

Special Section:

The COVID-19 pandemic:
linking health, society and
environment

Key Points:

- Disease control measures turn out to be the most important factors, followed by climatic factors and demographic characteristics
- Most of the variation can be accounted for by the spatiotemporal heterogeneity in the transmission trajectories among the countries
- Evaluation of mobility-transmission relationship and people's response to mobility-restrictions are vital in combating transmission

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Y. Cao and N. C. Stenseth,
yihan.cao@outlook.com;
n.c.stenseth@mn.uio.no

Citation:

Cao, Y., Whittington, J. D., Kausrud, K., Li, R., & Stenseth, N. C. (2022). The relative contribution of climatic, demographic factors, disease control measures and spatiotemporal heterogeneity to variation of global COVID-19 transmission. *GeoHealth*, 6, e2022GH000589. <https://doi.org/10.1029/2022GH000589>

Received 14 FEB 2022

Accepted 30 JUN 2022

Author Contributions:

Conceptualization: Yihan Cao

Data curation: Yihan Cao

© 2022. The Authors. GeoHealth published by Wiley Periodicals LLC on behalf of American Geophysical Union. This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs License](#), which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.



The Relative Contribution of Climatic, Demographic Factors, Disease Control Measures and Spatiotemporal Heterogeneity to Variation of Global COVID-19 Transmission

Yihan Cao¹ , Jason D. Whittington¹, Kyrre Kausrud², Ruiyun Li¹, and Nils Chr. Stenseth¹

¹Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway,

²Norwegian Veterinary Institute, Oslo, Norway

Abstract Despite a substantial number of COVID-19 related research papers published, it remains unclear as to which factors are associated with the observed variation in global transmission and what are their relative levels of importance. This study applies a rigorous statistical framework to provide robust estimations of the factor effects for a global and integrated perspective on this issue. We developed a mixed effect model exploring the relative importance of potential factors driving COVID-19 transmission while incorporating spatial and temporal heterogeneity of spread. We use an integrated data set for 87 countries across six continents for model specification and fitting. The best model accounts for 70.4% of the variance in the data analyzed: 10 fixed effect factors explain 20.5% of the variance, random temporal and spatial effects account for 50% of the variance. The fixed effect factors are classified into climatic, demographic and disease control groups. The explained variance in global transmission by the three groups are 0.6%, 1.1%, and 4.4% respectively. The high proportion of variance accounted for by random effects indicated striking differences in temporal transmission trajectories and effects of population mobility among the countries. In particular, the country-specific mobility-transmission relationship turns out to be the most important factor in explaining the observed global variation of transmission in the early phase of COVID-19 pandemic.

Plain Language Summary We have observed substantial variation in global transmission trajectories of COVID-19. Using statistical analysis, this study aims to investigate the factors that are associated with the observed variation in global transmission and what are their relative levels of importance. We conclude that the variation in transmission trajectories in various countries is mostly accounted for by spatiotemporal heterogeneity in transmission. In particular, disease control policies and population response to COVID-19 transmission make the largest contribution and demographic features have the least importance. Climatic factors also play a role but turn to be much less important than disease control policies. The mobility-transmission relationship is country-specific and turns out to be the most important factor in explaining the observed global variation of transmission. The complexity of COVID-19 transmission is also demonstrated through the wide range of estimated effects of population mobility on transmission between countries.

1. Introduction

COVID-19 continues to pose a considerable threat to the global public health and economy. We also see a resurgence of COVID-19 cases in the early winter of 2021–2022, with the epidemic trajectories varying in different parts of the world. Most scientists agree that it is near-impossible to eliminate the virus worldwide in 2022 or even beyond, but the future impacts will depend on the duration of the immune response after vaccination or recovery from infection, social mixing patterns now and in the future, and what preventative measures we put in place (Scudellari, 2020).

Similar to other endemic human coronaviruses, the transmission of COVID-19 are associated with various factors, including but not limiting to climatic factors (Alam & Sultana, 2021; McClymont & Hu, 2021), such as temperature, humidity (Adhikari et al., 2020; Fu et al., 2020), air pollution (Zheng et al., 2021), UV radiation (Karapiperis et al., 2020) and wind speed (Islam et al., 2020); demographic variables such as population size (Baker et al., 2020), density (Metelmann et al., 2021), age structure (Monod et al., 2021); socio-economic factors such as GDP (Sarmadi et al., 2020). These factors not only determine how SARS-CoV2 spread in the pandemic but also how effective each country's response is against the virus (Arsalan et al., 2020). Furthermore, we have seen that the timing and intensity of intervention actions against COVID-19 have varied considerably across

Formal analysis: Yihan Cao
Funding acquisition: Nils Chr. Stenseth
Methodology: Yihan Cao
Project Administration: Nils Chr. Stenseth
Supervision: Nils Chr. Stenseth
Writing – original draft: Yihan Cao
Writing – review & editing: Yihan Cao, Jason D. Whittington, Kyrre Kausrud, Ruiyun Li, Nils Chr. Stenseth

countries, and mobility patterns (Carteni et al., 2020; Li et al., 2021) were also associated with COVID-19 spread. To date, the published literature accessing the associations between transmissibility of SARS-CoV-2 and the climatic determinants, demographic factors and disease control variables either conducted analysis at national scales (Zheng et al., 2021), focused on meteorological factors (Sera et al., 2021), or ignored the temporal heterogeneity in transmission across countries (Metelmann et al., 2021). In addition, there is a scarcity of studies that compare and quantify the relative importance of different groups of factors in determining the variation of SARS-CoV-2 transmission. Moreover, since the epidemic trajectories differed by countries, it is important to understand the differences in the association between mobility and transmissibility of SARS-CoV-2.

In this study, we built a comprehensive statistical framework where the potential factors affecting COVID-19 transmission, as long as the data are accessible, are carefully accounted for. The spatiotemporal randomness in transmission, and the country-level random effects of population mobility are also included. The overall aim is to develop a statistical model in order to analytically determine and assess the relative contribution of climatic, demographic and disease control attributes to explain the variation of COVID-19 transmission across the world. The findings of this work could provide decision-makers new insights to coordinate global suppression efforts and resources against COVID-19.

2. Materials and Methods

2.1. Data Resource

We used a data set provided by Bhoopchand et al. (2020), which consolidates country-level COVID-19 data from multiple sources, and contains metadata on infection cases, tests, non-pharmaceutical interventions, mobility statistics, weather patterns, among many others, for up to 170 countries.

