

# Internet Search Effort on Covid-19 and the Underlying Public Interventions and Epidemiological Status

Aristides Moustakas (✉ [arismoustakas@gmail.com](mailto:arismoustakas@gmail.com))

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

---

## Short Report

**Keywords:** Data analytics, public interest, web search, epidemiology, Covid-19, public interventions

**Posted Date:** June 14th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-583289/v1>

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

[Read Full License](#)

---

# Abstract

This study aimed to quantify the relative importance of indices of personal freedom, economy, and epidemiology on the public interest on Covid-19 expressed by internet searches on the topic. The relationship between the effective reproduction rate  $R_t$ , news media cover, and web search effort was also quantified. Data of online search in Greece on Covid-19 topic for one year were analyzed using indices of social distancing, financial measures, and epidemiological variables using machine learning. Temporal autocorrelation of web search effort was quantified and control charts of web search,  $R_t$ , new cases and new deaths were employed. Results indicated that the trained model exhibited a fit of  $R^2 = 91\%$  between the actual and predicted web search effort. The top five variables for predicting web search effort were new deaths, the opening of international borders to non-Greek nationals, new cases, testing policy, and restrictions in internal movements. Web search had negligible temporal autocorrelation between weeks. Web search peaked during the same weeks that the  $R_t$  was peaking although new deaths or new cases were not peaking during those dates. The extent to which online searches may reflect the actual epidemiological situation is discussed.

## Introduction

Disease spread is a complex phenomenon and requires an interdisciplinary approach spanning from medicine to statistics and social sciences (Christakos et al. 2006). 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 (WHO 2020). Quantifying human psychology or motivation to be informed about facts is admittedly a complex and potentially polarized issue (Kreps and Kriner 2020). Health-related issues are particularly difficult, as humans in an effort to protect their own or other peoples' lives may need to compromise their daily habits, personal and social freedom or financial interests (Clinton et al. 2021). Public interest in a disease spread is a dynamical phenomenon, as different factors may dominate depending on the nature of a disease or the economic status, or the health system of the country they reside and these factors exhibit strong interaction effects (Green et al. 2020). In addition, intervention measures employed in order to control the disease are dynamically evaluated by the public (Bento et al. 2020).

Quantifying Covid-19 web search interest into quantifiable characteristics in a consistent and reproducible way, could facilitate explaining human preferences as well as divergences between intervention measures (Bento et al. 2020, Du et al. 2020, Zitting et al. 2021). In addition, it can provide a way to understand mass psychology during such events (Du et al. 2020, Zitting et al. 2021). To date there is no known anti-viral treatment, and vaccination against Covid-19 was implemented relatively recently (Molina et al. 2020). Therefore the available options against the virus were the immune system and health status of each individual, social distancing measures, and testing (Mack et al. 2007, Guo et al. 2020, Utsunomiya et al. 2020). Thus, there were relatively few options to be employed in order to diminish the spread of the disease (Haug et al. 2020).

Substantial analysis has been conducted regarding the efficacy of different intervention measures in the disease control (Haug et al. 2020). In addition, the impacts of social distancing and movement restrictions have been reported from a social and psychological perspective. However, it is hard to quantify what is the public risk perception and what is triggering the highest public attention around Covid-19 (Cori et al. 2020, Dryhurst et al. 2020). Public attention could be triggered by fear of getting sick or fear of death (Menzies and Menzies 2020, Pradhan et al. 2020). However, public interest could also be driven by fear of poverty, unemployment, or financial insecurity, or reciprocally financial aid or dept relief (Giordano 2020, Mamun and Ullah 2020, Esobi et al. 2021). Additional causes of public interest are disruption of individual and social freedom of movement and gathering, outgoing and entertainment, national and international travels, and school closures (Clinton et al. 2021). There clearly is a trade-off between freedom, safety, democracy, and economy (Landman and Splendore 2020, Simon 2020). Quantifying the public interest associated with the intervention measures comparable with the corresponding epidemiological parameters and ranking the relative contribution of each factor on the web search interest regarding Covid-19 may facilitate understanding elements of human nature.

In this study online search effort was used as a proxy of societal interest in Covid-19 (Effenberger et al. 2020, Szmuda et al. 2020) in Greece. The study addresses two questions: (a) what are the economic, epidemiological, and social distancing characteristics that better explain online searches on Covid? (b) How is online search effort related to the Covid-19 situation in the country as expressed by the disease reproduction rate  $R_t$ ? Web search effort and the corresponding epidemiological, social distancing, and economic data were analyzed using machine learning (Demertzis et al. 2020). Feature importance of each epidemiological, social distancing and economic explanatory variable on Covid-19 web search effort was quantified. The temporal autocorrelation of web search effort was quantified, and control charts of web search,  $R_t$ , new cases and new deaths were compared.

## Methods

### *Web search data*

Data mining (Moustakas and Katsanevakis 2018) from Google Trends (<https://trends.google.com/trends/>) was performed in order to connect these data with the current status at the time of search - see e.g. (Effenberger et al. 2020, Szmuda et al. 2020). Google Trends is a public repository of information on online search patterns of individuals that use Google as their search engine. Data regarding search effort on the topic 'Coronavirus disease 2019' (thereby 'Covid-19') was retrieved from 26 February 2020 to 14 February 2021. Throughout the analysis week 1 (first week) is the week 24 Feb 2020 - 1 March 2020, while week 52 (last week) is the week 8 Feb 2021 - 14 Feb 2021. The Covid-19 search topic includes searches in all major languages (covered by Google translator) conducted through a Greek IP address or internet provider. Data are normalized by Google Trends so as to represent search interest relative to the highest search term in Greece for that date. For example a value of 33% means that the term is one third as popular than the most popular searched topic on that date. The temporal resolution of the data is one week.

### *Epidemiological data*

New confirmed Covid-19 cases (*New cases*) per day in Greece were used as a proxy of disease spread as a scale variable. Similarly, new confirmed Covid-19 deaths (*New deaths*) per day (scale variable) were used as a proxy of mortality. New cases and deaths are the epidemiological variables most commonly reported in the media. The original temporal resolution of the data was one day. As the temporal resolution of the web search effort data is one week, new cases and new deaths were averaged per week and weekly average values were employed throughout the analysis. Data regarding new Covid-19 cases and deaths per day were retrieved from the database 'Our world in data' (Roser et al. 2020) publicly available at: (Our\_World\_in\_Data 2021b). The original data derive from the European Centre for Disease Prevention and Control (ECDC), and from John Hopkins University (Our\_World\_in\_Data 2021b). New cases and new deaths were averaged per week to match the temporal resolution and dates of the web search effort data.

### *Public intervention data*

All data were recorded at a daily temporal resolution as time series as ordinal factor variables with different factor levels (unless otherwise explicitly stated) and averaged per week to match the resolution and dates of web search effort weekly data. These are publicly available from the database 'Our world in data' (Roser et al. 2020) at: (Our\_World\_in\_Data 2021a). A detailed description of the variables with different levels of restrictions is provided in the supplementary material.

### *Containment and closure policies (C)*

The policies examined were: school closure, workplace closure, cancelation of public events, gathering restrictions, stay at home requirements, internal movement restrictions, and international travel restrictions.

### *Economic interventions (E)*

The economic interventions examined included: income support, debt relief, and fiscal support measures (scale variable).

### *Health system policies (H)*

The health system policies investigated included: testing policy, contact tracing, facial coverings, and vaccination policy.

