

---

Drivers and forecasts of multiple waves of the coronavirus disease 2019 pandemic: a systematic analysis based on an interpretable machine learning framework

Zicheng Cao<sup>1,2\*</sup>, Zekai Qiu<sup>1,2\*</sup>, Feng Tang<sup>1,3</sup>, Shiwen Liang<sup>1,4</sup>, Yinghan Wang<sup>1,5</sup>, Haoyu Long<sup>1,2</sup>, Cai Chen<sup>1</sup>, Bing Zhang<sup>1,2</sup>, Chi Zhang<sup>1,2</sup>, Yaqi Wang<sup>1,2</sup>, Kang Tang<sup>1,2</sup>, Jing Tang<sup>1,2</sup>, Junhong Chen<sup>1,2</sup>, Chunhui Yang<sup>1,2</sup>, Yuzhe Xu<sup>1,2</sup>, Yulin Yang<sup>1,2</sup>, Shenglan Xiao<sup>1,2</sup>, Dechao Tian<sup>1,2</sup>, Guozhi Jiang<sup>1,2</sup>, Xiangjun Du<sup>1,2,6</sup>

1. School of Public Health (Shenzhen), Sun Yat-sen University, Guangzhou 510275, P.R. China

2. School of Public Health (Shenzhen), Shenzhen Campus of Sun Yat-sen University, Shenzhen 518107, P.R. China

3. Foshan Center for Disease Control and Prevention, Foshan 528010, P.R. China

4. Fujian Provincial Center for Disease Control and Prevention, Fuzhou 350001, P.R. China

5. Clinical research center, Second Affiliated Hospital of Kunming Medical University, Kunming 650033, P.R. China

6. Key Laboratory of Tropical Disease Control, Ministry of Education, Sun Yat-sen University, Guangzhou 510030, P.R. China

\* Contributed equally

Correspondence to:

Dr. Xiangjun Du, School of Public Health (Shenzhen), Sun Yat-sen University, Guangzhou 510275, P.R.

China. Email: duxj9@mail.sysu.edu.cn

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/tbed.14492](#).

This article is protected by copyright. All rights reserved.

---

## Abstract

Coronavirus disease 2019 (COVID-19) has become a global pandemic and continues to prevail with multiple rebound waves in many countries. The driving factors for the spread of COVID-19 and their quantitative contributions, especially to rebound waves, are not well studied. Multidimensional time-series data, including policy, travel, medical, socioeconomic, environmental, mutant and vaccine related data, were collected from 39 countries up to June 30, 2021, and an interpretable machine learning framework (XGBoost model with Shapley Additive explanation interpretation) was used to systematically analyze the effect of multiple factors on the spread of COVID-19, using the daily effective reproduction number as an indicator. Based on a model of the pre-vaccine era, policy-related factors were shown to be the main drivers of the spread of COVID-19, with a contribution of 60.81%. In the post-vaccine era, the contribution of policy-related factors decreased to 28.34%, accompanied with an increase in the contribution of travel-related factors, such as domestic flights, and contributions emerged for mutant-related (16.49%) and vaccine-related (7.06%) factors. For single-peak countries, the dominant ones were policy-related factors during both the rising and fading stages, with overall contributions of 33.7% and 37.7%, respectively. For double-peak countries, factors from the rebound stage contributed 45.8% and policy-related factors showed the greatest contribution in both the rebound (32.6%) and fading (25.0%) stages. For multiple-peak countries, the Delta variant, domestic flights (current month) and the daily vaccination population are the three greatest contributors (8.12%, 7.59% and 7.26%, respectively). Forecasting models to predict the rebound risk were built based on these findings, with accuracies of 0.78 and 0.81 for the pre- and post-vaccine eras, respectively. These findings quantitatively demonstrate the systematic drivers of the spread of COVID-19, and the framework proposed in this study will facilitate the targeted prevention and control of the ongoing COVID-19 pandemic.

---

## Keywords

COVID-19; spread; rebound; machine learning; forecasting

## Introduction

An emerging infectious disease, coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2, has rapidly spread globally. On March 11, 2020, the World Health Organization declared COVID-19 as a global pandemic. Until February 11, 2022, there were 405,688,274 COVID-19 cases globally (Roser, Ritchie, Ortiz-Ospina, & Hasell, 2020; Looi, 2020). The number of new cases continues to increase every day, and COVID-19 remains a global concern after 2 years of circulation.

Since July 2020, there has been a second wave of infections in many countries (Cacciapaglia, Cot, & Sannino, 2020; Wise, 2020; Xu & Li, 2020; Zhu, Zhang, Wang, Li, & Tan, 2020), especially in Europe and Asia. Vaccine development was also ongoing in many countries during that period. In early 2021, countries around the world began to implement extensive vaccination programs with the aim of ending the COVID-19 pandemic (Awadasseid, Wu, Tanaka, & Zhang, 2021). However, new virus variants with greater transmissibility have emerged, combined with control measures, and caused multiple waves of record-breaking numbers of new infections (Sonabend et al., 2021; Mohapatra et al., 2022). The ongoing pandemic has become a long-term battle, not just for public health, as governments are required to balance economic recovery, political votes and control of the COVID-19 pandemic (Di Domenico, Pullano, Sabbatini, Boëlle, & Colizza, 2020; Ferrante et al., 2020; Leung, Wu, Liu, & Leung, 2020). As a result, decision-makers have adjusted intervention policies dynamically, while considering many types of factors, including the status of the COVID-19 pandemic (Panovska-Griffiths et al., 2020; Qiu, Chen, & Shi, 2020; Sebhatu, Wennberg, Arora-Jonsson, & Lindberg,

---

2020), the characteristics of new virus variants, vaccination status, socioeconomic conditions (Kandel, Chungong, Omaar, & Xing, 2020) and medical resources (Brauner et al., 2020), to contain the pandemic.

The combined effects of various control measures are complicated, and the effect could be context-dependent for a specific country. Those understandings are important for precise control and more quantitative studies are needed. Various non-pharmaceutical interventions (Flaxman et al., 2020; Haug et al., 2020; Hsiang et al., 2020; Russell et al., 2021; Sun, Shi, & Xu, 2020) implemented by many countries have been shown to be useful to contain and control the spread of the pandemic. Without restrictions, population migration, which may increase the risk of contact with infected people, may lead to the rapid spread of the disease (Cao et al., 2020; Kubota, Shiono, Kusumoto, & Fujinuma, 2020; van Oosterhout, Hall, Ly, & Tyler, 2021). This has been extensively studied based on data from many countries (Duhon, Bragazzi, & Kong, 2020). Soft power at the national level, including social and economic support and the reasonable optimization and mobilization of medical resources, has the potential to slow down and affect the spread of COVID-19 (Li et al., 2020; Haider et al., 2020). In addition, many studies have found that climate and other environmental factors may also play roles in the spread of COVID-19 (Price, Graham, & Ramalingam, 2019; Qu, Li, Hu, & Jiang, 2020). Overall, the spread and control of COVID-19 is a complex process involving various complex heterogeneous factors that are intertwined (Han, Zhao, & Pereira, 2022). This nonlinear system becomes even more complex with the emergence of new virus variants, the uncertain effect of vaccination on the human population (Sowa et al., 2021) and the heterogeneity of regional developmental levels. A well-developed interpretable machine learning framework (Murri et al., 2021; Ayoub, Yang, & Zhou, 2021) may adapt to this complex scenario and provide reasonable interpretation and predictive capabilities, which are critical for a greater understanding of the mechanisms underlying the effectiveness of various COVID-19 control measures. Ultimately, this will facilitate

---

the development and implementation of targeted and precise prevention and control strategies for the ongoing COVID-19 pandemic.

