



OPEN

# Harnessing Artificial Intelligence to assess the impact of nonpharmaceutical interventions on the second wave of the Coronavirus Disease 2019 pandemic across the world

Sile Tao<sup>1,5</sup>, Nicola Luigi Bragazzi<sup>2,5</sup>, Jianhong Wu<sup>2</sup>, Bruce Mellado<sup>3,4</sup> & Jude Dzevela Kong<sup>2✉</sup>

In the present paper, we aimed to determine the influence of various non-pharmaceutical interventions (NPIs) enforced during the first wave of COVID-19 across countries on the spreading rate of COVID-19 during the second wave. For this purpose, we took into account national-level climatic, environmental, clinical, health, economic, pollution, social, and demographic factors. We estimated the growth of the first and second wave across countries by fitting a logistic model to daily-reported case numbers, up to the first and second epidemic peaks. We estimated the basic and effective (second wave) reproduction numbers across countries. Next, we used a random forest algorithm to study the association between the growth rate of the second wave and NPIs as well as pre-existing country-specific characteristics. Lastly, we compared the growth rate of the first and second waves of COVID-19. The top three factors associated with the growth of the second wave were body mass index, the number of days that the government sets restrictions on requiring facial coverings outside the home at all times, and restrictions on gatherings of 10 people or less. Artificial intelligence techniques can help scholars as well as decision and policy-makers estimate the effectiveness of public health policies, and implement “smart” interventions, which are as efficacious as stringent ones.

Since late December 2019, an emerging viral pathogen belonging to the *Coronaviridae* family, termed as “Severe Acute Respiratory Syndrome-related Coronavirus type 2” (SARS-CoV-2), has been isolated as the infectious agent responsible for an outbreak of pneumonia cases of unknown etiology. The initial outbreak has spread from the epicenter of the metropolitan city of Wuhan, province of Hubei, mainland China, to neighboring countries, gradually becoming a global pandemic. SARS-CoV-2 causes a generally asymptomatic or mild, but sometimes severe and even life-threatening respiratory infection, named as “Coronavirus Disease 2019” (COVID-19)<sup>1</sup>. On March 11, 2020, the World Health Organization (WHO) announced that the COVID-19 disease had developed from a “Public Health Emergency of International Concern” (PHEIC) into a pandemic<sup>2</sup>, which is still ongoing and is representing a major public health challenge, due to the highly contagious, quickly spreading nature of the virus<sup>3</sup>. The current scenario is further complicated by the circulation of mutant strains of SARS-CoV-2, known as variants of concern (VoCs)<sup>4</sup>, against which currently licensed and available COVID-19 vaccines appear to be less effective<sup>5</sup>. Moreover, vaccination against COVID-19, despite being safe and efficacious, appears to confer protection that tends to decay after a period of 6 months<sup>6</sup>.

The infectious agent has been overwhelming healthcare settings and facilities worldwide: these are facing a shortage of personnel and medical equipment, which has further increased the strain they are bearing. Vaccines

<sup>1</sup>Quartic.ai, Toronto, ON, Canada. <sup>2</sup>Africa-Canada Artificial Intelligence and Data Innovation Consortium, Department of Mathematics and Statistics, York University, Toronto, ON M3J 1P3, Canada. <sup>3</sup>School of Physics, Institute for Collider Particle Physics, University of the Witwatersrand, Johannesburg, South Africa. <sup>4</sup>iThemba LABS, National Research Foundation, Somerset West, South Africa. <sup>5</sup>These authors contributed equally: Sile Tao and Nicola Luigi Bragazzi. ✉email: jdkong@yorku.ca

have been licensed and approved only recently, which due to the lack of effective drugs, has resulted in the implementation of non-pharmaceutical interventions (NPIs).

According to the definition of the US “Centers for Disease Control and Prevention” (CDC), NPIs can be conceived as “actions, apart from getting vaccinated and taking medicine, that people and communities can take to help slow the spread of illnesses like pandemic flu”. NPIs include actions implemented at the individual level (like enhanced hygiene practices, wearing of face masks<sup>7</sup>, practicing of social/physical distancing<sup>8</sup>, self-isolation, shelter-in-place/stay-at-home requirements<sup>9</sup>, and self-quarantine). They also include interventions implemented at the community level (such as partial/total lockdown<sup>10</sup>, bans on mass gathering events<sup>11</sup>, school<sup>12</sup> and workplace closure<sup>13</sup>, internal movements, and international traveling restrictions<sup>14</sup>, among others).

Some groups are particularly vulnerable and prone to SARS-CoV-2, including the frail elderly and those with underlying co-morbidities, who are at higher risk for contracting the virus and developing complications. This has suggested the implementation of ad hoc smart<sup>15</sup> or local<sup>16</sup> lockdown/quarantine<sup>17</sup>, known also as targeted<sup>18</sup>/precision shielding<sup>19</sup> measures.

Based on their stringency, NPIs can be classified into eradication *versus* mitigation strategies<sup>20</sup>.

Stringent and drastic measures like nation-wide/global lockdowns have been effective to contain the COVID-19 spreading but are economically and socially unsustainable and highly disruptive<sup>21</sup>.

For this reason, countries and public health authorities have been striving to find the best trade-off possible between COVID-19 induced strictures and relaxing/lifting of NPIs where and when data and epidemiological trends allow to do so. This<sup>22</sup>, together with other factors, such as seasonality, or meteorological/climatic parameters, has resulted in a series of relapses/waves<sup>23</sup>. Due to the emerging nature of the pathogen, with the population being immunologically naïve to SARS-CoV-2, and given the enforcement of NPIs, a significantly large proportion of the population has been kept susceptible during the first wave of COVID-19. Until the achievement of herd immunity, due to the cyclical relaxing and reinstatement of NPIs, several waves of COVID-19 have occurred and further ones are expected to occur until the disease extinction or its transition to endemicity<sup>22</sup>.

