



U.S. county level analysis to determine if social distancing slowed the spread of COVID-19

Tannista Banerjee¹ and Arnab Nayak²

Suggested citation Banerjee T and Nayak A. A county level analysis to determine if social distancing slowed the spread of COVID-19 in the United States. *Rev Panam Salud Publica*. 2020;44:e90. <https://doi.org/10.26633/RPSP.2020.90>

ABSTRACT

Objective. To analyze the effectiveness of social distancing in the United States (U.S.).

Methods. A novel cell-phone ping data was used to quantify the measures of social distancing by all U.S. counties.

Results. Using a difference-in-difference approach results show that social distancing has been effective in slowing the spread of COVID-19.

Conclusions. As policymakers face the very difficult question of the necessity and effectiveness of social distancing across the U.S., counties where the policies have been imposed have effectively increased social distancing and have seen slowing the spread of COVID-19. These results might help policymakers to make the public understand the risks and benefits of the lockdown.

Keywords

Coronavirus; pandemics; behavior, social; quarantine; United States

In March 2020, the World Health Organization (WHO) named a new category of coronavirus (SARS-CoV-2) that has started a global pandemic, causing the disease later named COVID-19. This virus is said to have existed in bats for a long time, before it transferred from bats to humans and then from human-to-human sometime by the end of 2019 in Wuhan, China. The disease spreads very rapidly from human-to-human as infected people transmit the virus when contaminating surfaces by touch or via droplets from coughing and sneezing [1-3]. According to the US Center for Disease Control and Prevention (CDC), the virus often spreads unbeknownst of a host with mild or no symptoms and yet kills many in its wake.

After COVID-19's emergence and spread in China, it started spreading globally via international air-travel and then through community spread within its new host countries. The WHO declared COVID-19 a global pandemic in early March 2020, and many countries' health organizations began warning about the extreme contagiousness of the disease. Given that there is no effective pharmaceutical intervention against the virus and as socialization in common spaces, including the workplace, is

the main source of infection, medical researchers worldwide advised early intervention in the form of strict social distancing as the most definitive tool to slow the virus' rapid spread and save thousands of lives [4]. Compelled by such warnings and some early validation of the effectiveness of the lockdown in China and a few other countries, governments across the globe have been forced to resort to extreme economy-wide lockdowns of all but the most essential services.

Economists and data scientists around the world have already started thinking about the economic and social effects of the COVID-19. In the online book *Economics in the Time of COVID-19* edited by Baldwin and Mauro [5] many different economic questions have been discussed by some leading economists. The book analyzes possible economic effects of COVID-19, including macroeconomic effects, financial effects and travel and trade sectors effects. However, the important question that we concentrated on in this paper is: How the US consumers' decisions to adhere to the social distancing regulations are affecting the spread of the virus? For example, it has been shown that non-pharmaceutical interventions like school

¹ Auburn University, Auburn, United States of America ✉ Tannista Banerjee, tz0018@auburn.edu

² Mercer University, Atlanta, United States of America

closures could lower the peak mortality of influenza pandemics [6]. The objective was to analyze different measures of social distancing using consumer movements from their home census blocks and to their work census blocks. This generated a good measure of social distancing for populations in the United States by each county for the last two months and allowed us to analyze the effects of the COVID-19 lockdown measures across and within these counties.

MATERIALS AND METHODS

Data

We connected several databases to create a complete database for our model design. Social distancing measures were created from Safegraph [7] “social distancing” database. Safegraph’s unique database provides daily mobile devices data for the U.S. and Canada. This database is collected by census block group level (12 digit FIPS codes). The period analyzed was from February 1, 2020 to March 31, 2020. The mobile device data tracks each consumer’s mobile device and provides raw device counts. The population covered in the database, and in this study, includes thousands of anonymous mobile devices’ customers from all across the U.S. states and territories. The number of total devices residing in homes in the census block by 12-digit FIPS code defines the total number of devices. Home is defined where the device spent last 6 weeks between 6 p.m. and 7 a.m. The total number of devices that did not leave their home location (geohash-7 measure) during the day defines the number of completely stay-at-home devices.