2.2. Data Processing

We removed the countries and days having missing values for the variables that we needed, a data set for 87 countries was eventually retained. This data set covers approximately the first half of 2020, and data on climatic variables are averaged across a list of representative major cities, a weighted average was then computed proportionally to population size associated with individual data points. Greater details can be found on Bhoopchand et al. (2020). The extent of transmission greatly varied across the study countries worldwide, as shown by the colorful dots on Figure 1. Our study area covers 87 countries across six continents excluding Antarctica. European countries have a good amount of data, so that most of them were retained in our data set. The Chinese mainland is not included in the data set, since it uses a different standard to measure the population mobility.

We assembled a set of covariates that potentially explain variation in global COVID-19 transmission and classify them into three broad categories: climatic, demographic and disease-control related. Below is a full list of the covariates that we considered as candidate explanatory variables. Factor descriptions can be reviewed on https://rs-delve.github.io/data_software/global-dataset.html. Further details on the retained covariates data, including the correlation among the covariates, are given in the Support Information. The weekly average values were then calculated for all the variables and for each country, where the first week starts on the first Monday in 2020 (the weeks are defined as ISO week, and calculated with R package “aweek”). By taking such smoothing for climatic variables, we are aware that this might mask the true effects of climatic factors at finer scales, since the regional climate varies significantly within large countries like Australia and America. The number of confirmed cases is the outcome variable, and the number of tests is assumed to offset the effect of under-reporting of confirmed cases. The number of confirmed cases in a previous period is included as a covariate to compensate for the effects of any missing factors. In the raw data set, there are other non-pharmaceutical variables that can be classified into disease control measures, but they are statistically significant correlated with the variables we retained. For example, we remove relative humidity due to its high correlation ($\text{cor} \approx 0.86$) with temperature. We also removed intervention stringency index due to its high correlation ($\text{cor} \approx -0.72$) with population mobility, less model fit improvement than population mobility, and itself is an aggregation of other variables including debt relief, health investment and contact tracing. The variables we eventually considered in the data analysis include:

1. Response variable: number of confirmed cases
2. Climatic variables: temperature, wind speed, UV radiation;

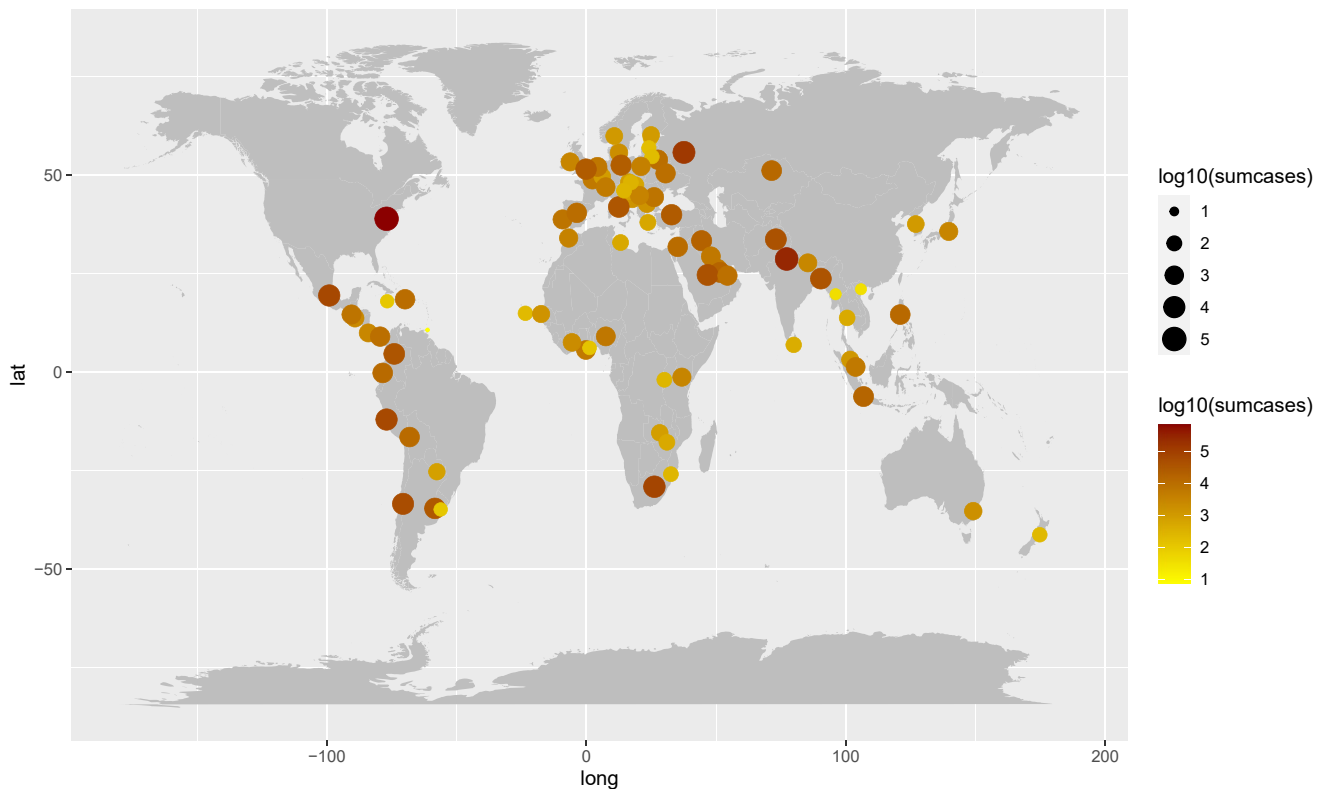


Figure 1. Mapped locations of all 87 countries examined. Dot color indicates the value of log-transformed accumulative number of cases within the study period (vary from country to country, from 6 to 21 weeks). A base 10 log-transformation was conducted in order to better graphically visualize the countries with small number of cases. Darker and larger dots indicating more confirmed cases. Dots are centered on the capital of each country.

3. Demographic variables: population size, population density, population median age;
4. Disease control measures: number of COVID-19 tests, population mobility restriction (Google COVID-19 Community Mobility, data stream “retail and recreation”), debt relief (larger debt relief index means that government freeze more financial obligations, e.g., stopping loan repayments, preventing services like water from stopping, or banning evictions), health investment (short term spending on healthcare system, e.g., hospitals, masks, etc), contact tracing;
5. Temporal variables: number of days since the first case confirmed, number of confirmed cases in previous week.