In this study, factors and datasets related to the spread and control of COVID-19 were collected as much as possible from all over the world and used for further analysis. The effects of various factors on the spread of COVID-19 were studied based on a nonlinear interpretable machine learning framework. Contributing factors were identified and characterized for different phases of the pandemic, before and after vaccine implementation, and for countries with or without rebound waves. The driving factors and potential mechanisms were explored, and forecasting efforts were made.

## **Materials and Methods**

### **Study design**

This study was designed as a quantitative study of the contribution of multi-dimensional factors to the spread of COVID-19, using the effective reproductive number as an indicator, and to the occurrence of multiple waves of infections. Specifically, this study focused on the following aspects: (1) the quantification of the overall contribution of multi-dimensional factors, before and after vaccine implementation; (2) a comparative analysis of the factors influencing the rising and fading stages of the pandemic, in countries with or without rebound waves, and before and after vaccine implementation; and (3) rebound risk forecasting for the spread of COVID-19.

### **Data**

Daily numbers of COVID-19 cases were collected from 39 countries up to June 30, 2021. These 39 countries were selected based on the following principles: (1) they represented countries from five continents, (2) the

total number of cases in these countries was among the highest during the study period and (3) data for multi-dimensional factors during the same period were available. The COVID-19 pandemic was divided into two phases—the pre-vaccine era and the post-vaccine era—using the date when vaccination began in each country as the dividing date. Of the 39 countries selected, 7 single-peak countries without a rebound wave and 13 double-peak countries with a rebound wave were identified in the pre-vaccine era (Figure S1), while 36 multi-peak countries with multiple rebound waves (excluding China, Australia and New Zealand, which had no rebound wave) were identified in the post-vaccine era. A rebound wave was identified when the number of new COVID-19 cases exceeded the average value of the peak in the pre-vaccine era for 3 consecutive weeks.

The factors analyzed were divided into the following groups: policy-related factors (including 13 specific non-pharmaceutical interventions in the previous week and at the current time point), travel-related factors (including domestic and foreign air passenger flow in the previous month and the current month), socioeconomic factors (population, population density, per capita income, aging ratio and sex ratio), environmental factors (maximum temperature, minimum temperature, average temperature, relative humidity and absolute humidity) and medical factors (number of doctors per 1,000 residents, number of nurses, number of pharmacists, number of beds, medical expenditure quota, cardiovascular death rate and diabetes prevalence). Additionally, the number of daily vaccinations was considered as a vaccine-related factor, and information regarding existing virus variants was included in the mutant-related factor category. All of these data were analyzed at the national level.

**Effective reproductive number:** The daily effective reproductive number ( $R_t$ ) was used as an indicator of the spread of COVID-19.  $R_t$  indicates the average number of secondary cases caused by one infected person per day (Fernández-Naranjo et al., 2021), and it was calculated based on the Susceptible-Infected-Recovered

---

model (Arroyo-Marioli, Bullano, Kucinskas, & Rondón-Moreno, 2021; Kermack & Mckendrick, 1927). When  $R_t \leq 1$ , maintaining the current prevention and control measures is expected to gradually control an infectious disease epidemic. In contrast, when  $R_t > 1$ , the infectious disease will continue to spread, suggesting that the prevention and control measures need to be optimized and strengthened. The  $R_t$  values for COVID-19 in the 39 countries included in the analysis were obtained from Our World in Data (<https://ourworldindata.org/covid-cases>).

Policy-related factors: These data were obtained from the Oxford COVID-19 Government Response Tracker (OxCGRT, <https://github.com/OxCGRT/covid-policy-tracker>), which includes data on the level of specific measures (strictness of implementation) taken by governments during the outbreak. Thirteen specific measures were analyzed, as detailed in Table S1. These classification variables were converted into a score in the range of [0,100]. The details of this calculation are presented in the Materials and Methods section of the Appendix.

Travel-related factors: Data on air passenger flow, based on monthly statistical measures at the national level, were collected from the Official Aviation Guide. These data included global passenger flow to the countries (*foreign flights*) and intra-country air passenger flow (*domestic flights*). A 1-month lag time was also included (Cao et al., 2020) for these data.

Socioeconomic factors: The number of residents (*population*), population density (*density*), per capita national income (*GDP*), percentage of the population aged  $\geq 65$  years (*old%*) and sex ratio (the ratio of men to woman, *sex ratio*) were collected from Our World in Data (<https://ourworldindata.org/covid-cases>).

Medical factors: Health resource data were obtained from the World Bank (<https://data.worldbank>). These data included the number of doctors (*doctor*), nurses (*nurse*), pharmacists (*pharmacist*) and beds (*beds*) and the

---

health expenditure per 1,000 people (*health expenditure*). Additional health data, such as the cardiovascular death rate (*cardiovascular death rate*) and diabetes prevalence (*diabetes prevalence*), were obtained from Our World in Data (<https://ourworldindata.org>).

Environmental factors: The daily maximum temperature, minimum temperature, average temperature, relative humidity and absolute humidity during the study period were collected from the National Oceanic and Atmospheric Administration (<https://www.noaa.gov/>).

Mutant-related factors: The five major virus strains (Alpha, Beta, Delta, Gamma and Lambda) were considered, and the temporal proportions referred to Alpha for Beta, Delta, Gamma and Lambda in each region were included as free variables in the analysis. These proportions were based on the weekly time scale in 39 countries obtained from the Global Initiative on Sharing Avian Influenza Data (<https://www.gisaid.org/hcov19-variants/>).

Vaccine factor: The number of daily COVID-19 vaccinations (*daily vaccination population*) in the 39 countries was obtained from Our World in Data (<https://ourworldindata.org/covid-cases>).

All time scales were unified as daily scales using the pandas time-series resampling method (<https://pandas.pydata.org/>). Moreover, considering the latent period of COVID-19, lag effects for policy- and travel-related factors were also considered and incorporated into the model (see the Materials and Methods section in the Appendix for more details).

