

Ranking the explanatory power of factors associated with worldwide new Covid-19 cases

Aristides Moustakas (✉ arismoustakas@gmail.com)

Natural History Museum of Crete <https://orcid.org/0000-0002-6334-747X>

Short Report

Keywords: Epidemiology and statistics, computational statistics, Covid-19, disease spread

Posted Date: July 19th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-702119/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Disease spread is a complex phenomenon requiring an interdisciplinary approach. Covid-19 exhibited a global spatial spread in a very short time frame resulting in a global pandemic. Data of new Covid-19 cases per million were analysed worldwide at the spatial scale of a country and time replicated from the end of December 2019 to late May 2020. Data driven analysis of epidemiological, economic, public health, and governmental intervention variables was performed in order to select the optimal variables in explaining new Covid-19 cases across all countries in time. Sequentially, hierarchical variance partitioning of the optimal variables was performed in order to quantify the independent contribution of each variable in the total variance of new Covid-19 cases per million. Results indicated that from the variables available new tests per thousand explained the vast majority of the total variance in new cases (51.6%) followed by the governmental stringency index (15.2%). Availability of hospital beds per 100k inhabitants explained 9% extreme poverty explained 8.8%, hand washing facilities 5.3%, the fraction of the population aged 65 or older explained 3.9%, and other disease prevalence (cardiovascular diseases plus diabetes) explained 2.9%. The percentage of smokers within the population explained 2.6% of the total variance, while population density explained 0.6%.

Introduction

Patterns of infectious diseases across spatial and temporal scales are fundamental for understanding their dynamics and for designing eradication strategies [1, 2]. Disease spread is a complex phenomenon and requires an interdisciplinary approach spanning from medicine to statistics and social sciences [3]. Covid-19 exhibited a global spatial spread in a relatively very short time frame resulting in being characterized as a pandemic by the World Health Organisation [4].

To date there is no known anti-viral treatment or vaccine against Covid-19 [5]. Therefore the available options against the virus are the characteristics of the immune system and health status of each individual, the health system of the country that the individual has access to, the social behaviour of the other individuals forming the society, public interventions of movement or public campaigns, and testing [6-8]. Thus, there are relatively few options to be employed in order to diminish the spread of the disease. This scarcity of options makes quantifying the factors associated with disease spread more important and ranking the relative contribution of each factor on Covid-19 spread may facilitate diminishing it [9].

Analysis so far has often been dominated either on a one-at-a-time factor analysis (e.g. new cases per testing effort or total cases per age structure) or on a country basis analysis (e.g doubling time in one country in comparison to other countries). While such comparisons are straightforward and comprehensive, from a statistical perspective they are examining one factor at a time and masking underlying characteristics within countries under a variable 'Country' [10]. This results in a hidden burden of the underlying factors regarding disease spread and causality is often discussed in a speculative manner. Admittedly the problem is complex due to the large differences across the potential underlying factors across countries. A small *variance* implies that the *mean* contains virtually all of the *information*,

while a large *variance* implies that *more information than the mean* is present [11]. When examining several countries together and across time, quantifying the variance may be more informative than the mean [12]. Variance is introduced both by space, there are large e.g. climatic differences between locations on the same date, as well as by temporal e.g. climatic or behavioural changes within locations in seasons [13-15].

The potential effect of economic, epidemiological and public health, or governmental interventions may become clearer when the contribution of these factors into new Covid-19 cases are analysed accounting for the fact that they derive from different countries as well as in different time snapshots [16] but in a way that the effect of each factor can be quantified in conjunction with the effects of other factors. To that end methods that can account for both spatial and temporal autocorrelation [17] in the data of new Covid-19 cases but can quantify the effect of each epidemiological, economic, public health, and governmental intervention are key to our understanding of how the disease spreads in populations worldwide [18, 19].

In this study spatio-temporal worldwide data of new Covid-19 cases from the “Our world in data” database [20] were analyzed using computational statistics. The spatial replicate of the dataset included over 150 countries time replicated for each country across a period of \approx five months. Data driven analysis was performed in order to quantify the optimal variables in cases where several candidate variables were available. During the data driven variable selection, the fact that data derived from different countries and were time replicated was accounted for by nesting the variance of time within the variance of country and treated as random effects [21, 22]. The percentage of the total variance explained by each epidemiological, economic, public health, and governmental intervention variables associated with Covid-19 new cases was quantified using hierarchical variance partitioning [23, 24] thereby ranking the relative contribution of the variance of each factor on new Covid-19 cases worldwide.