### *Covid-19 situation*

In an effort to compare the actual epidemiological situation in Greece against the web search effort, the daily score of the *effective reproduction number*  $R_t$  was used. The  $R_t$  is the average number of secondary infections produced when one infected individual is introduced into a susceptible host population

(Anderson and May 1992). It indicates the average number of people who will contract a contagious disease from one person with that disease (Pandit 2020). The  $R_t$  is defined as:

$$\mathcal{R}_t = \mathcal{R}_0^{(t)} \times (S_{t-1}/N), \quad [1]$$

where  $S$  is the number of susceptible individuals in a total population of  $N$  individuals, while  $R_0^{(t)}$  is the average number of individuals infected by a single infected individual when everyone else is susceptible (Arroyo-Marioli et al. 2021). In practice  $R_t > 1$  indicates a growing disease spread while  $R_t < 1$  a disease spread that is diminishing, with larger values indicating higher spread.  $R_t$  was calculated per day as part of the 'Our world in data' (Dong et al. 2020, Roser et al. 2020) epidemiological dataset publicly available at: (Our\_World\_in\_Data 2021b). Daily values of  $R_t$  were used averaged per week to match the temporal resolution and dates of web search effort data.

### *News cover data*

Greek news media cover in time for the same dates as web search effort was retrieved from the Media Cloud database ([mediacloud.org](https://mediacloud.org)). News media cover is estimated by counting the number of articles mentioning Covid-19 topic nationally or locally in Greece divided by the total of news media articles published. This resulted in the proportion (%) of Covid-19 news media articles per day. This was done in order to examine the correlation between web search effort and news cover on the topic and quantify the extent that web search is related with news (Bento et al. 2020, Lampos et al. 2021).

### *Analysis*

#### *Artificial Neural Networks (ANN)*

ANN were used to quantify the complex relationship between web search effort (dependent variable) and the 16 explanatory variables. ANN are machine learning computing algorithms that can solve complex problems imitating animal brain in a simplified manner and can handle correlated independent variables (Rojas 1996, Hasson et al. 2020). Perception-type neural networks consist of artificial neurons or nodes, which is information processing units arranged in layers and interconnected by synaptic weights (connections); (Rojas 1996, Hasson et al. 2020). Neurons can filter and transmit information in a supervised fashion in order to build predictive model that clarifies data stored in memory (Rojas 1996, Hasson et al. 2020).

The Multilayer Perceptron (MLP) module was used to build the ANN and test its accuracy (Salgado et al. 2020). The MLP in ANN was trained with a back-propagation learning algorithm which uses the gradient descent to update the weights towards minimizing the error function (Salgado et al. 2020). The data were randomly assigned to 60% training, and 40% testing subsets. The training dataset is used to build the ANN model (Rojas 1996). The testing data is used to find errors and validate the model. Before training,

all covariates were normalized using the formula  $(x-\min)/(\max-\min)$ , which returns values between 0 and 1.

For the hidden layer the hyperbolic tangent was used as activation function (Zamanlooy and Mirhassani 2013). The activation function  $O_j$  for each neuron of the  $j$ th output neuron takes real numbers as arguments and returns real values between -1 and 1. For the output layer, the identity function was used as activation function. Gradient descent optimization with the batch algorithm was used. The batch algorithm uses all records in the training dataset to update the synaptic weights between neurons (IBM 2016). The scaled conjugate gradient method was used for the batch training of the ANN (Marwala 2010). Before each iteration, the synaptic weights in the training dataset are updated. The algorithm finds the global error minimum by minimizing the total error made in the previous iteration (Møller 1993, IBM 2016).

Four parameters - initial lambda, initial sigma and interval center and interval offset - determine the way the scaled conjugate gradient algorithm builds the model. Lambda controls if the Hessian matrix is negative definite (Møller 1993). Sigma controls the size of weight change that affects the estimation of Hessian through the first order derivatives of error function (Rojas 1996). The parameters interval center  $a_0$  and  $a$  force the simulated annealing algorithm to generate random weights that iteratively minimize the error function (IBM 2016). Initial lambda was set to 0.0000005, initial sigma to 0.00005. Interval center was defined as 0 and interval offset was set to  $\pm 0.5$  (IBM 2016).

#### *Variable importance in ANN*

Variable importance was quantified using the outputs of the trained ANN model, in order to evaluate the effect of each input variable on the web search effort by using the variance based method (de Sá 2019, Ju et al. 2019). The input variables are ranked according to the sensitivity formula defined as:

$$S_i = \frac{V_i}{V(Y)} = \frac{V(E(Y|X_i))}{V(Y)}, \quad [2]$$

where  $V(Y)$  is the unconditional output variance,  $E$  is the integral over  $Y|X_i$ , while the variance operator  $V$  implies a further integral over  $X_i$ . Variable importance is then computed as the normalized sensitivity.  $S_i$  is the appropriate measure of sensitivity to rank the variables in order of importance for any combination of interactions and non-orthogonality among variables (Ju et al. 2019). The total sum of the overall  $V(I)$  of the ANNs is 1 (Ju et al. 2019).

#### *Temporal autocorrelation*

The correlation of web search in time was calculated, as this can provide information whether a high search effort week is likely to be followed by another high search effort one, or a low search effort is followed by a low search effort one, or if the search effort of a week is not indicative of the search effort

of the following weeks (Moustakas and Evans 2017). To do so the temporal autocorrelation function for web search effort per week as a time unit lag was calculated (Reynolds and Madden 1988).

### *Comparing web search and the Covid-19 situation*

Inquiry-Charts, (I-charts) also termed as Shewart charts; (Montgomery 2020), were computed. An I-chart is a type of control chart used to monitor the process mean when measuring individuals at regular intervals from a process (Montgomery 2020). Each point on the chart represents the value of an individual observation (Montgomery 2020). The center line is the process mean (the mean of the individual observations). The control limits (upper and lower confidence intervals) are a multiple (k) of three sigma ( $k=3\sigma$ ) above and below the center line. The process sigma is the standard deviation of the individual observations. I-charts **display** individual data points and monitor mean and shifts in the process when the data points collected at regular intervals of time (Montgomery 2020). I-Charts may facilitate identifying the common and assignable causes in the process, if any (Rigdon et al. 1994). The green line on each chart represents the mean, while the red lines show the upper and lower control limits. An in-control process shows only random variation within the control limits (Montgomery 2020). An out-of-control process has unusual variation, which may be due to the presence of special causes (Montgomery 2020). I-charts of web search effort,  $R_t$ , new cases, and new deaths were calculated.

Linear mixed effects models (LME); (Pinheiro and Bates 2000) were used to analyze the relationship between web search (dependent variable) and news media cover as well as  $R_t$  as fixed effects (independent variables). The model structure included also the random effects of time in terms of week number, accounting for the temporal autocorrelation in the data. This analysis was conducted in order to quantify and compare the effect of news on web searches in conjunction with the Covid-19 actual situation as expressed by the  $R_t$ .

## **Results**

The trained ANN model with the 16 epidemiological and intervention variables exhibited a relative error of 0.058 (Sum of Squares error = 0.978) in the training dataset and a relative error of 0.237 in the test data (Sum of Squares error = 1.130). All error computations are based on the testing sample. The predictive accuracy of the trained ANN was  $R^2 = 91\%$ ,  $p < 0.001$  between actual weekly values of web search effort of Covid-19 in Greece and the ANN model outputs (Fig. 1a). Model residuals exhibited a minor negative trend against the predicted value (Fig. 1b) but this deviance could not be differentiated from zero (confidence intervals were always crossing the zero line,  $R^2 = 5\%$ ,  $p < 0.01$ ).

