

## **The Causal Effect of Social Distancing on the Spread of SARS-CoV-2**

Patrick Sharkey, PhD. Professor of Sociology and Public Affairs, Woodrow Wilson School of Public and International Affairs, Princeton University

George Wood, PhD. Postdoctoral Fellow Institute for Policy Research, Northwestern University

### **Abstract**

To what degree does social distancing have a causal effect on the spread of SARS-CoV-2? To generate causal evidence, we show that week to week changes in weather conditions provided a natural experiment that altered daily travel and movement outside the home, and thus affected social distancing in the first several weeks when Covid-19 began to spread in many U.S. counties. Using aggregated mobile phone location data and leveraging changes in social distancing driven by weekly weather conditions, we provide the first causal evidence on the effect of social distancing on the spread of SARS-CoV-2. Results show that a 1 percent increase in distance traveled leads to an 8.1 percent increase in new cases per capita in the following week, and a 1 percent increase in non-essential visits leads to a 6.9 percent increase in new cases per capita in the following week. Results are stronger in densely populated counties and close to zero in less densely populated counties.

## **The Causal Effect of Social Distancing on the Spread of SARS-CoV-2**

In the absence of a vaccine or therapeutic treatment, social distancing has become the primary non-pharmaceutical intervention (NPI) to control the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Cities, counties and states in the United States have adopted a range of efforts designed to limit travel, large groups, and face-to-face interaction in schools, workplaces, business establishments, and public spaces.

The staggering scale of behavioral change is unprecedented in modern American history. Considering the economic and social costs of social distancing and the disruption to everyday life, it is crucial to generate rigorous, causal evidence assessing the effect of social distancing in limiting the spread of SARS-CoV-2. Existing evidence comes from well-established epidemiological models of infection<sup>1-4</sup> and innovative observational evidence from historical pandemics.<sup>5-9</sup> However, we are aware of no research that relies on data from this pandemic and provides convincing causal evidence on a basic yet essential question: To what degree does social distancing have a causal effect on the spread of SARS-CoV-2?

This study develops a method designed to generate causal evidence to estimate the impact of social distancing on the spread of SARS-CoV-2. We draw on mobile phone location data to measure changes in total distance traveled and visits to non-essential establishments within counties, which we merge with county-level information on confirmed cases of SARS-CoV-2.

Simply documenting the association between changes in social distancing and the spread of SARS-CoV-2 will generate biased results because the greatest changes in behavior have taken place in the areas hit hardest by the virus, leading to a spurious positive association between

distancing and growth in SARS-CoV-2. To generate causal evidence we rely on an exogenous source of variation in social distancing: daily weather conditions. We show that week to week changes in precipitation provided a natural experiment that altered daily travel and movement outside the home, and thus affected social distancing in the early weeks of the Covid-19 outbreak in the U.S. By leveraging changes in social distancing driven by weekly precipitation, this method allows us to identify the causal effect of social distancing on the spread of SARS-CoV-2.

## Data and Methods

We combine data sources on confirmed cases of SARS-CoV-2, social distancing, weather conditions, and population characteristics in U.S. counties. Due to inconsistencies in data sources, we include only the contiguous 48 states.

**SARS-CoV-2 cases.** We measure growth in confirmed cases using data from the Coronavirus Resource Center run by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University.<sup>10</sup> Data from U.S. counties run from January 22 through April 23, 2020. The resource center defines confirmed cases to include presumptive positive cases and probable cases, consistent with guidelines established by the CDC. We include all measures of confirmed cases with a county identifier.

**Social distancing.** Data on social distancing were provided by Unacast at the level of U.S. counties. Unacast analyzes human mobility through location data available via mobile phones. The dataset begins on February 4 and runs through April 23, 2020. All measures of social distancing capture change in movement relative to a baseline period covering the four weeks prior to March 8, 2020. The dataset is based on 15 to 17 million unique “identifiers” per day, aggregated to the level of US counties.