Methods

Data

The objective was to quantify epidemiological, public health, economic, and governmental intervention factors associated with Covid-19 spread worldwide. New Covid-19 cases per million per country per time step were used a proxy of disease spread. New cases per million in each country was chosen instead of new cases per country as this estimator is less biased by the total population of each country - countries with higher populations are more likely to produce higher new cases or total but the number may be relatively low in comparison to the total population pool. Data regarding new Covid-19 cases per million from the “Our world in data” database were analyzed. The dataset was last accessed on 25/05/2020 and the download location is <https://github.com/owid/covid-19-data/tree/master/public/data>. The data derive from the European Centre for Disease Prevention and Control (ECDC), an EU agency with the aim to strengthen Europe’s defense against infectious diseases. The ECDC collects and aggregates data from countries around the world. The most up-to-date data for any particular country is therefore typically

available earlier via the national health agencies than via the ECDC. This lag between nationally available data and the ECDC data is not very long as the ECDC publishes new data daily; typically this time lag is at the level of some hours and less than a day. The ECDC collects, compiles and harmonizes data from around the world in a consistent way which allows us to compare what is happening in different countries. The spatial replicate of the dataset comprised of 160 countries while the temporal replicate spans from 31/12/2019 to (including) 25/05/2020. The variables included:

Table 1. Description of the variables employed in the analysis and their source.

| Column | Description | Source |
|---------------------------------|--|--|
| iso_code | ISO 3166-1 alpha-3 - three-letter country codes | International Organization for Standardization |
| location | Geographical location (Country) | Our World in Data |
| date | Date of observation | Our World in Data |
| new_cases_per_million | New confirmed cases of COVID-19 | European Centre for Disease Prevention and Control |
| total_tests | Total tests for COVID-19 | National government reports |
| new_tests | New tests for COVID-19 | National government reports |
| new_tests_smoothed | New tests for COVID-19 (7-day smoothed). For countries that don't report testing data on a daily basis, we assume that testing changed equally on a daily basis over any periods in which no data was reported. This produces a complete series of daily figures, which is then averaged over a rolling 7-day window | National government reports |
| total_tests_per_thousand | Total tests for COVID-19 per 1,000 people | National government reports |
| new_tests_per_thousand | New tests for COVID-19 per 1,000 people | National government reports |
| new_tests_smoothed_per_thousand | New tests for COVID-19 (7-day smoothed) per 1,000 people | National government reports |
| tests_units | Units used by the location to report its testing data | National government reports |
| stringency_index | Government Response Stringency Index: composite measure based on response indicators including school closures, workplace closures, and | Oxford COVID-19 Government Response Tracker, |

| | | |
|--------------------|---|---|
| | national and international travel bans, canceling public events and exiting home rescaled to a value from 0 to 100 (100 = strictest response) | Blavatnik School of Government Reference:[25] |
| population | Population in 2020 | United Nations, Department of Economic and Social Affairs, Population Division, World Population Prospects: The 2019 Revision |
| population_density | Number of people divided by land area, measured in square kilometers, most recent year available | World Bank - World Development Indicators, sourced from Food and Agriculture Organization and World Bank estimates |
| median_age | Median age of the population, UN projection for 2020 | UN Population Division, World Population Prospects, 2017 Revision |
| aged_65_older | Share of the population that is 65 years and older, most recent year available | World Bank - World Development Indicators, based on age/sex distributions of United Nations Population Division's World Population Prospects: 2017 Revision |
| aged_70_older | Share of the population that is 70 years and older in 2015 | United Nations, Department of Economic and Social Affairs, Population Division (2017), World Population |

| | | |
|---------------------|---|--|
| | | Prospects: The 2017 Revision |
| gdp_per_capita | Gross domestic product at purchasing power parity (constant 2011 international dollars), most recent year available | World Bank - World Development Indicators, source from World Bank, International Comparison Program database |
| extreme_poverty | Share of the population living in extreme poverty, most recent year available since 2010 | World Bank - World Development Indicators, sourced from World Bank Development Research Group |
| cvd_death_rate | Death rate from cardiovascular disease in 2017 | Global Burden of Disease Collaborative Network, Global Burden of Disease Study 2017 Results |
| diabetes_prevalence | Diabetes prevalence (% of population aged 20 to 79) in 2017 | World Bank - World Development Indicators, sourced from International Diabetes Federation, Diabetes Atlas |
| female_smokers | Share of women who smoke, most recent year available | World Bank - World Development Indicators, sourced from World Health Organization, Global Health Observatory Data Repository |
| male_smokers | Share of men who smoke, most recent year available | World Bank - World Development Indicators, sourced from World Health Organization, |

