



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Short Communication

Examining the effect of social distancing on the compound growth rate of COVID-19 at the county level (United States) using statistical analyses and a random forest machine learning model

J.S. Cobb ^{a,*}, M.A. Seale ^b^a Biomedical Materials Science, University of Mississippi Medical Center, 2500 N State St, Jackson, MS 39216, USA^b Information Technology Laboratory, US Army Engineer Research and Development Center, 3909 Halls Ferry Rd, Vicksburg, MS 39180, USA

ARTICLE INFO

Article history:

Received 10 April 2020

Accepted 13 April 2020

Available online 28 April 2020

Keywords:

Shelter-in-place

Social distancing

COVID-19

SARS-CoV-2

Machine learning

Statistics

ABSTRACT

Objectives: The goal of the present work is to investigate trends among US counties and coronavirus disease 2019 (COVID-19) growth rates in relation to the existence of shelter-in-place (SIP) orders in that county.

Study design: This is a prospective cohort study.

Methods: Compound growth rates were calculated using cumulative confirmed COVID-19 cases from January 21, 2020, to March 31, 2020, in all 3139 US counties. Compound growth was chosen as it gives a single number that can be used in machine learning to represent the speed of virus spread during defined time intervals. Statistical analyses and a random forest machine learning model were used to analyze the data for differences in counties with and without SIP orders.

Results: Statistical analyses revealed that the March 16 presidential recommendation (limiting gatherings to ≤ 10 people) lowered the compound growth rate of COVID-19 for all counties in the US by 6.6%, and the counties that implemented SIP after March 16 had a further reduction of 7.8% compared with the counties that did not implement SIP after March 16. A random forest machine learning model was built to predict compound growth rate after a SIP order and was found to have an accuracy of 92.3%. The random forest found that population, longitude, and population per square mile were the most important features when predicting the effect of SIP.

Conclusions: SIP orders were found to be effective at reducing the growth rate of COVID-19 cases in the US. Counties with a large population or a high population density were found to benefit the most from a SIP order.

© 2020 The Royal Society for Public Health. Published by Elsevier Ltd. All rights reserved.

Novel coronavirus (severe acute respiratory syndrome coronavirus 2 [SARS-CoV-2] or coronavirus disease 2019 [COVID-19]) originated in the province of Wuhan, China, in December 2019, after which it spread rapidly across the globe owing to infected persons exhibiting little to no symptoms within the first five days of contracting the virus.¹ The devastation and infection rate triggered by the virus caused the World Health Organization to declare it a global pandemic. There are currently more than 170 countries infected with COVID-19, and all 50 states in the United States (US) have confirmed cases according to the Centers for Disease Control and Prevention. In the US, community transmission has become the

prominent mode of transmission of the virus.² It has therefore become imperative that the effectiveness of the primary forms of limiting social contact used by local and national governments be evaluated. Enough data at the county level are now available to provide a fair assessment of the efficacy of the presidential guidelines issued on March 16, 2020, instituting a form of ‘social distancing’ by limiting gatherings to 10 or fewer people. Data are additionally sufficient to assess the efficacy of county-level shelter-in-place (SIP) orders versus counties that did not issue any SIP orders after the guidelines issued on March 16.

County metrics were obtained from the US Census Bureau, USA Counties (2011) data sets from the 2010 census.³ The data included in this study are latitude, longitude, population, median age, number of physicians, median income, population per square mile, and water use per capita. Counties were placed into one of the two

* Corresponding author. Tel.: +1 601 984 6170; fax: +1 601 984 6087.

E-mail address: jscobb@umc.edu (J.S. Cobb).

bins: (1) counties that had confirmed cases of COVID-19 before issue of guidelines on March 16 and experienced a SIP order on or after March 19 (186 counties, referred to as wSIP); and (2) counties that had confirmed cases before March 16 and experienced no SIP order (60 counties, referred to as noSIP). A Student t-test was used to compare two groups for significance. Analysis of variance with the Tukey post hoc test was used to compare multiple groups. Significance was defined as $P < 0.05$. All data are reported as mean \pm 95% confidence interval. There were no statistically significant differences in the US census data between the wSIP and noSIP groups apart from latitude ($P < 0.0001$) and the number of physicians ($P = 0.04$). The wSIP group had a latitude of 39.47 ± 0.75 , which places it in the northern US, compared with the noSIP group with a latitude of $34.6 \pm 1.16^\circ$, placing it further south. The difference in number of physicians is a function of latitude with a lower mean number of physicians in the south (1697 ± 500) compared with the north (2677 ± 538).