### Machine learning framework

Interpretable machine learning frameworks were used to determine the nonlinear contribution of systematic factors to the  $R_t$  value. Specifically, multiple candidate nonlinear regression models, including random forest,



support vector machine and extreme gradient boosting (XGBoost) (Chen & Guestrin, 2016) models, were used, and important factors were selected using the sequential floating selection (SFS) algorithm (Fj, Pp, Mh, & Jk, 1994; Pudil, Novovi?Ov, & Kittler, 1994). The overall framework is shown in Figure S2. SFS is a sequential feature selection method based on greedy search that can remove functions from all feature sets and evaluate error functions. If the error reaches the optimal level, the combination of the remaining features is regarded as the best feature combination. A Bayesian search algorithm was used to identify the root mean square error (RMSE) of the parameter sets by fitting the model. The effect of generalization was evaluated by 10-fold cross-validation (S. M. Lundberg et al., 2020). Two temporally non-overlapping datasets were built based on pre-vaccine era (a training set up to October 31, 2020, and a testing set from November 1 to November 30, 2020) and post-vaccine era (a training set from January 2021 to May 31, 2021, and a testing set from June 1 to June 30, 2021) datasets. The model with the lowest fitting error was selected based on two training sets, and this model was used in subsequent analyses.

### Ranking contribution of the factors

Shapley Additive explanation (SHAP) is an interpretable analysis method (S. M. Lundberg et al., 2020) for the outputs of machine learning models. It treats all features as “contributors” and generates a predictive value (SHAP value). For each predicted sample, the value of each influencing factor, known as the feature value, for the sample was uniformly assigned. The SHAP value (S. Lundberg & Lee, 2017) was calculated as follows:

$$y_i = y_{base} + f(x_{i1}) + f(x_{i2}) + \dots + f(x_{iN}),$$

where the  $i$ -th sample of the  $N$ -th feature is defined as  $x_{iN}$ , the reference value of the model (the mean value of the sample target variable) is  $y_{base}$  and  $f(x_{iN})$  is the SHAP value of  $x_{iN}$ . A positive value indicates that the

feature has a positive contribution. The predicted value of the effective reproductive number  $y_i$  here is the linear combination of  $y_{base}$  and  $f(x_{iN})$ . The SHAP value was obtained based on the best model, and  $|f(x_{ij})|$  was used to measure the contribution of factors selected by the SFS algorithm, which was further converted into a percentage contribution value (Pan et al., 2020).

### Model evaluation and short-term pandemic rebound risk prediction

A rebound risk model was built by considering a daily increasing ( $R_t > 1$ ) or decreasing ( $R_t \leq 1$ ) trend as the subject (binary) and all factors selected as covariables. The RMSE and  $R^2$  values were used to evaluate the model fit to  $R_t$  using 10-fold cross-validation for the training set. The accuracy, precision, recall, F1 score and area under the curve were used to evaluate the rebound risk.

## Results

### Contribution of systematic factors to the spread of COVID-19 in pre- and post-vaccine eras

Based on the designed framework (Figure S2), the XGBoost model was selected (Table S2) and 1-week lag was used (Table S3) for subsequent analyses, as it showed the best fit with the training set (Table S2). Therefore, XGBoost with SHAP analysis (see Materials and Methods section for further details) was chosen to identify important factors contributing to the spread of COVID-19 ( $R_t$ ) in 39 countries, before and after vaccine implementation, and only top 20 factors were shown and discussed. Overall, the RMSE and  $R^2$  values, as evaluation measures of the model, were 0.109 and 0.725, respectively, in the pre-vaccine era, and 0.021 and 0.988, respectively, in the post-vaccine era. In the pre-vaccine era, policy-related (60.81%) and travel-related factors (18.84%) were the primary contributors to the spread of COVID-19. Socioeconomic and medical factors accounted for 9.31% and 6.58% of the spread of COVID-19, but environmental factors only accounted for 4.48%

---

(Figures 1A and S3A). In the post-vaccine era, the contribution pattern changed significantly, with policy-related factors reduced to 28.34% and travel-related, environmental, medical and socioeconomic factors reaching 21.67%, 9.62%, 8.64% and 8.20%, respectively. The contributions of mutant-related and vaccine factor reached 16.49% and 7.06%, respectively (Figures 1B and S3B).

Of the travel-related factors assessed in the pre-vaccine era, only population movement caused by foreign flights was identified as an important factor, contributing 18.84% (Figure 1A). Of the policy-related factors, the three greatest contributors were income support (previous week), stay-at-home requirements (previous week) and workplace closure (previous and current week), with contributions of 10.23%, 8.29%, 7.70% and 6.59%, respectively. Of the socioeconomic factors, GDP was the dominant contributor, with 5.80%. The contribution of single medical and environmental factors was low, with the highest percentages of 3.95% for the number of nurses and 2.69% for relative humidity, respectively. Although foreign flights (previous month) remained as the dominant factor with the highest contribution in the post-vaccine era (7.75%), the contributions from domestic flights in the current and previous months were 7.56% and 6.36%, respectively. Mutant-related factors contributed significantly to the dynamics of COVID-19 in the post-vaccine era, with 7.56%, 4.69% and 4.24% contributions for Delta, Gamma and Beta variants. Vaccine factor also made a significant contribution to the spread of COVID-19 in the post-vaccine era, with a contribution of 7.06%. As a result, the contributions from policy-related factors were significantly reduced in the post-vaccine era, with the top three factors—workplace closure in the previous week and the cancellation of public events in the current and previous weeks—contributing 5.63%, 5.63% and 4.66%, respectively.

At the country level, the overall pattern in the pre-vaccine era was the same as the pattern observed at the global level, with policy- and travel-related factors making the greatest contributions (Table S4).

This article is protected by copyright. All rights reserved.

Nevertheless, heterogeneity existed among countries, with the contribution of policy-related factors in several countries, such as Sweden, Mexico and Bulgaria, being as high as 70%. However, the contribution of travel-related factors in many countries, such as Spain, the United States and Cameroon, was 30%. Moreover, the contributions of socioeconomic, environmental and medical factors were low in all countries evaluated. Of note, there were noticeable differences in the contributions of these factors in different countries. For example, the contribution of socioeconomic factors in Colombia, Kenya and Chile was 16.17%, 14.89% and 14.64%, respectively; the contribution of environmental factors in Argentina, Ecuador and Colombia was 15.67%, 13.60% and 13.21%, respectively; and the contribution of medical factors in Algeria and Canada was 11.27% and 9.89%, respectively. In the post-vaccine era, a significant decrease in the dominance of policy-related factors was seen (Table S4), and the contribution of vaccine factor was relatively high in some countries, such as Bangladesh, Canada and South Africa, which showed contributions of 15.03%, 12.54% and 10.66%, respectively. At the same time, the contribution of variant viral strains in these countries was also prominent, reaching 32.67%, 25.92% and 28.16%, respectively.

#### **Factors related to the rebound of COVID-19**

As COVID-19 continued to spread, different countries responded differently, leading to different outcomes, with rebound waves occurring in many countries. In the pre-vaccine era, 13 of the 39 countries analyzed had a clear rebound wave with double peaks in the number of new cases, whereas 7 countries had no rebound wave, but a clear single peak (Figure S1). In addition, the first peak was further divided into the rising and fading stages, and the second peak was classified as the rebound stage. In the post-vaccine era, 36 of the 39 countries analyzed experienced multiple rebound waves, with more than one peak. Models were then built for the 7 double-peak countries with rebound waves, 13 single-peak countries without a rebound wave in the

---

pre-vaccine era and 36 countries with multiple rebounds in the post-vaccine era (Table S5). These models had RMSE values of 0.634, 0.298 and 0.021 and  $R^2$  values of 0.824, 0.712 and 0.988, respectively (Figure 2).

