics/coronavirus (accessed Aug. 23, 2020).

[4] "WHO | Middle East respiratory syndrome coronavirus (MERS-CoV)," *WHO*. http://www.who.int/emergencies/mers-cov/en/ (accessed Aug. 23, 2020).

[5] "Severe Acute Respiratory Syndrome (SARS)." https://www.who.int/westernpacific/health-topics/severe-acute-respiratory-syndrome (accessed Aug. 23, 2020).

[6] J. F.-W. Chan *et al.*, "A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster," *The Lancet*, vol. 395, no. 10223, pp. 514–523, Feb. 2020, doi: 10.1016/S0140-6736(20)30154-9.

[7] M. Coccia, "Two mechanisms for accelerated diffusion of COVID-19 outbreaks in regions with high intensity of population and polluting industrialization: the air pollution-to-human and human-to-human transmission dynamics," Infectious Diseases (except HIV/AIDS), preprint, Apr. 2020. doi: 10.1101/2020.04.06.20055657.

[8] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," *Lancet Infect. Dis.*, vol. 20, no. 5, pp. 533–534, 2020, doi: 10.1016/S1473-3099(20)30120-1.

[9] "Coronavirus (COVID-19) Deaths - Statistics and Research," *Our World in Data*. https://ourworldindata.org/covid-deaths (accessed Aug. 23, 2020).

[10] M. N. K. Boulos, "Descriptive review of geographic mapping of severe acute respiratory syndrome (SARS) on the Internet," *Int. J. Health Geogr.*, vol. 3, no. 1, p. 2, Jan. 2004, doi: 10.1186/1476-072X-3-2.

[11] A. Y. Li *et al.*, "Multivariate Analysis of Factors Affecting COVID-19 Case and Death Rate in U.S. Counties: The Significant Effects of Black Race and Temperature," Infectious Diseases (except HIV/AIDS), preprint, Apr. 2020. doi: 10.1101/2020.04.17.20069708.

[12] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," *Lancet Infect. Dis.*, vol. 20, no. 5, pp. 533–534, 2020, doi: 10.1016/S1473-3099(20)30120-1.

[13] H. S. Badr, H. Du, M. Marshall, E. Dong, M. M. Squire, and L. M. Gardner, "Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study," *Lancet Infect. Dis.*, p. S1473309920305533, Jul. 2020, doi: 10.1016/S1473-3099(20)30553-3.

[14] H. Y. Chu *et al.*, "Early Detection of Covid-19 through a Citywide Pandemic Surveillance Platform," *N. Engl. J. Med.*, vol. 383, no. 2, pp. 185–187, Jul. 2020, doi: 10.1056/NEJMc2008646.

[15] K. R. Mahtani, C. Heneghan, and J. K. Aronson, "What is the evidence for social distancing during global pandemics? A rapid summary of current knowledge," p. 9.

[16] CDC, "Coronavirus Disease 2019 (COVID-19) – Prevention & Treatment," *Centers for Disease Control and Prevention*, Jul. 31, 2020. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html (accessed Aug. 23, 2020).

[17] J. Gadziński, "Perspectives of the use of smartphones in travel behaviour studies: Findings from a literature review and a pilot study," *Transp. Res. Part C Emerg. Technol.*, vol. 88, pp. 74–86, Mar. 2018, doi: 10.1016/j.trc.2018.01.011.

[18] N. W. Ruktanonchai, C. W. Ruktanonchai, J. R. Floyd, and A. J. Tatem, "Using Google Location History data to quantify fine-scale human mobility," *Int. J. Health Geogr.*, vol. 17, no. 1, p. 28, Jul. 2018, doi: 10.1186/s12942-018-0150-z.

[19] "Variation in government responses to COVID-19." https://www.bsg.ox.ac.uk/research/publications/variation-government-responses-covid-19 (accessed Aug. 08, 2020).

[20] G. J. Musa *et al.*, "Use of GIS Mapping as a Public Health Tool–-From Cholera to Cancer," *Health Serv. Insights*, vol. 6, p. HSI.S10471, Jan. 2013, doi: 10.4137/HSI.S10471.

[21] A. K. Lyseen *et al.*, "A Review and Framework for Categorizing Current Research and Development in Health-Related Geographical Information Systems (GIS) Studies," *Yearb. Med. Inform.*, vol. 23, no. 1, pp. 110–124, 2014, doi: 10.15265/IY-2014-0008.

[22] C. O. Buckee *et al.*, "Aggregated mobility data could help fight COVID-19," *Science*, vol. 368, no. 6487, pp. 145–146, 10 2020, doi: 10.1126/science.abb8021.

[23] "COVID-19 Community Mobility Report," *COVID-19 Community Mobility Report*. https://www.google.com/covid19/mobility?hl=en (accessed Aug. 23, 2020).

[24] R. Chowdhury *et al.*, "Dynamic interventions to control COVID-19 pandemic: a multivariate prediction modelling study comparing 16 worldwide countries," *Eur. J. Epidemiol.*, vol. 35, no. 5, pp. 389–399, May 2020, doi: 10.1007/s10654-020-00649-w.

[25] G. Bonaccorsi *et al.*, "Economic and social consequences of human mobility restrictions under COVID-19," *Proc. Natl. Acad. Sci.*, p. 202007658, Jun. 2020, doi: 10.1073/pnas.2007658117.

[26] B. R. G. M. Couto *et al.*, "MOBILITY RESTRICTIONS FOR THE CONTROL OF COVID-19 EPIDEMIC," preprint, Jun. 2020. doi: 10.1590/SciELOPreprints.717.

[27] B. Oliveiros, L. Caramelo, N. C. Ferreira, and F. Caramelo, "Role of temperature and humidity in the modulation of the doubling time of COVID-19 cases," Public and Global Health, preprint, Mar. 2020. doi: 10.1101/2020.03.05.20031872.

[28] J. Zhang *et al.*, "Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study," *Lancet Infect. Dis.*, vol. 20, no. 7, pp. 793–802, Jul. 2020, doi: 10.1016/S1473-3099(20)30230-9.

[29] D. M. Abramov and S. C. Gomes Junior, "Multivariate Prediction Network Model for epidemic progression to study the effects of lockdown time and coverage on a closed community on theoretical and real scenarios of COVID-19.," Epidemiology, preprint, May 2020. doi: 10.1101/2020.05.04.20090712.

[30] W. C. Roda, M. B. Varughese, D. Han, and M. Y. Li, "Why is it difficult to accurately predict the COVID-19 epidemic?," *Infect. Dis. Model.*, vol. 5, pp. 271–281, Jan. 2020, doi: 10.1016/j.idm.2020.03.001.

[31] "Daily confirmed COVID-19 cases," *Our World in Data*. https://ourworldindata.org/grapher/daily-cases-covid-region (accessed Aug. 08, 2020).