| | | |
|------------------------|---|---|
| | | Global Health Observatory Data Repository |
| handwashing_facilities | Share of the population with basic handwashing facilities on premises, most recent year available | United Nations Statistics Division |
| hospital_beds_per_100k | Hospital beds per 100,000 people, most recent year available since 2010 | OECD, Eurostat, World Bank, national government records and other sources |

In particular regarding the governmental stringency index [25], the methodology is explained here:

<https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/codebook.md>

and here:

https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/index_methodology.md

From the available variables male and female smokers were averaged as 'smokers'.

Data analytics

We employed generalised linear mixed effects models (LME; [26]) with new Covid-19 cases per million as the dependent variable. As the dataset contained several potential indexes of testing, population density, or age structure within each country, initial analysis was conducted in order to select the most informative index of each.

We initially sought to quantify the most parsimonious data driven index of testing which included the fixed effects of (i) news tests (ii) total tests (iii) new tests per thousand (iv) total tests per thousand (v) new tests smoothed, (vi) new tests smoothed per thousand. This was achieved by fitting six LMEs with new cases per million as the dependent variable and six LMEs with i - vi as the single independent variable. The random effect structure of each LME included the nested variance of time within each country (Random~Country/Time). Doing so the fitted LMEs accounted for both temporal and spatial autocorrelation in the time replicated data deriving from different geographic locations [21, 22]. LMEs were fitted with Maximum Likelihood (ML) estimation to allow comparisons between models with different fixed effects and selecting the LME that exhibited the lowest Akaike (AIC) value [27, 28]. Here and throughout the analysis, there were 19,709 data points in the analysis but there were variables with missing values at some time steps or at some countries. Missing values were omitted from the statistical analysis. Therefore AIC values are compared between models fitted with different fixed effects but also with potentially different sample sizes.

Similarly, LMEs with new cases per million as the dependent variable and the fixed effects of (i) population or (ii) population density, and the nested random effects of time within country were fitted with ML and compared against AIC values to select the optimal data driven population index.

Regarding age structure of the population within each country, the available variables were (i) median age of the population, (ii) the percentage of the population aged 65 or older, and (iii) the percentage of the population aged 70 or older. The analysis proceeded by selecting the ML fitted LME with the lowest AIC between the three available age population structure variables. All three fitted LMEs contained the random effects of time nested within country.

Regarding economic status of the population within each country, the available variables were (i) gdp per capita, and (ii) percentage of the population under extreme poverty. The analysis proceeded by selecting the ML fitted LME with the lowest AIC between the two available economy status variables. The two fitted LMEs contained the random effects of time nested within country.

Having selected the optimal index of testing, population density and age structures the analysis proceeded with the following variables: (1) population density, (2) new tests per thousand, (3) governmental stringency index, (4) percentage of the population aged > 65, (5) percentage of the population under extreme poverty, (6) cvd death rate, (7) diabetes prevalence, (8) percentage of smokers, (9) percentage of the population with access to hand washing facilities, and (10) hospital beds per 100k inhabitants within each country as independent variables.

Hierarchical Variance Partitioning (HVP) statistical modelling was implemented to account for the contribution of each data driven epidemiological, economic, public health, and governmental intervention explanatory variable to the total variance of new Covid-19 per million cases [29, 30]. HVP is a statistical framework that is capable of handling correlated independent variables, whilst providing a reliable ranking of predictor importance of each variable [29]. Variance partitioning is calculated from the Akaike (AIC) weights of each explanatory variable and it is based upon the number of times that a variable was significant in all possible combinations of the explanatory variables. The HVP function produces a minor rounding error for hierarchies constructed from more than nine variables [31] - the available data driven variables were 10. To check if this error affects the inference from an analysis, the analysis was repeated several times with the variables entered in a different order [31]. The analysis resulted in changes in the derived results when the order of the variables was changed. The analysis proceeded by creating a new variable that merged together other disease related variables: other diseases variable= (cvd_death_rate + diabetes_prevalence) plus the other remaining eight variables resulting in a total of nine variables. There is no known statistical bias in HVP when 9 or fewer variables are used [31].