To prevent the effect of one factor dominating the others, and our interest is the relative effects instead of absolute effects, therefore, all covariates are standardized to be on the same scale (have a mean of 0 and a standard deviation of 1). In addition, it is worth mentioning that some factors might have lagged effect on confirmed cases. Therefore, we also considered 1 week and 2 weeks lagged population mobility, contact tracing, debt relief, health investment, as well as lagged climatic factors in the model. The response variable, weekly average number of cases, is base 10 logarithmic transformed to be closer to Gaussian distribution. Gaussian distribution assumption is made for the response variable instead of Poisson or Negative-binomial distribution (usually used for count numbers) due to the high auto-correlation, low mean and high variance in response. In addition, we found that the geometric shape of posterior likelihood with Poisson and Neg-binomial are extremely challenging for sampling probably due to light tails and thus potentially leads to an incomplete exploration of the target distribution and model convergence problems.

2.3. Data Analysis

A mixed-effect statistical model is constructed in this study. In addition to the fixed effects mentioned in Section 2.2, we also add country-level random intercepts and random slopes on the number of days since the

first case confirmed, and on the quadratic number of days, to account for the temporal correlation between observations in a given country and capture the various curvilinear patterns in transmission trajectories among the countries (see Figure S3 in Supporting Information S1). Random effects of population mobility are also included to see if country-specific effect of population mobility on COVID-19 transmission exists. Temperature can also be assumed to random effect, but it is correlated with time (number of days since first confirmed case) and contribute less in model fit improvement than the latter, therefore, it is assumed to be fixed effect instead. Besides, in order to account for time-varying volatility and temporal auto-correlation in the response variable, three assumptions on the types of errors are tested: (a) Gaussian i.i.d distributed; (b) autoregressive conditional heteroscedasticity process with one lag (ARCH(1) (Bollerslev, 1986); (c) autoregressive conditional heteroscedasticity process with two lags (ARCH(2)). Let ϵ_t denotes the error terms at time point t in a given country, and σ_t is the corresponding standard deviation. In a ARCH(q) model, the series σ_t^2 is modeled by

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_q \epsilon_{t-q}^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2, \quad (1)$$

where parameter $\alpha_0 > 0$, $\alpha_i \geq 0$, $i \geq 0$, and q is the length of ARCH lags. When the parameters α_1 to α_q are set to 0, the errors reduce to be Gaussian distributed with variance α_0 .

The statistical model formulated above (full model) is one of the models we implemented with the data set. There is no R package ready to be used for fitting such a model with such a complex structure (mixed effects together with ARCH errors), we instead define the model with R package TMB (Kristensen et al., 2016), which was developed for fitting linear, non-linear and complex statistical latent variable models. The model is formulated in C++ to have greater flexibility. We used marginal AIC (mAIC) based on information theory to assess the data support for the candidate models. Following Cao et al. (2019), we constructed the candidate models by adding the covariates to the model one by one, retaining the covariate if mAIC value was lowered, otherwise the covariate was removed from the model. The fixed effects entered into the model before random effects, and either lagged factors or non-lagged factors were kept in the model, depending on how much they improve model fit. The best model with the lowest mAIC value was selected through two-way mAIC comparison. To avoid the impact of the order that the variables entered into the model, we adjusted the selected model by adding or removing variables from it. By checking the model fit of such neighbor models (Table S1 in Supporting Information S1), we are certain that the selected model was the best one among the candidate models we have tested. The estimates of the parameters in each model, together with their standard errors were provided by the package.

We did several sets of prespecified secondary analyses to gain more insights into the results. For comparison purpose, we also used backward elimination, as an alternative method to conduct model selection. It turned out that the selected model and parameter estimates by the two methods are close. The details are given in Supporting Information. Additionally, we also fitted a model where the outcome variable is the daily growth rate of confirmed cases, as did in Kronfeld-Schor et al. (2021), but no statistically significant factors were detected except for population mobility (see Supporting Information S1 for details).

3. Results

Compared with the saturated model with a full set of covariates, the selected model removed humidity, wind speed, GDP, population density, and healthcare investment from the saturated model, since they did not improve the model fit (see model N3, model N4, N7, N15 and model N16, N17, N18 in Table S1 in Supporting Information S1). Furthermore, all of the random effects turned out to improve the model fit, and the model reports Gaussian i.i.d errors. The selected mixed-effect model is given by:

$$\begin{aligned} \logcases_i = & \beta_0 + \mu_{j(i)}^{(0)} + \beta_{temp}temp_i + \beta_{uv}uv_i + \beta_{populations}populations_i \\ & + \beta_{mage}mage_i + \beta_{newtests}newtests_i + \left(\beta_{mobility} + \mu_{j(i)}^{(mobility)} \right) mobility_i \\ & + \beta_{contatracing}contatracing_i + \beta_{debrelief}debrelief_i + \beta_{pweekcases}pweekcases_i \\ & + \left(\beta_{days} + \mu_{j(i)}^{(days)} \right) days_i + \mu_{j(i)}^{(sqdays)} sqdays_i + \epsilon_i, \end{aligned} \quad (2)$$

Table 1
Estimates of the Parameters in the Selected Model

Parameter	Estimate \pm s.d.	Description
$\hat{\beta}_0$	2.1 \pm 0.075	Fixed intercept
$\hat{\beta}_{temp}$	-0.058 \pm 0.028	Effect of temperature
$\hat{\beta}_{uv}$	0.047 \pm 0.013	Effect of UV radiation
$\hat{\beta}_{populations}$	0.204 \pm 0.081	Effect of population size
$\hat{\beta}_{mage}$	-0.127 \pm 0.076	Effect of population medium age
$\hat{\beta}_{newtests}$	0.125 \pm 0.058	Effect of number of new tests
$\hat{\beta}_{mobility}$	-0.128 \pm 0.015	Effect of population mobility
$\hat{\beta}_{contatracing}$	-0.051 \pm 0.02	Effect of contact tracing
$\hat{\beta}_{debtrelief}$	0.029 \pm 0.018	Effect of debt relief
$\hat{\beta}_{pweekcases}$	0.232 \pm 0.017	Effect of no. cases of previous week
$\hat{\beta}_{days}$	0.296 \pm 0.042	Fixed temporal effect
$\hat{\sigma}_{intercept}^2$	0.47 \pm 0.077	Variance of country-level random intercepts
$\hat{\sigma}_{days}^2$	0.11 \pm 0.021	Variance of country-level random slopes on days (number of days since first case confirmed)
$\hat{\sigma}_{days_sq}^2$	0.043 \pm 0.008	Variance of country-level random slopes on quadratic days
$\hat{\sigma}_{mobility}^2$	0.034 \pm 0.01	Variance of country-level random slopes on population mobility
$\hat{\sigma}_{noise}^2$	0.035 \pm 0.0015	Variance of Gaussian noise

where i ($i = 1, 2, \dots, 1476$) indicates all of the data points in our analyzed data set and j is the index for the country, $j = 1, 2, \dots, 87$. In this model, random intercept, random slope of days, quadratic days and mobility for country j are denoted by $\mu_{j(i)}^{(0)}$, $\mu_{j(i)}^{(days)}$, $\mu_{j(i)}^{(sqdays)}$, $\mu_{j(i)}^{(mobility)}$ and they are assumed to be normal distributed with variance being $\sigma_{intercept}^2$, σ_{days}^2 , $\sigma_{days_sq}^2$, $\sigma_{mobility}^2$ respectively. The meaning of other parameters can be seen in Table 1.