Full-time work location is specified if a device spent at least 6 hours a day between 8 a.m.- 6 p.m., at a location other than their home location for at least 6 weeks. Total number of devices at full-time work per day is provided by Safegraph.

Next, we obtained information on distance and time spent outside the home during the time. Median distance traveled from home is provided as the median distances travelled in meters by the devices from the home locations within a day (distances > 0). The database calculates the median across all of the devices (detailed description is available in [8]). *Median_home_time_{it}* is presented in minutes for all devices included in the total number of devices during the time period. Safegraph calculates this variable per devices by summing the observed minutes at home across the day. Then the database calculates the median of all these devices.

We obtained county level non-pharmaceutical intervention (NPIs) data from New York Times “See Which States and Cities Have Told Residents to Stay at Home” and Keystone “County level COVID-19 Non-pharmaceutical Database” [9]. The NPIs include local government imposed social distancing regulations, including social distancing regulations for vulnerable persons, social distancing of the general population, gathering size limitations, closure of public venues, closure of schools and universities, non-essential services closure and lock down between January 21, 2020 and March 31, 2020. The NPIs data is also defined by county.

County level COVID-19 infections data was obtained from the Centers for Disease Control and Prevention Coronavirus updates [10, 11]. The rural-urban characteristics of the county was obtained from the United States Department of

Agriculture Economic Research Service, Rural-Urban Continuum Codes [12].

Inclusion and exclusion criteria

We dropped duplicate observations from counties in Wyoming and some counties from Alaska for which county information was not available for any other data sources. We also dropped one single FIPS code for which there was some problem with data collection reported by Safegraph. Median distance traveled from home is measured in meter and excluded zero values.

Variable construction

We used the daily census block-level database to create daily county level data. For the *Completely_home_devices_{it}* and *Full_time_work_{it}* variables for the county *i* on day *t*, we took total sum of the variables by the county level.

For both *Median_home_time_{it}* and *Median_distance_traveled_{it}* variables we calculated the median weighted total number of devices in the census block *f* for county *i* in *t*. We used STATA function [aweight] in STATA16. Small geographical bias is possible as acknowledged by the data collectors. They tested the reporting bias and calculated it to be less than one percent and, therefore, the data is very accurate for this study. The weighted measures control for the effect of more populous counties.

Completely_home_devices_{it}, *Full_time_work_{it}*, *Median_home_time_{it}* and *Median_distance_traveled_{it}* variables were our social distancing measures. That is, social distancing in this study was measured by what proportion of a county’s population is staying home completely; how much time they were spending indoors, versus outdoors, in public spaces like working full-time, which is critical for this analysis. Further, we concentrated on social distancing measures at individual county level because we matched the above social distancing measures with county level COVID-19 infections data and NPIs data.

Analysis design

The complete data is detailed enough to help us to measure each consumer’s physical movement between counties and to different places of work. This allowed us to create a panel measure of social distancing by each county and enabled to design a difference-in-difference analysis of the impact of the lockdowns on the rate of spread of COVID-19 after controlling for all county, time and county-time fixed effects. Difference-in-difference analysis estimates the effect of NPIs through social distancing for the counties where NPIs were enacted compared to non-NPIs enacted counties. Thus we were able to filter out many unknown factors present, such as the numbers of tests done, availability of local test centers, general difference in demographic, and political and public health infrastructures across these counties, among others.

All the counties with NPIs enacted between February 1 and March 31, 2020 served as our treatment counties. If a county did not have any NPI then we considered that as a sample in the control group. We created a dummy variable, NPIs, which equals one if county *i* ordered NPIs on, or after day *t* (where day 1 starts February 1, 2020), and zero otherwise. Our treatment

counties were in states including New York and California, which have been extremely affected by COVID-19.