Overall, factors in the rising and fading stages had similar effects on the spread of COVID-19 in single-peak countries (Figure 2A). Policy- and travel-related factors were dominant during the rising stage (contributions of 33.7% and 18.7% in the rising and fading stages, respectively), but policy-related factors were predominant in the fading stage, reaching a contribution of 37.7% (Figure 2A). However, approximately half (45.8%) of the contribution to the spread of COVID-19 in double-peak countries was from factors in the rebound stage. Policy-related factors showed contributions of 32.6% and 25.0% during the rebound and fading stages, respectively (Figure 2B). In the post-vaccine era, the results of the model including countries with multiple rebounds were similar to the results of the model that included all 39 countries (Figure 1B), with policy- and travel-related factors accounting for more than half of the contribution, although the contribution from policy-related factors significantly decreased (Figure 2C). Mutant-related and vaccine factors contributed 17.7% and 7.3%, respectively, to the spread of COVID-19 in the post-vaccine era. The contributions from environmental and medical factors increased to 9.6% and 8.7% in the post-vaccine era.

In single-peak countries (Figure 2D), the contribution of domestic flights (previous month) in the rising stage was 18.71%, with effective SHAP values of 0.23, indicating a positive contribution to the spread of COVID-19. Other important factors in single-peak countries were policy-related factors, such as contact tracing (contribution, 9.90%; effective SHAP value, -1.14) and the cancellation of public events (contribution, 7.11%; effective SHAP value, -0.64) in the rising stage and school closure in the previous week (contribution, 8.03%; effective SHAP value, -0.79) and international travel controls (contribution, 7.34%; effective SHAP value, -0.32) in the fading stage. All of these factors decreased the spread of COVID-19 (Figure 2D).

This article is protected by copyright. All rights reserved.

However, a more diverse range of factors from different categories was involved in double-peak countries, many of which were present in the rebound stage (12/20, Figure 2E). In particular, foreign flights in the rising stage (contribution, 9.09%) and domestic flights in the fading stage (contribution, 7.66%) increased the spread of COVID-19, with effective SHAP values of 1.43 and 1.15, respectively. The cancellation of public events in the fading stage (contribution, 8.58%) and testing policies (previous week) in the rebound stage (contribution, 7.73%) reduced the spread of COVID-19, with effective SHAP values of -2.69 and -0.04, respectively. The contribution of foreign flights continued to decrease in the post-vaccine era (contribution, 6.39%; effective SHAP value, -1.56; negative effect). The Delta variant became the most influential contributor to the spread of COVID-19 in the post-vaccine era (contribution, 8.12%; SHAP value, 9.18). Other virus variants, such as Gamma and Beta, also had a promoting effect on the spread of COVID-19 in the post-vaccine era, with SHAP values of 1.37 and 5.37, respectively. The strong inhibitory effect of vaccine was evident (contribution, 7.26%; SHAP value, -9.28). Unexpectedly, most policies had a boosting effect on the spread of COVID-19, with SHAP values  $> 0$ , for five of the six factors evaluated (Figure 2F).

### Rebound risk forecasting

After understanding the different contributions from various factors, we sought to determine whether the rebound risk of COVID-19 could be predicted. Predictive models were built based on data from 20 selected countries in the pre-vaccine era and 36 countries in the post-vaccine era. A high risk of rebound was defined as  $R_t > 1$ , and a low risk of rebound was defined as  $R_t \leq 1$ . Daily forecasting was then performed for the following 30 days, based on actual data for all of the contributing factors in the model. COVID-19 rebound was defined as  $R_t > 1$  for more than half of the 30 days. The accuracy of the risk forecast was 0.78 and 0.81 in the pre-vaccine and post-vaccine eras, respectively (Tables 1 and S6).

This article is protected by copyright. All rights reserved.

**Table 1: Rebound risk forecasting for 39 countries**

Country	Rebound (%)	Pre-vaccine era			Post-vaccine era	
		Rebound prediction (%)	Correct prediction	Rebound (%)	Rebound prediction (%)	Correct prediction
Afghanistan	100 (30/30)	100 (30/30)	✓	93.33 (28/30)	100 (30/30)	✓
Algeria	–	–	–	100 (30/30)	100 (30/30)	✓
Argentina	–	–	–	0 (0/30)	53.33 (16/30)	✗
Austria	–	–	–	0 (0/30)	0 (0/30)	✓
Australia	36.67 (11/30)	3.33 (1/30)	✓	–	–	–
Bangladesh	–	–	–	100 (30/30)	86.67 (26/30)	✓
Bolivia	–	–	–	40 (12/30)	100 (30/30)	✗
Bulgaria	–	–	–	0 (0/30)	0 (0/30)	✓
Cameroon	–	–	–	0 (0/30)	0 (0/30)	✓
Canada	100 (30/30)	100 (30/30)	✓	0 (0/30)	0 (0/30)	✓
Chile	26.67 (8/30)	3.33 (1/30)	✓	26.67 (8/30)	63.33 (19/30)	✗
China	73.33 (22/30)	86.67 (26/30)	✓	–	–	–
Colombia	–	–	–	93.33 (28/30)	86.67 (26/30)	✓
Denmark	100 (30/30)	100 (30/30)	✓	6.67 (2/30)	0 (0/30)	✓
Ecuador	–	–	–	0 (0/30)	0 (0/30)	✓
Finland	100 (30/30)	100 (30/30)	✓	43.33 (13/30)	0 (0/30)	✓
France	26.67 (8/30)	100 (30/30)	✗	0 (0/30)	0 (0/30)	✓
Germany	100 (30/30)	100 (30/30)	✓	0 (0/30)	0 (0/30)	✓
Greece	–	–	–	16.67 (5/30)	0 (0/30)	✓
Iceland	–	–	–	0 (0/30)	0 (0/30)	✓