Table 1 gives a summary of the estimates of parameters retained in the selected model. The effects of the covariates could be compared through the parameter estimates since all of the covariates are standardized. The positive estimates of β_{UV} (0.047 \pm 0.013), $\beta_{population}$ (0.204 \pm 0.081), $\beta_{newtests}$ (0.125 \pm 0.058) and $\beta_{debtrelief}$ (0.029 \pm 0.018) suggest positive association between the number of confirmed cases and UV level, population size, number of new tests and debt relief. The negative estimated effects of temperature (-0.058 \pm 0.028), median age of population (-0.127 \pm 0.076), population mobility (-0.128 \pm 0.015) and contact tracing (-0.051 \pm 0.02) imply negative associations with weekly reported number of cases. Lagged effects are not detected for the covariates (model N1, N2, N10, N11, N13, N14, and N16, N17 in Supporting Information report worse model fit).

As mentioned in the last section, curvilinear temporal patterns of transmission dynamics are observed in most of the study countries (see the blue curves in Figure S3 in Supporting Information S1, implying that a quadratic coefficient of days should be considered). In addition, the curvilinear patterns in the study countries show significant differences, so that country-level random slopes on the quadratic days are included. Figure 2b displays the estimated random effects on quadratic days for each study country. The positive effect implies convex curve in transmission dynamics, for example, the log-transformed weekly confirmed cases in Luxembourg (the upper right plot, black dots are observations, blue line is the estimated value and gray band is the corresponding 95% confidence interval). In contrast, the negative effect suggests concave pattern, as shown for Belarus in the lower right corner. Thus, the estimated random temporal effects allow us to describe the country-specific transmission patterns without looking at the data. The heterogeneity in the transmission patterns (up-down or down-up, e.g.,) likely reflects differences in the interventions each country has implemented.

The selected model also reports a significant variance of random slopes on population mobility ($\hat{\sigma}_{mobility}^2$ equals to 0.034 \pm 0.01). By adding the estimated fixed slope of mobility to the estimated random slope for each country respectively, we can see from Figure 2a that, most of the study countries report statistically non-significant

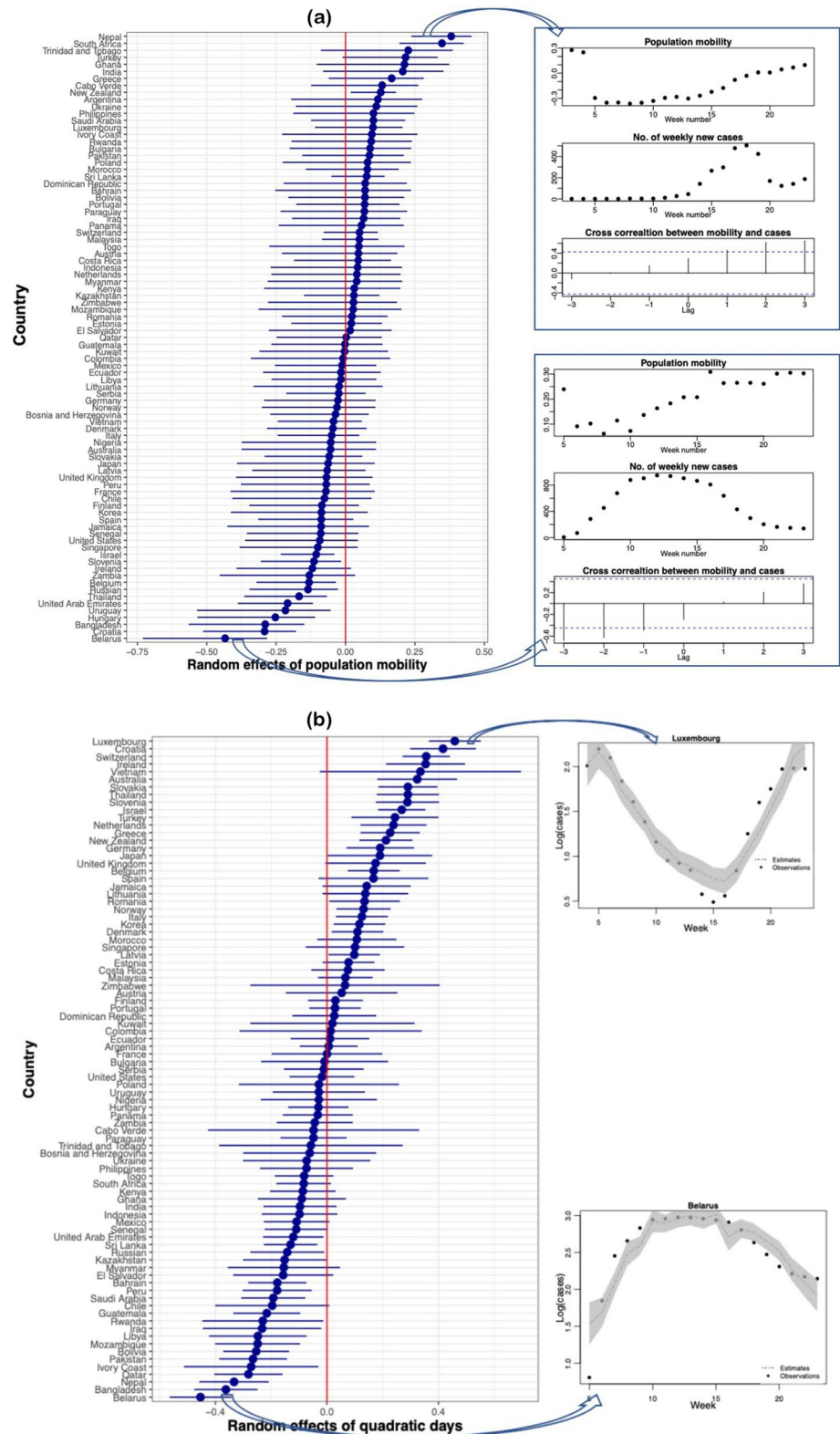


Figure 2. Plot (a) shows the estimated random slopes on population mobility for each study country. The two boxes on the right present the data on population mobility, number of weekly cases and cross correlation between mobility and cases for Nepal and Belarus respectively. Plot (b) displays the estimated random slopes on quadratic days (the number of days since first confirmed case) for each study country. The two boxes on the right indicate the log-transformed weekly number of cases for Luxembourg and Belarus respectively.