We use two measures of social distancing. First, change in the average distance traveled per user, measured as the percentage change relative to the baseline period. Second, change in visitation to non-essential venues, including (but not limited to) restaurants, department and clothing stores, jewelers, consumer electronics stores, office supply stores, spas and hair salons, fitness centers, car dealerships, hotels, craft and hobby shops. Change in non-essential visits is also measured as the percentage change relative to the baseline period. Weekly measures used in all models represent the average change in distance traveled and the average change in non-essential visits, respectively, over the calendar week. Aggregating by week is important because of substantial day of the week effects on social distancing. By averaging over a full calendar week we eliminate any variation due to the specific days of the week in which distancing is measured.

Weather. Daily precipitation, relative humidity, and temperature data were obtained from the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction. Precipitation (mm) is measured at 0.25 resolution (0.25 degree longitude by 0.25 degree latitude) across the contiguous United States. Our main analyses use a dichotomous measure of any precipitation in a given day rather than total amount of precipitation—results are the same using either measure. Relative humidity (%) is measured at 2.5 resolution and temperature (C) at 0.25 resolution. We rasterized the weather data into cells and calculated the total precipitation, mean relative humidity, and maximum temperature per day within each county. As with all variables, each measure is averaged over calendar weeks. Precipitation thus represents the percentage of days in the week with any precipitation, high temperature and relative humidity measure the average high temperature and average level of relative humidity over the week.

## Empirical strategy

Our analysis builds on a common approach to modeling growth in infectious diseases within areas through the use of time series methods that adjust for contagion and autocorrelation across time periods.<sup>11,12</sup> However, we adapt these approaches to generate causal evidence on the impact of social distancing.

Identifying the causal effect of social distancing on the spread of SARS-CoV-2 is challenging for two reasons. First, counties where residents make more substantial changes to their daily patterns of movement in response to the pandemic likely differ in unobserved ways from counties where residents make less dramatic changes. For this reason, analyses comparing counties with more or less social distancing will generate biased estimates even when matching or controlling for observable characteristics of counties.

A second problem is that changes in behavior are driven, in part, by the spread of the virus. In the language of causal inference, social distancing is endogenous. This means that residents in areas where the virus is spreading more rapidly are more likely to make more substantial changes in behavior, leading to a spurious positive association between social distancing and the spread of SARS-CoV-2.

To deal with both of these challenges we carry out an analysis of change within counties using county fixed effects, and we exploit an exogenous source of variation in social distancing arising from weather patterns.<sup>13,14</sup> We utilize two stage least squares (2SLS) to carry out the analysis, with precipitation and temperature over a given week used as an instrumental variable that creates exogenous variation in social distancing in the same week.<sup>15,16</sup> Further, we control

for the existing cases in the county, for relative humidity and for time trends. eFigure 1 in the supplement to the article describes the logic of the analysis in a causal diagram.

The analysis is represented with the following system of equations.

$$(1) \text{Distancing}_{c,t-1} = \alpha + \text{Rain}_{c,t-1} + \text{Temp}_{c,t-1} + \text{Humidity}_{c,t-1} + \text{Cases}_{c,t-1} + C_c + \text{Week}_t + \eta_{ct}$$

$$(2) \text{Cases}_{ct} = \alpha + \text{Distancing}_{c,t-1} + \text{Humidity}_{c,t-1} + \text{Cases}_{c,t-1} + C_c + \text{Week}_t + e_{ct}$$

In Equation 1,  $\text{Distancing}_{ct}$  is the measure of social distancing in county  $c$  in week  $t-1$ , the week prior to the measurement of the outcome;  $\text{Rain}_{c,t-1}$  represents precipitation, measured as the percentage of days in the prior week ( $t-1$ ) with any precipitation in county  $c$ ;  $\text{Temp}_{c,t-1}$  represents the high temperature, measured as the average high temperature in the prior week ( $t-1$ ) in county  $c$ . Rainfall and temperature are used as instruments that generate exogenous variation in social distancing in week  $t-1$ .  $\text{Humidity}_{c,t-1}$  represents the relative humidity, measured as the average in the prior week ( $t-1$ ) in county  $c$  and used as a control variable;  $\text{Cases}_{c,t-1}$  represents the natural log of the number of existing cases in the county in the prior week ( $t-1$ ) per capita in county  $c$ ;  $C_c$  is a vector of dummy indicators for each county;  $\text{Week}_t$  is a vector of dummy indicators for the calendar week; and  $\eta_{ct}$  is an idiosyncratic error term for county  $c$  and calendar week  $t$ .

