Data-Driven Estimation of Effectiveness of COVID-19 Non-pharmaceutical Intervention Policies

1st Yash Mahajan
Computer Science and Software Engineering
Auburn University
Auburn, USA
yzm0034@auburn.edu

2nd Sheikh Rabiul Islam
Department of Computing Sciences
University of Hartford
Connecticut, USA
shislam@hartford.edu

3rd Mohammad Ruhul Amin
Computer and Information Science
Fordham University
New York, USA
mamin17@fordham.edu

4th Shubhra Kanti Karmaker ("Santu")
Computer Science and Software Engineering
Auburn University
Auburn, USA
sks0086@auburn.edu

***Abstract*—Non-pharmaceutical Interventions (NPIs), such as Stay-at-Home, and Face-Mask-Mandate, are essential components of the public health response to contain an outbreak like COVID-19. However, it is very challenging to quantify the individual or joint effectiveness of NPIs and their impact on people from different racial and ethnic groups or communities in general. Therefore, in this paper, we study the following two research questions: 1) How can we quantitatively estimate the effectiveness of different NPI policies pertaining to the COVID-19 pandemic?; and 2) Do these policies have considerably different effects on communities from different races and ethnicity? To answer these questions, we model the impact of an NPI as a joint function of *stringency* and *effectiveness* over a duration of time. Consequently, we propose a novel *stringency function* that can provide an estimate of how strictly an NPI was implemented on a particular day. Next, we applied two popular tree-based discriminative classifiers, considering the change in daily COVID cases and death counts as binary target variables, while using stringency values of different policies as independent features. Finally, we interpreted the learned feature weights as the *effectiveness* of COVID-19 NPIs. Our experimental results suggest that, at the country level, restaurant closures and stay-at-home policies were most effective in restricting the COVID-19 confirmed cases and death cases respectively; and overall, restaurant closing was most effective in hold-down of COVID-19 cases at individual community levels such as *Asian, White, Black, AIAN and, NHPI*. Additionally, we also performed a comparative analysis between race-specific effectiveness and country-level effectiveness to see whether different communities were impacted differently. Our findings suggest that the different policies impacted communities (race and ethnicity) differently.**

***Index Terms*—COVID-19, Non-Pharmaceutical Interventions (NPIs), Communities, Policy Stringency**

I. Introduction

Non-pharmaceutical Interventions (NPIs) are essential components of the public health response to control and contain an outbreak like COVID-19. NPIs include policies like *Stay-at-Home*, *Face-Mask-Mandate*, *Closing-and-Reopening-Businesses*, *Travel-Ban*, among others. These measures help suppress the spread of the virus, thus as an outcome, reducing the infected population size and buying more time for healthcare professionals to better handle the pandemic. Not to mention that during the COVID-19 pandemic, those interventions also allowed us ample time to create potential vaccines and drugs to fight the virus. Recent research has shown that large-scale implementation of joint NPIs is effective in containing the virus [1], however, the impact of individual NPI is relatively under-explored. A reasonable way to estimate the effectiveness of an individual NPI is to adopt a data-driven approach to model the decay in COVID-19 cases/deaths as a joint impact function of multiple NPIs and then disentangle the weights of each individual NPI by fitting the target variable with NPIs impacts as independent variables and subsequently performing a detailed feature analysis. An additional challenge associated with this task is that there is always a lag between policy implementation and its effect on the targeted population [2]. Policy lag is well studied in economics [3], according to which we propose that a delay between NPI implementation and its consequence is expected and thus should be considered for NPI impact modeling. Therefore, we also incorporate a lag effect in our estimation of effectiveness.

In terms of research questions, we investigate the following two questions in this paper:

1) How can we quantitatively estimate the effectiveness of NPI policies pertaining to the COVID-19 pandemic?
2) Do these policies have significantly different effects on communities of different races and ethnicity?

In order to answer these questions, we considered the impacts of following five major policies related to COVID-19 NPIs implemented by the US government during the pandemic period from March'2020 to April'2021, i.e., *Stay-at-Home* [4], *Face-Mask-Mandate* [5], *Closing-and-Reopening-Restaurants* [6], *Closing-and-Reopening-Businesses* [7] and *Travel-Ban* [8].

Technically, we model the impact of an NPI over a duration of time as the product of the stringency and effectiveness of the corresponding NPI over that particular time period. Consequently, we propose a novel stringency function for NPIs over a duration of time, which can provide an estimate of how strictly an NPI was implemented on a particular day. This stringency function has been designed by incorporating the Policy Stringency Index Value (PSIV) [9] provided by The Oxford COVID-19 Government Response Tracker (OxCGRT) and subsequently, assuming an exponential decay period after the policy is lifted.

Next, we trained two popular discriminative classifiers, i.e., Random Forest [10] and Gradient Boosted Trees [11], with the change in daily COVID cases and death counts as binary target variables and stringency values of different policies as features. We conducted this analysis with data from both country-level (Whole US) and community-level (refers to six races, *–White, Black or African American, Asian, American Indian and Alaska Native (AIAN), Native Hawaiian and other Pacific Islander (NHPI)*). Finally, we interpreted the learned feature weights as the effectiveness of COVID-19 NPIs. Our experimental results suggest that, on a country level, restaurant closures and stay-at-home was most effective in restricting the COVID-19 confirmed cases and death cases, respectively, and overall, restaurant closing was most effective in hold-down of COVID-19 cases at individual community levels.

A. Detailed Literature Review

According to a news article in Nature by Gibney, 2020 [12], *"working out the effectiveness of the measures implemented worldwide to limit the coronavirus's spread is now one of the scientists' most pressing questions"*. In our paper, we aim to answer this pressing question by estimating the effectiveness of NPI (Non-Pharmaceutical Intervention) policies over a particular period. By far, to constrain the spread of the virus, governments from all over the world have responded with multiple NPIs such as *Face Mask Mandate*, *Social Distancing*, and many more. However, the effectiveness of individual policies is still under-studied. Recent works focused on estimating the joint effect of policies using Reproduction number (R_t) [13], [14] or the mobility rate [15], [16] or the Ordinary Differential Equation (ODE) [17]–[20].