Results regarding variable importance indicate that web search effort on Covid-19 in Greece was primarily featured by new deaths per week (relative importance of 100%), followed by international travelling restriction measures applicable to non-Greeks with relative importance of 95%, while new infections were ranked third with feature importance relative to other variables of 82% (Fig. 2). Testing policy was ranked fourth while restrictions on internal movement fifth, both with relative importance of 77 and 76%

respectively (Fig. 2). School closing exhibited a relative importance of 71%, facial covering requirements 67%, staying at home requirements 66%, while restrictions on social gatherings 59% (Fig. 2). Work place closures featured a relative importance of 56%, cancelling public events 56%, fiscal measures 55%, while dept/contract relief 54% (Fig. 2). Ultimately, vaccination policy had a relative importance of 34%, income support 30%, and contact tracing 12% (Fig. 2).

The temporal autocorrelation of web search effort indicated no significant time lags between weeks indicating that the search effort of one week is not exhibiting significant correlation with the search effort of the next or previous weeks (Fig. 3).

The I-chart of web search effort indicated a negative deviation from confidence intervals during the first week in the analysis (week 1, week ending in 1 March 2020) as well as during week 14 (31 May 2020), while exceeded the upper confidence interval for two weeks in a row during weeks 24 & 25 (2 August 2020 to including the week starting at 16 August 2020), as well as also exceeding the upper confidence intervals during weeks 34 to 36 (18 October 2020 to 1 November 2020; Fig. 4a).

The I-chart of  $R_t$  indicated that values were above the upper confidence interval during the first two weeks of available data (Fig. 4b; weeks 3 & 4; the  $R_t$  needs some prior data before it can be calculated and thus may not be calculated during the first week that a case was found). Values of  $R_t$  were below the lower confidence interval between weeks 8 to 13 (19 April 2020 to including the week starting at 24 May 2020), above the confidence interval the weeks 23 to 25 (2 August 2020 to including the week starting at 16 August 2020), and above the upper confidence interval the weeks 35 to 37 (25 October 2020 to including 8 November 2020; Fig. 4b). Values of  $R_t$  were below the lower confidence interval the during the weeks 41–45 (6 December 2020 to 3 January 2021; Fig. 4b).

The I-chart values of new cases were always below the lower confidence interval till week 24 (9 August 2020), within the confidence interval till week 35 (25 October 2020), and above the higher confidence interval until week 43 (20 December 2020), within the confidence interval between weeks 44–49 (27 December 2020–31 January 2021), and above the confidence interval for the remaining two weeks of data (Fig. 4c).

The I-chart values of new deaths were below the lower confidence interval till week 33 (11 October 2020, week 31 was marginally within the confidence interval; Fig. 4d), within the confidence interval till week 37 (8 November 2020), and above the higher confidence interval until week 47 (17 January 2021), and within the confidence interval for the remaining four weeks of data (Fig. 4d).

Results from LME between web search as dependent variable and news media cover and  $R_t$  as fixed effects and time as a random effect had a model fit of AIC = 389.01, BIC = 398.57, logLik = -189.50. The fixed effect model structure had an intercept of -61.4 (standard error = 18.48), an  $R_t$  coefficient of 56.23 (standard error = 8.62) and a news media coefficient of 3.22 (standard error = 0.72); (Fig. 5a) with all fixed

effects having a p-value  $< < 0.001$ . Web search consistently increased with increases of  $R_t$ , across news media cover values (Fig. 5b).

## Discussion

Results derived here studying indicated that fear of, or interest in death was the variable with the highest explanatory power in predicting internet search effort, implying that the first concern is either to stay alive, or interest and compassion to the individuals who passed way (Du et al. 2020, Lu and Reis 2021). This interest was higher than any financial related variable, or personal freedom of movement and entertainment, or every day habits. Internet searches on Covid were able to track Covid-19 spread (Azad and Devi 2020). This work indicates that a trained and validated data-driven model may provide useful insights regarding societal endeavors of future measures (Gozzi et al. 2020, Kar and Dwivedi 2020).

Interestingly the second variable for predicting Covid-19 web search effort was the opening of international borders applicable to non-Greek citizens and thus a variable directly related with tourism (Mariolis et al. 2020, Sharun et al. 2020). Opening international borders for non-Greeks and therefore the potential income generated by tourism featured considerably higher than any financial aid or facilitation derived by the state, as both debt/contract relief and fiscal measures were amongst the lower variables for predicting web search effort. This is in spite of the financial crisis driven by Covid-19 partially due to obligatory closures in a wide range of enterprises (Biskanaki et al. 2020, Mariolis et al. 2020). Interest in opening the international borders to non-Greek nationals has a potential two-fold implication: individuals may be searching about which foreign nationals are eligible for visiting Greece and technicalities of arrivals regarding potential testing requirements from the point of view of tourism-related interests or from an epidemiological perspective of how safe opening the international borders was (Mariolis et al. 2020, Sharun et al. 2020). The analysis conducted here may not differentiate between the two potential groups.

Another plausible option is that tourists that visited Greece during those dates (the peak season is 1–15 August) were extensively searching on Covid-19 to learn about the situation back in their home countries or to investigate the current situation in Greece. Foreign nationals searching for Covid-19 while in Greece are recorded within the Google the web search effort for Greece so long as they used a Greek IP address or connecting/roaming through a Greek internet network provider (Anderegg and Goldsmith 2014). However, data on web search effort are normalised by the other searches during those dates (see methods) and thus simply increasing the number of people searching would not result by default in higher web search effort on the topic (Ficetola 2013, McCallum and Bury 2014, Burivalova et al. 2018); the interest in the topic increased. It is also interesting to note that web search effort was not temporally correlated between weeks and thus there is no evidence that individuals are searching on the topic by habit or fear of the situation during the past few weeks, but rather dynamically tuning with the current weekly situation.

During the first half of August 2020, both  $R_t$  and web search effort were peaking despite the fact that new deaths or new cases were not peaking soon after. Greece was topping the list of tourist-imported Covid-19 cases in the UK followed up the 2020 summer season among all other examined countries (Aggarwal et al. 2021). It therefore seems plausible that infected visiting individuals were never recorded and returned back into their home countries. It was however widely discussed whether the opening of international borders to tourism with low testing frequencies or without a negative PCR Covid-19 test of incoming individuals was a safe practice (Pavli et al. 2020, Rocklöv et al. 2020, Sharun et al. 2020). Thus, the interest in the opening of international borders may have an interaction effect with the interest in testing policy in the total web search effort. The temporal coincidence of web search effort peak with the peak of  $R_t$  is interesting given that the  $R_t$  is not something regularly reported in the news, media, or non-scientific websites as more often than not new cases, or deaths, or intervention measures are reported (Liu et al. 2020). Web search effort peak was also coinciding with  $R_t$  peaks in October, from 27 October 2020, the date when new cases exceeded the threshold of 1,000 per day. Thus findings reported here confirm both that web search effort can be used to predict peaks in COVID-19, as well as that peaks of online searches precede the reported confirmed cases and deaths (Lampos et al. 2021).

Testing policy appeared as the fourth most predictive variable for predicting web search effort and above other variables such as exiting home restrictions, restrictions on gatherings, movements in other regions in Greece, obligatory use of masks, or school closures. This may seem as a surprising result given that testing policy is not affecting or restricting everyday life. However, testing frequency is directly related both with disease spread control through quarantining infected individuals (Wells et al. 2021). In addition, Greece has been reported not to test as frequent as other countries as exhibited by a high ratio of deaths per cases in comparison to other countries (Fouda et al. 2020). Greece and people could potentially worry regarding new cases as these could be either under-reported or disease spread via untested individuals. Frequent testing could control the disease and reduce the duration of quarantine (Peto 2020, Wells et al. 2021) and it is known to de-synchronize disease spread in other diseases too (Moustakas et al. 2018).