To analyze the effect of these social distancing measures on COVID-19 cases and how this effect is working on the treatment counties comparing to the control counties, we estimated the following difference-in-difference (DID) model:

$$\ln(\text{cases})_{it} = a_1 NPI_{it} + \lambda X_{it} + C_i + t + U_i + \varepsilon_{it} \quad (1)$$

where t represents the day starting from February 1, 2020 and i represents the county. The $\ln(\text{cases})$ presents the natural logarithm of the number of confirmed COVID-19 cases. We added one to the raw cases, before taking the logs, to control for zeros. NPI_{it} is our treatment variable. NPI_{it} is a dummy variable, which equals one if the county imposed NPIs on or before date t , and zero otherwise. The parameter a_1 measures the average effect of NPIs for county i after it was imposed on date t . X_{it} is a vector of interactions of the social distancing variables with the NPIs. The parameter λ measures the mean effect of NPIs as a result of the social distancing in county i comparing to control counties. We also included separate county, county-time and state fixed effects in C_i , time fixed effect (t) and a binary factor variable U_i as the urban-rural dummy variable that is equal to one if the county is an urban county. ε_{it} is the county and time specific error term.

Time invariant factors of a county or state, including geographical variables, political outlook, local public health and demographic differences, state infrastructure differences, etc., were controlled in our modelling design for by the county and state fixed effects. The time fixed effects captured the time varying pan-USA variables. More importantly, the county-time fixed effects were included to take account for any local county level time varying factors, such as local temperature variances as well as the number of test centers set up and the number of tests that were being administered, etc. As a further robustness measure, we clustered the standard errors at the county level. We re-estimated equation 1 with various lags.

RESULTS

We present the daily change in the natural logarithm of the confirmed number of cases [$\ln(\text{cases})$], daily number of completely stay at home devices and daily median time spent at home (minutes) for each day during the duration of our study, in Figure 1 (A, B and C). Figure 1A shows the highest affected states from the database, and it represents a sharp increase in the number of cases after March 4, 2020 with California and New York seeing the highest number of cases. Figure 1B depicts that, in these highly affected states, there was a sharp increase in percentage of devices stationed completely at home after March 15, 2020 and people started spending more time indoors after March 17, 2020. Together, figures 1B and 1C show that an increasing number of people started spending extended times at home after March 11. California observed the highest number of people staying at home till the 3rd week of March, but after that New Jersey was the state leading this measure. Table 1 presents the summary statistics of all the variables for all counties and also by treated and control groups.

As noted above, we estimated equation (1) with the right hand side treatment and interaction variables included at t , with a five days' and a fifteen days' lag. Column 1 of Table 2 shows

that after controlling for county, state, time and county-time fixed effects, counties where NPIs were enacted, full time work and distance-travelled-from-home increased the COVID-19 cases by 54% (p-value 0.001) and 13% (p-value 0.001), respectively. This might be because the first counties to have enacted an NPI are also those which were fast becoming the infection hot spots, combining with an artefact of the nature of the contagion that it can start an infection as early as within hours of contact. Column 2 of Table 2 shows that at the five-day lag of the interaction variables, distance-travelled-from-home increased the COVID-19 cases by 16% (p-value 0.001). Full-time-work variables after NPIs were imposed, were no longer significant. This is an interesting finding, as it might indicate exposure to the risks from full-time work not being significant anymore, as awareness of the virus increases, and most people who are sick are staying at home or quarantining. On the other hand, distance travelled from home now might indicate visits to the stores and other points of interest where likely the virus was spreading through droplets found in the air, or the other forms of spread discussed in the introduction.

After running the regressions with more days of lags for the interaction variables, we found significant negative effects of *Time-spent-at-home* at the fifteen-day lag for the counties with NPIs, as well as significant positive effects of $NPI^*full\text{-time-work}$ and $NPI^*distance\text{-travelled-from-home}$ in the treatment counties compared to control counties. Time spent at home decreased COVID-19 cases by 49%, 15 days after NPIs were enacted compared to control counties (column 3). After 15 days of enactment of the NPIs, the effects of full time work and distance travelled from home on COVID-19 infection increased to 84% and 25% (compared to 54% and 13% immediate effects). We have repeated the estimation with further days of lags, and found similar results till the 17th day lag. But given that much of the NPIs were enacted towards the end of our sample period, means we lose samples very quickly and the significance of the estimates disappears after the 17th day.

Figure 2 presents the effect on NPIs on the COVID-19 cases across two counties. The red line presents the change in cases for the county (e.g., LA county, CA) where NPI was enacted earlier and the blue line represents the change in COVID-19 cases for the county (e.g., Jefferson county, AL) where NPI was enacted later in the March.