Many studies have been conducted to analyze the joint impact of NPI policies through mathematical or statistical modeling of the severity of the spread of the COVID-19 virus. The most common method used to quantify the impacts of the policies on the model is the spread through Ordinary Differential Equations (ODE) [17]–[20]. For example, Johndrow et. al. [19] used simple SIR (Susceptible-Infected-Recovered) model to estimate the true count of confirmed cases with NPIs in consideration. They proposed the SIR model to estimate the disease spread using likelihood and suggested that the actual number of cases was 6-10 times higher than the reported cases and also accounted for the lag in time from infection to death and the infection fatality rate. Coughlin et.al. [20] used the SEIR model for Wuhan, China, and pointed out that if the strictness of social mixing policies were prolonged till April'2020, then the peak of the spread of the virus could have been delayed further. [21] proposed an enhancement in the traditional SEIR model by incorporating a micro-simulation modeling framework. This proposed framework estimates the transmission effect between susceptible and infectious individuals by approximating the impact of NPIs on the population of the USA and the United Kingdom. However, these simulation-based models are very complex in nature and may become biased in prediction tasks. The complex model may learn some complex behaviors of diseases but these behaviors are difficult to validate. Therefore, we rely primarily on a data-driven approach to estimate the effectiveness of the NPIs.

Several other methods were also proposed to estimate the efficacy of the policies such as the Change Point Detection (CPD) model [15], [22], or the Bayesian models [13]. The CPD model finds the abrupt changes in the time series data by observing the change in mean and variance of the distribution of infection. Previously, this model was mostly used for stock market analysis, genomics data modeling, or segmentation. Later, researchers found similar abruptly changing trends in the distribution of COVID-19 infection like in stock markets. Bian et.al. [15] used the CPD model to estimate the impact of NPIs on the transportation system using mobility data and incorporated the lag time in the reported cases. Dass et.al. [16] assessed the impact of social gathering policies using mobility data on the spread of the virus using the CPD method. Mbuvha et.al. [22] combined the Bayesian inference method with the simple SIR model to estimate the rate of spread of the virus due to the travel ban policy in South Africa. Brauner et.al. [13] used a bayesian hierarchical model to link the dates to cases and deaths and modeled each NPI effect as a multiplicative reduction in reproduction number (R) and estimated mean reduction in R across the countries. They suggested that a limit in gathering size up to 10 people, school closure, and high-exposure business closure were more effective in reducing the spread of the virus than the stay-at-home policy. Moreover, Cowling et.al. [14] estimated the NPI's impacts using the behavior change in population using reproduction number (R) and observed a decline in cases after the social distance and school closure policies were imposed.

However, all these approaches do not estimate how the policy's behavior changes during its active time. Therefore, OxCGRT (Oxford Coronavirus Government Response Tracker) [9] provides an intuitive way to track the policy behavior over the duration by estimating the government responses to the spread of COVID-19 and intervention across a standardized series of indicators for over 180 countries including the USA. The [9] contains eight categories on containment and closures, four categories on economic policies, and eight categories on health system policies. The categories reflect the continuous scale of stringency index values which allow us to conduct a quantitative analysis of government response and behavior of each policy during its active time. The stringency index values soon became notable due to their veracity and were thus used by many studies. For instance, [23] proposed a method that combines the growth rate of the virus before the vaccination rollout with the stringency index and reported that around 20-60% decline was observed after the first vaccination rollout. [24] examined the containment of stringency with a cumulative incidence of cases and reported the mitigation of the COVID-19 virus across 28 European countries as an effect of containment measures in the first wave. [25] examine the relationship between government response level, response time, and the epidemic trajectory using OxCGRT stringency value and Group-based trajectory modeling method. They reported that the early start of a high-level response from the government correlates highly with the early arrival of the peak number of daily new cases. Similarly, [26] examined the chaotic behavior of COVID-19 which depict the change in asymptotic behavior and trajectory of increased/decreased cases which resulted in easing and tightening of restriction.

In this paper, we analyze the effect of NPIs for the USA using the stringency index values as this country has been a special case in terms of COVID-19 response as the US government gave authority to sub-national states governments to impose policies according to their will [27], [28]; in contrast to other countries like China, India, South Korea, where the implementation of each policy was centralized to the country. Moreover, the US government had a delayed response to handling the pandemic and, therefore, became the epicenter [28], [29]. To quantitatively estimate the effectiveness of the individual NPI policies on the country-level and individual race/community levels, we first computed how strictly a policy was implemented at a particular timestamp using our proposed novel stringency function. Next, we trained two tree-based discriminative classifiers on daily COVID-19 cases and death counts as targets and stringency values of different policies as features. Finally, we interpreted the learned feature weights as the effectiveness of COVID-19 NPIs and performed a qualitative analysis of our findings.

II. Problem Formulation and Effectiveness Model

During the pandemic, several NPI policies have been imposed and lifted to control the spread of COVID-19; as a consequence, we saw multiple ups and downs in the number of recorded cases. We assume that these variations primarily occurred due to the following two factors:

1) **Effectiveness Index:** How effective an NPI policy is in restricting the spread of the virus if implemented properly. This is essentially our primary research question, let's call it *Effectiveness Index* and denote it as α. We assume that the *Effectiveness Index* of an NPI does not change with time, which is a reasonable assumption.
2) **Stringency Index:** How strictly the NPI policy was implemented by the government and law-enforcement agencies. Let's call it *Stringency Index* and denote it as β^t. Parameter t denotes a discrete timestamp (e.g., day as a unit) and captures the fact that the *Stringency Index* can and does change over time. For example, the government can sometimes be more lenient/strict towards the implementation of a particular law.

The impact of an NPI policy P at timestamp t is defined as $I_t(P) = \alpha\beta^t$. Next, we define the target as a temporal binary random variable $X(t)$, where, $X(t) = 1$ means decay in the number of positive cases at timestamp t and $X(t) = 0$ means otherwise. Finally, when multiple policies (P_1, P_2, ..., P_k,) are considered jointly, the target variable can be defined as the following function:

$$X(t) = f\left(I_t(P_1), I_t(P_2), ..., I_t(P_k)\right) = f\left(\alpha_1\beta_1^t, \alpha_2\beta_2^t, ..., \alpha_k\beta_k^t\right) \quad (1)$$

As our primary goal is to estimate α's (effectiveness index), we first computed $X(t)$ from the publicly available CDC infection database. We then proposed a novel stringency function for NPIs to compute values for $\beta_k(t)$. Next, we trained two popular discriminative classifiers, i.e., Random Forest and Gradient Boosted Trees, with daily COVID cases and death counts as targets and stringency values of different policies as features. Finally, the learned feature weights from the training process are interpreted as our estimates of α_1, α_2, ..., α_k.

A. Proposed Stringency Function

Devising an accurate Stringency Index (how strictly a policy is implemented at a particular moment) is very challenging as it can change with time and depends on various social and external factors. To provide a comprehensive definition of *Stringency*, we considered the following possible cases. For notations, we use t as the current timestamp, P_{Start} as the timestamp when the policy was imposed, and P_{Lifted} as the timestamp when the policy was lifted.