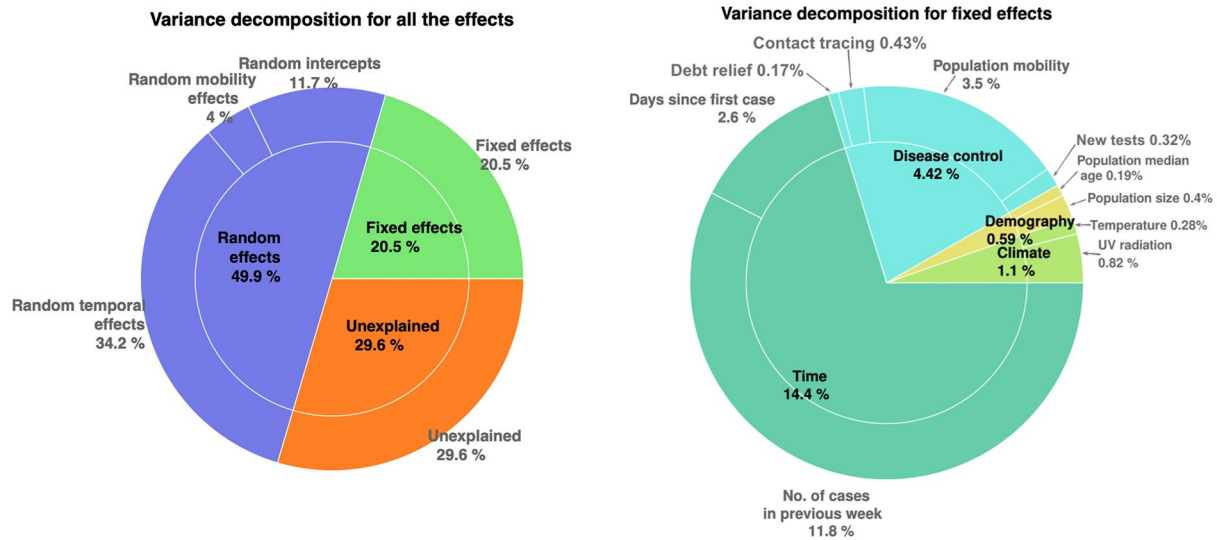


Figure 3. The left pie chart denotes collective proportion of variance explained by all the components (fixed effects, random effects, and unexplained variance). The right pie chart zooms out the variance explained by each group of factors (climate, demography and disease control measures), together with individual covariate in each group.

effects, indicating by the horizontal bars (95% confidence interval) for the 87 countries. Only Nepal and South Africa show a statistically positive effect of population mobility on transmission, whereas Belarus, Croatia, and a few other countries report negative effects of mobility. We also present the data on population mobility, number of weekly cases and cross correlation between mobility and cases for Nepal and Belarus respectively, as can be seen on the right of Figure 2a, to convey an intuition that why opposite effects of mobility are reported for different countries. Such difference likely reflects the heterogeneities in the relative reductions in mobility required to achieve COVID-19 control across the countries.

The pie charts in Figure 3 graphically display the proportion of explained variance by each component in the selected model. The left pie chart shows collective proportions of explained variance from all the components in the model, and the right pie chart focus on the fixed effects in the selected model. The explained variation by each component in the selected model is estimated with the third measure introduced in Xu (2003), which is based on the Kullback-Leibler information gain and proposed by Kent (1983). More details on variance decomposition method used in this study can be found in Supporting Information.

The blue part on the left pie chart indicates that the random effects make largest contribution to accounting for the response variance (49.9%), fixed effects contribute 20.5% and unexplained variance 29.6%. The large proportion of variance accounted for by random effects reflects the apparently random trajectories of infection dynamics across countries. The random temporal effects (aggregation of random effects of days and random effects of quadratic days) can be understood as each country having its own temporal transmission pattern and the proportion of variance in the response variable accounted for by random temporal effects is as large as 34.2%. We have seen in the last section that some countries report a negative relationship between mobility and infection, other countries (e.g., Nepal, South Africa) report a positive association, such variation accounts for 4% of the variance in global transmission. Among the fixed effects factors shown on the right chart in Figure 3, climate variables only explain 1.1% of the variance in global transmission of COVID19, followed by demographic variables (0.59%). In contrast, disease control measures explain much higher proportion of the variance (4.42%). It is not surprising that the number of confirmed cases in the previous week is the most important fixed effect factor (explain 11.8% of the variance), followed by population mobility (4%) and the number of days since first confirmed case (2.6%). Temperature and population median age are the least important factors, only explains 0.28% and 0.19% of the variance in global transmission respectively.

4. Discussion

As of 29 December 2021, over eight billion vaccine doses have been administered in the world (Ritchie et al., 2020), but a lot of countries, especially the most European countries, are still experiencing flare-ups in COVID-19 infections. As we have shown in this study, half of the variance in global transmission across the countries can only be accounted for by the spatiotemporal heterogeneity in transmission, making prediction of virus transmission patterns across the world in short term challenging. In the early stage of the pandemic, the transmissions shows no regional characteristics. The down-up trend in transmission might be spotted in any country, regardless of geographic location, including northern-hemisphere countries such as Luxembourg, Switzerland, or southern-hemispheric countries, like Australia and New Zealand. The substantial spatiotemporal stochasticity and variability across countries in transmission patterns are characterized and quantified by the random country effects in our study, turned to account for half of the global variation in transmission.