India	–	–	-	0 (0/30)	0 (0/30)	✓
Indonesia	–	–	-	100 (30/30)	26.67 (8/30)	✗
Ireland	0 (0/30)	0 (0/30)	✓	23.33 (7/30)	0 (0/30)	✓
Italy	60 (18/30)	53.33 (16/30)	✓	0 (0/30)	0 (0/30)	✓
Japan	100 (30/30)	100 (30/30)	✓	0 (0/30)	0 (0/30)	✓
Kenya	60 (18/30)	100 (30/30)	✓	83.33 (25/30)	70 (21/30)	✓
Mexico	–	–	-	100 (30/30)	6.67 (2/30)	✗
New Zealand	50 (15/30)	6.67(2/30)	✗	–	–	-
Norway	70 (21/30)	100 (30/30)	✓	6.67 (2/30)	0 (0/30)	✓
Pakistan	100 (30/30)	100 (30/30)	✓	6.67 (2/30)	36.67 (11/30)	✓
Panama	–	–	-	100 (30/30)	100 (30/30)	✓
Philippines	–	–	-	43.33 (13/30)	16.67 (5/30)	✓
Portugal	–	–	-	100 (30/30)	100 (30/30)	✓
Romania	–	–	-	0 (0/30)	0 (0/30)	✓
South Africa	83.33 (25/30)	60 (18/30)	✓	100 (30/30)	100 (30/30)	✓
South Korea	100 (30/30)	100 (30/30)	✓	50 (15/30)	0 (0/30)	✓
Spain	40 (12/30)	100 (30/30)	✗	73.33 (22/30)	0 (0/30)	✗
Sweden	–	–	-	0 (0/30)	0 (0/30)	✓
United States	100 (30/30)	83.33 (25/30)	✓	20 (6/30)	0 (0/30)	✓

*Note: “–” indicates that the country did not meet the selection requirements at this stage. Period of risk forecasting: November 1 to 30, 2020 (pre-vaccine era) and June 1 to 30, 2021 (post-vaccine era).*



---

## Discussion

Based on the available datasets, the contribution of systematical factors to the spread of COVID-19 was studied through an interpretable machine learning framework. Globally, policy-related and travel-related factors (Chinazzi et al., 2020; Jia et al., 2020; Bielecki et al., 2021; Wells et al., 2022) play essential roles in the spread of the COVID-19 epidemic during the stage when vaccine was not yet widely available (Figure 1A). Compared with double-peak countries, spread of COVID-19 in single-peak countries was mostly driven by domestic migration and policy-related factors (Figure 2D). Neglecting travel restrictions (both foreign and domestic flights) in the earlier stage and relaxing intervention policies in the later stages may increase the likelihood of pandemic resurgence (Figure 2E). This remained true for the post-vaccine era, during which unfavorable policy implementation and new virus variants with increased infectivity caused multi-peak rebounds, even after vaccine implementation (Figure 2F). A forecasting model based on this knowledge gave reasonable predictions for the risk of rebound. Our study sheds light on quantitatively understanding the mechanisms underlying the global spread of COVID-19 and will facilitate the precise and cost-effective control of the ongoing COVID-19 pandemic (Reddy et al., 2021).

We found that foreign flights played a dominant role in the global spread of COVID-19 in the pre-vaccine era, with a combined contribution of 18.84% (Figure 1A) (Bo et al., 2021; Liu et al., 2020; Middelburg & Rosendaal, 2020; Wells et al., 2020; Wong, Huang, Teoh, & Wong, 2020). The contribution gradually shifted from foreign migration to domestic flights with time (Figure 1B). The most important policy-related factor was income support (previous and current week), with a combined contribution of 12.59%, implicating the importance of economic support and the complex behaviors of human society during an outbreak of an emerging infectious disease. The contributions of socioeconomic and medical factors were less critical than the dominant policy-

This article is protected by copyright. All rights reserved.

and travel-related factors. However, their effects cannot be neglected. For example, countries with the greatest contributions from socioeconomic factors (Colombia, Kenya and Chile; Table S4) were underdeveloped regions, implying that the socioeconomic impact of the pandemic was more pronounced in these regions. During a pandemic, the effect of environmental factors is difficult to tease out due to the greater impact of other factors, but environmental factors have been shown to contribute to transmission (Jia et al., 2020; Malki et al., 2020; Yang et al., 2020). Some countries, such as Argentina, showed a relatively large contribution (15.76%) from environmental factors, although this was not a universal finding (Table S4). Overall, the differential implementation of control policies combined with the emergence of variants with greater infectivity has created additional uncertainties regarding recurrent global outbreaks (Figure 2E/F). The dynamic contribution of different factors together with their time-series data provides a visual description of the drivers and specific mechanisms of pandemic rebound (Figures 2 and S4). In the case of Japan, the rebound in the pandemic (June to August, 2020) was the result of weak policy implementation, whereas Finland showed a rebound because of a combination of policy relaxation and weakened travel restrictions (Figure S4).

A comparative analysis of countries with and without rebound indicated that strict contact tracing and effective control of population movement during the rising stage, while maintaining control during the fading stage, were crucial factors for effective control of COVID-19 in single-peak countries (Figure 2A, 2D). However, countries that experienced a rebound of COVID-19 showed relaxed control policies in the later stage (Figure 2B, 2E), along with the promoting effect of population movement (both domestic and international) and relaxed testing policies and public event cancellation. The relaxation of restrictions should be carefully evaluated (The, 2020), especially after the implementation of a vaccination program, and restrictions are essential when considering additional reinforcement factors such as new mutants (Figure 2F). Effectively controlling the

---

COVID-19 pandemic and preventing a rebound require a systematic approach (Coccia, 2020; Sarkodie & Owusu, 2020). Therefore, the quantitative estimation of effective measures and their contributions informs precise prevention and control measures and provides important suggestions for policymakers, which have been shown to be crucial for controlling the pandemic (Ferrante et al., 2021; Greer, King, da Fonseca, & Peralta-Santos, 2020).

The findings of this study also facilitate rebound risk forecasting. For this purpose, 1-month predictions for the selected countries in the pre- and post-vaccine eras were made with an accuracy of 0.78 and 0.81, respectively. The large amount of heterogeneity may explain the incorrect predictions in some countries, as the key contributing factors may not be comprehensive for these countries and their relationships may not be simple and universal. For example, the rebound risk prediction for November 2021 was incorrect for France, likely due to an excess positive contribution of foreign flights in the model, which increased the likelihood of a rebound (detailed in Table S7).

This result may have been influenced by more specific travel policies and the implementation of screening at the airport. Risk misidentification was most common in South American countries in the post-vaccine era (e.g., Argentina, Bolivia and Chile), which may be related to the unique circulation pattern of COVID-19 in these countries, in which Gamma and Lambda variants were dominant at a time when the Delta variant was dominant in other countries.

There are several limitations to our study. First, although the representation of the countries included in the analysis was considered, the limited number of countries included in the study due to data quality requirements may have affected the generalization of some of our conclusions. Data from more countries should be collected and analyzed in future studies. Second, multiple machine learning algorithms coupled with an interpretable

This article is protected by copyright. All rights reserved.

---

framework were used to explore the complex interactions between various influencing factors, to try to improve the reliability and robustness of the optimized models. Although machine learning models are well developed, show advantages in handling nonlinear problems and are broadly used in other fields, they are less frequently used in the field of policy evaluation. In the future, more comprehensive investigations are needed, coupled with studies based on mechanistic models. A more general framework for the effective evaluation and optimization of control policies is also needed and should be the next priority. Finally, the daily basic reproductive number was the only indicator of the spread of COVID-19 used in this study. This metric captures only one aspect of the disease. There may be more relevant disease burden indices for COVID-19, such as death or hospitalization rates. These outcomes and their influencing factors should be comparatively explored in future studies. The framework proposed in this study was helpful in quantitatively understanding the contributions of various factors to the spread of COVID-19. The patterns observed in this study provide insights to facilitate the development of improved prevention and control strategies for the ongoing COVID-19 pandemic and for future pandemics.