Case 1: When the policy has not yet been implemented, i.e., current timestamp *t* is smaller than the timestamp when the policy was imposed, i.e., $t < P_{Start}$ We assume that there is no stringency as well as the impact of a policy, i.e., $\beta_t = 0$. Therefore, $I_t(P) = 0$. In simple words, it means that we are looking at the days before the policy was imposed. Refer to figure 1, all straight vertical lines reflect the day on which policy was imposed, and before that, we observe no effect of a policy.

Case 2: When the current timestamp *t* lies in-between the policy starts day and policy end day, i.e., $P_{Start} + w_1 < t < P_{Lifted}$, we assume the stringency index (β_t) of a policy to be equal to the Policy Stringency Index Value (PSIV) [9] provided by the Oxford COVID-19 Government Response Tracker (OxCGRT). The PSIV ranges between 0 and 1 where 0 means no strictness of a policy and 1 means very strict policy administration. Note that, we assumed a lag of w_1 days before using PSIV as the stringency index, which

Algorithm 1: Computation of Stringency Index

Data:
(1) t: input timestamp (discrete timestamp)
(2) P: a particular policy
(3) $PSIV$: Policy Stringency Index Value (PSIV) [9] provided by the Oxford COVID-19 Government Response Tracker (OxCGRT)
(4) P_{Start}: timestamp when policy P was imposed
(5) P_{Lifted}: timestamp when policy P was lifted
(6) w_1: Assumed delay effect in PSIV (set to 5 days)
(7) w_2: Aftermath effect of a policy after lift (set to 10 days)
(8) γ: decaying rate (set to 0.01)

Result: β_k^t (Stringency Index of a Policy k at timestamp t);

```
if t < P_Start + w_1 then                              // Case 1
    β_k^t ← 0;
else if P_Lifted ≠ ∞ then
    if P_Start + w_1 < t < P_Lifted then               // Case 2
        β_K^t ← PSIV_P[t] ;
    else if P_Lifted < t < P_Lifted + w_2 then         // Case 3
        β_k^t ← PSIV_P[P_Lifted];
    else                                               // Case 4
        β_k^t ← PSIV_P[P_Lifted] × exp^{-γ×[t-w_2-P_Lifted]}
else                                                   // Case 5
    β_k^t ← PSIV_P[t]
```

captures the fact that the impact of a policy is not immediately observed. In other words, any cases reported on the current timestamp are a result of the infection received w_1 days ago. w_1 essentially models the incubation period [30] which is set to 5 days as per CDC guidelines.

Case 3: When the current timestamp is in between the Policy lift date and w_2 days after lifting,, i.e., $P_{Lifted} < t < P_{Lifted} + w_2$. This case captures the fact that the effect of a policy does not immediately disappear when it is lifted, rather it vanishes gradually and the vanishing effect is often observed after a certain lag period. Therefore, we assumed a lag period of w_2 before observing the decay of stringency. During this period, we assumed the stringency index of a policy to be equal to the same PSIV value on the lift day.

Case 4: More than w_2 days since the policy was lifted, i.e., $t > P_{Lifted} + w_2$. As the state officials lift the policy, we assume that a policy would still have a diminishing impact over time. This effect is expressed as a exponential decay function (β^t = PSIV[P_{Lifted}] $\times \exp^{-\gamma\times[t-w_2-P_{Lifted}]}$). Note that, here we are multiplying the PSIV of the policy lifted day, which remains the same till the end of the analysis period.

Case 5: When the state official didn't lift the policy before the analysis period, i.e. the policy has not been lifted during the analysis period. Thus, the policy effect would be equal to the PSIV on that particular day. Refer to figure 1-*face mask policy* in red. In all six states shown, the face-mask policy was imposed but was not lifted. Therefore, we did not see the decay effect of the face-mask policy.

Algorithm 1 combines these five cases into a single function which takes as input a particular timestamp t and returns the *Stringency Index* at timestamp t. The stringency indexes of different NPI policies are then used as features to represent the joint impact of NPIs on a particular day and further classified by a discriminative classifier like Random Forest, Gradient Boosted Trees, etc. into a binary output which represents whether there is a decay in the COVID-19 positive cases/deaths or not.

Figure 1 demonstrates the stringency of five NPI policies in 6 states [Alabama (AL), Arizona (AZ), California (CA), Florida (FL), Connecticut (CT), and Hawaii (HI)] of USA during the analysis period from March'14'2020 to April'10'2021. Here, the x-axis represents the number of days and the y-axis shows the stringency of individual NPI policies. Note that, we have evaluated the stringency of the policies across 48 states of the USA, and Figure 1 shows six states only due to lack of space. To interpret Figure 1, day 1 starts from March'14'2020 and last day represents the 410^{th} day i.e. April'10'2021. First, if we observe the CA state, we can notice that out of 5 policies, 4 policies were imposed except for the "Travel Ban" policy. The CA state government did not impose the "Travel Ban" policy during the analysis period and therefore, the policy did not have any impact (hence no stringency as well) in refraining from the COVID-19 infection in CA. Second, if we observe the "Face Mask Mandate" Policy in the states of CA and MD, we will notice that the state officials did not lift the policy after imposing it, therefore, we consider the policy stringency value until the last day of the analysis period without any exponential decay. In-state HI, state officials procrastinated the implementation of the "Face Mask Mandate" policy. As a result, the Face Mask policy has very less impact on the Hawaiian population. Third, many of the states imposed Stay-At-Home, Closing Restaurants, and Business Closure on the same day, for instance, state CA and AL had imposed "Closing Restaurants" and "Business Closure" policies on the same day, and thus, we see similar trends for those NPI policies.