**Aim of the study.** Given the variety of NPIs that can be implemented and the different possibilities of integrating/incorporating them into packages of public health measures, it is of crucial importance to track and monitor their effectiveness using real-world data generated by public health policies at the global level<sup>24</sup>. Artificial intelligence (AI) and big data can help in this<sup>25</sup>, assisting public health decision- and policy-makers in the complex decision-making process concerning the optimal implementation, enforcement, and timing of lifting and reinstatement of the most effective NPIs.

In the present paper, we will explore the impact of NPIs on the second wave of COVID-19 utilizing AI techniques, taking into account pre-existing country-specific characteristics (for example, economic-financial, socio-demographic, and environmental parameters).

## Materials and methods

All codes are available on a Github repository <https://github.com/sit836/covid>. For the main objective of this paper, the dependent variable is the effective reproduction number and the covariates are the NPIs and climatic, environmental, clinical, health, economic, pollution, social, and demographic (CECHEPSD) variables that can explain the epidemiological trends of the COVID-19 pandemic (Tables S1–S5).

**Spreading rates of COVID-19 during the first and second waves across the globe.** To get the values of the spreading rates of COVID-19 during the first ( $r_1$ ) and second ( $r_2$ ) waves for each country, we use Python’s SciPy curve fit function to fit the rate of change in cumulative cases of a logistic growth model to daily confirmed cases<sup>26</sup>. A statistical model was used because a mechanistic model would require a complex parameterization procedure. This would be characterized by a high degree of uncertainty, especially during the early phases of an outbreak, due to the lack of detailed data. Statistical models are data-driven, and thus do not suffer from such shortcomings. Among the statistical models, we were encouraged by the work of Ref.<sup>27</sup> to choose a logistic model. Ma et al.<sup>27</sup> compared four commonly used statistical models (namely, exponential, Richards, logistic, and delayed logistic models), and found out that the logistic model outperforms the others in estimating the growth of epidemics. Moreover, the logistic models have been extensively utilized to provide reliable estimations of the upper and lower bounds of COVID 19 related scenarios<sup>28,29</sup>. In the logistic model, the cumulative number of cases  $c(t)$  satisfies:

$$c(t) = \frac{K}{\left(1 + \left[\frac{K}{c(0)} - 1\right] e^{-rt}\right)}, \quad (1)$$

where  $K$  is the total number of people infected at the end of the outbreak,  $r$  the speed of the epidemic growth, and  $c(0)$  the initial number of cases. The change in cumulative cases that is fitted to the 7-day rolling mean of daily confirmed cases is given as  $I(t) = c(t+1) - c(t)$ , where  $t$  is a small increment in time (taken to be a day). We fit the change in the cumulative cases rather than the cumulative cases, because observations drawn from the same cumulative curve are correlated. Most curve fitting algorithms assume that the errors in individual observations are statistically independent; this is not true with cumulative data where each observation contains all of the cases from previous observations. For this reason, to avoid such assumptions and shortcomings, we utilized the least square fitting algorithm. We truncated all COVID-19 reported daily case time series within the window of the first and second waves, to the day with the highest daily count, because some countries have lingered near peak daily count for much longer than a logistic growth model would predict, which would pull the model peak to later than the actual date of peak incidence and thereby underestimate the spreading rate. We

manually checked each time series and ensured that the highest daily count only occurred during a peak. We only consider countries that experienced at least two waves. Also, we include only countries that were at least 6 days into a period (for both the first and second wave periods) with at least 30 daily cases as of July 29, 2020, after truncating at the peak. The time when countries observed their first 30 daily case count was considered the initial time. The first wave is based on a fitting window from the initial time until peak time and that for the second wave is based on a fitting window from the time at which a country records the lowest number of daily cases between the first and second peak to the second peak.

We eliminated countries whose logistic growth model has  $R^2$  less than 0.95 for any of the fits (first and second waves). This is to ensure that we only include countries that our model can explain at least 95% of the variations in their spreading rates.

Some countries do not report COVID-19 cases on a daily basis; some countries have variable reporting delays, and some may have changed reporting methods resulting in dramatic spikes in cases for particular dates. To circumvent this inaccuracy in date, we used the 7-day rolling average (right-aligned) for daily cases.

$R_e$  can serve as a baseline expectation for estimating how fast COVID-19 would spread if all interventions were prematurely lifted prior to the start of the second wave, given that the percentage of the population susceptible to COVID-19 was still relatively high at that time.

Using these growth rates, we calculated the basic reproduction number of COVID 19 ( $R_0$ ) (first wave) and the effective reproduction number of the second wave ( $R_e$ ) as follows<sup>30</sup>:

$$R_0 = e^{r_1 T}, R_e = e^{r_2 T},$$

where  $T$  is the serial interval of COVID-19 (time delay between the symptom onset of a primary case and the secondary cases). The value  $T$  for COVID-19 lies in the interval<sup>4,8,31–33</sup>.

**Covariates.** Next, we compiled data on the covariates.

*Non-pharmaceutical interventions (NPIs).* We compiled data on 18 common policy responses that governments across the globe have taken to respond to the pandemic. These include school and workplace closures, cancellation of public events and gatherings, stay-at-home orders, and international and domestic travel restrictions: these have been extracted from <https://github.com/OxCGRT/covid-policytracker/blob/master/documentation/codebook.md>. Each NPI is an indicator recorded on an ordinal scale where the larger the index, the stricter the policy (Tables S2, S3). The dataset records governmental responses implemented in the year 2020 for several countries.

*CECHEPSD variables.* Data on several climatic, environmental, clinical, health, economic, pollution, social, and demographic variables were obtained from publicly available databases (see Table S4 for the full list of variables and references).

**Pre-processing data.** We kept only countries having both first and second waves. Then we filtered out covariates if the missing ratio is greater than 10% and replaced the missing values (2% of data) with the mode. Next, we removed variables, such as country name and cumulative cases per million population, whose value either does not add any information to the model or would not actually be available at the time we want to make a prediction. In the end, we converted categorical variables into integers.