Results

The optimal data driven index for explaining new cases per million from the ones available regarding testing was new cases per thousand as fitted by LMEs and selected against the lowest AIC value (Table 2a). New tests per thousand smoothed could not be fitted as the LME did not converge (Table 2a). The

optimal data driven population index for new cases per million was population density (Table 2b). The optimal data driven index for age structure within the population was the percentage of the population within each country aged 65 or older (Table 2c). The optimal AIC selected LME regarding the economic status of the population within each country in relation to new Covid-19 cases per million was extreme poverty (Table 2d).

Table 2. Data driven variable selection in cases where several candidate variables were available. Variable selection was conducted by fitting LMEs with new Covid-19 cases per million as dependent variable and the candidate explanatory variables as single fixed effect variables and the variance introduced time nested within the variance of countries as random effects. Models were fitted with ML to allow comparisons between LMEs fitted with different fixed effects. The optimal model is bolded and italicized in each case.

Table 2a. Selecting the optimal variable for age structure

| Age variables | df | AIC |
|-----------------------------|----------|-----------------|
| median_age | 5 | 166614.9 |
| <i>aged_65_older</i> | 5 | 164586.0 |
| aged_70_older | 5 | 165940.0 |

Table 2b. Selecting the optimal variable for testing

| Testing index | df | AIC |
|--------------------------------------|----------|-----------------|
| total_tests | 5 | 49962.97 |
| new_tests | 5 | 44250.93 |
| total_tests_per_thousand | 5 | 49977.97 |
| <i>new_tests_per_thousand</i> | 5 | 43826.88 |
| new_tests_smoothed | 5 | 53489.32 |
| new_tests_smoothed_per_thousand | 5 | NC |

Table 2c. Selecting the optimal variable for population

| Population index | df | AIC |
|----------------------------------|----------|-----------------|
| Population | 5 | 211833.0 |
| <i>Population_density</i> | 5 | 187662.3 |

Table 2d. Selecting the optimal financial status variable

| Financial index | df | AIC |
|-------------------------------|----------|-----------------|
| GDP_per_capita | 5 | 172222.0 |
| <i>Extreme_poverty</i> | 5 | 104654.4 |

Results from HVP indicated that total tests per thousand explained 51.6% of the total variance of new cases per million, while governmental stringency index explained 15.2% (Fig. 1, Table 3). Availability of hospital bed per 100k inhabitants explained 9% (Fig. 1, Table 3). Extreme poverty explained 8.8% of the total variance of new cases per million, hand washing facilities 5.3%, the fraction of the population aged 65 or older explained 3.9%, other disease prevalence (cardiovascular diseases plus diabetes) explained 2.9% (Fig. 1, Table 3). The percentage of smokers within the population explained 2.6% of the total variance of new Covid-19 cases per million, while population density explained 0.6% (Fig. 1, Table 3).

Discussion

The best model fit regarding new Covid-19 cases per million and the economic status of the country where the new cases are recorded indicated that extreme poverty was a better predictor of new cases than gdp per capita. It is thus the poorest individuals within each country impacted rather than poor countries. From the data available, the fraction of the population aged 65 or older explained optimally new cases per million and not median population age. Total tests per thousand and not new tests or new tests smoothed or other available indexes is a better predictor of new cases per million, perhaps unsurprisingly as the number of new cases is already normalized by the population and thus the number of tests also normalized by the population explains better the pattern. The latter also applies for population density instead of total population as the best available predictor of new cases per million. Summing up, from the data-driven analysis it is evident that new Covid-19 cases per million are best explained by extreme poverty prevalence within each country as well as by the fraction of the population older than 65, thereby indicating association of the spread of the disease with the poor and older.

Results from variance partitioning of the data-driven selected 9 epidemiological, public health, economic, and governmental intervention variables explaining Covid-19 new cases per million across countries through time, indicated that the vast majority of new cases per million are explained by the number of tests conducted. The number of new tests per thousand explains over 50% of the total variance through time and countries and thus the message regarding the efficacy of testing against Covid-19 spread is strong, at least from the results derived here. The efficacy of testing has been highlighted as the best strategy against other diseases too across humans, agriculture, and wildlife [1, 24, 32, 33]. It therefore seems that the optimal strategies against Covid-19 spread should include high number of tests both to suspicious cases as well as random population testing.