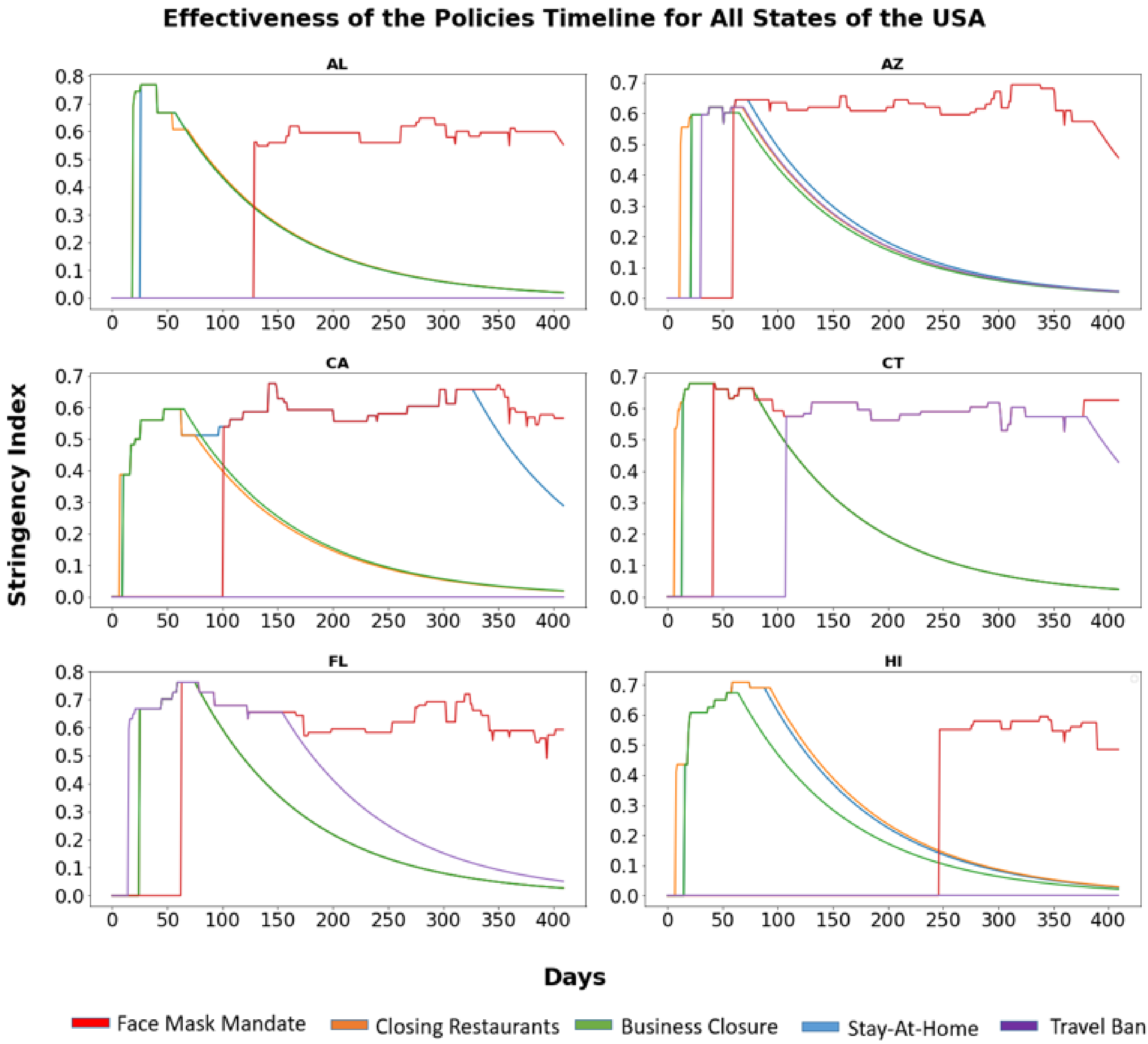
B. Estimation of NPI Policy Effectiveness

Target Labels: Unfortunately, we do not have ground truth labels for the joint impact of NPIs at any given timestamp. To address this limitation, we applied the following heuristics:

Given a particular timestamp t, the change in daily confirmed cases and death counts can serve as a good indicator of the joint impact of NPIs at that particular timestamp.

This is a reasonable assumption because if one or more NPIs are effective, they must help decrease the number of confirmed cases and death counts. Otherwise, they will be deemed ineffective. According to CDC, the average incubation

Fig. 1. The Figure demonstrates the effect of policies on six states during the span of the analysis period i.e. March 14, 2020, to April 10, 2021. The six states are Alabama (AL), Arizona (AZ), California (CA), Connecticut (CT), Florida (FL), and Hawaii (HI). AL, CA, and HI implemented four policies; whereas, AZ, CT, and, FL implemented all 5 policies. The decay curve begins the next day of a policy lift and showcases the aftermath effect of a policy. Among all six states shown, CT shows a longer hold of policies and therefore, more strictness of policies in the city.



period is around 5.6 days [30] and in the work, we captured this incubation period with the lag period w_1 which was set to 5 days. This means that any person reporting a positive case today could have gotten the infection around 5 days ago (as per CDC). Consequently, we labeled the data in the following way: on a particular timestamp t, if we observe a decline in the number of cases in terms of the moving average of the preceding 5 days from the current timestamp t, then timestamp t is assigned label 1, otherwise 0. Label 1 indicates that the cases are reducing on average for the past 5 days, whereas Label 0 indicates that cases are non-decreasing. We created two different target variables, i.e., decay in confirmed cases and deaths, separately using the same strategy and formulated a binary classification problem which was trained using tree-based methods.

NPI Policy Effectiveness: After constructing the feature vectors using algorithm 1 and creating the binary target variables as described above, the data set is ready to be used for the classification task. Since our data is non-linear, we used tree-based methods to perform the binary classification task and then ranked the policies based on the learned weights (α_i). Two tree-based methods we experimented which include 1) Random Forest and 2) Gradient Boosted Trees. The learned feature weights are finally interpreted as the effectiveness of individual NPI policies.

Policy Effectiveness on Different Races: Generally, a policy is implemented to benefit all communities equally; however, during the pandemic era, it has been reported that different communities and races were impacted differently [31]. Indeed, the impact of the COVID-19 pandemic in the Country has shed light on inequities among different races and ethnicity. The official report by Commonwealth Fund analysis [32] suggests that there are high disparities in COVID-19 cases and deaths in communities. Therefore, to quantify the disparities in policy impacts across different races, we computed the effectiveness index of each policy at individual race levels. Next, we computed the squared difference between the race-specific effectiveness index and the country-level effectiveness index normalized by the country-level effectiveness index. This squared difference signifies how differently a particular race/community was impacted by a policy in comparison to the whole country-level impact.

$$Squared\_Difference = \left\{ \frac{1}{\alpha_k^C} * \left(\alpha_k^C - \alpha_k^r \right) \right\}^2 \qquad (2)$$

where, k is a specific policy; r is a specific race, α_k^C = effectiveness index of policy k on Country-Level and α_i^r = effectiveness of each policies on race r.