The number of confirmed COVID-19 cases in each county was collected from local health department data and county/state press releases from January 21 to March 25. Confirmed cases from March 26 to 31 were obtained from The New York Times coronavirus data repository.⁴ The two data sets were compared to ensure consistency between the collected values. Data collection stopped on March 31 because the mean number of days with confirmed cases was approximately the same as that before issue of guidelines on March 16, after March 16 but before the institution of a SIP order, and after March 16 with a SIP order. This allowed for comparison of these time intervals with an equal number of days (7.62 ± 0.35).

Compound growth was calculated using the following equation: (final confirmed cases/first confirmed cases)^(1/number of days). Fig. 1 shows the compound growth rate for the wSIP (1.39 ± 0.044) and noSIP (1.30 ± 0.059) groups before the issue of presidential guidelines on March 16, the compound growth rate after March 16 for the wSIP (1.30 ± 0.023) and noSIP (1.21 ± 0.016) groups, and the compound growth rate for the wSIP group (1.19 ± 0.011) after the SIP orders went into effect. The lower compound growth rates seen in noSIP data are due to the difference in latitude between the two

data sets and suggest that southern states experienced a slower spread of the virus at the onset. This makes sense, given that before March 16, the hot spots for COVID-19 were northern states such as Washington, New York, and Illinois. The noSIP compound growth rates were normalized to the compound growth rate of wSIP data before March 16 to account for geographical differences. The normalized compound growth rates after March 16 were shown to be statistically similar between the wSIP and noSIP groups ($P > 0.05$, Fig. 1). This indicates that the presidential guidelines had the same magnitude of effect on reducing the compound growth rate by $6.6 \pm 1.4\%$ between the wSIP and noSIP groups before the wSIP group instituted a SIP. After instituting a SIP order, compound growth rate of the wSIP group decreased an additional 7.8%, for a total decrease of $14.4 \pm 1.6\%$ from the compound growth rates before March 16. This indicates that the effects from the presidential guidelines and SIP orders were additive in the US. This is reasonable, considering the virus is thought to spread by virus-containing airborne droplets and orders for social distancing limit the interaction of people who could potentially be infected.⁵ A study modeling the effect of social distancing from China indicated a strong association with both a decrease in the rate the virus spreads and the implementation of social distancing.⁶

A random forest machine learning model was trained to predict the compound growth rate after a SIP order was given in a county. The random forest was chosen as it has been shown to have the highest accuracy in disease prediction.⁷ The model achieved a mean absolute percentage error of 92.3% in the test data set. The three most important features were population, longitude, and population per square mile in predicting the compound growth rate after a SIP order was instituted in a county. The data for these features were split into four equal groups to explain how these features matter for the model in predicting the compound growth rate after a SIP order was issued. Counties that instituted a SIP order with a longitude between -79.7102° and -97.2363° had the largest decrease in compound growth rate at 10.4% compared with 8.2% for counties outside of that longitudinal range. Counties with the highest populations between 143,962 and 984,8011 saw the largest