To identify the effect of social distancing in week  $t-1$  on new cases in week  $t$  we estimate Equation 2 using two stage least squares, where social distancing is instrumented with precipitation and temperature. The outcome, growth in confirmed cases of SARS-CoV-2, is

measured as the natural log of new cases per capita, in week  $t$ . We estimated standard county fixed effects and first difference models, both of which effectively deal with unobserved, time-invariant aspects of counties. In models shown in the text we use first differences because they are more efficient and align with the model of contagion more intuitively. We use population weights to improve the precision of the estimates. All estimates adjust for heteroskedasticity, for autocorrelation, and for other forms of clustering of residuals within counties.

Two central assumptions must be met in order to interpret the coefficient from the 2SLS estimate as causal. First, weather conditions must induce a change in social distancing behavior. To check this assumption, we report results from Equation 1 and demonstrate the strong relationship between weather conditions and the two measures of social distancing.

Second, there must be no other pathway by which the instrumental variable (weather conditions) affects the outcome of interest (new cases of SARS-CoV-2) other than through its effect on social distancing. This is referred to as the exclusion restriction (Fig. S1). In our case, one concern is that precipitation and temperature are associated with other weather conditions, most notably relative humidity, that may directly affect the transmission of SARS-CoV-2. Although the available evidence on the relationship between humidity and the spread of this virus is mixed and uncertain,<sup>17-19</sup> it is nonetheless important to address this possibility directly. To do so we control for relative humidity in the same week as precipitation, eliminating this pathway from our instrumental variable to the outcome. As we demonstrate, controlling for humidity has no impact on our results.

## Results

The sample consists of 788 US counties observed over four weeks, beginning in the first week in which the particular county reported at least 10 confirmed cases and continuing for the subsequent three weeks. Counties were included in the sample if they reported at least 10 confirmed cases by March 31, 2020, which enables us to measure growth in confirmed cases for at least three weeks following the tenth case. By construction, the sample is composed of larger counties where SARS-CoV-2 appeared relatively early in March 2020 (Table 1).

We estimate the effect of social distancing in a given week on growth in confirmed cases in the following week. As noted previously, the central challenge in identifying the effect of distancing on new cases is that distancing is endogenous to the spread of the virus. For this reason, a basic regression analysis with a cross-section of counties would show that counties with greater social distancing have more confirmed cases, a spurious relationship. Even considering change within counties over time, a naïve analysis would conclude that distancing has no effect on the growth of confirmed cases. This conclusion is incorrect and is driven by the fact that distancing behavior is changing the most in places where the virus is spreading fastest.

To deal with this problem we exploit an exogenous source of variation in distancing behavior arising from daily weather conditions. In the first stage of the model, we use precipitation and average high temperature in the given week to predict social distancing in the same week, while controlling for the log of existing cases per capita, relative humidity and the calendar week of the year (Table 2).

We find that a change from no precipitation over the week to precipitation every day of the week leads to a 1.9 percent reduction in distance traveled ( $b = -1.918$ , 95% CI  $-3.458$  to  $-.377$ ,  $p=.015$ ), and a 1 degree increase in average high temperature leads to a .22 percent increase

in distance traveled ( $b=.215$ , 95% CI .086 to .345,  $p=.001$ ). Because data on non-essential visits is not available for 21 counties, results from the model using non-essential visits as the measure of social distancing are slightly different. In this model, precipitation is not strongly related to non-essential visits, but a one degree increase in average high temperature leads to a .32 percent increase in non-essential visits ( $b = .315$ , 95% CI .165 to .464,  $p <.001$ ). Tests for “weak instruments” typically consider an F-statistic below 10 to indicate a weak instrument.<sup>20</sup> The F-statistic from the first stage is over 10 in both models, indicating that precipitation is strongly linked with social distancing behavior.