III. Experiments and Results

Data-set: We used four different publicly available data sets dated from March 14' 2020 to April 10' 2021, to analyze the effects of NPI policies on the COVID-19 confirmed cases and deaths across 48 states of the USA. The datasets are:

1) Daily reported cases and death counts by each state [33]
2) Daily reported cases of communities [34].
3) Policy implementation records [35]
4) Policy Stringency Index value (PSIV) [9] records

The analysis is first performed at the country level and then, at the individual race level. The community refers to six races, *–White, Black or African American, Asian, American Indian and Alaska Native (AIAN), Native Hawaiian and other Pacific Islander (NHPI).*

Model Performance: To evaluate the effectiveness of a policy we trained two tree-based models; Random Forest (RF) and Gradient Boosting (GB) on stringency index values as a dependent feature. The independent feature is the manually annotated binary labels which depict the increase/decrease of the number of confirmed cases and death counts on a particular day (also refer to Target Label in the previous section). Since a stringency index value of the policies depicts the non-linearity in the dataset with no ground truth labels, therefore, we decide to use the tree-based classifier and compare the model performance. To train the classifier, first, we generate the data using algorithm 1 as stringency index values. Next, we trained

TABLE I

This table shows models accuracy scores (%) on **country-level** when trained with stringency index values. We compare the accuracy value of both models (Random Forest or RF; and Gradient Boost or GB) using confirmed cases and death counts observed per day. Here, the accuracy depicts that all models can predict the increase/decrease in the number of cases/death on a particular day with at least 70% accuracy on the country level.

Models	**Accuracy Metric**	
	Cases	**Deaths**
RF	79.67	71.21
GB	80.10	72.76

TABLE II

This table demonstrates the model accuracy scores (%) on **community-level**. These models (Random Forest or RF; and Gradient Boost or GB) were trained with stringency index with increase/decrease confirmed cases as labels. The accuracy metric shows that models can predict the increase/decrease in confirmed cases on a particular day with at least 70% accuracy.

Communities	**Accuracy Metric**	
	RF	**GB**
Asian	72.18	73.22
White	71.1	72.34
Black	70.12	71.85
AIAN	75.8	72.3
NHPI	77.21	77.48

each classifier with the stringency value and target label. Next, we performed extensive parameter search on both the models with search space as 1) *'max_depth' : [3, 5, 8, 15, 25, 30]*, 2) *'n_estimator' : [1, 10, 25, 50, 100, 300, 500, 800, 1000, 1200]*. Next, we used the best-valued parameters for both models; for *RF : ['max_depth': 15, 'n_estimator': 800]* and *GB : ['max_depth': 15,'n_estimator': 1000]* and computed the feature importance vector using best-fitted parameters. These feature importance scores reflect the effectiveness of the policies with are shown in table III and IV. Both the models well fitted to the data and demonstrate a comparable score for policy effectiveness estimation at the country level. The accuracy in table I demonstrates the model's ability to predict the increase/decrease of the number of confirmed cases on a particular day with 79.76% and 80.10% accuracy with RF and GB respectively. Similarly, prediction of death counts with 71.21% and 72.76% accuracy. We believe this work is very crucial as it can predict the trend of death counts with such high accuracy.

Further, extend the experiment to the community level where we utilize the same set of best-fitted hyper-parameter from the country level on both models respectively. The model performance on community-level cases is demonstrated in

TABLE III

EFFECTIVENESS OF THE POLICIES (α) ON CONFIRMED CASES AND DEATH ON THE COUNTRY-LEVEL ESTIMATED AFTER TRAINING RANDOM FOREST (RF) AND GRADIENT BOOSTING (GB). TABLES (A) AND (B) SHOW THE EFFECTIVENESS OF POLICY (α) WHICH IS LEARNED DURING THE TRAINING OF THE MODELS WITH STRINGENCY INDEX. TABLE (A) REPRESENT THE EFFECTIVENESS OF A POLICY WHEN TRAINED ON CONFIRMED CASES, WHEREAS, TABLE (B) REPRESENTS THE EFFECTIVENESS OF A POLICY WHEN TRAINED WITH DEATH COUNTS.

Method	Stay Home	Closes Restaurant	Closes Business	Face Mask	Travel Ban
RF	0.221	0.257	**0.258**	0.177	0.087
GB	0.207	**0.285**	0.216	0.198	0.093

(A) Policy Effectiveness (α) on Confirmed Cases

Method	Stay Home	Closes Restaurant	Closes Business	Face Mask	Travel Ban
RF	**0.280**	0.173	0.204	0.199	0.144
GB	**0.307**	0.158	0.184	0.216	0.134

(B) Policy Effectiveness (α) on Death cases

table II. We can observe that both models' performance aligns with the country-level model performance which depicts that the data generated from the proposed algorithm is consistent with country-level and community-level confirmed cases. Table II shows that models can predict the change in cases on the community level as well. With nearly more than 70% accuracy from both models, we can predict which community will suffer the most. And, thus, we believe this will help authoritarians to make careful discussions on the community level.

Country-Level Analysis: Table III showcases the effectiveness indexes estimated by two tree-based methods, i.e., 1) Random Forest and 2) Gradient Boosting, using country-level confirmed cases and death counts as target variables separately. As we observe from table III that nearly both methods suggest that the policy *Closing Restaurants* was most effective in refraining from the spread of COVID-19 cases (number of confirmed cases). According to Forbes [36] and Stanford University [37] report, if restaurants were allowed to open, then they would have been responsible for more than 600K infections in major cities. Additionally, in retrospect, it was concluded that 10% of leniency in restaurant opening could cause more than 85% of the cases according to the same reports. Therefore, we believe that *Closing Restaurants* policy was vital in hindering the spread of COVID-19 at the country level.

On the other hand, if we observe the death cases, both methods confidently suggest that the *Stay-At-Home* policy has reduced the risk of death in the country the most. Although the effectiveness indexes from both methods differ slightly in Random Forest (0.280) and Gradient Boosting (0.307), they are pretty close, suggesting that the *Stay-At-Home* policy has distinctly contributed to reducing the death toll in the country. It is very intuitive that, to reduce the number of death, the infected people need to be isolated and therefore, the *Stay-At-Home* policy would be the right action. According to a report from the University of Alabama, Birmingham(UAB) [38], with the absence of a *Stay-At-Home* policy, the death rate could have been 22% higher as opposed to if the policy was implemented nationwide. Additionally, according to the same reports, a quick stringent lockdown was very effective in controlling the early death toll due to its immediate effect. On the other hand, other policies have contributed a fair share of the amount in controlling the death toll. It can be observed that the *Face Mask* policy was also very effective following the *Stay-At-Home* policy. *Closing Restaurant* policy was not found to be as effective for death count reduction as it was for containing the spread of the virus. This finding is supported by CDC [39] as well which stated the following about *Closing Restaurant* policy: *"0.7 percentage point decrease (p=0.03) in the daily death growth rates 1-20 days after the implementation, a decrease of 1.0, 1.4, 1.6 and 1.9 percentage points 21-40, 21-60, 61-80 and 81-100 days respectively"*. Additionally, *Restaurant Reopening* did not have a significant impact on death counts until day 60 of the policy lift. It was observed that after day 60, CDC [39] saw an increase of only 2-3% points in death cases.