DISCUSSION

COVID-19 has quickly made us realize that each of us, as socially responsible agents, have a major role to play in this difficult time. Yet, we also realize that we know very little about the nature of the control of the COVID-19 so far within the U.S. and globally. Active research is going on around the world to find a cure and vaccination for this deadly virus. However, it will probably be months before we could see any viable and effective vaccine for the COVID-19. Therefore, understanding the control measures of the viral disease is the most important question for the world at this current time. Non-pharmaceutical control measures against infectious diseases have been used throughout mankind's documented history. These measures have included school closure, as discussed and implemented for seasonal influenza pandemics [6]. The earliest literature studying the combined effect of quarantine, school closure, and workplace distancing on COVID-19 infections include the

FIGURE 1. (A) COVID-19 cases in natural logarithms (B) Proportion of population completely staying home (Represented here at state level for graphing. Analysis at the county level) (C) Median time spent at home (aggregated to state level for graphing. Estimation done at the county level)

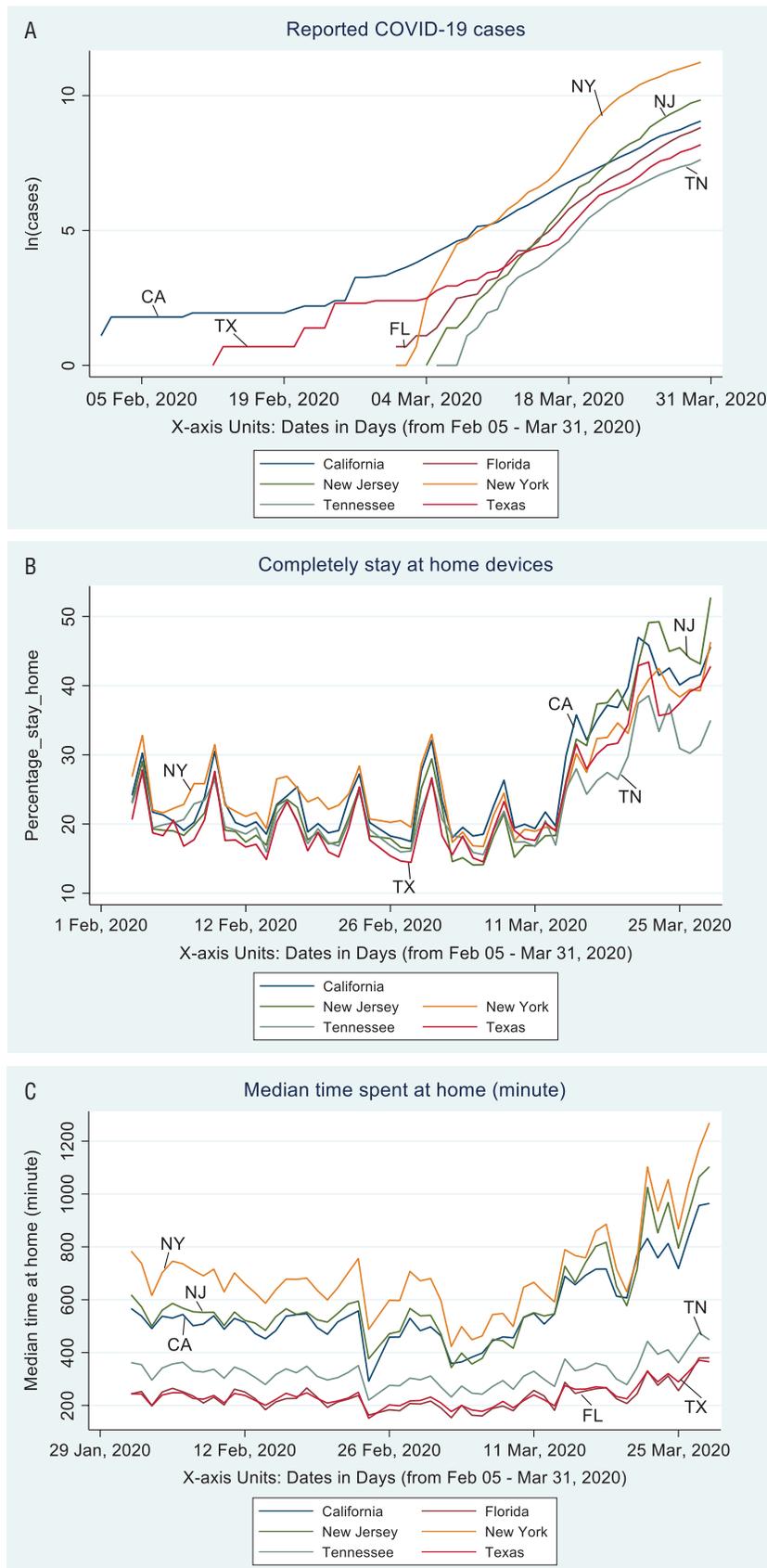


TABLE 1. Statistics summary of all variables

	All counties		Treated Counties		Control Counties	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
COVID-cases (number)	2.92	8.40	38.57	8.52	1.95	8.30
Completely-stay-at-home (number of devices)	24.02	7.70	24.38	8.25	24.01	7.68
Distance-traveled-from-home (meter)	29082.22	5459.11	17666.88	13018.8	29395.60	5534.98
Full_time_work (number)	17.76	4.89	17.38	4.73	17.77	4.89
Median_home_time (minute)	644.68	365.10	718.15	446.69	642.66	362.39
Total Observations	177029.00		4332.00		172697.00	

Note: Table 1 presents the summary statistics of different social distancing measures using the Safegraph data.