We represented temporal policy responses as time-independent numerical values. First, for every country and policy, we ignored cells with no measures and counted the number of days lasting for each possible action. For example, regarding the policy of canceling public events, assume Canada recommended canceling public events for 30 days and required canceling for 60 days. Then we used the tuple (30, 60) to represent the information of such a policy. Next, we imputed the missing values (0.03% of data) with the mode.

In the end, we have 55 countries and 35 covariates when we studied the association between  $R_0$  and CECHPSD variables; 53 countries and 73 covariates when we regressed  $\hat{R}_e - R_e$  on NPIs; 53 countries and 108 covariates when we regressed the growth rate of the second wave on NPIs and CECHPSD variables.

*Evaluation metrics.* We adopted the mean squared error (MSE) and the coefficient of determination  $R^2$  as the evaluation measures. MSE measures the average of the squares of the errors.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

where  $n$  is the sample size,  $y_i$  is the observed value and  $\hat{y}_i$  is the predicted one;  $R^2$  measures how well a model performs compared to naïve average forecasting

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2},$$

where  $\bar{y}_i$  is the average of the observed values. It is worth pointing out that  $R^2$  in our definition can be negative if the model predictions are far away from the actual values.

**Random forest regression analysis of the association between CECHEPSD, NPIs and  $R_e$ .** We used random forests (RF), an “off-the-shelf” machine learning algorithm, to predict  $R_e$  based on 18 CECHEPSD, and 8 NPIs variables. Suppose the inputs pairs are  $(x_1, y_1), \dots, (x_n, y_n)$ , where  $x_i \in \mathbb{R}^p$  and  $y_i \in \mathbb{R}$ . Every decision tree in a forest forms a step function over a partition  $R_1, R_2, \dots, R_M$ :

$$f(x) = \sum_{m=1}^M c_m I_{R_m}(x),$$

where  $c_m$  are model parameters and  $I_{R_m}$  is an indicator function:

$$I_{R_m}(x) = \begin{cases} 1, & x \in R_m \\ 0, & \text{otherwise} \end{cases}$$

RF builds a large collection of de-correlated trees and then averages them. Differently from generalized additive models (GAMs), RF is a decision tree-based method, which can capture the interactions among covariates. Therefore, in practice, we expect RF to almost always outperform GAM. In the implementation, we used the “RandomForestRegressor” module in the Python Scikit-learn library.

For regression, we grew and combined 500 decision trees. Each tree is grown with the randomly selected square root of the total number of covariates when making splits<sup>34</sup> and it has the maximum depth found via tenfold cross-validation. A large number of trees was used to stabilize feature importance measures. If the number of trees is not large enough, then some covariates will not be given a chance to play a role in each tree. For low complexity trees, increasing the number of trees will not cause over-fitting<sup>35</sup>.

To understand how covariates are contributing to the model fitting, we used Breiman’s<sup>35</sup> permutation-based measures, which assess the importance of a feature by calculating the increase in the model’s error after permuting the feature. A feature is important if shuffling its values increases the model error, because, in this case, the model relied on the feature. A feature is unimportant if shuffling its values does not change the model’s error much.

**Comparing the growth rate of the first wave and second waves.** Here we compare the spreading rates of the first wave and second waves of COVID-19 across countries. To account for susceptible population, we instead compared the  $R_e$  and  $\hat{R}_e$ , which is the product of the predicted value  $R_0$  and the fraction of susceptible population at the start of the second wave.

We would like to know whether the centers of the observations  $R_e$ ’s and  $\hat{R}_e$ ’s are statistically different. If there is no statistical difference between the centers of  $\hat{R}_e$  and  $R_e$ , it is likely that partial lifting of NPIs led to the spreading rate of the second wave of COVID-19 to be statistically similar to what it was during the first wave (after accounting for those with partial immunity). One explanation for this could be that the NPIs that had the greatest impact on the spreading rate were lifted.

We investigated the shape of distributions for  $R_e$  and  $\hat{R}_e$  via Kolmogorov–Smirnov tests. Both P-values are extremely small: One for  $R_e$  is  $5.7 \times 10^{-47}$  and the other for  $\hat{R}_e$  is  $3.1 \times 10^{-54}$ . Also, the side-by-side boxplots of  $R_e$  and  $\hat{R}_e$  (Fig. S3) show that data in either group are not symmetric. Therefore, we decided to use Wilcoxon signed-rank test, a non-parametric analogy to the classical paired  $t$  test to test whether the two populations have the same distribution. If the null hypothesis is rejected, then we have evidence that the centers of the two populations differ.

