Multi-Stage Covid Prevention Dynamics

Hongsen Qin, Serena Yan, Hanzhi Zheng, Pengda Liu

February 22, 2021

1 Executive Summary

The evolution of a novel pandemic never obeys a single model or trend. Covid-19, in our case, leaves a different trail in each region and country it touches, depending on the people's response and the country's inherent population dynamics such as travel habits and medical infrastructure. To address the complex challenges of combating Covid, we ask ourselves what interventions should be taken to contain Covid at each stage of the evolution of the pandemic. We divided our Covid prevention analysis into two stages: 1) early stage, where a susceptible population is being introduced to Covid, and 2) late stage, when there is a sizable proportion of individuals in the population which already are infected. We recognize in both that the deciding factor for survivability will be the amount of treatment available, which is a function of the number of active cases and whether medical facilities are overwhelmed.

To summarize, Our research direction will focus on how to best control the spread of the virus to reduce the strain on the healthcare system in early and late stages(i.e. flatten the curve). For early stage, we estimate how various quarantine measures under a SEIR model can provide actionable predictions when little precedence is available. For late stages, we analyze the effectiveness of various public health interventions (lockdown, masking, ...) that countries have already taken, and demonstrate that public health measures vary in effectiveness based on levels of development, median age, and other factors. We show that there is no one size fits all solution, and that every country should decide on interventions according to their own status.

For countries like China, where the pandemic broke out, it is virtually impossible to predict how covid will spread when it has no precedence. In light of this, we take a first principles approach to modeling the spread of Covid using a SEIR model. Using SEIR, we model disease spread under various levels of quarantine assumptions and provide a numeric estimate of confirmed cases in California by fitting our model on US travel data. We find that our SEIR model provides insightful actionables on the surge of covid cases and its output can be used to suggest government response when no other data is available.

In the later stages where Covid is already spreading, public health response become essential. We analyze how different countries Covid situation responds differently to the quarantine measures, such as mandatory mask and types of lockdowns, and how effective each intervention was as a function of their wealth level, health level, .. etc. We note prominently that countries' clustered on their wealth level share similar effective interventions, and that interventions with a focus on restricting a certain type of economic activity impacts countries with different developmental indexes different. Furthermore, we also note that population health indicators like age also influence what types of intervention are effective. Finally, we address some of the challenges of enforcing these interventions, and how that influences the nuances of our analysis. With these points, we outline a general framework for each country to consider their own situation and what types of intervention is most appropriate.

As a result, we are able to recommend the most effective methods to prevent Covid to different countries in its current stage of Covid development.

2 Covid-19 Data in Light of Demographic and Socioeconomic Factors

Upon initial visualization of the dataset provided, we noticed that within each country, the indicators for Covid-19 over time behave similarly (Fig. 1) while the indicators for different countries behave differently over time. Thus, we think it's more interesting to conduct analysis on country level data instead of aggregated continental or global data.

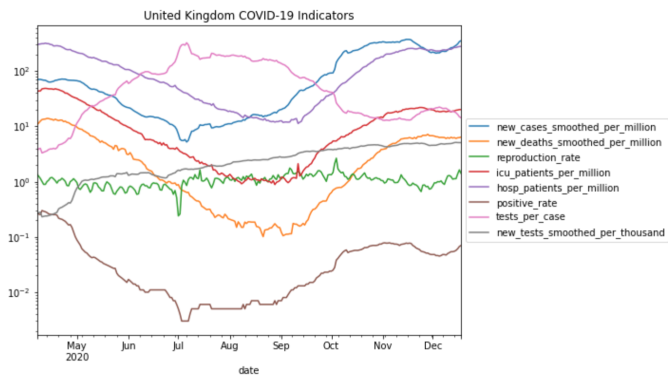

Figure 1: Covid-19 indicators of the United Kingdom plot over time.

In this section, we used Principal Component Analysis (PCA) to identify the most important demographic and socioeconomic factors differentiating between European countries, and K-Means Clustering to group them according to results of PCA. Without the use of any geographic or Covid-19 data, we surprisingly found that the countries in the same group are geographically clustered and have similar trends in Covid-19 indicators over time.

2.1 Principal Component Analysis

We performed PCA on the owid dataset. We selected demographic and socioeconomic indicators such as population, population density, median age, senior distribution, per capita GDP, disease death/prevalence rate, hospital resource per thousand, life expectancy, and human development index for all available European countries. All data are standardized to zero mean and unit variance.

We chose the optimal number of Principal components by picking the minimum number of components that explain 95% of the original data's variance. we determine this number to be 5 after producing a scree plot. With PCA we reduce our dimensionality from 11 to 5 (Fig. 15).

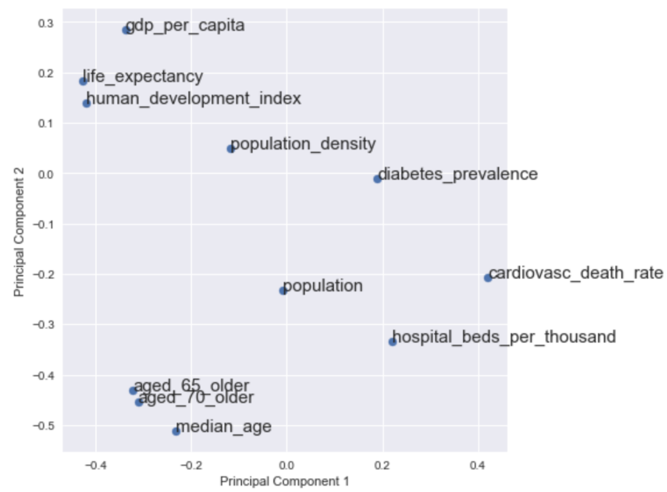

Figure 2: Principal Components of European countries demographic and socioeconomic indicators.