## Conclusion

Factors that affect the spread of COVID-19 in 39 countries were systematically evaluated and prioritized using an interpretable machine learning framework. Policy- and travel-related factors were found to be the main drivers, with policy-related factors being more dominant (more than 60% of the overall contribution). Travel-related factors played an important role in the earlier stage, whereas policy-related interventions were dominant contributors at the later stage, especially in countries that experienced a rebound. Care should be taken when deciding to relax non-pharmaceutical interventions, even after the implementation of a vaccination program, and new mutants with greater infectivity may worsen the situation and cause recurrent outbreaks. A

---

reliable prediction model was built based on the findings of the study, and this model may be used to evaluate potential control strategies. A quantitative understanding of the combinatorial effect of various control measures is needed for precisely and effectively controlling the spread of COVID-19.

### **Supplementary Materials**

Supplementary material is available in the online section of *Transboundary and Emerging Diseases*.

### **Funding**

This work was supported by National Key Research and Development Projects of the Ministry of Science and Technology of China under grant 2021YFC2301300; the Guangdong Frontier and Key Tech Innovation Program under grants 212020012620600006, 2019B020228001 and 2019B111103001; the Natural Science Foundation of Guangdong Province, China, under grant 2021A1515011592; and the Shenzhen Science and Technology Program under grant KQTD20180411143323605.

### **Authors' contributions**

XD and ZC designed the study. ZC, ZQ, FT and SL collected and analyzed the data. XD, SX, DT and GJ interpreted the data. XD, ZC and ZQ prepared the manuscript. XD, ZC, SX, DT and GJ edited the paper. All authors reviewed and approved the submitted manuscript.

### **Acknowledgments**

The authors would like to thank the many thousands of Centers for Disease Control and Prevention staff, health workers and data scientists who continuously collect and publicly share data and are dedicated to containing the spread of COVID-19.

---

## Data Availability Statement

The data that support the findings of this study are openly available at the Oxford Covid-19

Government Response Tracker

(<https://github.com/OxCGRT/covid-policy-tracker/>) and Our World in Data

(<https://ourworldindata.org/>).

## Conflicts of Interest

All authors declare no competing interests.

## References

- Roser, M., Ritchie, H., Ortiz-Ospina, E., & Hasell, J. (2020). Coronavirus pandemic (COVID-19). *Our world in data*.
- Looi, M. K. (2020). Covid-19: Is a second wave hitting Europe? *Bmj*, *371*, m4113. doi:10.1136/bmj.m4113
- Cacciapaglia, G., Cot, C., & Sannino, F. (2020). Second wave COVID-19 pandemics in Europe: a temporal playbook. *Sci Rep*, *10*(1), 15514. doi:10.1038/s41598-020-72611-5
- Wise, J. (2020). Covid-19: Risk of second wave is very real, say researchers. *Bmj*, *369*, m2294. doi:10.1136/bmj.m2294
- Xu, S., & Li, Y. (2020). Beware of the second wave of COVID-19. *Lancet*, *395*(10233), 1321–1322. doi:10.1016/s0140-6736(20)30845-x
- Zhu, N., Zhang, D., Wang, W., Li, X., & Tan, W. (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine*, *382*(8).
- Awadasseid, A., Wu, Y., Tanaka, Y., & Zhang, W. (2021). Current advances in the development of SARS-CoV-2 vaccines. *Int J Biol Sci*, *17*(1), 8–19. doi:10.7150/ijbs.52569
- Sonabend, R., Whittles, L. K., Imai, N., Perez-Guzman, P. N., Knock, E. S., Rawson, T., . . . Cori, A. (2021). Non-pharmaceutical interventions, vaccination, and the SARS-CoV-2 delta variant in England: a

mathematical modelling study. *Lancet*, 398(10313), 1825–1835.

doi:10.1016/s0140-6736(21)02276-5

- Mohapatra, R. K., Tiwari, R., Sarangi, A. K., Sharma, S. K., Khandia, R., Saikumar, G., & Dhama, K. (2022). Twin combination of Omicron and Delta variant triggering a Tsunami wave of ever high surges in COVID-19 cases: a challenging global threat with a special focus on Indian sub-continent. *J Med Virol*. doi:10.1002/jmv.27585
- Di Domenico, L., Pullano, G., Sabbatini, C. E., Boëlle, P. Y., & Colizza, V. (2020). Impact of lockdown on COVID-19 epidemic in Île-de-France and possible exit strategies. *BMC Med*, 18(1), 240. doi:10.1186/s12916-020-01698-4
- Ferrante, L., Steinmetz, W. A., Almeida, A. C. L., Leão, J., Vassão, R. C., Tupinambás, U., . . . Duczmal, L. H. (2020). Brazil's policies condemn Amazonia to a second wave of COVID-19. *Nat Med*, 26(9), 1315. doi:10.1038/s41591-020-1026-x
- Leung, K., Wu, J. T., Liu, D., & Leung, G. M. (2020). First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment. *Lancet*, 395(10233), 1382–1393. doi:10.1016/s0140-6736(20)30746-7
- Panovska-Griffiths, J., Kerr, C. C., Stuart, R. M., Mistry, D., Klein, D. J., Viner, R. M., & Bonell, C. (2020). Determining the optimal strategy for reopening schools, the impact of test and trace interventions, and the risk of occurrence of a second COVID-19 epidemic wave in the UK: a modelling study. *Lancet Child Adolesc Health*, 4(11), 817–827. doi:10.1016/s2352-4642(20)30250-9
- Qiu, Y., Chen, X., & Shi, W. (2020). Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. *J Popul Econ*, 1–46. doi:10.1007/s00148-020-00778-2
- Sebhatu, A., Wennberg, K., Arora-Jonsson, S., & Lindberg, S. I. (2020). Explaining the homogeneous diffusion of COVID-19 nonpharmaceutical interventions across heterogeneous countries. *Proc Natl Acad Sci U S A*, 117(35), 21201–21208. doi:10.1073/pnas.2010625117
- Kandel, N., Chungong, S., Omaar, A., & Xing, J. (2020). Health security capacities in the context of COVID-19 outbreak: an analysis of International Health Regulations annual report data from 182 countries. *Lancet*, 395(10229), 1047–1053. doi:10.1016/s0140-6736(20)30553-5
- Brauner, J. M., Mindermann, S., Sharma, M., Johnston, D., Salvatier, J., Gavenčiak, T., . . . Kulveit, J. (2020). Inferring the effectiveness of government interventions against COVID-19. *Science*. doi:10.1126/science.abd9338

- Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., . . . Bhatt, S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, *584*(7820), 257–261. doi:10.1038/s41586-020-2405-7
- Haug, N., Geyrhofer, L., Londei, A., Dervic, E., Desvars-Larrive, A., Loreto, V., . . . Klimek, P. (2020). Ranking the effectiveness of worldwide COVID-19 government interventions. *Nat Hum Behav*, *4*(12), 1303–1312. doi:10.1038/s41562-020-01009-0
- Hsiang, S., Allen, D., Annan-Phan, S., Bell, K., Bolliger, I., Chong, T., . . . Wu, T. (2020). The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*, *584*(7820), 262–267. doi:10.1038/s41586-020-2404-8
- Russell, T. W., Wu, J. T., Clifford, S., Edmunds, W. J., Kucharski, A. J., & Jit, M. (2021). Effect of internationally imported cases on internal spread of COVID-19: a mathematical modelling study. *Lancet Public Health*, *6*(1), e12–e20. doi:10.1016/s2468-2667(20)30263-2
- Sun, J., Shi, Z., & Xu, H. (2020). Non-pharmaceutical interventions used for COVID-19 had a major impact on reducing influenza in China in 2020. *J Travel Med*, *27*(8). doi:10.1093/jtm/taaa064
- Cao, Z., Tang, F., Chen, C., Zhang, C., Guo, Y., Lin, R., . . . Du, X. (2020). Impact of Systematic Factors on the Outbreak Outcomes of the Novel COVID-19 Disease in China: Factor Analysis Study. *J Med Internet Res*, *22*(11), e23853. doi:10.2196/23853
- Kubota, Y., Shiono, T., Kusumoto, B., & Fujinuma, J. (2020). Multiple drivers of the COVID-19 spread: The roles of climate, international mobility, and region-specific conditions. *PLoS One*, *15*(9), e0239385. doi:10.1371/journal.pone.0239385
- van Oosterhout, C., Hall, N., Ly, H., & Tyler, K. M. (2021). COVID-19 evolution during the pandemic – Implications of new SARS-CoV-2 variants on disease control and public health policies. *Virulence*, *12*(1), 507–508. doi:10.1080/21505594.2021.1877066
- Duhon, J., Bragazzi, N., & Kong, J. D. (2020). 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. doi:10.1016/j.scitotenv.2020.144325
- Li, M., Zhang, Z., Cao, W., Liu, Y., Du, B., Chen, C., . . . Wang, X. (2020). Identifying novel factors associated with COVID-19 transmission and fatality using the machine learning approach. *Sci Total Environ*, 142810. doi:10.1016/j.scitotenv.2020.142810



- Haider, N., Yavlinsky, A., Chang, Y. M., Hasan, M. N., Benfield, C., Osman, A. Y., . . . Kock, R. (2020). The Global Health Security index and Joint External Evaluation score for health preparedness are not correlated with countries' COVID-19 detection response time and mortality outcome. *Epidemiol Infect*, 148, e210. doi:10.1017/s0950268820002046
- Price, R. H. M., Graham, C., & Ramalingam, S. (2019). Association between viral seasonality and meteorological factors. *Sci Rep*, 9(1), 929. doi:10.1038/s41598-018-37481-y
- Qu, G., Li, X., Hu, L., & Jiang, G. (2020). An Imperative Need for Research on the Role of Environmental Factors in Transmission of Novel Coronavirus (COVID-19). *Environ Sci Technol*, 54(7), 3730–3732. doi:10.1021/acs.est.0c01102
- Han, Y., Zhao, W., & Pereira, P. (2022). Global COVID-19 pandemic trends and their relationship with meteorological variables, air pollutants and socioeconomic aspects. *Environ Res*, 204(Pt C), 112249. doi:10.1016/j.envres.2021.112249
- Sowa, P., Kiszkiel, Ł., Laskowski, P. P., Alimowski, M., Szczerbiński, Ł., Paniczko, M., . . . Kamiński, K. (2021). COVID-19 Vaccine Hesitancy in Poland–Multifactorial Impact Trajectories. *Vaccines (Basel)*, 9(8). doi:10.3390/vaccines9080876
- Murri, R., Lenkiewicz, J., Masciocchi, C., Iacomini, C., Fantoni, M., Damiani, A., . . . Valentini, V. (2021). A machine-learning parsimonious multivariable predictive model of mortality risk in patients with Covid-19. *Sci Rep*, 11(1), 21136. doi:10.1038/s41598-021-99905-6
- Ayoub, J., Yang, X. J., & Zhou, F. (2021). Combat COVID-19 infodemic using explainable natural language processing models. *Inf Process Manag*, 58(4), 102569. doi:10.1016/j.ipm.2021.102569
- Fernández-Naranjo, R. P., Váscquez-González, E., Simbaña-Rivera, K., Gómez-Barreno, L., Izquierdo-Condoy, J. S., Cevallos-Robalino, D., & Ortiz-Prado, E. (2021). Statistical data driven approach of COVID-19 in Ecuador: R0 and Rt estimation via new method. *Infect Dis Model*, 6, 232–243. doi:10.1016/j.idm.2020.12.012
- Arroyo-Marioli, F., Bullano, F., Kucinkas, S., & Rondón-Moreno, C. (2021). Tracking R of COVID-19: A new real-time estimation using the Kalman filter. *PLoS One*, 16(1), e0244474. doi:10.1371/journal.pone.0244474
- Kermack, W. O., & Mckendrick, A. G. A. (1927). A Contribution to the Mathematical Theory of Epidemics. *Proceedings of The Royal Society A Mathematical Physical and Engineering Sciences*, 115(772), 700–721.

- 
- Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. Paper presented at the the 22nd ACM SIGKDD International Conference.
- Fj, A., Pp, B., Mh, C., & Jk, C. (1994). Comparative study of techniques for large-scale feature selection – ScienceDirect. *Machine Intelligence and Pattern Recognition*, 16, 403–413.
- Pudil, P., Novovi?Ov, J., & Kittler, J. (1994). Floating Search Methods in Feature Selection. *Pattern Recognition Letters*, 15(11), 1119–1125.
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., . . . Lee, S. I. (2020). From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat Mach Intell*, 2(1), 56–67. doi:10.1038/s42256-019-0138-9
- Lundberg, S., & Lee, S. I. (2017). *A Unified Approach to Interpreting Model Predictions*. Paper presented at the Nips.
- Pan, P., Li, Y., Xiao, Y., Han, B., Su, L., Su, M., . . . Xie, L. (2020). Prognostic Assessment of COVID-19 in the Intensive Care Unit by Machine Learning Methods: Model Development and Validation. *J Med Internet Res*, 22(11), e23128. doi:10.2196/23128
- Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., . . . Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489), 395–400. doi:10.1126/science.aba9757
- Jia, J. S., Lu, X., Yuan, Y., Xu, G., Jia, J., & Christakis, N. A. (2020). Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature*, 582(7812), 389–394. doi:10.1038/s41586-020-2284-y
- Bielecki, M., Patel, D., Hinkelbein, J., Komorowski, M., Kester, J., Ebrahim, S., . . . Schlagenhauf, P. (2021). Air travel and COVID-19 prevention in the pandemic and peri-pandemic period: A narrative review. *Travel Med Infect Dis*, 39, 101915. doi:10.1016/j.tmaid.2020.101915
- Wells, C. R., Pandey, A., Fitzpatrick, M. C., Crystal, W. S., Singer, B. H., Moghadas, S. M., . . . Townsend, J. P. (2022). Quarantine and testing strategies to ameliorate transmission due to travel during the COVID-19 pandemic: a modelling study. *Lancet Reg Health Eur*, 100304. doi:10.1016/j.lanepe.2021.100304
- Reddy, K. P., Shebl, F. M., Foote, J. H. A., Harling, G., Scott, J. A., Panella, C., . . . Siedner, M. J. (2021). Cost-effectiveness of public health strategies for COVID-19 epidemic control in South Africa: a microsimulation modelling study. *Lancet Glob Health*, 9(2), e120–e129. doi:10.1016/s2214-109x(20)30452-6