**Ethics and consent.** All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

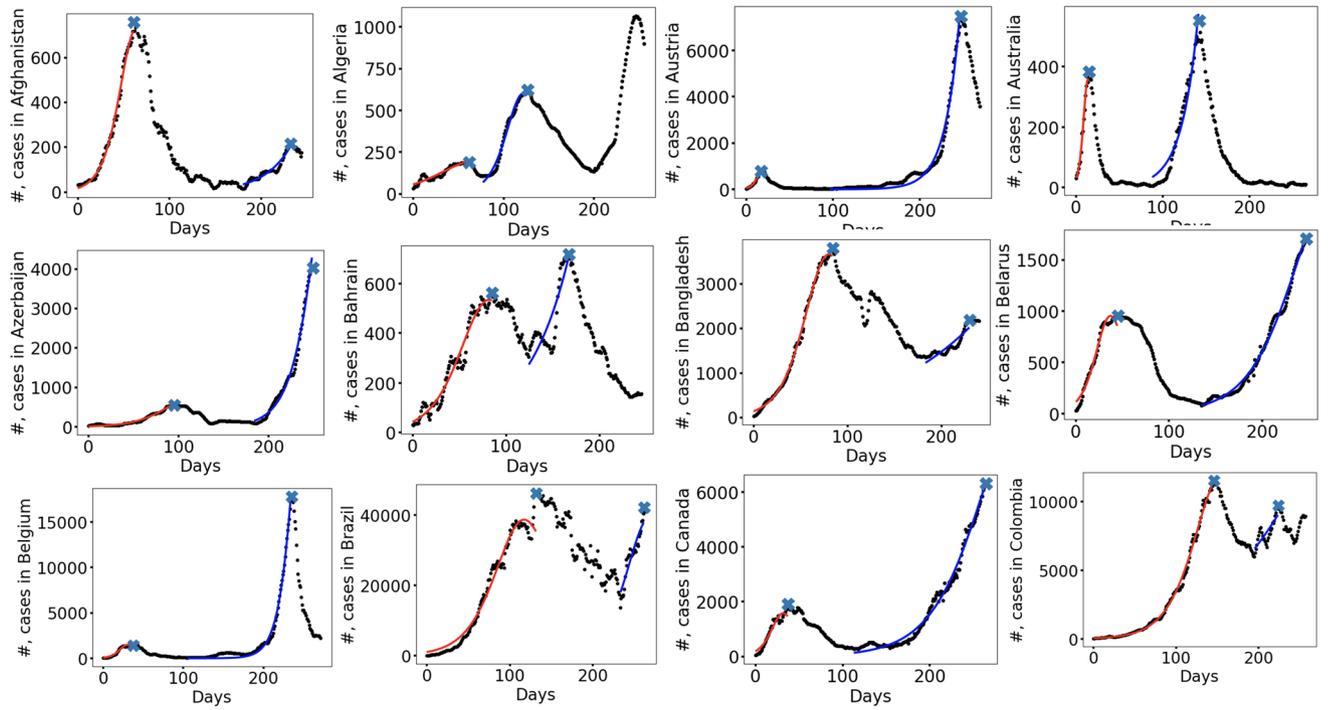
## Results

**Estimation of the spreading rates of COVID-19 during the first and second waves.** Figure 1, Figs. S1 and S2 show the growth curves fitted to the observed time-series of daily confirmed cases across countries. The plotted estimate for the first wave is based on a fitting window from the initial time until peak time and that for the second wave is based on a fitting window from the time between the first and the second peak with the lowest number of cases until the second peak time. Only countries whose logistic growth model had an  $R^2$  of or greater than 0.95 were considered.

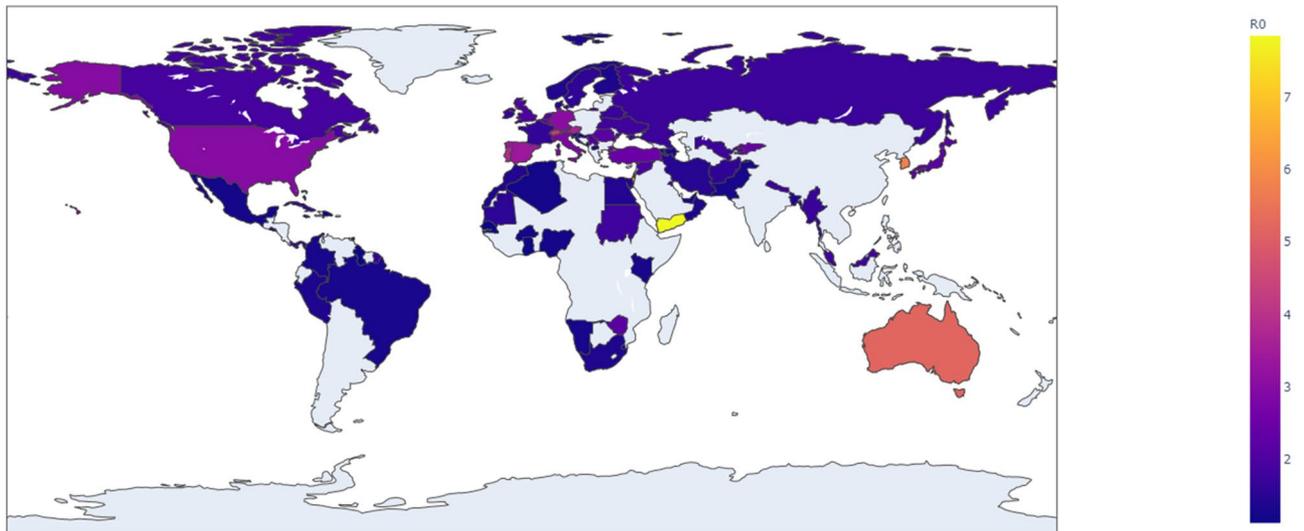
Figure 2 and the second column in Table S5 summarize the estimated basic reproduction number  $R_0$  across countries while Fig. 3 and the third column in Table S5 summarize the estimated effective reproduction number (second wave)  $R_e$  across countries. Figures 2 and 3 were created using Plotly.py 4.14.3, a Python open source library (<https://plotly.com/graphing-libraries/>).

$R_0$  and  $R_e$  are highest in Israel ( $R_0 = 6.93$ ) and Mexico ( $R_e = 3.08$ ) respectively. The lowest  $R_0$  and  $R_e$  were respectively estimated in Senegal ( $R_0 = 1.13$ ) and Bangladesh ( $R_e = 1.07$ ). Overall, the mean  $R_0$  and  $R_e$  were respectively 2.02 (S.D. 1.09) and 1.45 (S.D. 0.41). The United Kingdom ( $R_0 = 2.01$ ), Luxembourg ( $R_0 = 1.90$ ) and the Netherlands ( $R_0 = 2.17$ ) had  $R_0$  values that were closer to the mean  $R_0$  value, while the Netherlands ( $R_e = 1.45$ ), Oman ( $R_e = 1.50$ ), and Namibia ( $R_e = 1.39$ ) had  $R_e$  values closer to the mean  $R_e$ . We used the mean and standard deviation as descriptive statistics for  $R_e$  and  $R_0$  because we observed that they are normally distributed across countries.