We plotted the indicators along the first and second Principal components, and found that life expectancy, human development index, and cardiovascular death rate are the most important factors differentiating between the European countries.

2.2 K-Means Clustering

Using the results of PCA with 5 components, we performed K-Means Clustering to group the European countries. We determined the optimal number of k clusters using an Elbow plot, which plots distortion(sum of squared distances from cluster centers) against number of clusters. From the elbow plot we produce candidate values of k where the elbow plot bends sharply. Although both $k = 3$ and $k = 6$ can serve as the elbow (Fig 16), using 3 clusters is more reasonable considering the number of European countries with complete data.

In general, the algorithm grouped countries by European regions (eastern, mid-southern, and northern) although the input data does not contain geographic factors. The groupings are as follows:

1. Cluster 1 (Eastern Europe): Albania, Belarus, Bosnia and Herzegovina, Moldova, Montenegro, North Macedonia, Romania, Russia, Slovakia, Ukraine.

2. Cluster 2 (Northern Europe): Belgium, Cyprus, Denmark, Iceland, Ireland, Luxembourg, Malta, Netherlands, Norway, Sweden, Switzerland, United Kingdom.

3. Cluster 3 (Western Europe): Austria, Bulgaria, Croatia, Czechia, Estonia, Finland, France, Germany, Greece, Hungary, Italy, Latvia, Lithuania, Poland, Portugal, Slovenia, Spain.

We also visualize the country groups using a color-coded pair plot to see how good is the algorithm fit (Fig. 17). The cluster within each pairplot (feature i vs feature j) are generally correctly, confirming the validity of our PCA. The K-Means inertia, summarized by the sum of squared distances within the group, is 200.23.

2.3 Covid-19 Indicators by Cluster

We plot the new death per million population and the new positive rate indicators of countries in each cluster together and found that time series within each cluster are highly similar but the average time series between different clusters display different patterns.

For example, for new Covid-related deaths per million population (Fig. 3), all countries in the Eastern Europe cluster do not have the double bell shape pattern displayed in the other two clusters. Western Europe's first wave of deaths is notably smaller than the second wave (except 3 outlier countries) while the Northern Europe's first and second wave are relatively the same in magnitude.

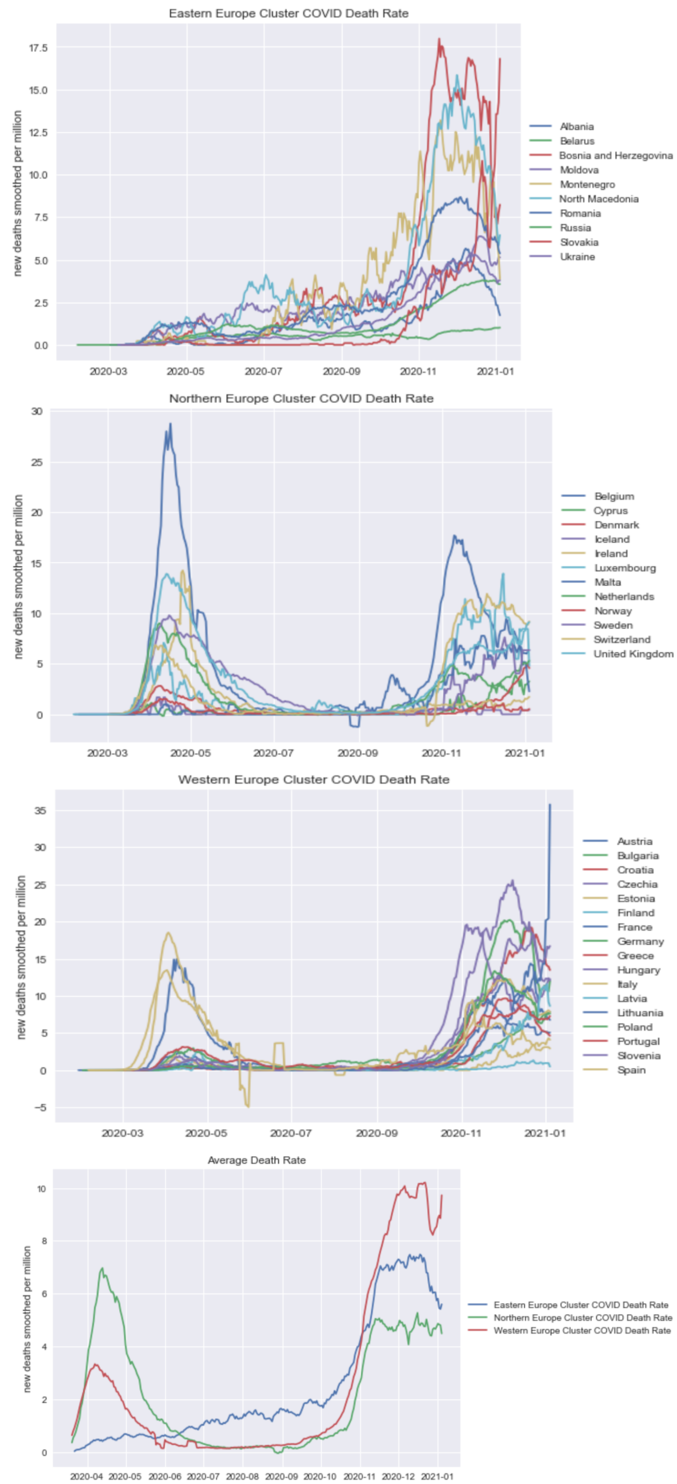

Figure 3: European clusters Covid-related death per million population.

