STEM Fellowship Big Data Challenge 2020

Policy, Demobilization, and Infection: A Multilateral Approach to Social Distancing

Alexander Y. Liebeskind[1], Amy MiHyun Jang[1], Thomas Tran[1], and Nikhil Mehta[1]

[1]Columbia University

July 9, 2020

Abstract

Over the course of the 2020 COVID-19 pandemic, social distancing has proven to be one of the most effective and prevalent methods of mitigating viral spread. Social distancing policies implemented in different countries, however, have widely diverged in strictness and area of focus and have therefore catalyzed varying responses in population mobility. The effects of specialized changes in social distancing policy are still largely unknown or unvalidated. This study aims to address this gap in comprehension by incorporating independently sourced data on COVID-19 infection, government action, and population mobility into a consolidated analysis of social distancing. Results indicated that social distancing policies targeting containment/health were correlated with rapid, significant decreases in population mobility, while policies directed at economic support failed to produce a quick or statistically significant change in population mobility. Furthermore, stricter and more robust social distancing policies were discovered to be more effective than less austere policies for some but not all areas of focus. These findings provide vital insight into the nuanced repercussions of individual changes in social distancing policy. This information is advantageous in the alteration and creation of future social distancing policies.

Keywords
social distancing, data visualization, health policy, epidemiology, biostatistics

1 Introduction

The COVID-19 pandemic has placed enormous strain on the global medical system, forcing many nations to adopt social distancing policies in an attempt to minimize vectors of disease transmission and slow the COVID-19 infection rate. These social distancing measures have been especially pronounced in countries with the largest number of COVID-19 cases, including Italy, China, and the United States, and have been largely confirmed effective at reducing the spread of the virus [1]. Additionally, social distancing policies within and among countries vary in rigor and take diverse approaches including limitations on public transportation, economic sanctions, and legal enforcement of shelter in place orders.

Despite the prevalence of social distancing, however, thorough measurement and projection of infection rate under conditions of social distancing remain elusive, especially given the numerous quantifiable metrics describing social distancing efficacy [2]. This deficit presents a hurdle in assessing the need for future social distancing and the consequences of reducing demobilization measures.

In this study, we present a multilateral approach to social distancing quantification, using several metrics derived from an involved analysis of social distancing policy and mobility data. More specifically, we investigate national social mobility policies executed with varying strictness, areas of focus, and time frames, and the resulting demobilization or lack thereof. The purpose of this analysis is to illuminate the driving mechanisms between social distancing policy and realized impact, including both intended and extraneous effects.

2 Materials & Methods

2.1 Overview

Research methodology was optimized to include multiple datasets addressing social distancing across varied metrics. The pipeline was divided into data acquisition, preprocessing, analysis, interpretation, and results, as shown in Fig. 1.

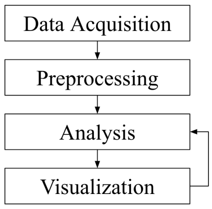

Figure 1: Overall research pipeline

2.2 Data Acquisition

Data used in the analysis was obtained from three sources: Johns Hopkins University, Google, and University of Oxford. The Johns Hopkins CSSE COVID-19 Data Repository [3] documents raw number of confirmed COVID-19 cases, deaths, and recoveries at daily intervals. The Google Maps Community Mobility Reports [4] dataset records daily percentage change in activity on a country level, as measured by geolocation, within a predetermined sample population. The baseline for percentage calculations is the median activity level for the corresponding day of the week between January 3 and February 6, 2020. These percentages are subdivided across six categories: residential areas, parks, grocery stores and pharmacies, retail and recreation, transit stations, and workplaces. The University of Oxford COVID-19 Government Response Tracker [5] provides indices tracking implemented policy changes across numerous categories, including but not limited to school and university closures, public gathering restrictions, and economic sanctions. These individual indices are consolidated into four primary indicator metrics: government response, containment/health, stringency, and economic support.