The idea that crowd wisdom might reflect the reality better than expert opinion or any single individual, has been considered as provocative in the past (Galton 1907, Prelec et al. 2017). Web search has the cultural, geographical, social, temporal diversity and a sample size that is hard to be ignored (Surowiecki 2005, Sunstein 2006). In Greece web search effort peaked during the same time periods with  $R_t$  indicating that web search may reflect the reality better than the recorded situation at least in the scarcity of testing data. Web search was increasing considerably more with higher values of  $R_t$  rather than when news media cover was high.

## References

1. Aggarwal D, Page AJ, Schaefer U, Savva GM, Myers R, Volz E, Ellaby N, Platt S, Groves N, Gallagher E, Tumelty NM, Le Viet T, Hughes GJ, Chen C, Turner C, Logan S, Harrison A, Peacock SJ, Chand M, Harrison EM (2021) An integrated analysis of contact tracing and genomics to assess the efficacy of

- travel restrictions on SARS-CoV-2 introduction and transmission in England from June to September, 2020. medRxiv:2021.2003.2015.21253590
2. Anderegg WRL, Goldsmith GR (2014) Public interest in climate change over the past decade and the effects of the 'climategate' media event. *Environmental Research Letters* 9:054005
  3. Anderson RM, May RM (1992) *Infectious diseases of humans: dynamics and control*. Oxford university press
  4. Arroyo-Marioli F, Bullano F, Kucinskas S, Rondón-Moreno C (2021) Tracking R of COVID-19: A new real-time estimation using the Kalman filter. *Public Library of Science One* 16:e0244474
  5. Azad S, Devi S (2020) Tracking the spread of COVID-19 in India via social networks in the early phase of the pandemic. *Journal of Travel Medicine* 27
  6. Bento AI, Nguyen T, Wing C, Lozano-Rojas F, Ahn Y-Y, Simon K (2020) Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. *Proceedings of the National Academy of Sciences* 117:11220–11222
  7. Biskanaki F, Rallis E, Andreou E, Sfyri E, Tertipi N, Kefala V (2020) Social-economic impact of COVID-19 pandemic on aesthetic centers in Greece. *J Cosmet Dermatol* 19:2165–2168
  8. Burivalova Z, Butler RA, Wilcove DS (2018) Analyzing Google search data to debunk myths about the public's interest in conservation. *Front Ecol Environ* 16:509–514
  9. Christakos G, Olea RA, Serre ML, Yu HL, Wang LL (2006) *Interdisciplinary Public Health Reasoning and Epidemic Modelling: The Case of Black Death: The Case of Black Death*. Springer
  10. Clinton J, Cohen J, Lapinski J, Trussler M (2021) Partisan pandemic: How partisanship and public health concerns affect individuals' social mobility during COVID-19. *Science Advances* 7:eabd7204
  11. Cori L, Bianchi F, Cadum E, Anthonj C (2020) Risk perception and COVID-19. *Multidisciplinary Digital Publishing Institute*
  12. de Sá CR (2019) *Variance-Based Feature Importance in Neural Networks*. Springer International Publishing, Cham, pp 306–315
  13. Demertzis K, Tsiotas D, Magafas L (2020) Modeling and Forecasting the COVID-19 Temporal Spread in Greece: An Exploratory Approach Based on Complex Network Defined Splines. *International Journal of Environmental Research Public Health* 17:4693
  14. Dong E, Du H, Gardner L (2020) An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases* 20:533–534
  15. Dryhurst S, Schneider CR, Kerr J, Freeman ALJ, Recchia G, van der Bles AM, Spiegelhalter D, S van der Linden (2020) Risk perceptions of COVID-19 around the world. *J Risk Res* 23:994–1006
  16. Du H, Yang J, King RB, Yang L, Chi P (2020) COVID-19 Increases Online Searches for Emotional and Health-Related Terms. *Appl Psychol Health Well Being* 12:1039–1053
  17. Effenberger M, Kronbichler A, Shin JI, Mayer G, Tilg H, Perco P (2020) Association of the COVID-19 pandemic with Internet Search Volumes: A Google Trends™ Analysis. *International Journal of Infectious Diseases* 95:192–197