In summary, if we compare the effectiveness of policies on reducing confirmed cases and death counts at the country level, the effects of different policies appear to be somewhat different. For containing the spread, closing restaurants/ businesses were found to be very effective; while for reducing death counts, stay-at-home was the most useful.

Individual Race-Level Analysis: We further extended our analysis to the individual races using the target variable "confirmed cases". It was observed that some races have been impacted differently compared to the overall population of the country. From the country-level analysis, it was found that policy *Restaurant Closing* has the highest effect in containing the COVID-19 confirmed cases, and from table IV it is observed that the country-level analysis results do not apply to all races. Indeed, for some races, *Closing Restaurant* didn't have the maximum impact. Our findings on the effectiveness

TABLE IV

EFFECTIVENESS OF THE POLICIES (α) ON INDIVIDUAL RACES ESTIMATED BY TRAINING RANDOM FOREST (RF) AND GRADIENT BOOSTING (GB) WITH STRINGENCY INDEX (DEPENDENT FEATURE) AND INCREASE/DECREASE IN CONFIRMED CASES IN EACH COMMUNITY ON A PARTICULAR DAY (TARGET LABEL). EACH SCORE REPRESENTS THE EFFECTIVENESS OF A POLICY ON A PARTICULAR RACE. HERE, SCORES IN BOLD REPRESENT THAT THE POLICY WAS MORE EFFECTIVE IN THAT COMMUNITY. FOR INSTANCE, CLOSING RESTAURANTS WAS MOST EFFECTIVE FOR THE MAJORITY OF THE COMMUNITIES.

Method	Race	Stay Home	Closes Restaurant	Closes Business	Face Mask	Travel Ban
RF	Asian	0.225	0.265	**0.282**	0.150	0.078
	White	0.210	**0.274**	0.264	0.158	0.094
	Black	0.227	**0.258**	0.255	0.166	0.094
	AIAN	0.191	**0.275**	0.235	0.209	0.090
	NHPI	0.217	0.239	**0.267**	0.183	0.095
GB	Asian	0.200	**0.324**	0.257	0.119	0.104
	White	0.191	**0.350**	0.213	0.162	0.085
	Black	0.239	**0.253**	0.244	0.168	0.097
	AIAN	0.128	**0.385**	0.162	0.256	0.069
	NHPI	0.248	0.232	0.**297**	0.160	0.064

of NPI policies on individual races and comparison against country-level analysis are presented below.

Asian: Table IV indicates that policy *Closing Business* (RF) and policy *Closing Restaurants* (GB) have the highest effectiveness indexes in terms of containing the spread with 0.282 and 0.324 scores, respectively, for the Asian race. These numbers are slightly different from the corresponding country-level scores, i.e., 0.258 and 0.285, respectively. This finding is consistent with the report published by August Census Survey [40] which states that Asian people were *"afraid to go or didn't want to go out and buy food"* implying that the Asian community was themselves resisting to visit restaurants. Additionally, according to a study by the National Bureau of Economic Research [41], around 22% decline was observed in small Asian vendor businesses nationwide after the policy was implemented. Therefore, the reports advocate that both policies *Closing Business* and *Closing Restaurants* worked together in reducing the spread of COVID-19 infection with better effectiveness than other policies. On the other hand, we observed a significant drop in the effectiveness index of *Face Mask* policy from the country level to the community-level. For the country-level (see table III) the effectiveness of *Face Mask* policy for GB is 0.198, whereas, on the community level, the score is 0.119, i.e., it dropped by 40%, which is significant. Therefore, we can say that *Closing Business/Restaurant* has impacted the Asian community more than the overall population, whereas, *Face Mask* policy has contributed lesser in comparison to the whole population. This is further substantiated by the bold numbers in the "Face Mask" column for the Asian race (Table V), where the squared differences (eqn. 2) for both methods (0.024 for RB and 0.162 for GB) show high values.

White: According to Table IV, both methods strongly suggest that the *Closing Restaurants* policy has been very effective for the white community to fight the infection. However, GB suggests a significant increase (23%) in the effectiveness index of *Closing Restaurants* policy for the White population, which is interesting. This is further substantiated by the bold number in the "Closing Restaurant" column for the White race from Table V, where the squared differences (equation 2) for GB method shows a high value, i.e., 0.051.

Black: From table IV, we observe that for the Black community, both methods suggest that *Closing Restaurants* has contributed the most in reducing the COVID-19 spread, which is consistent with the community-level results. However, findings for the GB method in Table V are more interesting, which suggests that "Stay-at-home" had a significantly different impact (squared difference of 0.025) on the Black Community compared to the same for the whole population. This is consistent with the report from the Economics Policy Institute [42] which suggests that, less than one in five people from the black community work on tele-platform (work-from-home meetings) which showcases a lack of technical jobs in the community. Moreover, during pandemics, many people from the community lost their job due to the shunt of in-person jobs. As a result, *Closing Restaurants* policy avoided customer-oriented jobs which result in helping the community to stay away from infection.

NHPI (Native Hawaiian and Other Pacific Islander): From table IV, we can observe that both methods suggest that *Closing Business* has worked for the NHPI community in reducing the infection. However, if we compare the scores of GB methods on both levels (table III and table IV), we will notice a significant difference in some policies' effectiveness. For instance, the difference in score of *Closing Restaurants* policy is 0.083, which is significant. On the other hand, scores from RF methods differ slightly but follow the same policy effectiveness trend. Therefore, the estimations by the two methods do not quite agree with each other.

TABLE V

THE TABLE REPRESENTS HOW THE EFFECTIVENESS OF THE POLICY HAS CHANGED FROM THE OVERALL COUNTRY LEVEL TO THE COMMUNITY LEVEL. THE BOLD VALUES DEPICT THAT THE POLICY HAS WORKED VERY DIFFERENTLY FOR THAT COMMUNITY AS COMPARED TO THE POLICY WORKED ON THE COUNTRY LEVEL. THESE VALUES ARE ESTIMATED USING SQUARED_DIFFERENCE METRIC (REFER EQ. 2). FOR INSTANCE, THE FACE MASK POLICY HAS A VERY DIFFERENT IMPACT OF 0.162 (GB) AND 0.024 (RF) ON THE ASIAN COMMUNITY AS COMPARED TO THE WHOLE COUNTRY. THIS DEMONSTRATES THAT THE FACE MASK POLICY HAS THE LEAST IMPACT ON THE ASIAN COMMUNITY.