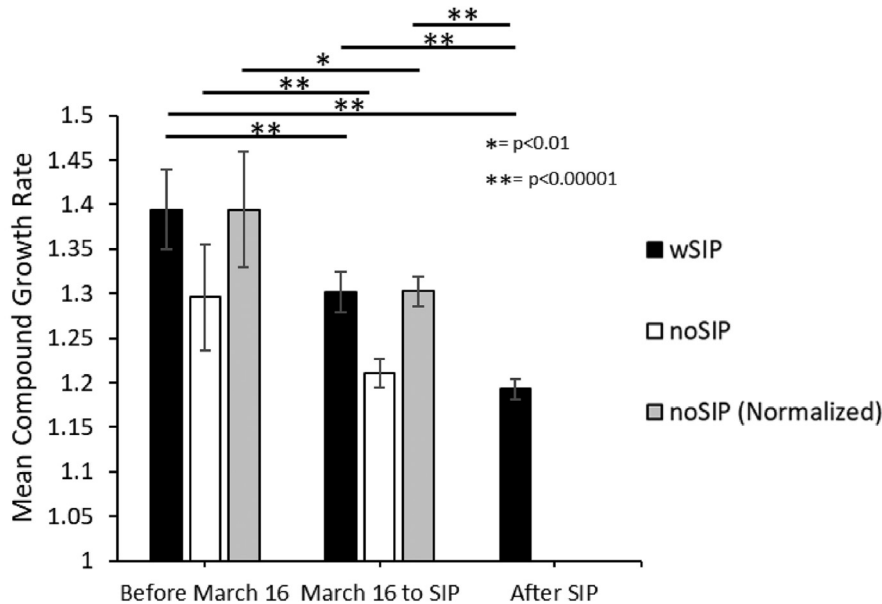


Fig. 1. Mean compound growth rate of counties that had confirmed cases (1) before the presidential guidelines issued on March 16, (2) from March 16 to issuance of SIP, and (3) from issuance of SIP to March 31 (black bars) compared with counties with confirmed cases (1) before March 16 and (2) after March 16 and noSIP (white bars). Statistical analyses indicated no differences in summary data for the two groups except for latitude, indicating that counties with a SIP order were further north and those without a SIP order were predominately located in the south. The counties without SIP were normalized to those with SIP before March 16 to account for this difference (gray bars). SIP = shelter-in-place; wSIP, with SIP; noSIP, no SIP.

percent reduction after instituting a SIP order at 10.5%, compared to counties with a lower population between 7457 and 142,151 at 8.2%. Similar to population, population per square mile showed the largest reduction in compound growth rate in counties with a population per square mile between 405.8 and 1755.5 at 11.6% compared with counties with a lower population density of 2.1–405.6 at 8.9%.

In conclusion, the data suggest that at a county level, in the US, the SIP order is effective at decreasing the compound growth rate of COVID-19 (Fig. 1). The counties that have the largest impact from a SIP order are ones with a large population or a high population density, as indicated by the random forest feature importance.

Author statements

Acknowledgments

The authors wish to thank William Glenn Bond and Meredith B. Cobb for their input and edits of the manuscript.

Ethical approval

Ethical approval was not required as this study made use of publicly available data.

Funding

None.

Competing interests

None declared.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.puhe.2020.04.016>.

References

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020. <https://doi.org/10.1056/NEJMoa2001316>.
2. Vukkadala N, Qian ZJ, Holsinger FC, Patel ZM, Rosenthal E. COVID-19 and the otolaryngologist - preliminary evidence-based review. *Laryngoscope* 2020. <https://doi.org/10.1002/lary.28672>.
3. US Census Bureau. *USA counties*. 2011. <https://www.census.gov/library/publications/2011/compendia/usa-counties-2011.html>. [Accessed 24 March 2020].
4. nytimes/covid-19-data <https://github.com/nytimes/covid-19-data> (accessed 27 March 2020).
5. Ong SWX, Tan YK, Chia PY, Lee TH, Ng OT, Wong MSY, Marimuthu K. Air, surface environmental, and personal protective equipment contamination by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from a symptomatic patient. *J Am Med Assoc* 2020. <https://doi.org/10.1001/jama.2020.3227>.
6. Chen X, Yu B. First two months of the 2019 coronavirus disease (COVID-19) epidemic in China: real-time surveillance and evaluation with a second derivative model. *Global Health Research and Policy* 2020;5(1). <https://doi.org/10.1186/s41256-020-00137-4>.
7. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inf Decis Making* 2019;19(1). <https://doi.org/10.1186/s12911-019-1004-8>.