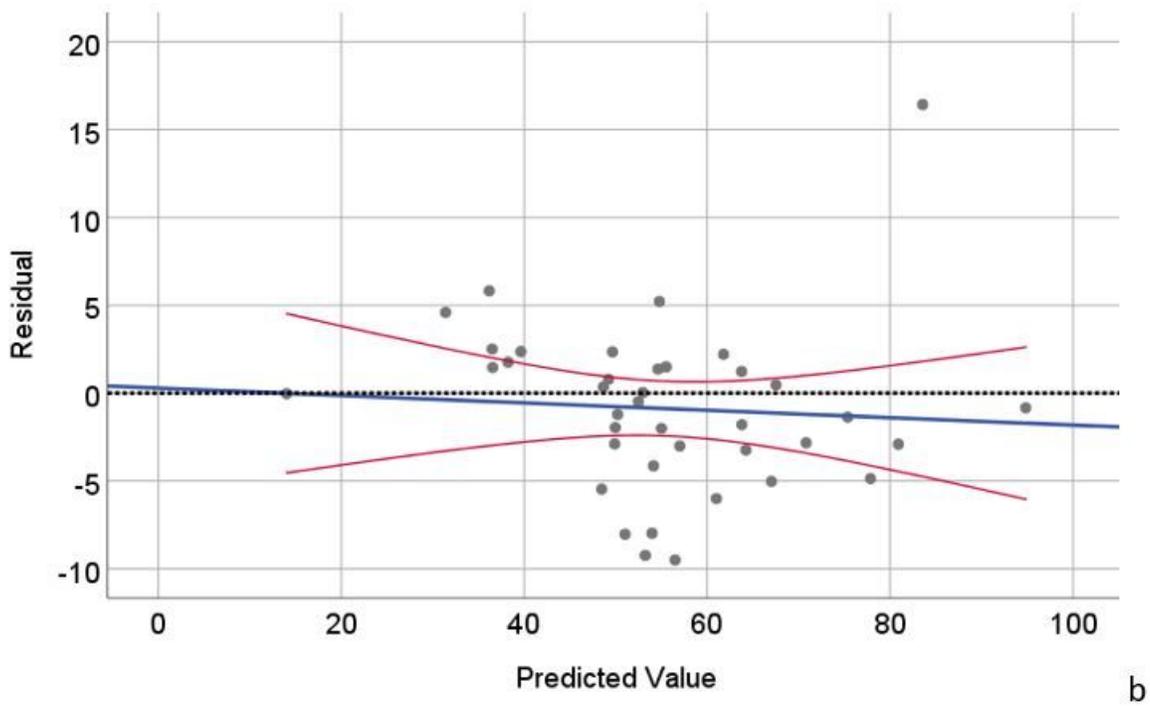
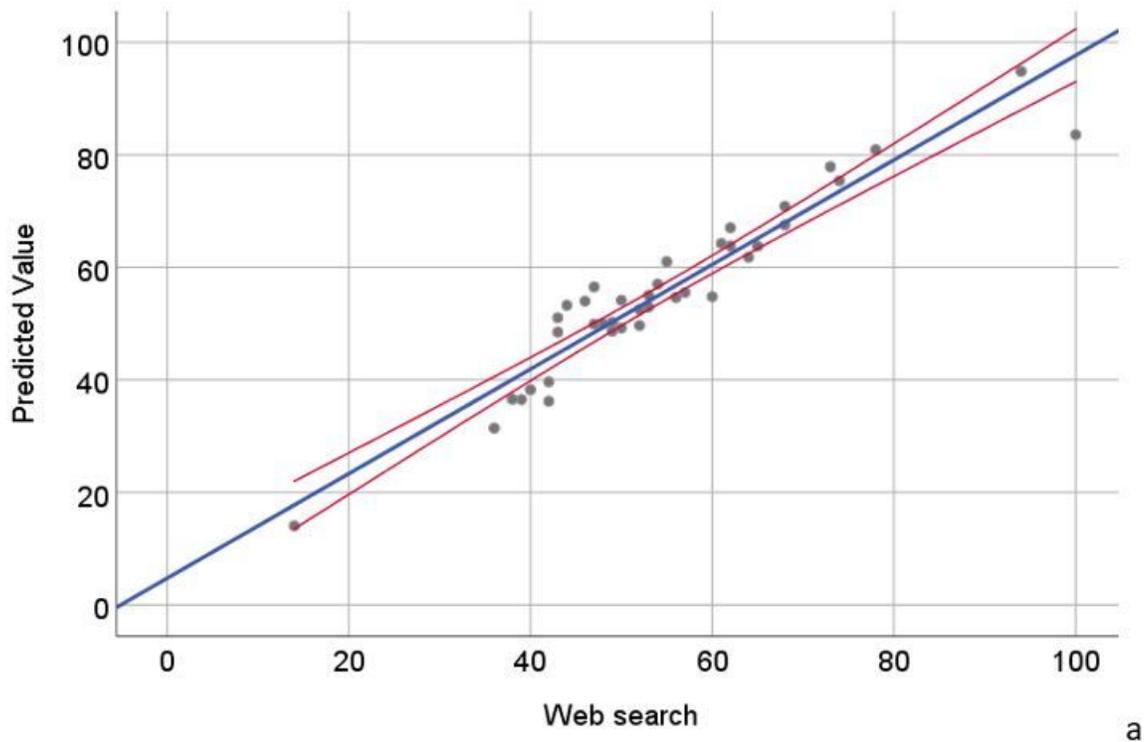
18. Esobi I, Lasode M, Anyanwu C, Barriguete MF, Okorie M, Lasode D (2021) Food Insecurity, Social Vulnerability, and the Impact of COVID-19 on Population Dependent on Public Assistance/SNAP: A Case Study of South Carolina, USA. *Journal of Food Security* 9:8–18
19. Ficetola GF (2013) Is interest toward the environment really declining? The complexity of analysing trends using internet search data. *Biodiversity Conservation* 22:2983–2988
20. Fouda A, Mahmoudi N, Moy N, Paolucci F (2020) The COVID-19 pandemic in Greece, Iceland, New Zealand, and Singapore: Health policies and lessons learned. *Health Policy Technology* 9:510–524
21. Galton F (1907) Vox populi (the wisdom of crowds). *Nature* 75:450–451
22. Giordano C (2020) Freedom or money? The dilemma of migrant live-in elderly carers in times of COVID-19. *Gender, Work, and Organization*
23. Gozzi N, Perrotta D, Paolotti D, Perra N (2020) Towards a data-driven characterization of behavioral changes induced by the seasonal flu. *PLoS Comput Biol* 16:e1007879
24. Green J, Edgerton J, Naftel D, Shoub K, Cranmer SJ (2020) Elusive consensus: Polarization in elite communication on the COVID-19 pandemic. *Science Advances* 6:eabc2717
25. Guo G, Ye L, Pan K, Chen Y, Xing D, Yan K, Chen Z, Ding N, Li W, Huang H, Zhang L, Li X, Xue X (2020) New Insights of Emerging SARS-CoV-2: Epidemiology, Etiology, Clinical Features, Clinical Treatment, and Prevention. *Frontiers in Cell and Developmental Biology* 8
26. Hasson U, Nastase SA, Goldstein A (2020) Direct fit to nature: An evolutionary perspective on biological and artificial neural networks. *Neuron* 105:416–434
27. Haug N, Geyrhofer L, Londei A, Dervic E, Desvars-Larrive A, Loreto V, Pinior B, Thurner S, Klimek P (2020) Ranking the effectiveness of worldwide COVID-19 government interventions. *Nature Human Behaviour* 4:1303–1312
28. IBM (2016) Knowledge Center. <http://goo.gl/SuuMHu>
29. Ju E, Kim KH, Chang KH, Shim JB, Kim CY, Lee NK, Lee S, Park CG (2019) Artificial Neural Network-Based Variable Importance Analysis of Prognostic Factors Related to Radiation Pneumonitis in Patients with Lung Cancer: Preliminary Study. *Journal of the Korean Physical Society* 75:277–282
30. Kar AK, Dwivedi YK (2020) Theory building with big data-driven research – Moving away from the “What” towards the “Why”. *Int J Inf Manage* 54:102205
31. Kreps SE, Kriner DL (2020) Model uncertainty, political contestation, and public trust in science: Evidence from the COVID-19 pandemic. *Science Advances* 6:eabd4563
32. Lampos V, Majumder MS, Yom-Tov E, Edelstein M, Moura S, Hamada Y, Rangaka MX, McKendry RA, Cox IJ (2021) Tracking COVID-19 using online search. *npj Digital Medicine* 4:17
33. Landman T, Splendore LDG (2020) Pandemic democracy: elections and COVID-19. *J Risk Res* 23:1060–1066
34. Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, Chu B, Zhu H, Akinwunmi B, Huang J (2020) Health communication through news media during the early stage of the COVID-19 outbreak in China: digital topic modeling approach. *J Med Internet Res* 22:e19118

35. Lu T, Reis BY (2021) Internet search patterns reveal clinical course of COVID-19 disease progression and pandemic spread across 32 countries. *npj Digital Medicine* 4:22
36. Mack A, Choffnes ER, Sparling PF, Hamburg MA, Lemon SM (2007) Ethical and legal considerations in mitigating pandemic disease: workshop summary. National Academies Press
37. Mamun MA, Ullah I (2020) COVID-19 suicides in Pakistan, dying off not COVID-19 fear but poverty?—The forthcoming economic challenges for a developing country. *Brain, behavior, and immunity*
38. Mariolis T, Rodousakis N, Soklis G (2020) The COVID-19 multiplier effects of tourism on the Greek economy. *Tourism Economics*:1354816620946547
39. Marwala T (2010) Finite Element Model Updating Using Computational Intelligence Techniques: Applications to Structural Dynamics. Springer Publishing Company, Inc
40. McCallum ML, Bury GW (2014) Public interest in the environment is falling: a response to Ficetola (2013). *Biodiversity & Conservation* **23**:1057–1062
41. Menzies RE, Menzies RG (2020) Death anxiety in the time of COVID-19: theoretical explanations and clinical implications. *The Cognitive Behaviour Therapist* 13:e19
42. Molina JM, Delaugerre C, Goff JL, Mela-Lima B, Ponscarne D, Goldwirt L, and N. de Castro. 2020. 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* **10**
43. Møller MF (1993) A scaled conjugate gradient algorithm for fast supervised learning. *Neural Netw* 6:525–533
44. Montgomery DC (2020) Introduction to statistical quality control. John Wiley & Sons
45. Moustakas A, Evans MR (2017) A big-data spatial, temporal and network analysis of bovine tuberculosis between wildlife (badgers) and cattle. *Stoch Env Res Risk Assess* 31:315–328
46. Moustakas A, Evans MR, Daliakopoulos IN, Markonis Y (2018) Abrupt events and population synchrony in the dynamics of Bovine Tuberculosis. *Nat Commun* 9:2821
47. Moustakas A, Katsanevakis S (2018) Editorial: Data Mining and Methods for Early Detection, Horizon Scanning, Modelling, and Risk Assessment of Invasive Species. *Frontiers in Applied Mathematics and Statistics* **4**: 5
48. Our\_World\_in\_Data (2021a) Covid-19 Data: Policy tracker. [https://github.com/OxCGRT/covid-policy-tracker/blob/master/data/OxCGRT\\_latest.csv](https://github.com/OxCGRT/covid-policy-tracker/blob/master/data/OxCGRT_latest.csv)
49. Our\_World\_in\_Data (2021b) Data on COVID-19 (coronavirus) by Our World in Data: the complete dataset. <https://github.com/owid/covid-19-data/tree/master/public/data>
50. Pandit JJ (2020) Managing the R0 of COVID-19: mathematics fights back. *Anaesthesia* 75:1643–1647
51. Pavli A, Smeti P, Hadjianastasiou S, Theodoridou K, Spilioti A, Papadima K, Andreopoulou A, Gkolfinopoulou K, Sapounas S, Spanakis N, Tsakris A, Maltezou HC (2020) In-flight transmission of COVID-19 on flights to Greece: An epidemiological analysis. *Travel Medicine Infectious Disease* 38:101882