**Association between NPIs, CECHEPSD variables and growth rate of the second wave.** Growing a RF with 500 trees and maximum depth = 2 gives MSE 0.08 and  $R^2$  0.51. We compared RF with the least absolute shrinkage and selection operator (LASSO) regression<sup>36</sup> with regularization parameter = 0.3. The value of the regularization parameter was found via tenfold cross-validation on the normalized covariates. LASSO



**Figure 1.** The time course dynamics of COVID-19. The red line corresponds to the first COVID-19 wave, the blue line to the second COVID-19 wave, the blue cross to the peak, and the dotted line to real data.

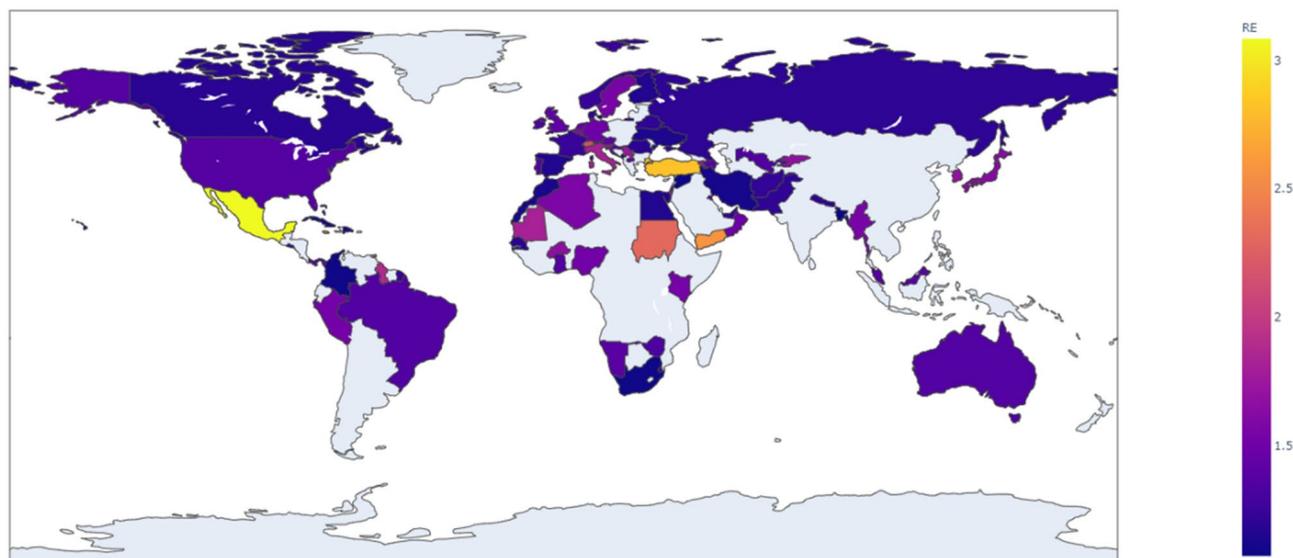


**Figure 2.** The basic reproduction number of COVID-19 across countries.

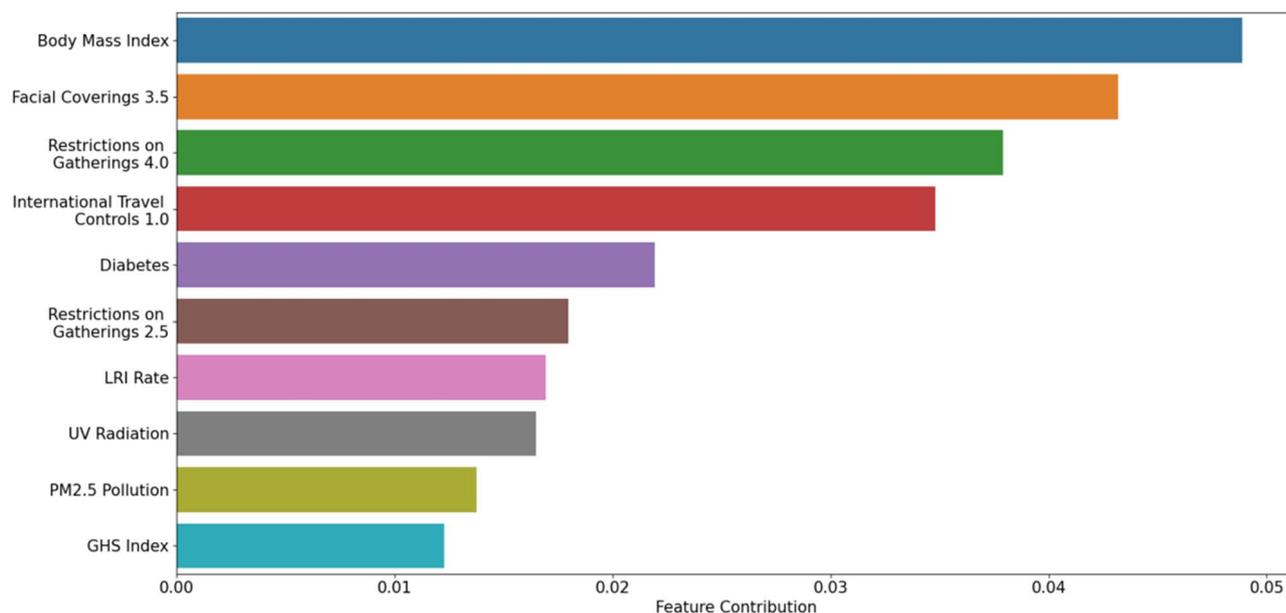
gives MSE 0.16 and  $R^2$  0.00 which are of several orders of magnitude worse than RE, since LASSO does not take nonlinearity into account.