2.3 Preprocessing

The Google Maps Mobility Community Mobility Reports dataset was cleaned for each of its six categories, for every country ($n = 132$) within the dataset. Countries with more than 75% of time series data missing for any one of the categories were removed from the pipeline. For the countries that remained ($n = 129$), the missing values were derived using linear interpolation. The mobility data was also fairly noisy. Given the importance of accuracy in determining the derivative with respect to time, we applied a Savitzky-Golay filter with a window size of 25 and polynomial degree of 3 to smooth the data based on local least-squares approximation [6]. We solved for the beta coefficients for each window using the following equation for each datapoint:

$$y_i = \beta_3 x_i^3 + \beta_2 x_i^2 + \beta_1 x_i + \beta_0 x_i + \epsilon_i$$

We then applied the following summation of squares calculation:

$$S = \sum_{i=1}^{n} \epsilon_i$$

After minimizing this sum, we arrived at the solution of least squares:

$$\beta = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

where β is our coefficient vector and $\mathbf{X}$ is the Vandermonde matrix. Our smoothed function at each window was then given by:

$$\hat{\mathbf{y}} = \mathbf{X}\beta$$

The University of Oxford COVID-19 Government Response Tracker calculates indices based on 17 indicators overtime for every country ($n = 168$) using the mean of the indicators, but contains missing values when there is not enough data to calculate the mean. This missing data was populated by taking the same value as the previous non-missing value earlier in time, similarly to howshown Hale et al. visualized the missing data. The Johns Hopkins CSSE COVID-19 Data Repository contained data at a country level scale ($n = 188$) and did not have any missing values.

After handling missing data, these three datasets were consolidated into a single dataset for ease of analysis. The datasets were merged by date and country name, and countries that were the same across datasets but named differently were manually concatenated. After aggregation, our dataset consisted of a sample size of $n = 114$ countries.

2.4 Analysis

To initially differentiate the effects of social distancing policy on mobility by area of focus, policy indicators were separately plotted with observed mobility data. Various parametric functions were tested to determine the best regression model for this data without overfitting for each scatterplot; this was empirically found to be a polynomial function of degree 2, as determined by the relationship:

$$\hat{\mathbf{y}} = \beta_2 \mathbf{x}^2 + \beta_1 \mathbf{x} + \beta_0$$

Corresponding 95% confidence intervals and p-values were also calculated. This data plot along with our parametric regression line is shown in Fig. 2.

We define response lag as the number of days it took for the country to reach the day of greatest decrease in social mobility. Social mobility includes the aggregate of Google mobility data, excluding data from changes in residential mobility. To find the day of greatest mobility, the aggregate mobility was found over time. The aggregate mobility over time followed a negative logistic trend as it exhibited a plateau, followed by a decrease, and then another plateau. Using the Savitzky-Golay filtered mobility data, the day of greatest decrease was derived by the index of the largest negative derivative. We found the response lag with respect to the government implementation of containment/health policies, i.e. the number of days it took to reach the day of greatest decrease in social mobility from the first day containment/health policies were announced, and the response lag with respect to economic support policies. As the containment health index and economic support index change over time, the level of policy strictness that each country implemented was calculated by finding the average of the non-zero values, as the non-zero values are the index values pre-policy implementation. As the time variable to implement government policy is implicitly present in the response lag, only the non-zero values were included to avoid double-counting of this time variable. The mobility response lag vs. the non-zero averages of the economic support index and the containment health index is shown in Fig. 3. A similar method to calculate the regression models as Fig. 2 was used to implement the regression models for Fig. 3, which also uses a degree 2 polynomial regression.

Analyzing the effectiveness of specific government containment, health, and economic policies, we performed a survival analysis on mobility grouped by countries that implemented certain degrees, or types, of policies. Because the Oxford dataset was a time series dataset, the countries were classified into various types based on the median of that particular policy implementation degree after the first case. The average mobility was calculated by taking the mean of all mobility indicators in the Google dataset, besides residential changes. We then defined the time to event as whether the average mobility of each country decreases by 30% for all times past the first case.

To approximate the survival function for each type within each policy, the Kaplan-Meier estimator was employed [7]. The resultant plots can be observed in Fig. 4. The survival functions for each policy compare time to events given some type of policy implementation, allowing us to understand the relationship between containment, health, and economic policy implementations and how effectively they encourage social distancing on a country level. The type for a given policy is defined as the level of strictness of each policy where a higher type number corresponds to a stricter implementation. To illustrate, C1 type 3 is a stricter C1 policy than C1 type 1.