Would increasing the number of tests result in detecting more Covid-19 cases? Lost Covid-19 cases are not uncommon [34, 35]. From a statistical perspective, variance partitioning does not provide information on the sign of the effects (positive or negative) it simply shows in how many cases this variable could not be excluded from the final optimal statistical model in explaining new Covid-19 cases. The slope between new tests per thousand inhabitants and new cases per million vary between countries (Fig. 2a). Indeed there are countries where the slope between new cases per million and new tests per thousand are positive indicating that testing more would actually result into identifying more cases (Fig. 2a); [32]. However, there are also countries with a negative slope between new tests – new cases' indicating that testing frequency is saturated (Fig. 2a). Overall, using data from all countries and time steps, the relationship between testing and new cases is positive indicating that worldwide more tests would result in identifying higher number of new cases (Fig. 2b). Therefore the efficacy of testing has not been saturated.

The second best variable in explaining new Covid-19 cases per million was the governmental stringency index. The governmental stringency index contains several measures taken by governments including school closures, national and international movement restrictions, public gathering and public events

restrictions, exiting home restrictions as well as testing policies and financial measures [25]. To that end testing policy is in part contained in the stringency index as a weighted percentage of the overall index. However the relationship between stringency index and new tests per thousand is very weak with both R^2 and slope close to zero: linear regression ($new_tests_per_thousand$) = $0.1959 + 0.002479$ ($stringency_index$), $S = 0.597$, $R^2 = 1.1\%$, $P < 0.001$. Therefore of the available governmental measures summarized in the stringency index, testing frequency can be treated independently.

In general, countries increased their level of stringency as their number of confirmed COVID-19 cases raised, however there is significant variation in the rate and timing of this relationship [25]. Another study indicated that in the early and accelerating stages of the pandemic, many citizens across 58 countries viewed their governments' response as insufficient [36]. In general the status of the infection spread and policy implementation influence restrictions uniformly across every countries [37]. Given the overall large effect of testing on new cases, it has been investigated whether there exists a testing frequency for Covid-19 such that the shutdown could have been avoided [38]. The study concluded that indeed there is an optimal testing frequency such that lockdown and thus governmental stringency may not be deemed necessary [38]. The test against Covid-19 is known to be imperfect but not precisely known [39] and testing strategies to surmount this problem have been proposed [40].

The availability of hospital beds per 100k inhabitants, hand washing facilities, the effect of Covid-19 in the older people as well as prevalence of other diseases and smokers have been highlighted [19, 41, 42] and this study confirms their importance. Environmental factors have also been reported to play an important role in Covid-19 dynamics [43] however this study did not explore their relative contribution.

Declarations

The author declares that no conflict of interest exists to the best of his knowledge.

References

1. Moustakas, A., et al., *Abrupt events and population synchrony in the dynamics of Bovine Tuberculosis*. Nature Communications, 2018. **9**(1): p. 2821.
2. Viboud, C., et al., *Synchrony, waves, and spatial hierarchies in the spread of influenza*. Science, 2006. **312**(5772): p. 447-451.
3. Christakos, G., et al., *Interdisciplinary Public Health Reasoning and Epidemic Modelling: The Case of Black Death: The Case of Black Death*. 2006: Springer.
4. WHO, *World Health Organization. Director-General's opening remarks at the media briefing on COVID-19-11 March 2020*. Geneva, Switzerland, 2020.
5. Molina, J.M., et al., *No evidence of rapid antiviral clearance or clinical benefit with the combination of hydroxychloroquine and azithromycin in patients with severe COVID-19 infection*. Med Mal Infect,

2020. **10**.