Figure 4 indicates that average body mass index (BMI) was the first most important variable associated with the growth rate of the second wave. The second variable in terms of importance is the number of days that the government sets restrictions on requiring facial coverings outside the home at all times regardless of location or presence of other people in some areas. Restrictions on gatherings of 10 people or less, and screened foreign travelers on international travel are the third and fourth most important variables associated with the growth rate of the second wave, respectively.

**Hypothesis testing for difference of medians between of  $R_e$  and  $\hat{R}_e$ .** The value of the statistical Wilcoxon test is 169.0 and P-value is  $4.8 \times 10^{-7}$ . The null hypothesis is rejected. It suggests that the actual observations and the estimates are unlikely from the same distribution. Therefore, a statistically significant difference exists between the two medians. Thus, it is likely that the partial lifting of NPIs did not cause the spreading rate



**Figure 3.** The effective reproduction number of COVID-19 across countries during the second wave.



**Figure 4.** Features importance (Top 10): (i) body mass index, (ii) the number of days that the government sets restrictions on requiring facial coverings outside the home at all times, (iii) restrictions on gatherings 4.0, (iv) screened foreign travelers on international travel, (v) diabetes prevalence, (vi) restrictions on gatherings 2.5, (vii) LRI rate, (viii) UV radiation, (ix) PM2.5 air pollution, and (x) GHS index.

of the second wave of COVID-19 to be statistically similar to what it was during the first wave (after accounting for those with partial immunity).

## Discussion

In the present investigation, we found that (i) body mass index, (ii) the number of days that the government sets restrictions on requiring facial coverings outside the home regardless of location or presence of other people in some areas, and (iii) restrictions on gatherings of 10 people or less are the three most important variables in the model. Among health-related variables, body mass index has been found to be associated with COVID-19. Sarmadi et al.<sup>37</sup> have performed an ecological study, utilizing global databases (from the WHO and the NCD Risk Factor Collaboration, or NCD-RisC), to dissect the correlation between age-standardized body mass index and the risk of contracting COVID-19 in terms of incidence and mortality ratio. Authors were able to find a positive correlation, which was stronger in nations and territories with younger populations (like developing countries). Such a correlation remained statistically significant after adjusting for confounding factors (such as

socio-demographic and economic parameters). This finding has epidemiological relevance, in that it has practical implications in terms of public health policies. Health decision- and policy-makers could devise and implement interventions aimed at monitoring and counteracting overweight and obesity, promoting health literacy and the adoption of healthy lifestyles, mitigating, in this way, the burden of disease imposed by high body mass index. The other two of the three most significant variables are NPIs. Quantifying the efficacy of mitigation strategies against an outbreak caused by an emerging pathogen is of paramount importance<sup>38</sup>, both to avoid further waves/relapses of the same outbreak and to guide future preparedness response plans<sup>39</sup>.

Despite the importance of tracking and monitoring the effectiveness of NPIs, there are few large-scale studies conducted at the global level. Exploring this topic is technically challenging because the variables under study are highly intercorrelated, exhibiting spatial, temporal, and spatio-temporal clustering patterns<sup>40</sup>. There exist, instead, several studies estimating the impact of single individual NPIs at the country-level or in a group of countries, whereas a comprehensive assessment of all the NPIs being implemented (enforced/lifted) is necessary. James and Menzies<sup>41</sup> have investigated changes in numerous aspects of COVID-19 related behaviors between the first and second waves, for example in terms of outbreak severity across the United States, where each state has individually responded to the pandemic. Authors have developed a formal definition and mathematical framework to properly classify COVID-19 surges/peaks, differentiating between a first and second wave, and have compared the various infectious trajectories across states to identify the most effective pandemic responses. In a second paper, James et al.<sup>42</sup> have extended their analytical techniques to incorporate European countries as well, demonstrating substantial heterogeneity within Europe and the United States. In a subsequent paper, James et al.<sup>43</sup> have compared three countries most hardly hit by the outbreak, namely the United States, India, and Brazil, assessing patterns of similarity and dissimilarity in the response to the pandemic.

In a previous study<sup>44</sup>, we analyzed the effects of the implementation of NPIs on the initial growth rate of COVID-19, taking into account as well CECHEPSD variables, using a multiple linear regression model and incorporating 29 parameters. Out of these 29 variables, ten (8 CECHEPSD characteristics and 2 NPIs) were found to correlate with the initial growth of COVID-19. In particular, the population residing in urban agglomerations (centers of more than 1 million inhabitants), atmospheric fine particulate matter (PM<sub>2.5</sub>) air pollution mean annual exposure, life expectancy, number of hospital/healthcare setting beds available, urban population, Global Health Security (GHS) index, and international movement restrictions were the parameters which had the most significant impact on the initial growth of COVID-19. Based on these findings, we concluded that, among NPIs being implemented, only one (namely, restrictions on international movements) was found to have a relative significance with respect to the initial growth rate of COVID-19, whilst CECHEPSD factors seemed to play a more prominent role in the initial growth rate of COVID-19 and its transmission dynamics.

A study<sup>39</sup> attempted to quantitatively assess the effects of NPIs enforced in several countries/territories in terms of changes in the COVID-19 effective reproduction number, employing an integrative modeling approach, combining classical inference, bio-statistics, and AI techniques. The authors utilized a training dataset of 6068 hierarchically coded NPIs from 79 countries and a validated external database merging two datasets, including 42,151 additional NPIs from 226 countries. Authors were able to find that a highly disruptive, costly, intrusive NPI like a national lockdown was as effective as a package of less drastic and stringent NPIs. In particular, the largest effects in terms of reduction in the effective reproduction number were found for NPIs like the ban of small gathering events, school closure, and border control/restrictions.