- Bo, Y., Guo, C., Lin, C., Zeng, Y., Li, H. B., Zhang, Y., . . . Lao, X. Q. (2021). 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. doi:10.1016/j.ijid.2020.10.066
- Liu, D., Clemente, L., Poirier, C., Ding, X., Chinazzi, M., Davis, J., . . . Santillana, M. (2020). Real-Time Forecasting of the COVID-19 Outbreak in Chinese Provinces: Machine Learning Approach Using Novel Digital Data and Estimates From Mechanistic Models. *J Med Internet Res*, *22*(8), e20285. doi:10.2196/20285
- Middelburg, R. A., & Rosendaal, F. R. (2020). COVID-19: How to make between-country comparisons. *Int J Infect Dis*, *96*, 477–481. doi:10.1016/j.ijid.2020.05.066
- Wells, C. R., Sah, P., Moghadas, S. M., Pandey, A., Shoukat, A., Wang, Y., . . . Galvani, A. P. (2020). Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proc Natl Acad Sci U S A*, *117*(13), 7504–7509. doi:10.1073/pnas.2002616117
- Wong, M. C., Huang, J., Teoh, J., & Wong, S. H. (2020). Evaluation on different non-pharmaceutical interventions during COVID-19 pandemic: An analysis of 139 countries. *J Infect*, *81*(3), e70–e71. doi:10.1016/j.jinf.2020.06.044
- Malki, Z., Atlam, E. S., Hassanien, A. E., Dagneu, G., Elhosseini, M. A., & Gad, I. (2020). Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches. *Chaos Solitons Fractals*, *138*, 110137. doi:10.1016/j.chaos.2020.110137
- Yang, J., Li, J., Lai, S., Ruktanonchai, C. W., Xing, W., Carioli, A., . . . Tatem, A. J. (2020). Uncovering two phases of early intercontinental COVID-19 transmission dynamics. *J Travel Med*, *27*(8). doi:10.1093/jtm/taaa200
- The, L. (2020). COVID-19 in India: the dangers of false optimism. *Lancet*, *396*(10255), 867. doi:10.1016/s0140-6736(20)32001-8
- Coccia, M. (2020). Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. *Sci Total Environ*, *729*, 138474. doi:10.1016/j.scitotenv.2020.138474
- Sarkodie, S. A., & Owusu, P. A. (2020). Impact of meteorological factors on COVID-19 pandemic: Evidence from top 20 countries with confirmed cases. *Environ Res*, *191*, 110101. doi:10.1016/j.envres.2020.110101

Ferrante, L., Duczmal, L., Steinmetz, W. A., Almeida, A. C. L., Leão, J., Vassão, R. C., . . . Fearnside, P. M. (2021). How Brazil's President turned the country into a global epicenter of COVID-19. *J Public Health Policy*, 42(3), 439–451. doi:10.1057/s41271-021-00302-0

Greer, S. L., King, E. J., da Fonseca, E. M., & Peralta-Santos, A. (2020). The comparative politics of COVID-19: The need to understand government responses. *Glob Public Health*, 15(9), 1413–1416. doi:10.1080/17441692.2020.1783340

## Figure Legends

Figure 1: Contribution of different factors to the  $R_t$  value of COVID-19 in 39 countries before (A) and after (B) vaccine implementation.

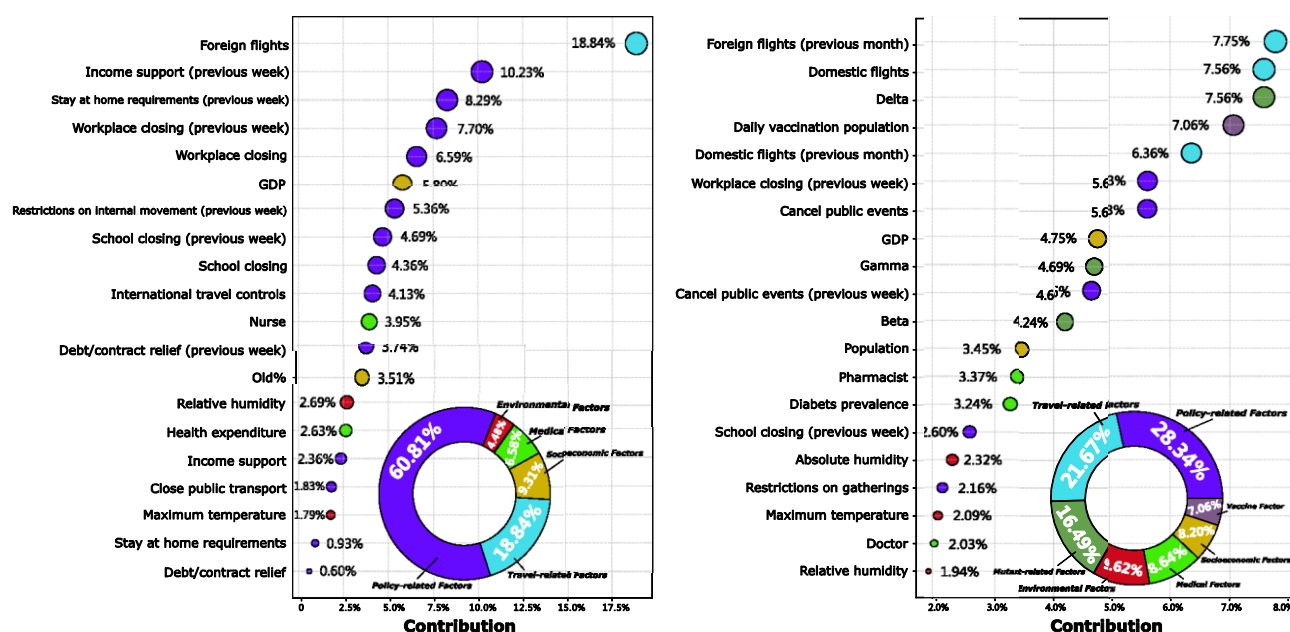


Figure 2: Contribution of the 20 most important factors and their effective SHAP values (sum of SHAP values)

for  $R_t$  at different stages of the pandemic in single-peak (A, D), double-peak (B, E) and multiple-peak (C, F)

countries. The first peak includes the rising and fading stages, and the additional peaks were defined as the rebound stage. The color for each group is the same as in Figure 1.