Results presented so far indicate that weather conditions generate substantial changes in social distancing. Our main interest is in using this source of exogenous variation driven by weather patterns to identify the effect of social distancing on the spread of SARS-CoV-2 (Table 3). We find that a 1 percent increase in distance traveled leads to an 8.1 percent increase in new cases per capita ( $b = .078$ ,  $\exp(b)=1.081$ , 95% CI .019 to .116,  $p=.005$ ) in the following week (Table 3, top panel). Results focusing on non-essential visits show that a 1 percent increase in non-essential visits leads to a 6.9 percent increase in new cases per capita ( $b = .067$ ,  $\exp(b)=1.069$ , 95% CI .019 to .116,  $p = .006$ ) in the following week (Table 3, bottom panel).

We carried out an additional analysis to assess whether the effect of social distancing varies depending on population density. We split the counties in our sample by the median population density and ran the same models shown previously. We find that precipitation is an equally strong predictor of social distancing in more and less densely populated counties within our sample (Table S1 and Table S2).

Second stage results focusing on distance traveled show no effect of social distancing in counties that are less densely populated ( $b = -.021$ , 95% CI  $-.072$  to  $.029$ ,  $p = .411$ ) (Table S1). In

more densely populated counties, we find an extremely strong effect. A one percent increase in distance traveled leads to a 10 percent increase in new cases per capita in the following week ( $b = .095$ ,  $\exp(b)=1.100$ , 95% CI .029 to .160,  $p = .004$ ) (Table S2). In results not shown we find the same pattern using non-essential visits as the measure of social distancing.

## Discussion

Existing evidence in support of social distancing as the primary non-pharmaceutical response to SARS-CoV-2 comes from sophisticated epidemiological models that rely on assumptions about the degree to which social distancing reduces interpersonal contact and infection, and from innovative observational studies of historical pandemics. No research that we are aware of provides direct empirical evidence identifying the causal effect of social distancing behavior on the spread of SARS-CoV-2. We draw on newly available mobile phone data and a method exploiting exogenous variation in distancing behavior within counties to generate such evidence.

We find that increasing distance traveled by 1 percent leads to an 8.1 percent increase in confirmed cases per capita in the following week. Similarly, increasing non-essential visits by 1 percent leads to a 6.9 percent increase in confirmed cases per capita in the following week. The findings are large in magnitude and robust to various adjustments to the model specification.

It is important to note that our results are limited in several ways. First, because we draw on data from counties where SARS-CoV-2 appeared relatively early, results are most applicable to larger and more densely populated counties. We found that the estimated effects are particularly large in counties with high population density within our sample, and were close to zero in less densely populated counties. Second, we are unable to distinguish between the effects

of particular NPIs designed to generate social distancing and limit movement and interaction.

Our results do not point to specific measures as being most effective, but rather indicate that any NPI that limits travel and non-essential visits will likely be effective in reducing the spread of SARS-CoV-2. Third, the analysis does not consider the costs of social distancing. Any decisions about NPIs must balance the benefits of reducing the spread of the virus with the costs associated with the intervention.

With those limitations in mind, the findings do provide strong support for efforts to control SARS-CoV-2 through policies intended to reduce non-essential travel and visits to non-essential establishments. Such policies have been justified by results from formal modeling exercises and by the experience of cities during the 1918 influenza pandemic. This study shows that the effectiveness of non-pharmaceutical interventions focusing on reducing movement and interventions is bolstered with causal evidence from the early period of the current pandemic.

## References

<sup>1</sup> N. M. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, I. Dorigatti, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, L.C. Okell, S. van Elsland, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang, P. G. T. Walker, C. Walters, P. Winskill, C. Whittaker, C. A. Donnelly, S. Riley, A. C. Ghani, “Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand.” (COVID-19 Response Team, Imperial College London, London, England, 2020; <https://doi.org/10.25561/77482>).

<sup>2</sup> N. M. Ferguson, D. A. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, D. S. Burke, Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437** (7056), 209- 214 (2005).

<sup>3</sup> Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. M. Leung, E. H. Y. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Y. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, Z. Feng, Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* **382**, 1199-1207 (2020).