Method	Race	Stay Home	Close Rest.	Close Buss.	Face Mask	Travel Ban	$\sum Race\_SD$
RF	Asian	0.000	0.001	0.009	**0.024**	0.011	0.045
	White	0.003	0.004	0.001	**0.012**	0.007	0.027
	Black	0.001	0.000	0.000	0.004	**0.006**	0.011
	AIAN	0.019	0.005	0.008	**0.033**	0.001	**0.066**
	NHPI	0.000	0.005	0.001	0.001	**0.007**	0.014
	$\sum SD$	0.023	0.015	0.019	**0.074**	0.032	-
GB	Asian	0.001	0.019	0.028	**0.162**	0.013	0.223
	White	0.006	**0.051**	0.000	0.033	0.008	0.098
	Black	**0.025**	0.013	0.016	0.024	0.001	0.079
	AIAN	**0.144**	0.122	0.064	0.084	0.069	**0.483**
	NHPI	0.040	0.035	**0.138**	0.039	0.102	0.354
	$\sum SD$	0.216	0.240	0.246	**0.342**	0.193	-

AIAN (American Indian/Alaska Native): From table IV, we observe that both methods suggest that *Closing Restaurants* (RF) policy has proven to be effective for the AIAN community. However, in terms of the squared difference in table V, *Stay-at-home* policy had a very different impact on the AIAN population compared to the whole population.

IV. CONCLUSION

In this work, we proposed a novel data-driven approach to estimate the effectiveness of different COVID-19-related NPI policies for fighting the spread (confirmed cases) and severity (death counts) of the virus. To achieve this, we proposed a novel stringency function to estimate the strictness of the NPI policies on a particular day (timestamp). Next, we trained two discriminative classifiers on confirmed cases and death counts as the binary target variable separately and stringency scores of NPI policies as features. The learned weights were then considered as the effectiveness of the NPI policies. We also performed a comparative analysis between race-specific effectiveness and country-level effectiveness to see whether different communities were impacted differently. Our findings suggest that NPI policies indeed affect different races differently; for example, *Closing Restaurants* and *Closing Businesses* policies were effective in refraining from the spread of the virus. Whereas, *Stay-at-home* was the most useful policy for decreasing the number of severe cases/deaths. Overall, we believe, in the future with similar scenarios the authoritarian could control the spread of infection and death counts by referring to this work.

Going forward, the work can be extended to incorporate new target variables like the change in hospitalization rate, vaccination rate, and many more features. One can also utilize the proposed method to estimate the effectiveness of other countries' NPIs policies considering their local issues.

REFERENCES

[1] N. Banholzer, E. van Weenen, B. Kratzwald, A. Seeliger, D. Tschernutter, P. Bottrighi, A. Cenedese, J. P. Salles, W. Vach, and S. Feuerriegel, "Impact of non-pharmaceutical interventions on documented cases of covid-19," *medRxiv*, 2020. [Online]. Available: https://www.medrxiv.org/content/early/2020/04/28/2020.04.16.20062141

[2] J. M. Culbertson, "Friedman on the lag in effect of monetary policy," *Journal of Political Economy*, vol. 68, no. 6, pp. 617–621, 1960.

[3] T. Jovanovski and M. Muric, "The phenomenon of lag in application of the measures of monetary policy," *Economic research-Ekonomska istraživanja*, vol. 24, no. 2, pp. 154–163, 2011.

[4] CDC, "Centers for disease control and prevention: Stay- at-home order by centers for disease control and prevention,." [Online]. Available: https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Stay-At-Home-Orders-Marc/y2iy-8irm

[5] CDC-FaceMask, "Centers for disease control and prevention: Face- mask order by centers for disease control and prevention,." [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/about-face-coverings.html#:~:text=People%20may%20choose%20to%20mask,about%20how%20to%20protect%20yourself.

[6] CDC-ClosingBusiness, "Centers for disease control and prevention: Closing and reopening of restuarants by centers for disease control and prevention." [Online]. Available: https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Orders-Closing-and-Reope/azmd-939x

[7] CDC, "Centers for disease control and prevention: Closing and reopening of business by centers for disease control and prevention." [Online]. Available: https://public4.pagefreezer.com/browse/CDC%20Covid%20Pages/11-05-2022T12:30/https://www.cdc.gov/coronavirus/2019-ncov/community/guidance-business-response.html

[8] CDC-TravelBan, "Centers for disease control and prevention: Travel ban by centers for disease control and prevention." [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/travelers/travel-during-covid19.html

[9] T. Hale, N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. Phillips, S. Webster, E. Cameron-Blake, L. Hallas, S. Majumdar *et al.*, "A global panel database of pandemic policies (oxford covid-19 government response tracker)," *Nature Human Behaviour*, vol. 5, no. 4, pp. 529–538, 2021.

[10] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, no. 2, pp. 197–227, 2016.

[11] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[12] E. Gibney, "Whose coronavirus strategy worked best? scientists hunt most effective policies." [Online]. Available: https://www.nature.com/articles/d41586-020-01248-1

[13] J. M. Brauner, S. Mindermann, M. Sharma, D. Johnston, J. Salvatier, T. Gavenčiak, A. B. Stephenson, G. Leech, G. Altman, V. Mikulik *et al.*, "The effectiveness of eight nonpharmaceutical interventions against covid-19 in 41 countries," *MedRxiv*, 2020.

[14] B. J. Cowling, S. T. Ali, T. W. Ng, T. K. Tsang, J. C. Li, M. W. Fong, Q. Liao, M. Y. Kwan, S. L. Lee, S. S. Chiu *et al.*, "Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in hong kong: an observational study," *The Lancet Public Health*, vol. 5, no. 5, pp. e279–e288, 2020.

[15] Z. Bian, F. Zuo, J. Gao, Y. Chen, S. S. C. P. Venkata, S. D. Bernardes, K. Ozbay, X. J. Ban, and J. Wang, "Time lag effects of covid-19 policies on transportation systems: A comparative study of new york city and seattle," *Transportation Research Part A: Policy and Practice*, vol. 145, pp. 269–283, 2021.