2.5 Visualization

Analysis outputs were assessed for interpretability, significance, and relevance. Visualization was performed to demonstrate dominant trends and patterns. As different metrics can be used to describe mobility, it was important to visualize the data to ensure that the mobility metrics used corresponded to visible trends within the dataset. Additionally, visualizations were used to discover data metrics with missing data, and these analyses were removed from the final results as it is difficult to interpret results from data with missing points. Regressions of the data were shown on top of our plots for ease of interpretability, but also to discover clearer trends within our analysis.

3 Results

Total mobility was disparately related to with changes in economic support and containment/health policy indices. As shown in Fig. 2, no significant correlation was found between total mobility and economic support index (Pearson correlation, $p > 0.05$). Analysis produced a statistically significant relationship between total mobility and containment/health policy index (Polynomial of Degree 2, $p < 0.05$).

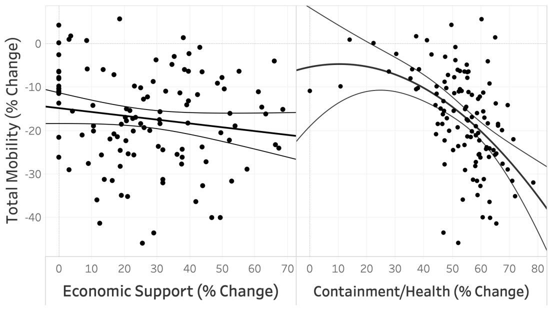

Figure 2: Average percentage change in total mobility by country vs. average change in policy index

The mobility response lag is the number of days it took for the country to reach the day of greatest decrease in social mobility with respect to the implementation of economic support policies (Fig. 3a) and with respect to the implementation of containment/health policies (Fig. 3b). The average index value is the non-zero average, as explained in the Analysis section above. Fig. 3 highlights how stricter containment/health policies lead to a quicker public response to decrease social mobility while changes while strict economic policies do not elicit a quicker public response to decrease social mobility.

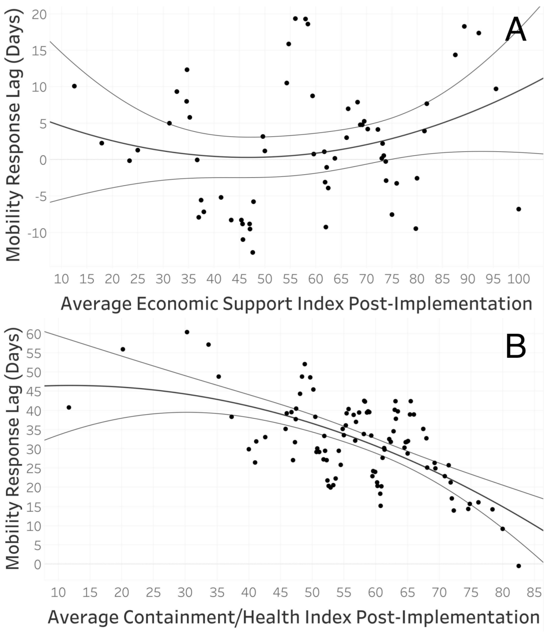

Figure 3: Average mobility response lag vs. mean of non-zero policy indices by country

The probability that mobility fails to decrease by 30% was plotted against the number of days since the first infection was reported for each strictness type for ten policy categories in Fig. 4. Each plot in Fig. 4 shows 3 to 5 curves corresponding to countries that exhibited a specific strictness type. A lower numbered type corresponds to a weaker implementation of the policy where as a higher type number would mean a stricter implementation.

4 Discussion

4.1 Interpretation/Significance

Overwhelmingly, stronger forms of containment/health policies correspond with higher likelihoods of mobility decrease by 30% of the baseline. We chose to use a decrease in mobility of 30% as our threshold value to since this boundary has been shown to keep infection rates at just 10% [8]. Plots C1 through C8 in Fig. 4 have type 0 and 1 curves as being generally higher than their stricter counterparts which dip faster. This equates to weaker policies exhibiting higher probabilities of failing to decrease mobility by 30% when compared to stricter versions of the same policy.