52. Peto J (2020) Covid-19 mass testing facilities could end the epidemic rapidly. *BMJ* 368
53. Pinheiro JC, Bates DM (2000) *Mixed-Effects Models in S and S-PLUS*. Springer Verlag, New York
54. Pradhan M, Chettri A, Maheshwari S (2020) Fear of death in the shadow of COVID-19: The mediating role of perceived stress in the relationship between neuroticism and death anxiety. *Death Studies*:1–5
55. Prelec D, Seung HS, McCoy J (2017) A solution to the single-question crowd wisdom problem. *Nature* 541:532–535
56. Reynolds K, Madden L (1988) Analysis of epidemics using spatio-temporal autocorrelation. *Phytopathology* 78:240–246
57. Rigdon SE, Cruthis EN, Champ CW (1994) Design strategies for individuals and moving range control charts. *Journal of Quality Technology* 26:274–287
58. Rocklöv J, Sjödin H, Wilder-Smith A (2020) COVID-19 outbreak on the Diamond Princess cruise ship: estimating the epidemic potential and effectiveness of public health countermeasures. *Journal of Travel Medicine* 27
59. Rojas R (1996) *Neural Networks: A Systematic Introduction*. Springer-Verlag, Berlin
60. Roser M, Ritchie H, Ortiz-Ospina E, Hasell J (2020) Coronavirus pandemic (COVID-19). *Our world in data*
61. Salgado C, Dam R, Salgado W, Werneck R, Pereira C, Schirru R (2020) The comparison of different multilayer perceptron and General Regression Neural Networks for volume fraction prediction using MCNPX code. *Appl Radiat Isot* 162:109170
62. Sharun K, Tiwari R, Natesan S, Yattoo MI, Malik YS, Dhama K (2020) International travel during the COVID-19 pandemic: implications and risks associated with ‘travel bubbles’. *Journal of Travel Medicine* 27
63. Simon R (2020) Is There a Trade-off Between Freedom and Safety? A Philosophical Contribution to a Covid-19 Related Discussion. *Philosophy* 10:445–453
64. Sunstein CR (2006) *Infotopia: How many minds produce knowledge*. Oxford University Press
65. Surowiecki J (2005) *The wisdom of crowds*. Anchor
66. Szmuda T, Ali S, Hetzger TV, Rosvall P, Słoniewski P (2020) Are online searches for the novel coronavirus (COVID-19) related to media or epidemiology? A cross-sectional study. *International Journal of Infectious Diseases* 97:386–390
67. Utsunomiya YT, Utsunomiya ATH, Torrecilha RBP, Paulan SdC, Milanesi M, Garcia JF (2020) Growth Rate and Acceleration Analysis of the COVID-19 Pandemic Reveals the Effect of Public Health Measures in Real Time. *Frontiers in Medicine* 7
68. Wells CR, Townsend JP, Pandey A, Moghadas SM, Krieger G, Singer B, McDonald RH, Fitzpatrick MC, Galvani AP (2021) Optimal COVID-19 quarantine and testing strategies. *Nat Commun* 12:356
69. WHO (2020) World Health Organization. Director-General's opening remarks at the media briefing on COVID-19-11 March 2020. Geneva, Switzerland

70. Zamanlooy B, Mirhassani M (2013) Efficient VLSI implementation of neural networks with hyperbolic tangent activation function. *IEEE Trans Very Large Scale Integr VLSI Syst* 22:39–48
71. Zitting KM, Lammers-van der HM, Holst RK, Yuan W, Wang SF, Quan, Duffy JF (2021) Google Trends reveals increases in internet searches for insomnia during the 2019 coronavirus disease (COVID-19) global pandemic. *J Clin Sleep Med* 17:177–184

## Figures



**Figure 1**

Accuracy of the Artificial Neural Network (ANN) model. The ANN model was trained with 75% of the data and tested against the 25% of the data. Error computations are based on the testing data. The blue line indicates the regression, while red lines indicate a 95% confidence interval of the regression. Actual values are plotted in grey circles. a. Linear regression between the trained ANN predicted values versus actual values of web search effort. b. Residuals of the regression between the trained ANN versus the

predicted values. The horizontal dotted black line indicates zero value. The 95% confidence intervals crossing zero indicate that residuals may not be differentiated from zero although there is a weak negative correlation between residuals and predicted values.