Liu et al.<sup>40</sup> obtained similar findings, utilizing hierarchical clustering and panel, longitudinal regression tools to quantify the efficacy of 13 NPI-related categories in the study period January–June 2020. The authors found that two NPIs (closure of educational institutions and internal movement restrictions) were particularly efficacious in decreasing time-varying reproduction numbers. Other NPIs (namely, workplace closure, debt/contract relief, income support, cancellation of public events, and gathering events ban/restrictions) were effective as well. Evidence concerning other mitigation strategies (such as shelter-in-place/stay-at-home orders, public information awareness campaigns, public transportation closure, travel restrictions, testing, and contact tracing) was, instead, contrasting.

Li et al.<sup>45</sup> conducted a modeling study on the effect of escalating/de-escalating NPIs in terms of variation of the COVID-19 reproduction number in the period January–July 20, 2020, collecting data from 131 countries. Authors found that NPIs like school closure, workplace closure, public events cancellation/ban, shelter-in-place/stay-at-home orders, and internal movement restrictions were able to curtail the spreading of the virus, with bans on public events achieving the statistical significance threshold. Lifting of bans on public gathering events and reopening of schools resulted in a significant increase in the COVID-19 reproduction number.

Bo et al.<sup>46</sup> analyzed 1,908,197 confirmed COVID-19 cases from 190 countries in the period January–April 2020, categorizing NPIs as mandatory face-mask use in public, self-isolation/quarantine, social/physical distancing, and traffic controls/restrictions. These resulted in a decrease in the COVID-19 reproduction number, which was more marked when a coherent, integrated package of public health interventions was implemented and enforced.

Our investigation confirmed the usefulness of NPIs implemented worldwide, complementing and adding to the existing literature. The strength of the present paper is, indeed, the fact that we used a quite large list of covariates and NPIs to discern their association with the growth rate of the second wave of COVID-19. However, this list is far from being exhaustive and other covariates could have been included, given that the literature on the determinants of COVID-19 is constantly under flux and continuously evolving. For example, there is evidence of causal correlations between COVID-19 and PM<sub>10</sub><sup>47</sup> as well as between COVID-19 and relative humidity<sup>48</sup>. Furthermore, the PM<sub>2.5</sub> characteristic analyzed in this paper is the mean annual exposure, while some papers have found correlations with the exceeding of daily thresholds<sup>49</sup>. This warrants further research exploring other covariates for which recent studies have shown causal associations.

In conclusion, extremely aggressive measures like nation-wide lockdowns, have significantly contributed to the containment of the COVID-19 pandemic, by curbing the SARS-CoV-2 transmission dynamics, and saving lives, but, on the other hand, have imposed a dramatically high societal and economic burden. Advanced data mining techniques, including approaches relying on Big Data and AI, can enable scholars as well as public health decision- and policy-makers to estimate the effectiveness of public health policies and mitigation strategies to counteract the toll of the outbreak in terms of infections and deaths, enforcing and implementing “smart” interventions, which are as efficacious as drastic and stringent ones.