TABLE 2. Effect of social distancing on daily COVID-19 cases reported

	Dependent Variable		
	Natural log of Daily Confirmed COVID-19 Cases by County		
	Unlagged (1)	5-Day Lag effect (2)	15-Day Lag effect (3)
NPIs	-0.52*** (0.09)	-0.03 (0.01)	-0.03*** (0.01)
NPIs*Completely Home Devices	0.54 (0.94)	0.49 (0.96)	-0.15 (0.215)
NPIs*Time Spent at Home	-0.08 (0.12)	-0.19 (0.14)	-0.49*** (0.03)
NPIs*Full-time work	0.54*** (0.16)	0.27 (0.18)	0.84*** (0.02)
NPIs* Distance Travelled From Home	0.13*** (0.01)	0.16*** (0.01)	0.25*** (0.00)
Day Fixed Effects	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes
County day Fixed effect	Yes	Yes	Yes
Observations	1,77,029	1,44,551	80,315

Note: Table 2 presents the effect of social distancing on log(COVID cases) by county. Explanatory variables are the interaction variables. Robust standard errors are shown in the parenthesis. ***, **, and * denote 99%, 95%, and 90% levels of confidence.

NPI, non-pharmacological interventions

cross-country study [13] and a study based specifically on Singapore [14]. Both these studies found a negative effect of NPIs on COVID-19 infections. China also proved that aggressive quarantine measures can reduce the spread of the virus, but literature [15] suggests that the same mechanism might not work for other countries. Therefore, the estimation of the effect of NPIs through social distancing for the U.S. offers crucial insights. The U.S. is currently facing an increasing threat of infection, combined with the dilemma of the process of reopening the economy and risking thousands of lives to COVID-19 infections. This paper contributes to this pandemic literature by analyzing the exact lagged effect of the social distancing on the COVID-19 spread in the U.S. population. The novelty of the study is that it analyzed the U.S. populations' social distancing decision at the individual consumer level and estimated the impact of social distancing on the COVID-19 spread for the county. This analysis