<sup>4</sup> J. T. Wu, K. Leung, G. M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* **395** (10225), 689–697 (2020).

<sup>5</sup> M. C. Bootsma, N. M. Ferguson, The effect of public health measures on the 1918 influenza pandemic in U.S. cities. *Proc. Natl. Acad. Sci. U.S.A.* **104** (18), 7588-7593 (2007).

<sup>6</sup> R. J. Hatchett, C. E. Mecher, M. Lipsitch, Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proc. Natl. Acad. Sci. U.S.A.* **104** (18), 7582-7587 (2007).

<sup>7</sup> H. Markel, H. B. Lipman, J. A. Navarro, A. Sloan, J. R. Michalsen, A.M. Stern, M. S. Cetron, Nonpharmaceutical interventions implemented by US cities during the 1918-1919 influenza pandemic. *JAMA* **298** (6), 644-654 (2007).

<sup>8</sup> R.J. Barro, Non-Pharmaceutical Interventions and Mortality In U.S. Cities During the Great Influenza Pandemic, 1918-1919, NBER 27049 (2020).

<sup>9</sup> J. E. Aledort, N. Lurie, J. Wasserman, S. A. Bozzette, Non-pharmaceutical public health interventions for pandemic influenza: an evaluation of the evidence base. *BMC Public Health* **7** (2007).

<sup>10</sup> E. Dong, D. Hongru, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 10.1016/S1473-3099(20)30120-1 (2020).

- 11 C. Imai, B.A. Chisato, P.M. Zaid Chalabi, H. Masahiro Hashizume, Time series regression model for infectious disease and weather, *Env Res* **142** (2015).
- 12 C. Imai, H. Masahiro Hashizume, Systematic review on methodology: time series regression analysis for environmental factors and infectious diseases, *Trop Med Health* (2014).
- 13 B. Jacob, L. Lefgren, E. Moretti, The Dynamics of Criminal Behavior: Evidence from Weather Shocks, *J Hum Resour* **42** (2007).
- 14 E. Miguel, S. Satyanath, E. Sergenti, Economic Shocks and Civil Conflict: An Instrumental Variables Approach. *J. Polit. Econ.* **112** (4), 725-753 (2003).
- 15 J. D. Angrist, G.W. Imbens, D. B. Rubin, Identification of Causal Effects using Instrumental Variables. *J Am Stat Assoc* **91** (434), 444-455 (1996).
- 16 J. D. Angrist, J. Pischke, Mostly Harmless Econometrics: An Empiricist's Companion. (Princeton University Press, Princeton, NJ, 2008).
- 17 M. Martinez-Alvarez, A. Jarde, E. Usuf, H. Brotherton, M. Bittaye, A. L. Samateh, COVID-19 pandemic in west Africa. *Lancet Glob Health* **8** (5), E631-E632 (2020).

<sup>18</sup> M. Lipsitch, Seasonality of SARS-CoV-2: Will COVID-19 go away on its own in warmer weather? (Center for Communicable Disease Dynamics, Harvard University School of Public Health, Cambridge, MA, 2020).

<sup>19</sup> J. Shaman, E. Goldstein, M. Lipsitch, Absolute Humidity and Pandemic Versus Epidemic Influenza. *Am J Epidemiol* **173** (2), 127–135 (2011).

<sup>20</sup> J. H. Stock, M. Yogo, “Testing for Weak Instruments in Linear IV Regression.” In: Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, D. W. K. Andrews, J. H. Stock, Eds. (Cambridge University Press, New York, 2005) pp. 80-108.