In this study, we highlight the relative importance of three groups of factor associated with COVID-19 transmission, in descending order, that are disease control measures (population mobility restrictions, contact tracing, new tests and debt relief), climatic factors (temperature and UV radiation) and demographic factors (population median age and population size). We also reported positive associations between weekly confirmed infection cases with UV radiation, population size, number of new tests and debt relief, and negative associations with temperature, population median age, population mobility and contact tracing. The overall findings of our study are qualitatively in line with the finding in prior studies investigating the impact of various factors that influence COVID-19 transmission, as summarized in Supporting Information, although most of these studies did not provide effect sizes associated with COVID-19 transmission that allow direct comparison. One exception is UV radiation, which has been shown to affect COVID-19 transmission though viral inactivation in lab study (Ratnesar-Shumate et al., 2020) and been reported to negatively associated with transmission (Carleton et al., 2021; Metelmann et al., 2021). While our inconsistent finding (positive association) is likely the result of the fact that most of the study countries in the Northern Hemisphere saw an increase in weekly confirmed cases when approaching summer, and UV rays get stronger meanwhile. This leads to a positive association between infection cases and UV radiation. In contrast, even though temperature is positively correlated with UV radiation ($\text{cor} \approx 0.51$), the estimated effect of temperature turned to be negative. The potential causal pathway between debt relief to transmission might be that lower financial obligations (e.g., banning evictions) lead to more social activities and contacts. Furthermore, even though the role of weather conditions, especially temperature and humidity, are recognised to affect the virus activation in lab studies (Morris et al., 2021; Riddell et al., 2020, e.g.) and in modulating the disease transmission in the real world (Mecenas et al., 2020; Majumder & Ray, 2021, e.g.), our study has shown that these environmental drivers were not able to curb transmission when the transmissibility is high and immunity is low in the early stage of COVID-19 pandemic. We found that non-pharmaceutical interventions have a stronger impact on transmission over space and time than any environmental driver. Identifying seasonality in COVID-19 infections, whether induced by environmental and physiological factors, requires careful investigation on the relative importance of drivers (Kronfeld-Schor et al., 2021) and will be more evident in subsequent years (Smit et al., 2020).

Our study also highlighted country-specific association between weekly number of cases and population mobility. Interestingly, in the majority of countries with statistically significant effects of mobility, we found clear evidence of a negative relationship between transmission and mobility, this likely suggests that the reduction in mobility have been insufficient to hold transmission. We also see that the control strategies were insufficient to reduce mobility so that both mobility and infections were growing in countries such as Nepal and South Africa during the study period. The different effects of mobility might be the result of the substantial heterogeneity between countries of mobility thresholds to interrupt transmission, which might reflect socio-cultural differences and/or different response to the intervention measures that each country implemented (Nouvellet et al., 2021). In addition, the country-level random effects that we included in the analysis not only account for a large portion of global transmission variation, but also enable us to straightforwardly depict the country-specific transmission trajectory characteristics (up and down, or down and up, e.g.).

Our study had several strengths. First, we assembled a long list of factors that might be associated with COVID-19 transmission, and the effects of missing variables are compensated by the number of confirmed cases in the previous period. Second, the integrated country-level random effects not only characterized the country-specific transmission trajectory, but also accounted for half portion of global variation in transmission, leaving only a

small portion (29.6%) of variation unexplained. Third, by pooling data from 87 countries across a period of more than 6 months, we were able to assess the association between each factor and COVID-19 transmission with substantial statistical power, and explore the global spatiotemporal heterogeneity in transmission simultaneously, which has been rarely done by previous studies. Finally, we incorporated two secondary analyses (one is using an alternative model selection method, another is using daily growth rate of confirmed cases as outcome variable) to reflect the statistical uncertainty around the selected model and parameter estimates, and gain more insights into the results.

However, our study also must balance between appropriate spatiotemporal resolution (Zeka et al., 2020) and data availability. Due to limited data availability, the geographic resolution in this study is at country-level instead of finer scales. For example, data on the number of daily new tests are only easily accessible at country-level, but it turned out to be an important covariate according to our preliminary analysis, and a similar situation applies to other factors, such as population median age. We are aware that this may lead to the result less reliable. We hope that the data analysis could be done when city-level data on all the variables are accessible without too much efforts for researchers. Unlike other studies where the region-specific R values were calculated for a time window and used the estimated R values as the outcome variable, we used weekly number of confirmed cases as the response. Even though this allows us to depict the temporal trend of transmission, meanwhile, we neglect the biases in the data due to the potential time lag between infection confirmation and data recording. Besides, this study analyzes six months of case data during the early stage of the epidemic with quality varying between countries, therefore, the interpretation of the estimated effect of climatic factors has to be conservative since the seasonality of COVID-19 transmission can only be accurately detected when longer time series data are available.

To date, we have observed that a vaccination programme, even though with a gradual increase in coverage to over 70% in some countries, could not fully contain resurgence, creating a chance for new SARS-CoV-2 variants to emerge. Additionally, the interaction between SARS-CoV-2 and other seasonal respiratory viruses is rarely understood, and seasonal characteristics could not effectively curb transmission in summer. All of these implied the important role of mobility restrictions and physical distancing in this pandemic combat, but governments must remain vigilant and evidence-informed in the effectiveness of mobility restrictions. Careful evaluation of mobility-transmission association should be ongoing in order to modulate prevention policies accordingly. Other disease interventions measures, including contact tracing, new tests, masks wearing, are also proved to be effective in holding transmission and should remain in force regardless of the weather or vaccination coverage. As the virus continues to evolve, it will be necessary to repeat our analysis using longer time series data on finer geographical scales from genomic surveillance of SARS-CoV-2 to understand the effects of various factors on virus transmission after vaccination.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

Source data for all of the variables used in the study can be accessed at https://github.com/rs-delve/covid19_datasets/tree/master/dataset. It is an openly licensed research-ready data set. All of the relevant information and details on the data set are available at https://rs-delve.github.io/data_software/global-dataset.html (Bhoopchand et al., 2020). The statistical model in this study is defined using C++ with R package TMB version 1.7.15 (Kristensen et al., 2016), available under the TMB license at <https://CRAN.R-project.org/package=TMB>. It was developed for fitting linear, non-linear and complex statistical latent variable models. Other analyses are made with software R version 3.6.0 (R Core Team, 2019), licenses and other info can be found at <https://www.R-project.org>. YC and JW have verified the data linkage of the publicly available data set and all the authors had full access to the linked data in this study. All the necessary data and source code to carry out the analyses in this study are available in the Github digital repository <https://github.com/shelly77/COVID19-global-analysis>.

Acknowledgments

We thank Dr. Boris Valentijn Schmid from CEES at University of Oslo for active discussions on issues including model selection and model evaluation techniques, and for his help with draft editing. We also thank Professor Ottar N. Bjornstad from Pennsylvania State University for comments that greatly improved the manuscript. This work was supported by the Research Council of Norway (RCN) with RCN project number 312740 and UiO project number 145021.