Received: 1 August 2021; Accepted: 23 December 2021

Published online: 18 January 2022

## References

- Hu, B., Guo, H., Zhou, P. & Shi, Z.-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **19**, 141–154 (2021).
- Bai, Y. *et al.* Advances in SARS-CoV-2: A systematic review. *Eur. Rev. Med. Pharmacol. Sci.* **24**, 9208–9215 (2020).
- Mallapaty, S. Why does the coronavirus spread so easily between people? *Nature* **579**, 183–184 (2020).
- Alem, A. & Slenker, A. K. *Emerging Variants of SARS-CoV-2 and Novel Therapeutics Against Coronavirus (COVID-19)* (StatPearls, 2021).
- Lopez Bernal, J. *et al.* Effectiveness of covid-19 vaccines against the B. 1617. 2 (Delta) variant. *N. Engl. J. Med.* **385**, 585–594 (2021).
- Thomas, S. J. *et al.* Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine through 6 months. *N. Engl. J. Med.* **385**, 1761–1773 (2021).
- Howard, J. *et al.* An evidence review of face masks against COVID-19. *Proc. Natl. Acad. Sci.* **118**, e2014564118 (2021).
- Gupta, S., Simon, K. I. & Wing, C. Mandated and voluntary social distancing during the covid-19 epidemic: A review. *Brook. Pap. Econ. Act.* **2020**, 269–326 (2020).
- Berry, C. R., Fowler, A., Glazer, T., Handel-Meyer, S. & MacMillen, A. Evaluating the effects of shelter-in-place policies during the COVID-19 pandemic. *Proc. Natl. Acad. Sci.* **118**, e2019706118 (2021).
- Perra, N. Non-pharmaceutical interventions during the COVID-19 pandemic: A review. *Phys. Rep.* **913**, 1–52 (2021).
- Ebrahim, S. H. & Memish, Z. A. COVID-19—The role of mass gatherings. *Travel Med. Infect. Dis.* **34**, 101617 (2020).
- Walsh, S. *et al.* Do school closures and school reopenings affect community transmission of COVID-19? A systematic review of observational studies. *BMJ Open* **11**, e053371 (2021).
- D'angelo, D. *et al.* Strategies to exit the COVID-19 lockdown for workplace and school: A scoping review. *Saf. Sci.* **134**, 105067 (2021).
- Grépin, K. A. *et al.* Evidence of the effectiveness of travel-related measures during the early phase of the COVID-19 pandemic: A rapid systematic review. *BMJ Glob. Health* **6**, e004537 (2021).
- Ibarra-Vega, D. Lockdown, one, two, none, or smart: Modeling containing COVID-19 infection. A conceptual model. *Sci. Total Environ.* **730**, 138917 (2020).
- Karatayev, V. A., Anand, M. & Bauch, C. T. Local lockdowns outperform global lockdowns on the far side of the COVID-19 epidemic curve. *Proc. Natl. Acad. Sci.* **117**, 24575–24580 (2020).
- Khan, S. D., Alarabi, L. & Basalamah, S. Toward smart lockdown: A novel approach for COVID-19 hotspots prediction using a deep hybrid neural network. *Computers* **9**, 99 (2020).
- Smith, G. & Spiegelhalter, D. Shielding from covid-19 should be stratified by risk. *BMJ* **369**, m2063 (2020).
- Ioannidis, J. P. Precision shielding for COVID-19: Metrics of assessment and feasibility of deployment. *BMJ Glob. Health* **6**, e004614 (2021).
- Ferguson, N. *et al.* Report 9: Impact of non-pharmaceutical interventions (npis) to reduce COVID19 mortality and healthcare demand. *Imperial Coll. Lond.* **10**, 491–497 (2020).
- Fezzi, C. & Fanghella, V. Real-time estimation of the short-run impact of COVID-19 on economic activity using electricity market data. *Environ. Resour. Econ.* **76**, 885–900 (2020).
- Engelbrecht, F. A. & Scholes, R. J. Test for Covid-19 seasonality and the risk of second waves. *One Health* **12**, 100202 (2021).
- Xu, S. & Li, Y. Beware of the second wave of COVID-19. *Lancet* **395**, 1321–1322 (2020).
- Suryanarayanan, P. *et al.* AI-assisted tracking of worldwide non pharmaceutical interventions for COVID-19. *Sci. Data* **8**, 1–14 (2021).
- Bragazzi, N. L. *et al.* How big data and artificial intelligence can help better manage the COVID-19 pandemic. *Int. J. Environ. Res. Public Health* **17**, 3176 (2020).
- Ritchie, H. *et al.* Coronavirus pandemic (COVID-19). *Our World in Data* (2020). <https://ourworldindata.org/coronavirus>. (Accessed 5 January 2021).
- Ma, J., Dushoff, J., Bolker, B. M. & Earn, D. J. Estimating initial epidemic growth rates. *Bull. Math. Biol.* **76**, 245–260 (2014).
- Bertozzi, A. L., Franco, E., Mohler, G., Short, M. B. & Sledge, D. The challenges of modeling and forecasting the spread of COVID-19. *Proc. Natl. Acad. Sci.* **117**, 16732–16738 (2020).
- Pelinovsky, E., Kurkin, A., Kurkina, O., Kokoulina, M. & Epifanova, A. Logistic equation and COVID-19. *Chaos Solitons Fractals* **140**, 110241 (2020).
- Wallinga, J. & Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. Biol. Sci.* **274**, 599–604 (2007).
- Du, Z. *et al.* Serial interval of COVID-19 among publicly reported confirmed cases. *Emerg. Infect. Dis.* **26**, 1341–1343 (2020).
- Park, M., Cook, A. R., Lim, J. T., Sun, Y. & Dickens, B. L. A systematic review of COVID-19 epidemiology based on current evidence. *J. Clin. Med.* **9**, 967 (2020).
- He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).
- Friedman, J., Hastie, T. & Tibshirani, R. *The Elements of Statistical Learning* (Springer, 2001).
- Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
- Tibshirani, R. Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc. Ser. B (Methodol.)* **58**, 267–288 (1996).
- Sarmadi, M., Ahmadi-Soleimani, S. M., Fararouei, M. & Dianatinasab, M. COVID-19, body mass index and cholesterol: An ecological study using global data. *BMC Public Health* **21**, 1712 (2021).
- Brauner, J. M. *et al.* Inferring the effectiveness of government interventions against COVID-19. *Science* **371**, 9338 (2021).
- Haug, N. *et al.* Ranking the effectiveness of worldwide COVID-19 government interventions. *Nat. Hum. Behav.* **4**, 1303–1312 (2020).
- Liu, Y., Morgenstern, C., Kelly, J., Lowe, R. & Jit, M. The impact of non pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. *BMC Med.* **19**, 40 (2021).
- James, N. & Menzies, M. COVID-19 in the United States: Trajectories and second surge behavior. *Chaos* **30**, 091102 (2020).

42. James, N., Menzies, M. & Radchenko, P. COVID-19 second wave mortality in Europe and the United States. *Chaos* **31**, 031105 (2021).
43. James, N., Menzies, M. & Bondell, H. Comparing the dynamics of COVID-19 infection and mortality in the United States, India, and Brazil. <http://arXiv.org/2108.07565> (2021).
44. Duhon, J., Bragazzi, N. & Kong, J. D. The impact of non-pharmaceutical interventions, demographic, social, and climatic factors on the initial growth rate of COVID-19: A cross-country study. *Sci. Total Environ.* **760**, 144325 (2021).
45. Li, Y. *et al.* The temporal association of introducing and lifting non pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: A modelling study across 131 countries. *Lancet Infect. Dis.* **21**, 193–202 (2021).
46. Bo, Y. *et al.* Effectiveness of non-pharmaceutical interventions on COVID-19 transmission in 190 countries from 23 January to 13 April 2020. *Int. J. Infect. Dis.* **102**, 247–253 (2021).
47. Pegoraro, V., Heiman, F., Levante, A., Urbinati, D. & Peduto, I. An Italian individual-level data study investigating the association between air pollution exposure and COVID-19 severity in primary-care settings. *BMC Public Health* **21**, 902 (2021).
48. Ganegoda, N. C., Wijaya, K. P., Amadi, M., Erandi, K. H. & Aldila, D. Interrelationship between daily COVID-19 cases and average temperature as well as relative humidity in Germany. *Sci. Rep.* **11**, 11302 (2021).
49. Rovetta, A. & Castaldo, L. Relationships between demographic, geographic, and environmental statistics and the spread of novel coronavirus disease (COVID-19) in Italy. *Cureus* **12**, e11397 (2020).

### Author contributions

J.D.K. designed research; J.D.K., S.T. and N.L.B. conducted literature search and data collection; and all authors analyzed the data and wrote the paper.

### Funding

This research is funded by Canada's International Development Research Centre (IDRC) and the Swedish International Development Cooperation Agency (SIDA) (Grant No. 109559-001).

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-04731-5>.

**Correspondence** and requests for materials should be addressed to J.D.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022