Table 1. Descriptive statistics for 788 counties included in sample

	<b>N</b>	<b>%</b>
First week with 10 cases		
week ending March 3	2	<1%
week ending March 10	11	1%
week ending March 17	86	11%
week ending March 24	265	34%
week ending March 31	424	54%
	<b>Mean</b>	<b>SD</b>
Population	334,989	661,278
Growth of new cases per 100,000		
week of 10th case	18.48	29.91
3 weeks after 10th case	58.56	131.94
Log growth of new cases per 100,000		
week of 10th case	2.31	1.10
3 weeks after 10th case	3.23	1.28
Percentage change in distance traveled		
week prior to 10th case	-16.76	12.41
2 weeks after 10th case	-40.31	9.03
Percentage change in non-essential visits		
week prior to 10th case	-33.31	20.43
2 weeks after 10th case	-60.12	10.37
Precipitation (% of days with any)		
week prior to 10th case	0.65	0.25
2 weeks after 10th case	0.58	0.24
High temperature (Celsius, average over week)		
week prior to 10th case	15.58	7.31
2 weeks after 10th case	17.81	6.26
Relative humidity		
week prior to 10th case	74.88	9.95
2 weeks after 10th case	68.96	10.88

Table 2. Effects of weather conditions on social distancing

---

First stage model outcome: Percentage change in distance traveled

	b	95% CI		p-value
Precipitation	-1.918	-3.458	-0.377	0.015
High temperature	0.215	0.086	0.345	0.001
Log cases per capita	-0.052	-0.687	0.582	0.871
Relative humidity	0.006	-0.07	0.083	0.868

F statistic: 10.95

N = 2,364 county-week observations from 788 counties.

---

First stage model outcome: Percentage change in non-essential visits

	b	95% CI		p-value
Precipitation	-0.687	-2.13	0.756	0.351
High temperature	0.315	0.165	0.464	<.001
Log cases per capita	-0.638	-1.102	-0.173	0.007
Relative humidity	0.047	-0.002	0.097	0.058

F statistic: 15.20

N = 2,310 county-week observations from 770 counties with non-missing data on visits

---

All variables enter the model as first differences within counties.

Models include calendar week fixed effects not shown in the table.

Models are weighted by county population, standard errors are clustered within counties.

Table 3. Effects of change in distance traveled on new cases of Covid-19.

---

Second stage model outcome: New cases in following week (natural log)

	b	95% CI		p-value
% change in distance traveled	0.078	0.024	0.132	0.005
Log cases per capita	0.076	-0.077	0.228	0.332
Relative humidity	-0.002	-0.011	0.007	0.662

---

Second stage model outcome: New cases in following week (natural log)

	b	95% CI		p-value
% change in non-essential visits	0.067	0.019	0.116	0.006
Log cases per capita	0.114	-0.028	0.256	0.115
Relative humidity	-0.006	-0.015	0.003	0.192

---

All variables enter the model as first differences within counties.

Models include calendar week fixed effects not shown in the table.

Models are weighted by county population, standard errors are clustered within counties.

## Online Supplement

Sample and robustness. Results are robust to several minor changes in the model that we document here. First, in alternative models we use only precipitation to instrument for social distancing, with no substantive differences in results. Second, alternative models did not include the lagged measure of existing cases. Again, findings are not sensitive to this change. Third, we run alternative models that exclude weeks where there were 0 new cases or a decline in cases, resulting in negative growth. Because we use the natural log of new cases as our outcome measure, we assigned a value of 1 case in this scenario in order to avoid missing data. If we instead treat these as missing, results do not change.

Causal diagram (Figure S1). Our aim is to estimate the effect of distancing behavior within county  $c$  in the previous week  $t-1$  on the number of cases in county  $c$  in week  $t$ . By focusing on within-county differences in distancing over time, we ensure that unobserved differences between counties do not bias the estimated effect. However, the extent of distancing behavior within a county is likely to increase contemporaneously with the frequency of recorded cases, leading to a spurious positive association between distancing and cases (represented by the dotted line connecting cases at  $t-1$  to distancing at  $t-1$ ).

To identify the effect of distancing, we require an exogenous source of variation in distancing behavior in week  $t-1$ . Furthermore, this source of variation—or instrument—must only influence the frequency of cases in week  $t$  through its effect on distancing during  $t-1$ . We use weather conditions (labeled “precipitation” in the diagram) in week  $t-1$  as the instrument, and show that weather conditions in week  $t-1$  affect social distancing in the same week. However,

one concern is that precipitation is associated with humidity, which may itself be related to the spread of SARS-CoV-2, thereby opening a “backdoor path” between precipitation and new cases in week  $t$ . To account for this pathway and ensure that precipitation is a valid instrument, our model conditions on relative humidity.