6. Mack, A., et al., *Ethical and legal considerations in mitigating pandemic disease: workshop summary*. 2007: National Academies Press.
7. Utsunomiya, Y.T., et al., *Growth Rate and Acceleration Analysis of the COVID-19 Pandemic Reveals the Effect of Public Health Measures in Real Time*. *Frontiers in Medicine*, 2020. **7**(247).
8. Guo, G., et al., *New Insights of Emerging SARS-CoV-2: Epidemiology, Etiology, Clinical Features, Clinical Treatment, and Prevention*. *Frontiers in Cell and Developmental Biology*, 2020. **8**(410).
9. Lai, S., et al., *Effect of non-pharmaceutical interventions to contain COVID-19 in China*. 2020.
10. Little, R.J., *Statistical analysis of masked data*. *Journal of Official statistics*, 1993. **9**(2): p. 407.
11. Stamp, M., *Introduction to machine learning with applications in information security*. 2017: CRC Press.
12. Gelman, A., *Analysis of variance—why it is more important than ever*. *The annals of statistics*, 2005. **33**(1): p. 1-53.
13. O'Reilly, K.M., et al., *Effective transmission across the globe: the role of climate in COVID-19 mitigation strategies*. *The Lancet. Planetary Health*, 2020.
14. Sajadi, M.M., et al., *Temperature and latitude analysis to predict potential spread and seasonality for COVID-19*. Available at SSRN 3550308, 2020.
15. Van Bavel, J.J., et al., *Using social and behavioural science to support COVID-19 pandemic response*. *Nature Human Behaviour*, 2020: p. 1-12.
16. Moerbeek, M., *The Consequence of Ignoring a Level of Nesting in Multilevel Analysis*. *Multivariate Behavioral Research*, 2004. **39**(1): p. 129-149.
17. Reynolds, K. and L. Madden, *Analysis of epidemics using spatio-temporal autocorrelation*. *Phytopathology*, 1988. **78**(2): p. 240-246.
18. Zhou, F., et al., *Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study*. *The Lancet*, 2020. **395**(10229): p. 1054-1062.
19. Li, X., et al., *Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan*. *Journal of Allergy and Clinical Immunology*, 2020.
20. *Coronavirus Pandemic (COVID-19) – the data*. 2020: <https://ourworldindata.org/coronavirus-data>.
21. Laird, N.M. and J.H. Ware, *Random-effects models for longitudinal data*. *Biometrics*, 1982: p. 963-974.
22. Baayen, R.H., D.J. Davidson, and D.M. Bates, *Mixed-effects modeling with crossed random effects for subjects and items*. *Journal of memory and language*, 2008. **59**(4): p. 390-412.
23. Chevan, A. and M. Sutherland, *Hierarchical partitioning*. *The American Statistician*, 1991. **45**(2): p. 90-96.
24. Moustakas, A. and M. Evans, *Coupling models of cattle and farms with models of badgers for predicting the dynamics of bovine tuberculosis (TB)*. *Stochastic Environmental Research and Risk Assessment*, 2015. **29**(3): p. 623-635.

25. Hale, T., et al., *Variation in government responses to COVID-19*. Blavatnik School of Government Working Paper, 2020. **31**.
26. Pinheiro, J.C. and D.M. Bates, *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing, ed. J. Chambers, et al. 2000, New York: Springer Verlag.
27. Burnham, K.P. and D.R. Anderson, *Model Selection and Multimodel Inference*. 2002, New York: Springer Verlag.
28. Moustakas, A., I.N. Daliakopoulos, and T.G. Benton, *Data-driven competitive facilitative tree interactions and their implications on nature-based solutions*. Science of The Total Environment, 2019. **651**: p. 2269-2280.
29. Mac Nally, R., *Multiple regression and inference in ecology and conservation biology: further comments on identifying important predictor variables*. Biodiversity & Conservation, 2002. **11**(8): p. 1397-1401.
30. Konstantopoulos, K., A. Moustakas, and I.N. Vogiatzakis, *A spatially explicit impact assessment of road characteristics, road-induced fragmentation and noise on birds species in Cyprus*. Biodiversity, 2020: p. 1-11.
31. Olea, P.P., P. Mateo-Tomás, and A. de Frutos, *Estimating and modelling bias of the hierarchical partitioning public-domain software: implications in environmental management and conservation*. PloS one, 2010. **5**(7): p. e11698-e11698.
32. Moustakas, A. and M.R. Evans, *Regional and temporal characteristics of bovine tuberculosis of cattle in Great Britain*. Stochastic Environmental Research and Risk Assessment, 2016. **30**(3): p. 989-1003.
33. Salathé, M., et al., *COVID-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation*. Swiss medical weekly, 2020. **150**(11-12): p. w20225-.
34. Lau, H., et al., *Internationally lost COVID-19 cases*. Journal of Microbiology, Immunology and Infection, 2020.
35. Pedersen, M.G. and M. Meneghini, *Quantifying undetected COVID-19 cases and effects of containment measures in Italy*. ResearchGate Preprint (online 21 March 2020) DOI, 2020. **10**.
36. Fetzer, T., et al., *Global Behaviors and Perceptions in the COVID-19 Pandemic*. 2020.
37. Morita, H., H. Kato, and Y. Hayashi, *International comparison of behavior changes with social distancing policies in response to COVID-19*. Available at SSRN 3594035, 2020.
38. Hlavacs, H., *How Often Should People be Tested for Corona to Avoid a Shutdown?* arXiv preprint arXiv:2004.14767, 2020.
39. Hutchison, R.L. *The accuracy of COVID-19 tests*. 2020.
40. Yi, G., et al., *COVID-19: Should We Test Everyone?* arXiv preprint arXiv:2004.01252, 2020.
41. Williamson, E., et al., *OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients*. medRxiv, 2020.
42. Yi, Y., et al., *COVID-19: what has been learned and to be learned about the novel coronavirus disease*. International Journal of Biological Sciences, 2020. **16**(10): p. 1753-1766.