[16] S. C. Dass, W. M. Kwok, G. J. Gibson, B. S. Gill, B. M. Sundram, and S. Singh, "A data driven change-point epidemic model for assessing the impact of large gathering and subsequent movement control order on covid-19 spread in malaysia," *PloS one*, vol. 16, no. 5, p. e0252136, 2021.

[17] S. He, Y. Peng, and K. Sun, "Seir modeling of the covid-19 and its dynamics," *Nonlinear dynamics*, vol. 101, no. 3, pp. 1667–1680, 2020.

[18] T. Dergiades, C. Milas, T. Panagiotidis, and E. Mossialos, "Effectiveness of government policies in response to the covid-19 outbreak," *SSRN Electronic Journal*, pp. 1–25, 2020.

[19] J. Johndrow, P. Ball, M. Gargiulo, and K. Lum, "Estimating the number of sars-cov-2 infections and the impact of mitigation policies in the united states," 2020.

[20] S. S. Coughlin, A. Yiğiter, H. Xu, A. E. Berman, and J. Chen, "Early detection of change patterns in covid-19 incidence and the implementation of public health policies: A multi-national study," *Public Health in Practice*, vol. 2, p. 100064, 2021.

[21] F. Spooner, J. F. Abrams, K. Morrissey, G. Shaddick, M. Batty, R. Milton, A. Dennett, N. Lomax, N. Malleson, N. Nelissen *et al.*, "A dynamic microsimulation model for epidemics," *Social Science & Medicine*, vol. 291, p. 114461, 2021.

[22] R. Mbuvha and T. Marwala, "Bayesian inference of covid-19 spreading rates in south africa," *PloS one*, vol. 15, no. 8, p. e0237126, 2020.

[23] D. M. Vickers, S. Baral, S. Mishra, J. C. Kwong, M. Sundaram, A. Katz, A. Calzavara, M. Maheu-Giroux, D. L. Buckeridge, and T. Williamson, "Stringency of containment and closures on the growth of sars-cov-2 in canada prior to accelerated vaccine roll-out," *International Journal of Infectious Diseases*, vol. 118, pp. 73–82, 2022.

[24] R. Mezencev and C. Klement, "Stringency of the containment measures in response to covid-19 inversely correlates with the overall disease occurrence over the epidemic wave," *medRxiv*, 2021.

[25] Y. Ma, S. R. Mishra, X.-K. Han, and D.-S. Zhu, "The relationship between time to a high covid-19 response level and timing of peak daily incidence: An analysis of governments' stringency index from 148 countries," *Infectious diseases of poverty*, vol. 10, no. 1, pp. 1–10, 2021.

[26] I. V. Necesito, J. M. S. Velasco, J. Jung, Y. H. Bae, J. H. Lee, S. J. Kim, and H. S. Kim, "Understanding chaos in covid-19 and its relationship to stringency index: Applications to large-scale and granular level prediction models," *PloS one*, vol. 17, no. 6, p. e0268023, 2022.

[27] M. E. Warner and X. Zhang, "Social safety nets and covid-19 stay home orders across us states: a comparative policy analysis," *Journal of Comparative Policy Analysis: Research and Practice*, vol. 23, no. 2, pp. 176–190, 2021.

[28] R. L. Haffajee and M. M. Mello, "Thinking globally, acting locally—the us response to covid-19," *New England journal of medicine*, vol. 382, no. 22, p. e75, 2020.

[29] J. G. Hodge, "Federal vs. state powers in rush to reopen amid the coronavirus pandemic," *Hodge JG. Federal vs. state powers in rush to reopen amid coronavirus pandemic. Just Security*, 2020.

[30] CDC, "Clinical questions about covid-19: Q/a." [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/hcp/faq.html#:~:text=Based%20on%20existing%20literature%2C,2%E2%80%9314%20days.

[31] T. C. Fund, "Beyond the case count: The wide-ranging disparities of covid-19 in the united states." [Online]. Available: https://www.commonwealthfund.org/publications/2020/sep/beyond-case-count-disparities-covid-19-united-states

[32] C. F. analysis, "Beyond the case count: The wide-ranging disparities of covid-19 in the united states." [Online]. Available: https://www.commonwealthfund.org/publications/2020/sep/beyond-case-count-disparities-covid-19-united-states

[33] CDC-Cases, "Covid-19 weekly cases and deaths per 100,000 population." [Online]. Available: https://covid.cdc.gov/covid-data-tracker/#trends_dailycases

[34] CDC, "Covid-19 weekly cases per 100,000 population by race/ethnicity." [Online]. Available: https://covid.cdc.gov/covid-data-tracker/#demographicsovertime

[35] CUSP, "Covid-19 us state policy database (cusp)." [Online]. Available: https://docs.google.com/spreadsheets/d/1zu9qEWI8PsOI_i8nI_S29HDGHlIp2lfVMsGxpQ5tvAQ/edit

[36] E. Dans, "There's no denying the evidence: Restaurants and bars are helping spread covid-19." [Online]. Available: https://www.forbes.com/sites/enriquedans/2020/11/15/theres-no-denying-the-evidence-restaurants-and-bars-are-helping-/spreadcovid-19/?sh=2d1f08d93353

[37] A. Lucas, "This chart shows the link between restaurant spending and new cases of coronavirus." [Online]. Available: https://www.cnbc.com/2020/06/26/this-chart-shows-the-link-between-restaurant-spending-and-new-/coronavirus-cases.html

[38] B. University of Alabama, "Study suggests stay-at-home orders reduced covid-19 infections and deaths - news — uab." [Online]. Available: https://www.uab.edu/news/research/item/11642-study-suggests-stay-at-home-orders-reduced-covid-19-/infections-and-deaths

[39] "Association of state-issued mask mandates and allowing on-premises restaurant dining with county-level covid-19 case and death growth rates — united states, march 1–december 31, 2020." [Online]. Available: https://www.cdc.gov/mmwr/volumes/70/wr/mm7010e3.htm

[40] U. S. of Census Bureau, "Asian population more likely to report fear of going out for food as reason they did not have enough to eat during covid." [Online]. Available: https://www.census.gov/library/stories/2021/08/asian-households-cite-fear-of-going-out-as-reason-for-food-/insufficiency-during-pandemic.html

[41] N. B. of Economic Research, "Racism targets asian food, business during covid-19 pandemic." [Online]. Available: https://www.pbs.org/newshour/nation/racism-targets-asian-food-business-during-covid-19-pandemic

[42] R. B. Freeman, "Planning for the "expected unexpected": Work and retirement in the u.s. after the covid-19 pandemic shock." [Online]. Available: https://www.nber.org/papers/w29653