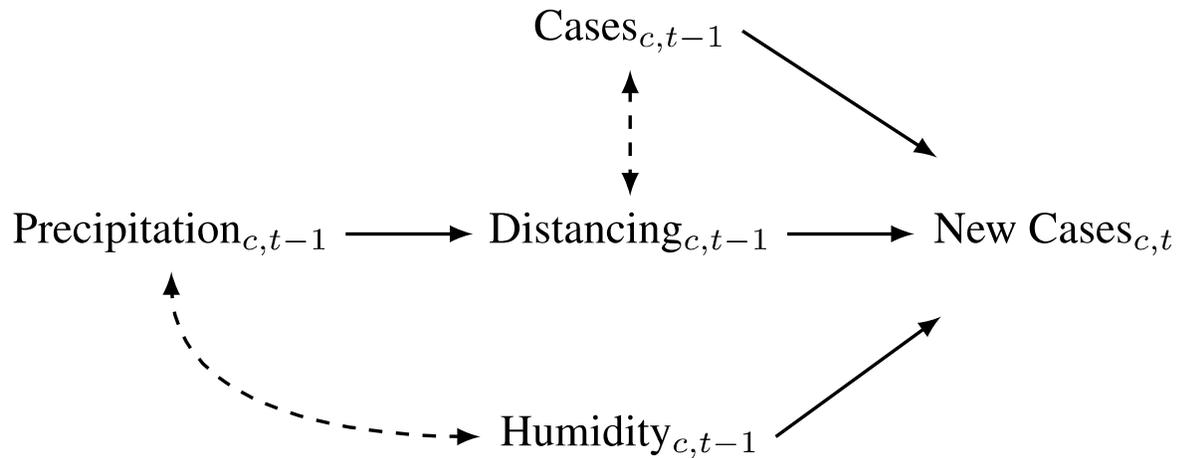


Figure S1: Directed Acyclic Graph (DAG) showing the causal identification strategy for the effect of distancing on cases.

Table S1. Effects of change in distance traveled on new cases of Covid-19 in low-density counties

First stage model outcome: Percentage change in distance traveled

	b	95% CI		p-value
Precipitation	-1.76	-3.046	-0.473	0.007
High temperature	0.215	0.069	0.36	0.004
Log cases per capita	-0.376	-1.323	0.571	0.437
Relative humidity	-0.112	-0.193	-0.031	0.007
F statistic: 12.11				

Second stage model outcome: New cases in following week (natural log)

	b	95% CI		p-value
% change in distance traveled	-0.021	-0.072	0.029	0.411
Log cases per capita	-0.113	-0.214	-0.011	0.03
Relative humidity	-0.014	-0.025	-0.002	0.02

N = 1,182 county-week observations from 394 counties.

First stage model uses precipitation and temperature to predict change in distance traveled.

Second stage model estimates causal effect of distance traveled on new cases of Covid-19.

All variables enter the model as first differences within counties.

Models include calendar week fixed effects not shown in the table.

Models are weighted by county population, standard errors are clustered within counties.

Table S2. Effects of change in distance traveled on new cases of Covid-19 in high-density counties

First stage model outcome: Percentage change in distance traveled

	b	95% CI		p-value
Precipitation	-2.032	-3.081	-0.261	0.025
High temperature	0.24	0.093	0.387	0.001
Log cases per capita	-0.048	-0.683	0.587	0.882
Relative humidity	0.051	-0.027	0.128	0.199
F statistic: 9.57				

Second stage model outcome: New cases in following week (natural log)

	b	95% CI		p-value
% change in distance traveled	0.095	0.029	0.16	0.004
Log cases per capita	0.094	-0.076	0.263	0.28
Relative humidity	-0.005	-0.016	0.005	0.34

N = 1,182 county-week observations from 394 counties.

First stage model uses precipitation and temperature to predict change in distance traveled.

Second stage model estimates causal effect of distance traveled on new cases of Covid-19.

All variables enter the model as first differences within counties.

Models include calendar week fixed effects not shown in the table.

Models are weighted by county population, standard errors are clustered within counties.