For Covid-19 positive rate (Fig. 4), similar patterns are shown and the differences are more apparent. In particular, the first wave for Western Europe arrives earlier than that for Northern Europe, and its magnitude is significantly smaller than that for Northern Europe. Both regions also display a small increase in positive rate between the first and second rate, with differing magnitude, duration, and start time.

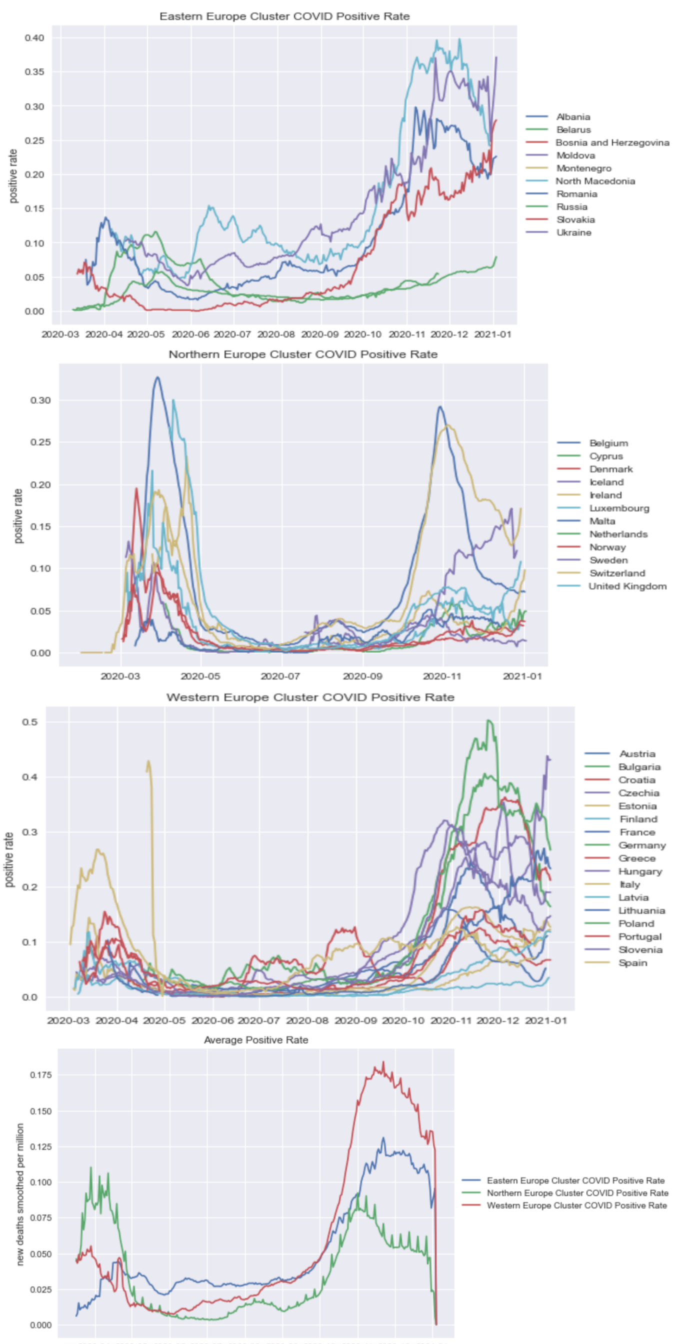

Figure 4: European clusters Covid-19 positive rate.

From the surprising patterns found, we concluded that demographic, economic, and possibly cultural factors affect the Covid-19 dynamics of a particular country or region. Thus, in our multi-stage analysis for Covid prevention, we try to analyze similar countries as a group and interpret the results in light of these factors.

3 How could quarantine help slow the spread of Covid-19 in early stages

In the very early stages of a pandemic outbreak, there is very little data that allows empirical modeling of infection dynamics. Hence we adopt a first principles modeling approach to illuminate infection dynamics. We will investigate how the state government could help prevent the spread of Covid-19 in their state by issuing a quarantine policy for travellers from other states. Concretely, we explore the relationship between the quarantine ratio and total infection cases. In order to do so, we simulate the development using the graph SEIR model.

3.1 Graph SEIR model

Let the states be $1, 2, ..., n$ and S_i(susceptible group), E_i(exposed group), I_i(infectious group), R_i(recovered group) of state i. We can model the states as nodes in a directed graph G where the weighted edges represent the number of travel from one state to another and let L be the graph Laplacian of G. Suppose the contact rate is β, the latency rate is α and the infectious rate is γ and let v be the normalization factor, then the graph SEIR model [1] describes the development of the disease with a set of differential equations:

$$\dot{S}_i = -\sum_{j=1}^{n} vL_{ij}S_j - \beta S_i I_i$$

$$\dot{E}_i = -\sum_{j=1}^{n} vL_{ij}E_j + \beta S_i I_i - \alpha E_i$$

$$\dot{I}_i = -\sum_{j=1}^{N} vL_{ij}I_j + \alpha E_i - \gamma I_i$$

$$\dot{R}_i = -\sum_{j=1}^{n} vL_{ij}R_j + \gamma I_i$$

3.2 Quarantine Model

Let the quarantine ratio ρ stand for the ratio of travellers that get quarantined upon entering the state. To keep things simple, we assume that every one in the quarantine group gets quarantined for 14 days and when they come up of quarantine they enter the susceptible group S_i.