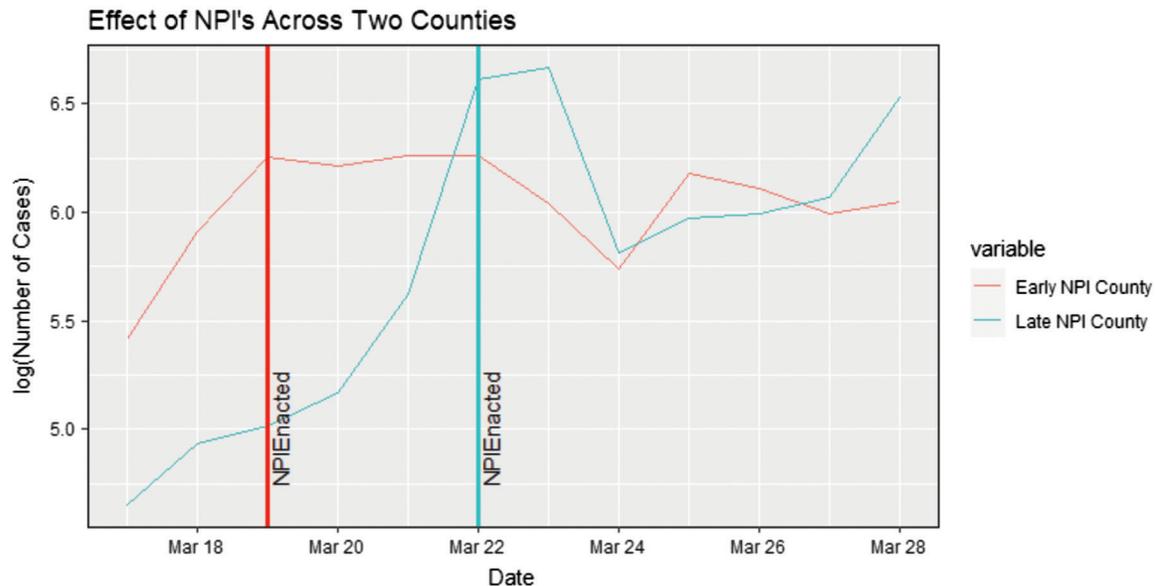
is disaggregated and an advancement over the literature which shows that social distancing, has a negative effect on the COVID-19 spread for the US population at the aggregate level [16].

This study had some limitations. Not all population use cell phones, and the social distancing data is collected from the cell phone data by Safegraph. A new survey conducted by the Pew research center shows that 95% the U.S. population own some kind of a cell phone and 81% own a smartphone [17]. However, the sample do not represent 100% of the population and if the cell phone is switched off when the person is at home or goes out of the house for a short duration and the cell phone does not ping, then the observation will be missing from the database. Given the vast coverage of the data, we are not concerned that the results will be influenced with this small bias. In future studies, it will be interesting to include additional measures of social distancing by the points of interest visits and the durations of these visits by consumers. For example, the visit to the grocery store, shopping malls, commercial establishments, airports or other locations affects the COVID-19 infections differently. This is outside the scope of this paper but, if undertaken, this analysis will provide a complete picture of how social distancing in different segments of the economy is affecting the public health.

Conclusion

This paper analyzes the U.S. consumers' decisions to adhere to the social distancing regulations and the effectiveness of social distancing on COVID-19. As people stay at home it can reduce the spread of the virus by 49% after two weeks of the social distancing decision, and as people start working full-time it increases the spread of the virus by 84% within two weeks. This result is close but more accurate to the prediction in the literature [18] that school closure and other NPIs reduce the COVID-19 cases by 60%. We conclude that as people spent more time at home; did not work full-time; and, traveled less distance from home it reduced COVID-19 infections for the county with about a two weeks lagged effect. Social distancing is important in controlling the infections and it is important to encourage these non-pharmaceutical intervention within each county.

Author contributions. TB received the original data. TB and AN planned the analysis, analyzed the data and interpreted the results. TB and AN wrote the paper. All authors critically revised the paper, reviewed and approved the final version. The

FIGURE 2. Effects of non-pharmacological interventions (NPI) across two counties with different NPI dates.

datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments. We thank Safegraph for generously providing the social distancing metrics database for this analysis, and Dr. Aditi Sengupta and Dr. Ayanangshu Nayak for their helpful comment on an earlier version of the manuscript.

Conflicts of interest. None declared.

Disclaimer. Authors hold sole responsibility for the views expressed in the manuscript, which may not necessarily reflect the opinion or policy of the RPSP/PAJPH and/or the Pan American Health Organization (PAHO).