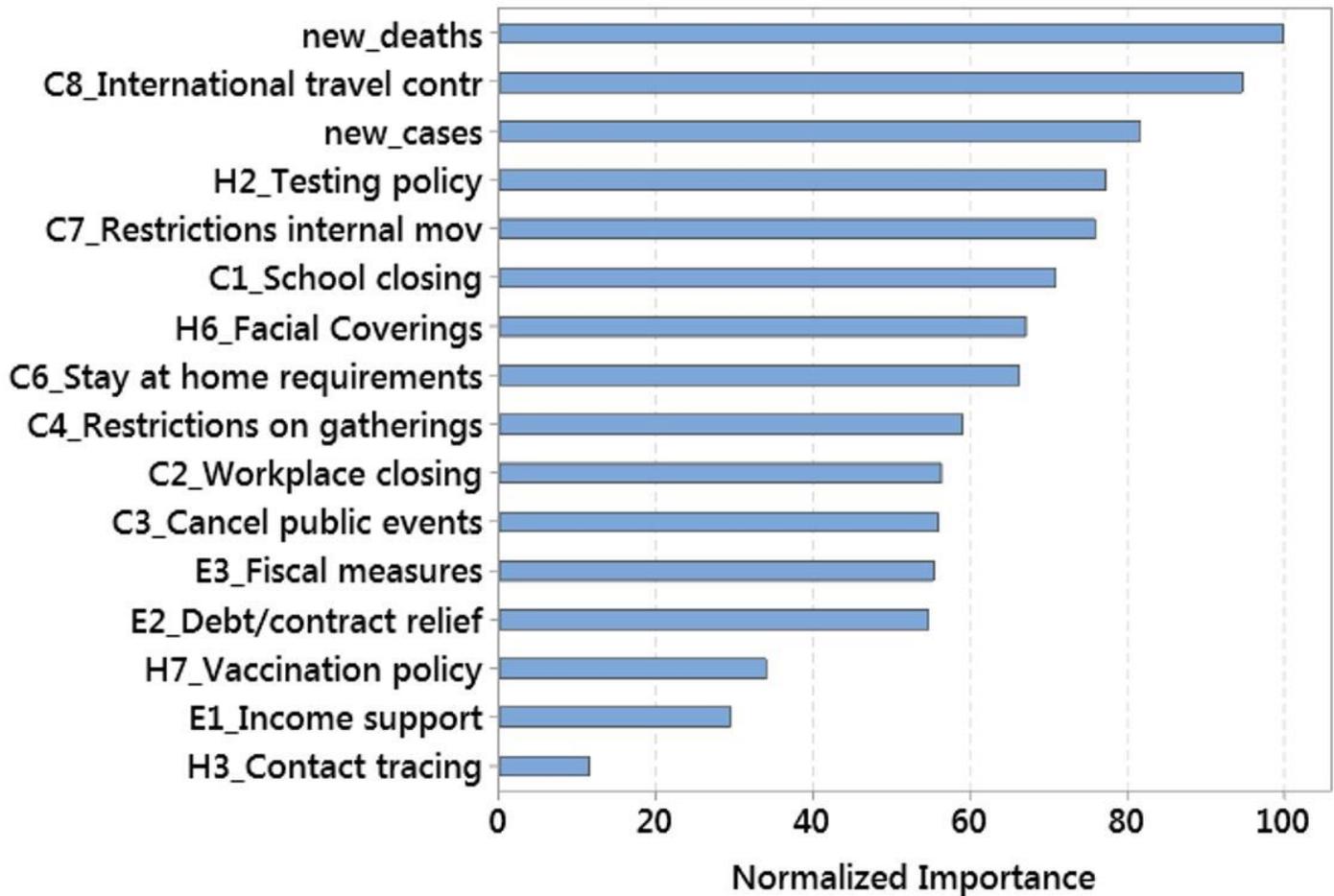


Figure 2

Normalised importance of the 16, containment and closure policies (C) health system policies (H), economic interventions (E), and epidemiological (new cases, new deaths) variables of the trained ANN model for predicting web search effort regarding Covid-19 as a topic in Greece from February 2020 to February 2021 on a weekly time scale.

### Autocorrelation Function for Web search

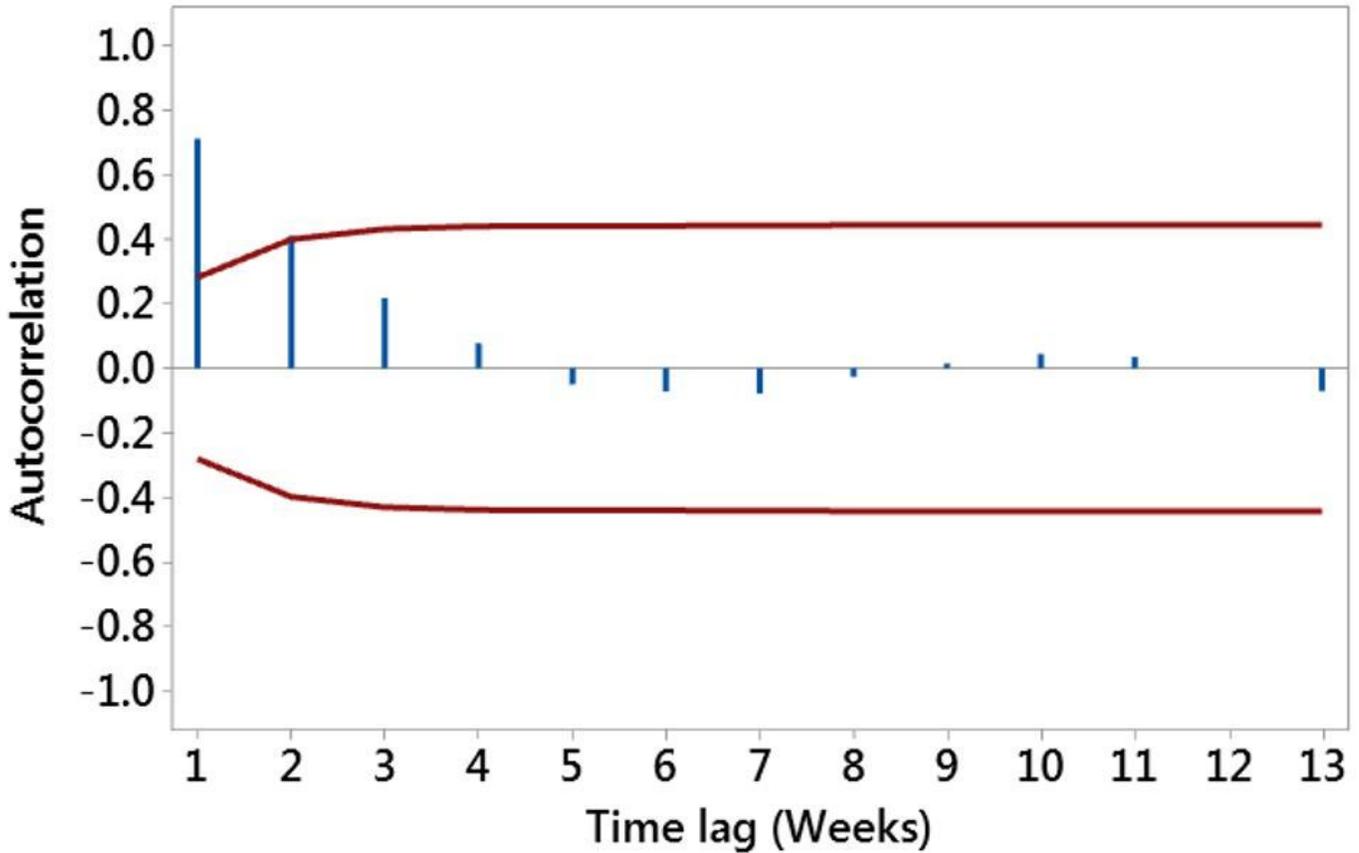
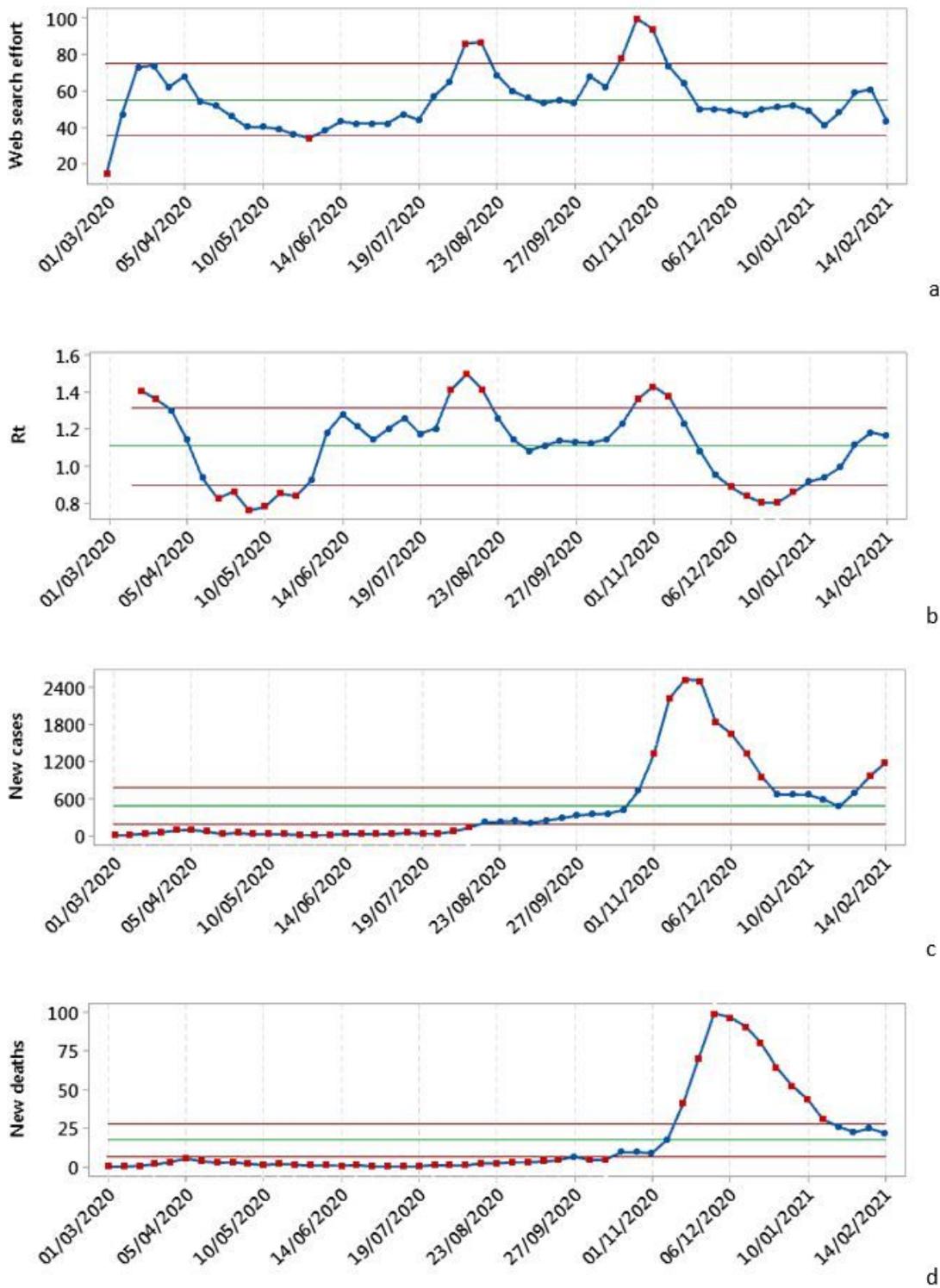