3.3 Result

We ran the simulation using explicit integration method for California and predicted 2 month forward-looking total confirmed cases of California. We plot with respect to the quarantine ratio $[0.0, 0.01, ...0.99, 1.0]$. The Covid-19 spread parameters(α, β, γ) are from [2] and the travelling inflow population data that we use to build the graph are from [3] and [4].

As we can see that the quarantine scheme is quite effective in reducing the confirmed cases and on average an 1% increase in quarantine can reduce 20000 Covid cases in California.

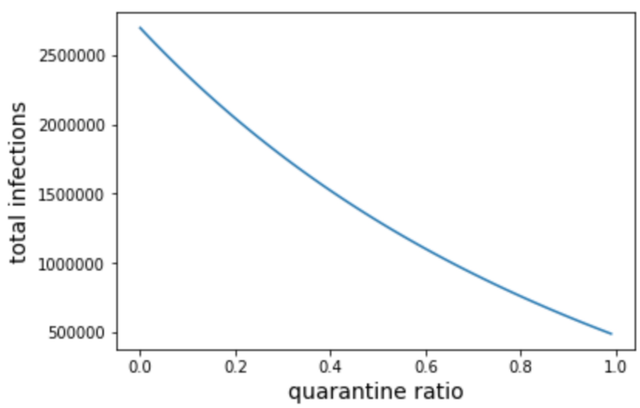

Figure 5: As we can see that the quarantine scheme is quite effective in reducing the confirmed cases and on average an 1% increase in quarantine can reduce 20000 Covid cases in California.

4 Comparison Analysis of Different Public Health Treatments

4.1 Causal Time-Series Analysis

During the long period that the world is fighting Covid, a number of government imposed measures have been adopted in order to control the spread of Covid. Some of these measures might appear to be effective in controlling the spread of Covid-19, but it is not always clear that such effectiveness is causal. In order to analyze the causal effect of various measures taken, in this section, we would like to conduct a counterfactual analysis of how the development of positive ratio of Covid-19 tests, ICU patients per million population, new deaths per million population and reproduction rate would develop if some controlling measure had not been implemented. However, it is not possible to perform such analysis on the data of a single region because for every single region, there is only one ground-truth of whether we implemented or did not implement a certain measure. Hence, we propose the following method that uses another region as the counterfactual situation of "had this region not implemented some measure" to conduct such analysis. Specifically, we do the following:

1. Using the dataset *2_ecdc/country_response_measures* [5], fix a public health treatment measure and find a country A that implemented this measure during a time period on which relatively few other countries implemented the same measure. Denote the date on which this measure has been implemented as day k (and suppose the data goes from day 1 to day $m > k$).

2. Use the Dynamic Time Warping (DTW) distance metric [6] to find another country B such that country B's certain Covid related statistics (positive ratio of Covid-19 tests, ICU patients per million population, new deaths per million population or reproduction rate) is the closest to that of country A during day 1, day $2, \cdots$, day k. Note that country B will also need to satisfy that it did not implement that measure when country A did.

3. Observe how the statistics evolve for country A and B after day k. If country A's statistics grew significantly slower (or declined significantly faster) than that of country B, then we have some evidence that the analyzed measure is effective in controlling the spread of Covid. If no such discrepancy is observed (or if country A's statistics grew even faster), then we will be prone to draw the conclusion that most likely the measure analyzed is not very effective.

In the following, we show some results coming from the above type of analysis.

4.1.1 Analysis of *StayHomeOrder*

One of the most stringent measures enforced by government is stay-at-home orders for the general population (enforced and also known as lockdown), hereby denoted *StayHomeOrder*.

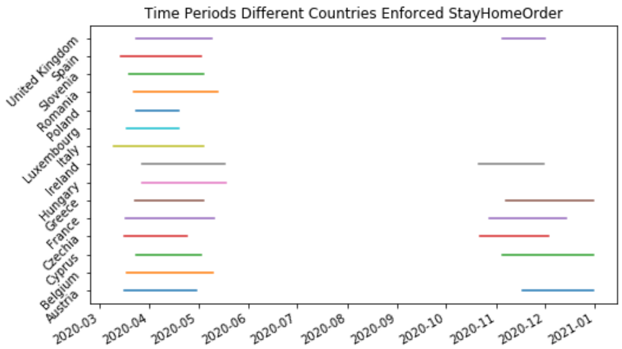

Figure 6: When *StayHomeOrder* was enforced by each country

As demonstrated in Fig. 6, there are two main periods during which some subset of European countries enforced *StayHomeOrder*, corresponding to the two waves of Covid in Europe. In order that we have enough data to ensure that two countries' Covid statistics are similar up until the intervention date, we pick the second period of *StayHomeOrder* as our object of analysis. Fig. 6 also shows that Ireland is the earliest to enforce *StayHomeOrder* during the second wave, and thus we choose Ireland as the objective country because we want to make sure that few other countries were enforcing this restriction when the objective country was.