This pattern is less pronounced in plots C4 and C8 of Fig. 4. In plot C4, the type 0, 1, and 2 curves do not exhibit the general pattern discussed above. We believe this is a result of having a small sample size of countries of types 1 and 2 leading to a larger confidence interval that largely overlaps. These results suggest that these less strict types are insignificant. In plot C8 the vast majority of countries (73 out of 113) have implemented the strictest type of international travel controls: type 4. The lack of a distinct difference between the different curves can be attributed to the widespread ban on international travel. As international travel bans affect all other countries on a global scale, the difference between a type 4 and a type 3 policy are fairly trivial from a practical standpoint.

As shown in plots E1 and E2 in Fig. 4, more robust economic policies show no greater likelihood of decreasing mobility by 30%. Here, type 2 policies are seen to correspond with higher probabilities of failure to demobilize as with a type 1 policy across time. In plot E1, the type 2 curve bears a large resemblance to the type 0 curve. This observation generally would agree with our results in Fig. 2 where greater levels of economic support do not coincide with greater decreases in mobility. In contrast, Fig. 2B shows a correlation between an increase in containment/health policies and a decrease in mobility.

4.2 Sources of Error

To construct the Kaplan-Meier estimators for each policy, we grouped countries by median strictness type following the first confirmed case of COVID-19. Although this method is theoretically valid, the way in which countries are aggregated into different types has a large effect on how the survival functions are estimated for each policy.

As policy changes may not happen not be publicized immediately, there may be gaps in the Oxford dataset. The University of Oxford has addressed this issue by making assumptions about this missing data and interpolated the missing points. However, these assumptions