References

- Adhikari, A., Ghosh, S., Sen, M. M., & Adhikari, R. (2020). New models of transmission of covid-19 with time under the influence of meteorological determinants. *medRxiv*. <https://doi.org/10.1101/2020.05.26.20113985>
- Alam, M. S., & Sultana, R. (2021). Influences of climatic and non-climatic factors on covid-19 outbreak: A review of existing literature. *Environmental Challenges*, 5, 100255. <https://doi.org/10.1016/j.envc.2021.100255>
- Arsalan, M., Mubin, O., Alnajjar, F., & Alsinglawi, B. (2020). Covid-19 global risk: Expectation vs. reality. *International Journal of Environmental Research and Public Health*, 17(15), 5592. <https://doi.org/10.3390/ijerph17155592>
- Baker, R. E., Yang, W., Vecchi, G. A., Metcalf, C. J. E., & Grenfell, B. T. (2020). Susceptible supply limits the role of climate in the early sars-cov-2 pandemic. *Science*, 369(6501), 315–319. <https://doi.org/10.1126/science.abc2535>
- Bhoopchand, A., Paleyes, A., Donkers, K., Tomasev, N., & Paquet, U. (2020). Delve global covid-19 dataset [dataset]. The Royal Society. Retrieved from https://rs-delve.github.io/data_software/global-dataset.html
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- Cao, Y., Visser, M. E., & Tufto, J. (2019). A time-series model for estimating temporal variation in phenotypic selection on laying dates in a Dutch great tit population. *Methods in Ecology and Evolution*, 10(9), 1401–1411. <https://doi.org/10.1111/2041-210X.13249>
- Carleton, T., Cornet, J., Huybers, P., Meng, K. C., & Proctor, J. (2021). Global evidence for ultraviolet radiation decreasing covid-19 growth rates. *Proceedings of the National Academy of Sciences*, 118(1), e2012370118. <https://doi.org/10.1073/pnas.2012370118>
- Carteni, A., Di Francesco, L., & Martino, M. (2020). How mobility habits influenced the spread of the covid-19 pandemic: Results from the Italian case study. *Science of the Total Environment*, 741, 140489. <https://doi.org/10.1016/j.scitotenv.2020.140489>
- Fu, S., Wang, B., Zhou, J., Xu, X., Liu, J., Ma, Y., et al. (2020). Meteorological factors, governmental responses and covid-19: Evidence from four European countries. *Environmental Research*, 194, 110596. <https://doi.org/10.1016/j.envres.2020.110596>
- Islam, N., Bukhari, Q., Jameel, Y., Shabnam, S., Erzurumluoglu, A. M., Siddique, M. A., et al. (2020). Covid-19 and climatic factors: A global analysis. *Environmental Research*, 193, 110355. <https://doi.org/10.1016/j.envres.2020.110355>
- Karapiperis, C., Kouklis, P., Papastratos, S., Chasapi, A., Danchin, A., & Ouzounis, C. A. (2020). Preliminary evidence for seasonality of covid-19 due to ultraviolet radiation. *F1000Research*, 9(658), 658. <https://doi.org/10.12688/f1000research.24774.1>
- Kent, J. T. (1983). Information gain and a general measure of correlation. *Biometrika*, 70(1), 163–173. <https://doi.org/10.1093/biomet/70.1.163>
- Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. J., & Bell, B. (2016). Tmb: Automatic differentiation and laplace approximation [software]. *Journal of Statistical Software*, 70(5), 1–21. <https://doi.org/10.48550/arXiv.1509.00660>
- Kronfeld-Schor, N., Stevenson, T., Nickbakhsh, S., Schernhammer, E., Dopico, X., Dayan, T., et al. (2021). Drivers of infectious disease seasonality: Potential implications for covid-19. *Journal of Biological Rhythms*, 36(1), 35–54. <https://doi.org/10.1177/0748730420987322>
- Li, Y., Wang, X., Campbell, H., Nair, H., & for COVID, U. N. (2021). The association of community mobility with the time-varying reproduction number (r) of sars-cov-2: A modelling study across 330 local UK authorities. *The Lancet Digital Health*, 3(10), e676–e683. [https://doi.org/10.1016/S2589-7500\(21\)00144-8](https://doi.org/10.1016/S2589-7500(21)00144-8)
- Majumder, P., & Ray, P. P. (2021). A systematic review and meta-analysis on correlation of weather with covid-19. *Scientific Reports*, 11(1), 1–10. <https://doi.org/10.1038/s41598-021-90300-9>
- McClymont, H., & Hu, W. (2021). Weather variability and covid-19 transmission: A review of recent research. *International Journal of Environmental Research and Public Health*, 18(2), 396. <https://doi.org/10.3390/ijerph18020396>
- Mecenas, P., Bastos, R. T. D. R. M., Vallinoto, A. C. R., & Normando, D. (2020). Effects of temperature and humidity on the spread of covid-19: A systematic review. *PLoS One*, 15(9), e0238339. <https://doi.org/10.1371/journal.pone.0238339>
- Metelmann, S., Pattni, K., Brierley, L., Cavalerie, L., Caminade, C., Blagrove, M. S., et al. (2021). Impact of climatic, demographic and disease control factors on the transmission dynamics of covid-19 in large cities worldwide. *One Health*, 12, 100221. <https://doi.org/10.1016/j.onehlt.2021.100221>
- Monod, M., Blenkinsop, A., Xi, X., Hebert, D., Bershan, S., Tietze, S., et al. (2021). Age groups that sustain resurging covid-19 epidemics in the United States. *Science*, 371(6536), eabe8372. <https://doi.org/10.1126/science.abe8372>
- Morris, D. H., Yinda, K. C., Gamble, A., Rossine, F. W., Huang, Q., Bushmaker, T., et al. (2021). Mechanistic theory predicts the effects of temperature and humidity on inactivation of sars-cov-2 and other enveloped viruses. *ELife*, 10, e65902. <https://doi.org/10.7554/eLife.65902>
- Nouvellet, P., Bhatia, S., Cori, A., Ainslie, K. E., Baguelin, M., Bhatt, S., et al. (2021). Reduction in mobility and covid-19 transmission. *Nature Communications*, 12(1), 1–9. <https://doi.org/10.1038/s41467-021-21358-2>
- Ratnesar-Shumate, S., Williams, G., Green, B., Krause, M., Holland, B., Wood, S., et al. (2020). Simulated sunlight rapidly inactivates sars-cov-2 on surfaces. *The Journal of infectious diseases*, 222(2), 214–222. <https://doi.org/10.1093/infdis/jiaa274>
- R Core Team. (2019). R: A language and environment for statistical computing [Software]. Retrieved from <https://www.R-project.org>
- Riddell, S., Goldie, S., Hill, A., Eagles, D., & Drew, T. W. (2020). The effect of temperature on persistence of sars-cov-2 on common surfaces. *Virology Journal*, 17(1), 1–7. <https://doi.org/10.1186/s12985-020-01418-7>
- Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., et al. (2020). Coronavirus pandemic (covid-19). *Our world in data*. Retrieved from <https://ourworldindata.org/coronavirus>
- Sarmadi, M., Marufi, N., & Moghaddam, V. K. (2020). Association of covid-19 global distribution and environmental and demographic factors: An updated three-month study. *Environmental Research*, 188, 109748. <https://doi.org/10.1016/j.envres.2020.109748>
- Scudellari, M. (2020). How the pandemic might play out in 2021 and beyond. *Nature*, 584(7819), 22–25. <https://doi.org/10.1038/d41586-020-02278-5>
- Sera, F., Armstrong, B., Abbott, S., Meakin, S., O'Reilly, K., von Borries, R., et al. (2021). A cross-sectional analysis of meteorological factors and sars-cov-2 transmission in 409 cities across 26 countries. *Nature Communications*, 12(1), 1–11. <https://doi.org/10.1038/s41467-021-25914-8>
- Smit, A. J., Fitchett, J. M., Engelbrecht, F. A., Scholes, R. J., Dzihvhuho, G., & Sweijd, N. A. (2020). Winter is coming: A southern hemisphere perspective of the environmental drivers of sars-cov-2 and the potential seasonality of covid-19. *International Journal of Environmental Research and Public Health*, 17(16), 5634. <https://doi.org/10.3390/ijerph17165634>
- Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine*, 22(22), 3527–3541. <https://doi.org/10.1002/sim.1572>
- Zeka, A., Tobias, A., Leonardi, G., Bianchi, F., Lauriola, P., Crabbe, H., et al. (2020). Responding to covid-19 requires strong epidemiological evidence of environmental and societal determining factors. *The Lancet Planetary Health*, 4(9), e375–e376. [https://doi.org/10.1016/S2542-5196\(20\)30169-8](https://doi.org/10.1016/S2542-5196(20)30169-8)