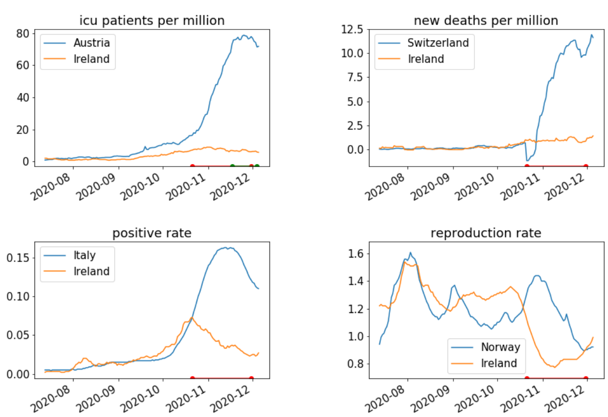

Figure 7: How Ireland & other selected (based on DTW) countries' different statistics evolved over time. The red interval in the bottom corresponds to when Ireland enforced *StayHomeOrder*. The green interval (if exist) corresponds to when the other countries enforced *StayHomeOrder*. New deaths per million is smoothed by 7-day moving average.

As we can see in Fig. 7, in all four statistics, the Ireland curve declined quite significantly immediately after the intervention date 10/21/2020, while the countries selected for counterfactual scenarios have increasing statistics for a period of time before starting to show decreasing trend. Specifically, in the plot of ICU patients per million, Austria's curve did not start decreasing until Austria had *StayHomeOrder* in place on 11/17/2020. This presents some evidence on the causal effect of a *StayHomeOrder* in place on controlling Covid in Ireland. We will see in the subsequent analysis that the *StayHomeOrder* will be one of the important factors in our regression analysis for a cluster of countries we identified with Kmeans earlier in the paper.

4.1.2 Analysis of Closure of Schools

Often times, it is not the case that one measure gets enforced at a time by a certain country's government - similar measures are commonly adopted simultaneously. For example, *ClosPrim*, *ClosSec*, *ClosHigh* (Closure of primary, secondary schools and higher education) are relatively frequently enforced next to each other. One such example is on 10/14/2020 in Czechia. We follow our regime to get the following plots in .

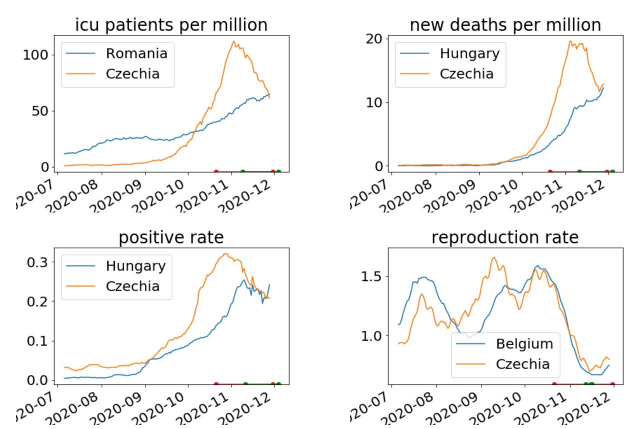

Figure 8: How Czechia & other selected (based on DTW) countries' different statistics evolved over time. The red interval in the bottom corresponds to when Czechia enforced at least one of *ClosPrim*, *ClosSec* and *ClosHigh*. The green interval (if exist) corresponds to when the other countries enforced these measures. New deaths per million is smoothed by 7-day moving average.

Notice in Fig. 8 that Czechia is earlier (and in fact also more thorough because Czechia closed all three types of school for a period of time while others mainly closed one type) in enforcing the closure of schools. We can see that in all plots but the reproduction rate plot, Czechia's curve shows a much earlier or a much stronger decreasing trend when compared with the countries representing the counterfactual scenarios. Hence, it might be the case that closing a more complete set of schools helps with controlling Covid. This presents some evidence on the causal effect of closure of schools on controlling Covid in Czechia.

4.2 How developmental indicators modify the effectiveness of masks, quarantine, and other interventions

Considering the exponential-like growth of Covid as determined by the reproduction rate, public health services can be overwhelmed quickly. From the governmental perspective it is imperative that all interventions be taken to reduce this constant by means of masking, social distancing, quarantining, ... etc. Our data shows that during a typical day during the pandemic from March 2020 - December 2020, a government may have its country under multiple interventions. to isolate the effect of each intervention, we run a ordinary least square regression (OLS) with the response variable as reproduction rate($r_{reproduction}$) [7] and the covariates as binary variables representing the existence and absence of an intervention (e.g. mandatory closing of public spaces). We define our model as

$$\begin{aligned} r_{reproduction} = \beta_0 + \beta_1 \cdot intervention_1 \\ + \beta_2 \cdot intervention_2 + ... \\ + \beta_n \cdot intervention_n \end{aligned} \quad (1)$$

We first tried aggregating the reproduction rate for all countries across the entire time span of the pandemic data we have. This method did not yield a very ideal regression as our adj. r^2 was around 0.25. Referring back to the K-means analysis in our preliminary analysis, we realized that it is logical that each country response differently given the same intervention. For instance, an intervention that restricts religious gatherings will disproportionately impact the reproduction rate of countries with higher proportion of population that were religious. With this goal in mind, we split the data-set by country, and we perform the regression only on interventions which the country has implemented. This was the correct approach as we observed that our adj. r^2 changed to the 0.7-0.93 range for more than 65% of the countries in question. We took great care in making sure the assumptions of OLS were satisfied. While fitting the regression, we noticed strong multicollinearity on some variables, this is because some measures such as "Stay home order" and "Teleworking" can reflect similar results in real life. We resolved this by removed some of the covariates which were correlated with each other and kept a representative one in each category (This analysis was peroformed by looking at covariance). Resolving multicollinearity ensured that our p-values on coefficients were accurate. We show a few examples in Table 1.