Table 3

Table 3 is not available in this version.

Figures

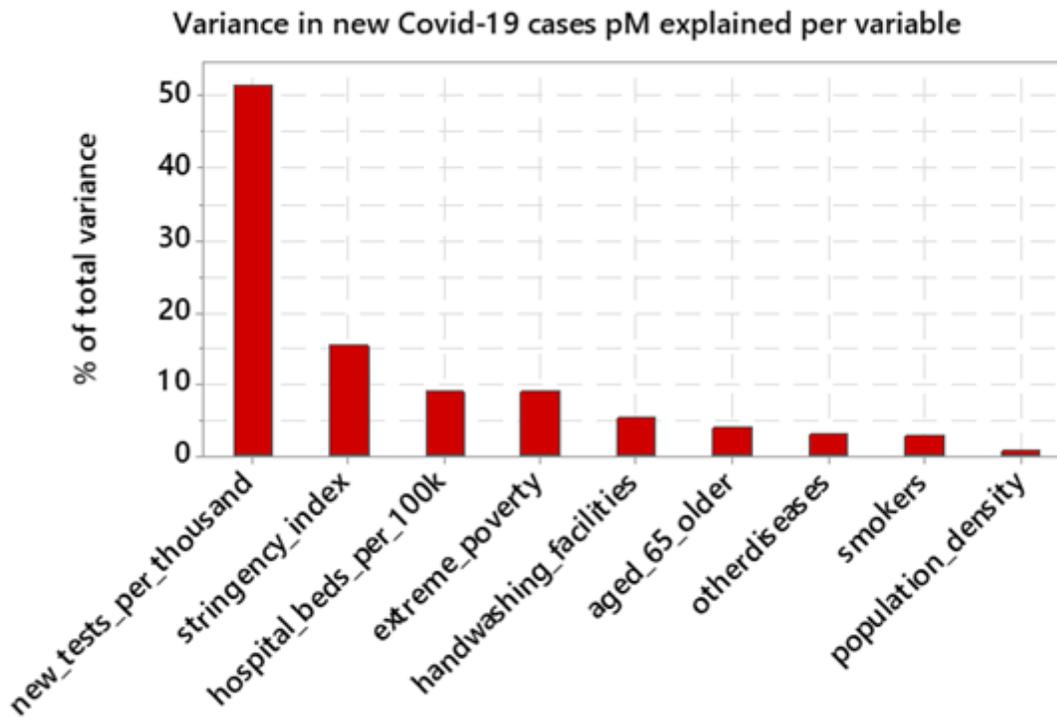
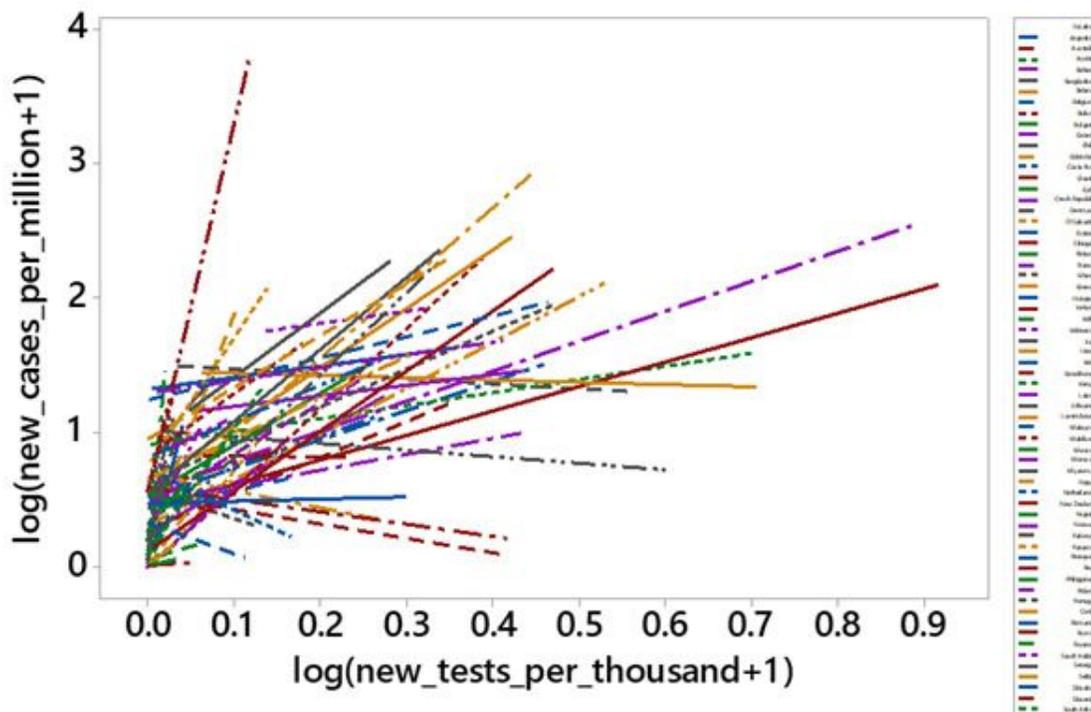
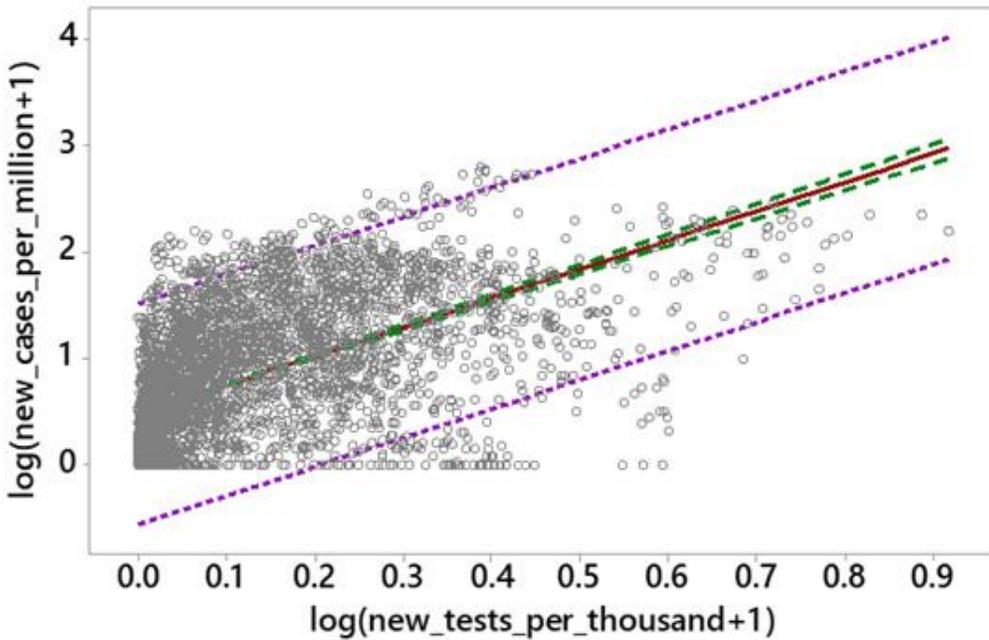


Figure 1

Percentage of the total variance in new Covid-19 cases per million explained by the nine data driven explanatory variables. Total values add to 100%. The variance explained by each variable derived from hierarchical variance partitioning (HVP) analysis.



a



b

Figure 2

a. Linear regression between new Covid-19 cases per million and new Covid-19 tests per thousand per country across all time steps. Both variables were $\log(x+1)$ transformed before regression in this graph. Each line indicates the slope of the regression between new cases and new tests on each country. b. Linear regression between new Covid-19 cases per million and new Covid-19 tests per thousand across all countries and time step (all data). Both variables were $\log(x+1)$ transformed before regression in this

graph. The solid red line indicates the regression, dashed green lines indicate the 95% confidence interval, and dotted red lines the 95% prediction interval. The regression equation is $\log(\text{new_cases_per_million}+1) = 0.4777 + 2.725 \log(\text{new_tests_per_thousand}+1)$. $S = 0.530$, $R^2 = 33.5\%$, $P < 0.001$.