Zheng, P., Chen, Z., Liu, Y., Song, H., Wu, C.-H., Li, B., et al. (2021). Association between coronavirus disease 2019 (covid-19) and long-term exposure to air pollution: Evidence from the first epidemic wave in China. *Environmental Pollution*, 276, 116682. <https://doi.org/10.1016/j.envpol.2021.116682>

References From the Supporting Information

- Buja, A., Paganini, M., Cocchio, S., Scioni, M., Rebba, V., & Baldo, V. (2020). Demographic and socio-economic factors, and healthcare resource indicators associated with the rapid spread of covid-19 in northern Italy: An ecological study. *PLoS One*, 15(12), e0244535. <https://doi.org/10.1371/journal.pone.0244535>
- Carlson, C. J., Gomez, A. C., Bansal, S., & Ryan, S. J. (2020). Misconceptions about weather and seasonality must not misguide covid-19 response. *Nature Communications*, 11(1), 1–4. <https://doi.org/10.1038/s41467-020-18150-z>
- Chang, S. L., Harding, N., Zachreson, C., Cliff, O. M., & Prokopenko, M. (2020). Modelling transmission and control of the covid-19 pandemic in Australia. *Nature Communications*, 11(1), 1–13. <https://doi.org/10.1038/s41467-020-19393-6>
- Gupta, D. (2020). Effect of ambient temperature on covid-19 infection rate. Available at SSRN 3558470. <https://doi.org/10.2139/ssrn.3558470>
- Kampf, G., Todt, D., Pfaender, S., & Steinmann, E. (2020). Persistence of coronaviruses on inanimate surfaces and their inactivation with biocidal agents. *Journal of Hospital Infection*, 104(3), 246–251. <https://doi.org/10.1016/j.jhin.2020.01.022>
- Jüni, P., Rothenbühler, M., Bobos, P., Thorpe, K. E., da Costa, B. R., Fisman, D. N., et al. (2020). Impact of climate and public health interventions on the covid-19 pandemic: A prospective cohort study. *Canadian Medical Association Journal*, 192(21), E566–E573. <https://doi.org/10.1503/cmaj.200920>
- Kretzschmar, M. E., Rozhnova, G., Bootsma, M. C., van Boven, M., van de Wiggert, J. H., & Bonten, M. J. (2020). Impact of delays on effectiveness of contact tracing strategies for covid-19: A modelling study. *The Lancet Public Health*, 5(8), e452–e459. [https://doi.org/10.1016/S2468-2667\(20\)30157-2](https://doi.org/10.1016/S2468-2667(20)30157-2)
- Mandal, M., Jana, S., Nandi, S. K., Khatua, A., Adak, S., & Kar, T. (2020). A model based study on the dynamics of covid-19: Prediction and control. *Chaos, Solitons & Fractals*, 136, 109889. <https://doi.org/10.1016/j.chaos.2020.109889>
- Meyer, A., Sadler, R., Faverjon, C., Cameron, A. R., & Bannister-Tyrrell, M. (2020). Evidence that higher temperatures are associated with a marginally lower incidence of covid-19 cases. *Frontiers in Public Health*, 8, 367. <https://doi.org/10.3389/fpubh.2020.00367>
- Müller, S., Scealy, J. L., & Welsh, A. H. (2013). Model selection in linear mixed models. *Statistical Science*, 28(2), 135–167. <https://doi.org/10.1214/12-STS410>
- Oliveiros, B., Caramelo, L., Ferreira, N. C., & Caramelo, F. (2020). Role of temperature and humidity in the modulation of the doubling time of covid-19 cases. *medRxiv*. <https://doi.org/10.1101/2020.03.05.20031872>
- Vaida, F., & Blanchard, S. (2005). Conditional akaike information for mixed-effects models. *Biometrika*, 92(2), 351–370. <https://doi.org/10.1093/biomet/92.2.351>
- Vandenberg, O., Martiny, D., Rochas, O., van Belkum, A., & Kozlakidis, Z. (2020). Considerations for diagnostic covid-19 tests. *Nature Reviews Microbiology*, 19(3), 1–13. <https://doi.org/10.1038/s41579-020-00461-z>
- Wang, J., Tang, K., Feng, K., Li, X., Lv, W., Chen, K., & Wang, F. (2020). High temperature and high humidity reduce the transmission of covid-19. arXiv preprint arXiv:2003.05003.