Country	adj. r^2
Denmark	0.878
Finland	0.742
Germany	0.959

Table 1: adj. r^2 of several countriesś $r_{reproduction}$ regression with resp. to interventions. Note that we have more than 30 countries, thus it is not feasible to show the regression results for every one

In order to determine which intervention was the most important for variable selection, we standardize the regression coefficients so that their magnitude may be used to compare importance; We also make sure that we only select the statistically significant covariates ($p < 0.05$). We show an example result of this new OLS for Austria in table 2. Note due to a limitation in pages, we cannot show all 30 countries in this report.

Intervention	coefficients	p-value
MasksMandatoryClosedSpaces	-1.29	0.000
PlaceOfWorship	-0.233	0.039
ClosPubAny	-0.186	0.024
...	...	...

Table 2: top three most important interventions for Austria (note that there are more than 50 covariates, so we cannot show them all here. We pick Austria since it was the first alphabetically)

Through this variable selection method, we are able to acquire a list of top k important features for each country. Recall that in the above section, we took note that the countries in the European union were clustered via PCA/K-means and separated by what appears to be development index or wealth. Using previously developed cluster we group the countries into wealthy and non wealthy, and we tally what their most effective interventions were in each group. We observed that in less developed countries such as

Romania and Ukraine, measures such as lockdown place of worship, primary education, hotels and other accommodations occurred most frequently, signaling their usefulness. Where as in more developed countries such as Germany and France had features such as closing restaurants, cafes, and places of higher education. We summarize the findings in table 3.

More Developed	Less Developed
EntertainmentVenues	PlaceOfWorshipPartial
ClosPubAnyPartial	StayHomeOrderPartial
ClosHigh	HotelsOtherAccommodationPartial
RestaurantsCafes	EntertainmentVenuesPartial
	ClosPrim
	RestaurantsCafes
	ClosPrimPartial
	NonEssentialShopsPartial

Table 3: In general, more developed countries seem to have a restriction on more commercial activities as the most important interventions. To explain a few of the categories that are not obvious,ClosePrim and CloseHigh are closing primary and higher education resp. Partial implies partial closing.

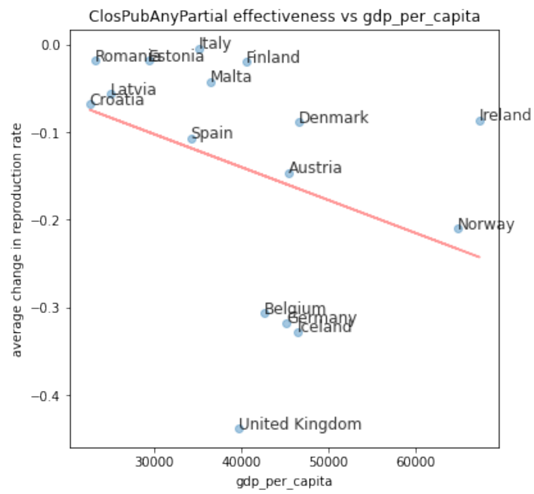

Figure 9: gdp-per-capital vs the change in reproduction rate after closing public spaces to suppress a Covid transmission vector

Noting the trend and differences in developed vs undeveloped countries, we seek to further confirm the findings of the clusters. We will be focusing on the parameters in the more developed cluster through the following analysis.

We wish to show a relationship between the developmental level of a country and certain intervention's effectiveness by comparing the average reduction of the rate of reproduction. For an intervention M implemented by country K, We take the average change in reproduction rate that occurs over M days, and we normalize this change in reproduction rate by all the other concurrently applied interventions (using regression coefficients). This yields a series of datapoints of the form ($\Delta_{r_{reproduction}}$, K). We then take each country K's GDP-per-capita and plot them with a line of best fit.

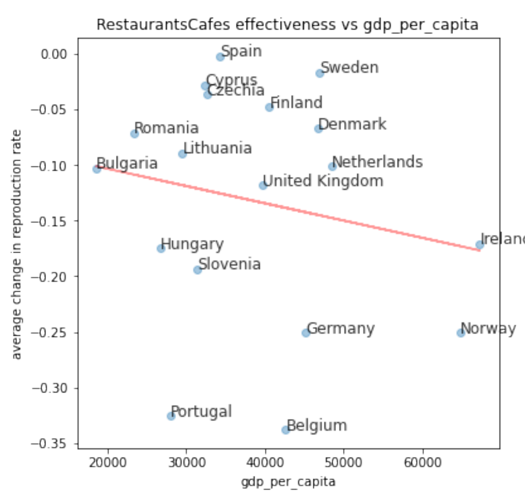

Figure 10: gdp-per-capital vs the change in reproduction rate after closing restaurants to suppress a Covid transmission vector