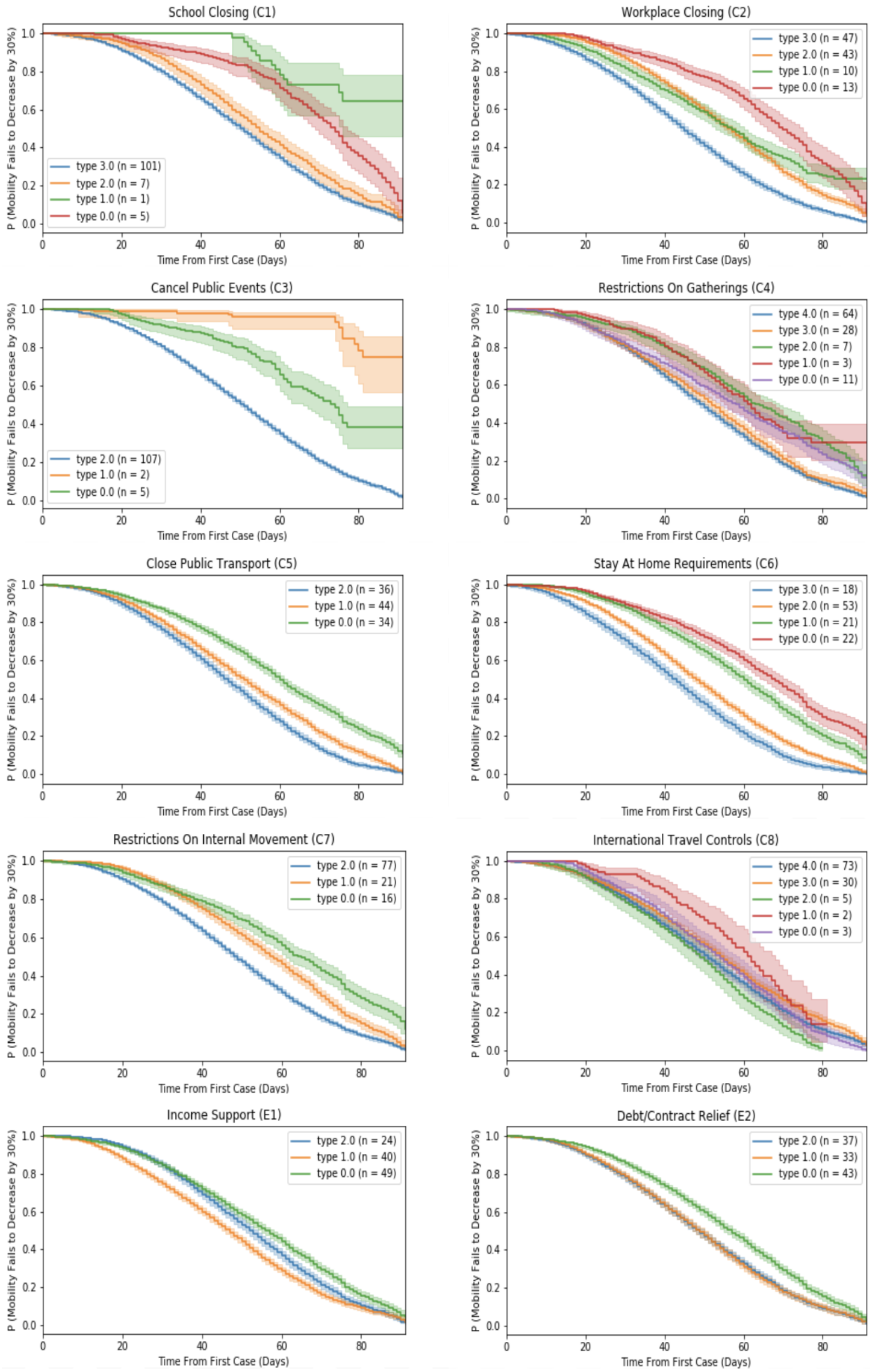

Figure 4: Probability total mobility fails to decrease by 30% in the time following initial case

may lead to errors in our analysis as we rely on these reports to make interpretations. Additionally, to find the derivative of the aggregate mobility trends, we used a filter to smooth the data. Although mobility trends predominantly follow a clear pattern, this usage of the filter assumed a predictable rate of mobility change.

5 Conclusions

Our analysis ultimately suggests that containment/health policies are more strongly correlated with decreases in mobility than economic support policies. Furthermore, increases in the relative strictness or robustness of policies in either category correspond to higher likelihoods of demobilization for the majority of areas of focus.

As many countries prepare to reopen, this insight into the efficacy of government action is crucial in optimizing policy strictness and more accurately predicting public behavior. Social distancing measures, while effective in limiting COVID-19 spread, are nonetheless inhibitors to economic growth and other aspects of public health [9]. Furthermore, as evidenced by South Korea [10] and Singapore [11], countries are prone to incur further outbreaks as they remove social distancing measures. These results on specific policy outcomes are therefore valuable as government agencies attempt to rapidly limit public mobility if and when necessary.

Future studies may ideally aim to generalize this analysis to longer time frames, which would allow for the consideration of recent data on easing social distancing restrictions. Additionally, this research may examine social distancing policy and demobilization during potential later waves of the COVID-19 pandemic.

References

[1] Michael Greenstone and Vishan Nigam. Does Social Distancing Matter? *SSRN Electronic Journal*, Mar 2020.

[2] Inga Holmdahl and Caroline Buckee. Wrong but Useful — What Covid-19 Epidemiologic Models Can and Cannot Tell Us. *New England Journal of Medicine*, May 2020.

[3] COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, 2020.

[4] Google COVID-19 Community Mobility Report. *Google LLC.*, 2020.

[5] Thomas Hale, Sam Webster, Anna Petherick, and Beatriz Kira. Oxford COVID-19 Government Response Tracker. *Blavatnik School of Government*, 2020.

[6] Ronald Schafer. What Is a Savitzky-Golay Filter? [Lecture Notes]. *IEEE Signal Processing Magazine*, 28(4):111–117, 2011.

[7] Jason T. Rich, J. Gail Neely, Randal C. Paniello, Courtney C. J. Voelker, Brian Nussenbaum, and Eric W. Wang. A Practical Guide to Understanding Kaplan-Meier Curves. *Otolaryngology–Head and Neck Surgery*, 143(3):331–336, Sep 2010.

[8] George J Milne and Simon Xie. The Effectiveness of Social Distancing in Mitigating COVID-19 Spread: a modelling analysis. *MedRxiv*, Mar 2020.

[9] Nuno Fernandes. Economic Effects of Coronavirus Outbreak (COVID-19) on the World Economy. *SSRN Electronic Journal*, Mar 2020.

[10] Joyce Lee. Coronavirus Outbreak at South Korea E-Commerce Warehouse Drives Spike in New Cases. *Reuters*, May 2020.

[11] Hannah Beech. Singapore Seemed to Have Coronavirus Under Control, Until Cases Doubled. *The New York Times*, Apr 2020.

Acknowledgements

We would like to express our gratitude to Priyanka Gogna for consulting on the project.