Figure 3

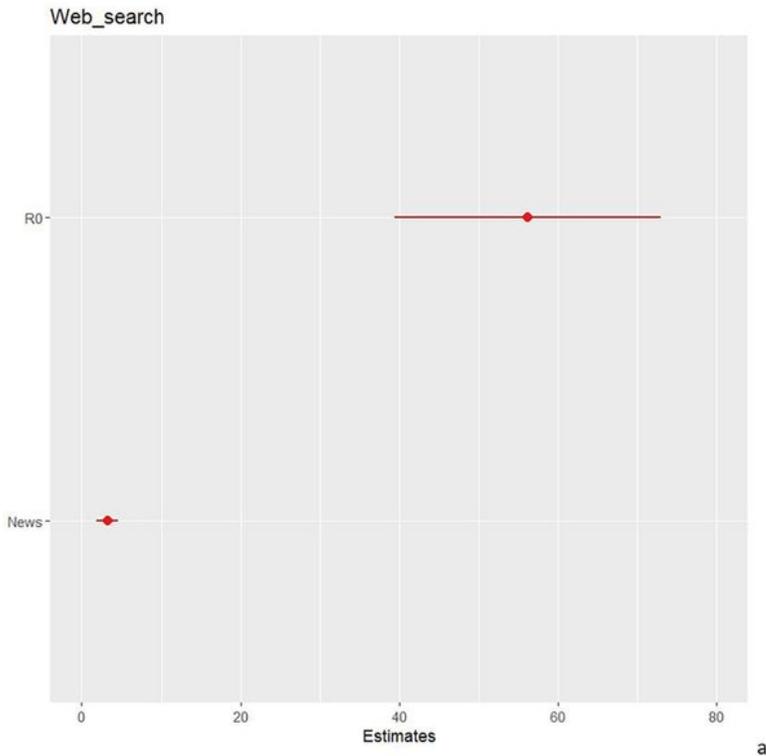
Temporal autocorrelation function of web search effort across weekly time lags. Values close to 1 or -1 indicate strong (positive or negative) correlation while values close to zero no correlation. The horizontal red lines indicate a 95% significance confidence intervals, while the vertical blue lines the actual correlation value for that time lag. Correlation values within the 95% confidence interval are not significant and they may not be differentiated from random correlations.



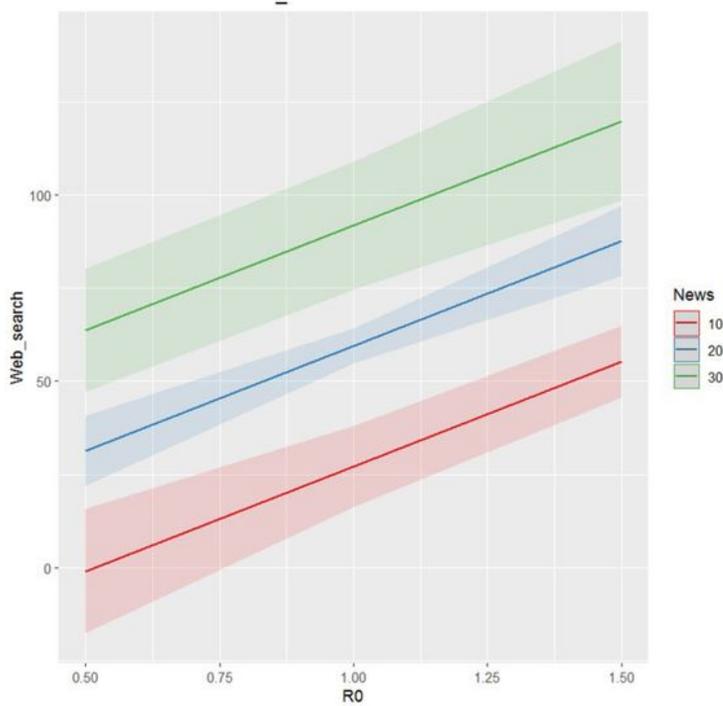
**Figure 4**

I-Charts of data in time on a weekly scale from the week 26 February to 1 March 2020 (week 1 when the first case was recorded), to 14 February 2021 (week 51). The horizontal green line indicates the grand mean, while the upper and lower red lines indicate three sigma above and below the mean respectively. The blue line indicates the actual values during that week. All values deviating the confidence intervals

are depicted in red symbols. a. I-chart of web search effort. b. I-chart of Rt. c. I-chart of new cases. d. I-chart of new deaths.



a



b

### Figure 5

Results from a linear mixed effects model (LME) between web search (dependent variable) and the independent fixed effects of news media cover and Rt. Time (week) was also included as a random effect. a. LME estimates of news cover and Rt on web search effort. Red circles indicate mean estimated

effect while whiskers indicate a 95% confidence interval. b. LME estimates of the effects of Rt across news media cover values. Higher news media cover values result in higher web search but web search is consistently increasing across Rt values.