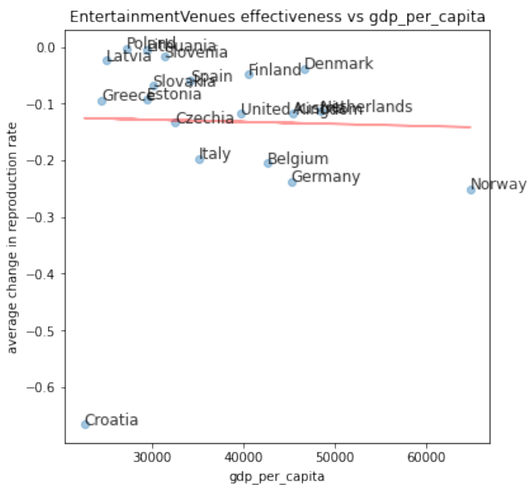

Figure 11: In figure 9, figure 10, and figure 11 We measure effectiveness as the average change of reproduction rate over the policy's implementation period (normed for duration) for both ClosePub, Closing RestaurantCafes, and closing EntertainmentVenues. For the first two we note a trend where increase in gdp correlates with more effective reduction of reproduction rate in those interventions. In the intervention on entertainment venue, we note that the trend is mostly flat as we vary gdp and only slightly negative.

Based on these results in figure 9, 10, and 11, we see that increasing gdp per capital introduces a more negative change in reproduction rate with the same interventions (ClosPubAny, close restaurantsCafes, close EntertainmentVenues); thus the interventions can be said to be relatively more "effective" in wealthier countries compared to less wealthy countries. We provide a theory to why this is. Since, gdp is correlated with disposable income, and that nations with higher gdp tend to have a more developed consumer economy, the interactions in the consumer economy such as shopping in public places, eating out at fine restaurants and spending time in entertainment venues will be how the majority of the population spends their time, and thus constitute the biggest vector for Covid spread in those countries. This is most likely why "richer" countries benefit from these types of Covid interventions a bit more. Contrastingly, less developed countries still benefit, just not as much from lockdowns in these specific areas since their population is less likely to spend time in these areas to begin with, thus areas such as restaurants are less likely of a Covid reproduction vector.

Another interesting trend in the effectiveness of mandating closing public spaces is its relationship with the inverse of the median age of the population. We plot the average change in reproduction rate for ClosPubAnyPartial in each country as a function of the median age of that country.

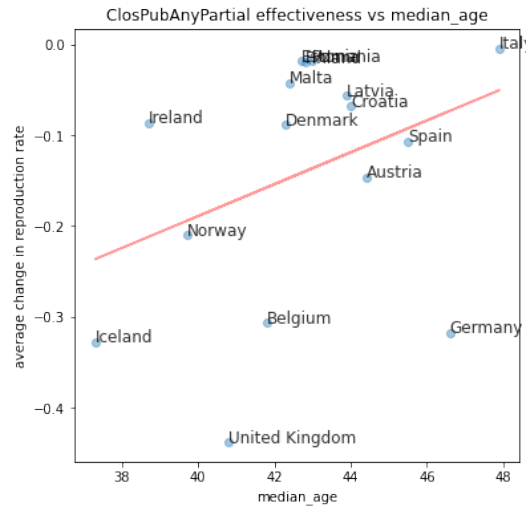

Figure 12: We take the effectiveness as a average change of reproduction rate over the policy's implementation period (normed for duration) for Closing public spaces (ClosePub). It appears that as the average person gets older in country K, the country's effectiveness in mandating closing of public spaces decreases, this might be due to older populations being aware of their higher risks of Covid, and voluntarily avoid public places.

We note that as age increases, closing public spaces (ClosePub) is less likely to affect the population as a primary vector of disease transmission. This may be due to changing life style from city focused to a suburbia focused lifestyle as people age, or it maybe that older aged population purposefully avoid public spaces since they know they are more susceptible to Covid. We cannot concretely establish the reason for this trend, though we note that it exists.

Finally, we also intuitively note that the longer the intervention is, the more effective it should be at decreasing the Covid reproduction rate. To quantify this, we looked at the net change in reproduction rate at the beginning and end of government intervention outlined in the category in developed country above. We then plot this as a function of the length of the intervention, producing one graph for each intervention.

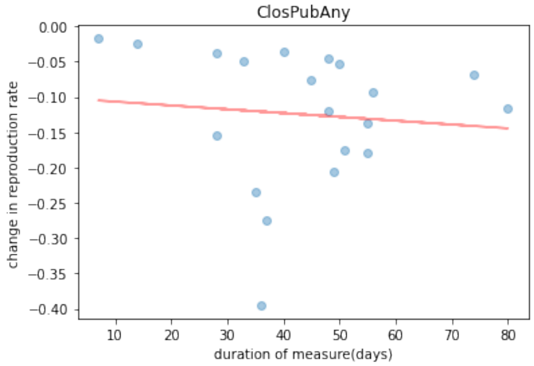

Figure 13: For Closing public spaces (ClosePub), it appears that as the length of intervention increases, the more effective it is at lower reproduction rate

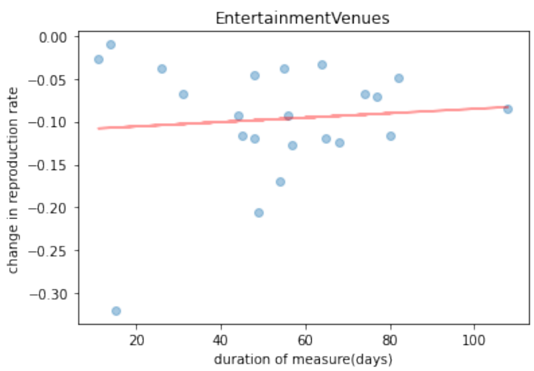

Figure 14: For Closing public spaces (ClosePub), it appears that as the length of intervention increases, the less effective it is at lower reproduction rate