REFERENCES

- Centers for Disease Control and Prevention. How COVID-19 Spreads. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html>. Accessed on March 10, 2020.
- Di W, Tiantian W, Qun L, Zhicong Y. The SARS-CoV-2 Outbreak: What We Know. *Int J Infect Dis.* 2020 March; 94:44-48.
- Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, Ji R, Wang H, Wang Y, Zhou Y. Prevalence of comorbidities in the novel Wuhan coronavirus (COVID-19) infection: a systematic review and meta-analysis. *Int J Infect Dis.* 2020 May; 94:91-95.
- Stephen SM. Global Infectious Disease Surveillance And Health Intelligence. *Health Aff.* 2007; 26(4):1069-77.
- Baldwin R, Mauro WB. Economics in the Time of COVID-19. 2020 March. Available from <https://voxeu.org/content/economics-time-covid-19>.
- Alexandra MS, Martin SC, Howard M. Closing The Schools: Lessons From The 1918-19 U.S. Influenza Pandemic. *Health Aff.* 2009;28.
- Safegraph. <https://www.safegraph.com/>.
- Safegraph detailed data manual. Available from: <https://docs.safegraph.com/docs/social-distancing-metrics>. Accessed on March 1, 2020.
- The New York Times Coronavirus stay at home order. Available from: <https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html?auth=login-google>. <https://www.keystonestrategy.com/coronavirus-covid19-intervention-dataset-model/>. Accessed on March 30, 2020.
- Centers for Disease Control and Prevention. Coronavirus updates. Available from <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html>. Accessed on March 30, 2020.
- USA facts, coronavirus location maps. Available from. <https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/>. Accessed on March 30, 2020.
- United States Department of Agriculture Economic Research Service, Rural-Urban Continuum Codes. Available from: <https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx>. Accessed on March 30, 2020.
- Dursun D, Enes E, Behrooz D. No Place Like Home: A Cross-National Assessment of the Efficacy of Social Distancing During the COVID-19 Pandemic. *JMIR Public Health Surveill.* 2020 May 20. Available from <https://preprints.jmir.org/preprint/19862/accepted>.
- Koo JR, Cook AR, Park M. Interventions to mitigate early spread of COVID-19 in Singapore: a modelling study. *Lancet Infect Dis.* 2020 March.
- Kupferschmidt K, Cohen J. China's aggressive measures have slowed the coronavirus. They may not work in other countries. *Science.* 2020 March. Available from: <https://www.sciencemag.org/news/2020/03/china-s-aggressive-measures-have-slowed-coronavirus-they-may-not-work-other-countries>.
- Charles C, Joseph G, Anh L, Joshua P, Aaron Y. Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth. *Health Aff.* 2020 May.
- Mobile fact sheet. <https://www.pewresearch.org/internet/fact-sheet/mobile/>. Accessed on March 30, 2020.
- Roy M A, Hans H, Don K, Déirdre HT. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet.* 2020 March; 395(10228):931-934.

Manuscript received on 23 April 2020. Revised version accepted for publication on 3 June 2020.

Análisis a nivel de condado para determinar si el distanciamiento social ralentizó la propagación de la COVID-19 en los Estados Unidos

RESUMEN

Objetivo. Analizar la efectividad del distanciamiento social en los Estados Unidos.

Métodos. Se empleó un método novedoso de contacto con teléfonos celulares (*ping*) para cuantificar las medidas de distanciamiento social de todos los condados de EE.UU.

Resultados. Usando un enfoque de diferencia en diferencias los resultados indicaron que el distanciamiento social ha sido efectivo para reducir la propagación de la COVID-19.

Conclusiones. A medida que los responsables de la formulación de políticas se enfrentan a la muy difícil cuestión de la necesidad y la eficacia del distanciamiento social en Estados Unidos, los condados en los que se han impuesto las políticas han aumentado efectivamente el distanciamiento social y en ellos se ha enlentecido la propagación de la COVID-19. Estos resultados pueden ayudar a los responsables de las políticas a hacer comprender a la población los riesgos y beneficios de las restricciones.

Palabras clave

Coronavirus; pandemias; conducta social; cuarentena; Estados Unidos