For figure 13, the change is as we expect, we see that the longer the public spaces are close the more impact this has on the reproduction number. However we also note a peculiar trend with closing entertainment venues, in that the longer they are closed, the less effective they become at decreasing the reproduction rate. We theorize that it could be due to people who used to be able to go entertainment venues no longer able to do so, and the longer they go without it, the more likely they will seek some alternative, which is making the intervention less effective. Further research is needed to test this hypothesis, but the existence of this possibility demonstrates a key challenge in intervening on public health through mandates: there's a limit to how well we can enforce mandates, and "quarantine fatigue" [8] impacts this measurement of enforcement.

5 Data Cleaning

To address the issue of missing data, we used linear interpolation for any gaps in the middle and the end of the dataset, and fill the beginning null values as zero since all Covid-related indicators should be zero before any records exists for a country.

We selected only the indicators that are representative of scale, meaning that we used rate data rather than a simple scalar number (such as death or new case number), and analyzed them considering the number of Covid tests given. We generally try to use positive rate or tests per case data to address this issue.

For PCA analysis, all data are standardized to zero mean and unit variance. For the comparison of different public health treatments, we used owid and ecdc datasets. The dates for owid are in 'yyyy-mm-dd' format while those for ecdc are in 'yyyy-WeekNum' format. We also created functions to convert them to each other, assuming that any weekly data is released on the first day of the week.

6 Conclusion

In our analysis we first investigate the relationship between Covid-19 with demographic and socioeconomic factors and find that life expectancy, human development index and cardiovascular death rate are the most important factors differentiating the Covid-19 status within European countries. Based on this discovery, we measure the effectiveness of prevention measures within countries with similar demographic and socioeconomic factors. We then tested the effectiveness of quarantine measures for travellers arriving in a new state using network SEIR model and find that quarantine is helpful in reducing the number of infections. Finally, we perform counterfactual analysis to study causal relationship between the prevention measures and the spread of the virus. We analyze the stay-home-order and the closure of schools and investigate the relationship between their effectiveness and the developmental indicators and we find that increasing gdp per capital introduces a more negative change in reproduction rate with the same interventions and we note that as age increases, closing publicspaces (ClosePub) is less likely to affect the popu-lation as a primary vector of disease transmission.

References

[1] Mathias Peirlinck et al. "Outbreak dynamics of COVID-19 in China and the United States". In: (Apr. 2020). DOI: 10.1101/2020.04.06.20055863.

[2] Statista. "Rt of COVID-19 in the U.S. as of January 23, 2021, by state". In: 2020. URL: https://www.statista.com/statistics/1119412/covid-19-transmission-rate-us-by-state/.

[3] Bureau of Transportation Statistics. "Passenger Travel of USA". In: URL: https://www.bts.gov/topics/passenger-travel.

[4] Stanford ME 233. "Travel Graph for USA States". In: URL: https://canvas.stanford.edu/files/6773280/download?download_frd=1.

[5] *Data on 14-day notification rate of new COVID-19 cases and deaths*. 20. URL: https://www.ecdc.europa.eu/en/publications-data/data-national-14-day-notification-rate-covid-19.

[6] T. K. Vintsyuk. "Speech discrimination by dynamic programming". In: *Cybern Syst Anal* 4 (1968). DOI: https://doi.org/10.1007/BF01074755.

[7] *Our World In Data*. 20. URL: https://ourworldindata.org/coronavirus.

[8] J. Zhao et al. "Quarantine Fatigue: first-ever decrease in social distancing measures after the COVID-19 outbreak before reopening United States". In: *ArXiv* abs/2006.03716 (2020).

Appendix

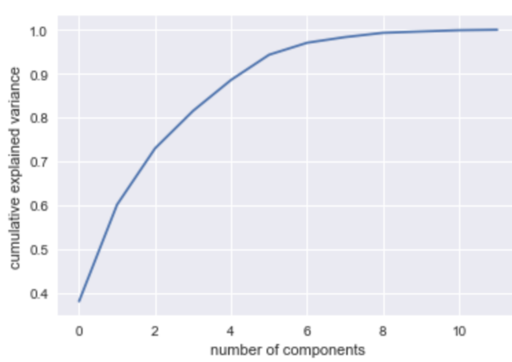

Figure 15: Cumulative summation of the explained variance versus the number of Principal components used.

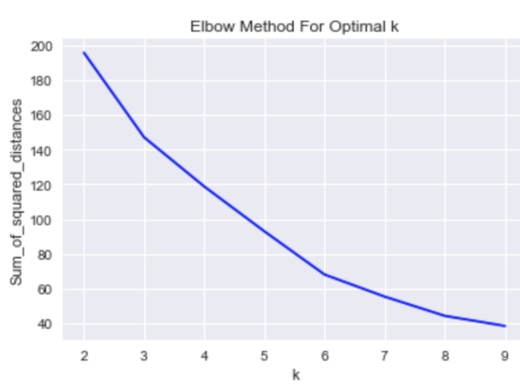

Figure 16: Sum of squared distances within the clusters versus the number of clusters.

Figure 17: Principal Component Analysis on European countries demographic and socioeconomic indicators.