Social Distancing Metrics and Estimates of SARS-CoV-2 Transmission Rates: Associations Between Mobile Telephone Data Tracking and R

Christopher P. Morley, PhD; Kathryn B. Anderson, MD, PhD, MSPH; Jana Shaw, MD, MPH, MS; Telisa Stewart, DrPH; Stephen J. Thomas, MD; Dongliang Wang, PhD

ABSTRACT

Background: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of coronavirus disease 2019 (COVID-19). In the absence of robust preventive or curative strategies, the implementation of social distancing has been a key component of limiting the spread of the virus.

Methods: Daily estimates of $R(t)$ were calculated and compared with measures of social distancing made publicly available by Unacast. Daily generated variables representing an overall grade for distancing, changes in distances traveled, encounters between individuals, and daily visitation, were modeled as predictors of average R value for the following week, using linear regression techniques for 8 counties surrounding the city of Syracuse, New York. Supplementary analysis examined differences between counties.

Results: A total of 225 observations were available across the 8 counties, with 166 meeting the mean $R(t) < 3$ outlier criterion for the regression models. Measurements for distance ($\beta = 1.002$, $P = .012$), visitation ($\beta = .887$, $P = .017$), and encounters ($\beta = 1.070$, $P = .001$) were each predictors of $R(t)$ for the following week. Mean $R(t)$ drops when overall distancing grades move from D+ to C−. These trends were significant ($P < .001$ for each).

Conclusions: Social distancing, when assessed by free and publicly available measures such as those shared by Unacast, has an impact on viral transmission rates. The scorecard may also be useful for public messaging about social distance, in hospital planning, and in the interpretation of epidemiological models.

KEY WORDS: COVID-19, epidemiology, infectious disease, SARS-CoV-2, social distancing

Author Affiliations: *Department of Public Health & Preventive Medicine (Drs Morley, Shaw, Stewart, and Wang), Department of Family Medicine (Dr Morley), Department of Psychiatry & Behavioral Sciences (Dr Morley), Department of Medicine (Drs Anderson and Thomas), Institute for Global Health & Translational Science (Drs Anderson and Thomas), Department of Microbiology and Immunology (Drs Anderson and Thomas), Department of Pediatrics, Division of Pediatric Infectious Disease (Dr Shaw), Department of Urology (Dr Stewart), and Department of Geriatrics (Dr Stewart), SUNY Upstate Medical University, Syracuse, New York.*

All work was performed with internal resources and funding. The authors are grateful to Unacast (https:// www.unacast.com/) for making COVID-19 resources freely available.

None of the authors have real or apparent conflicts of interest with the study presented in this article.

Correspondence: *Christopher P. Morley, PhD, Department of Public Health & Preventive Medicine, SUNY Upstate Medical University, 475 Irving Ave, Ste 200, Syracuse, NY 13210 (morleycp@upstate.edu).*

DOI: 10.1097/PHH.0000000000001240

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of coronavirus disease 2019 (COVID-19).[1] The disease was first recognized with an outbreak of idiopathic pneumonia in Wuhan City, China, at the end of December 2019. On March 11, 2020, the World Health Organization declared COVID-19 a global pandemic.[2] As of April 30, 2020, the virus has resulted in approximately 3.3 million COVID-19 cases globally and more than 230 000 deaths.[3]

Once infected, individuals appear capable of transmitting the virus whether they are asymptomatic, presymptomatic, or symptomatic, making nonpharmacologic public health interventions challenging.[4] Variance in global testing capacity makes identification and isolation of all infected individuals, as well as tracking and monitoring of all exposed individuals, extremely difficult, bordering on impossible. In the absence of a safe and efficacious vaccine solution or prophylactic medications, public health efforts have been focusing on strict social distancing and hand and respiratory hygiene.[5] As a pathogen spread largely by droplet transmission, reductions in human movement and reducing human contacts have been viewed as critical in reducing transmission. Furthermore, social distancing has a history of demonstrated effectiveness in other settings, such as during the H1N1 influenza pandemic in 2009.[6-8]

In an effort to mitigate the scale of the pandemic, in March 2020, many states in the United States implemented social distancing by rolling out stay home orders and closing nonessential business and schools to slow down the spread of the virus.[9] Some, but not all, states enacted "shelter in place" or "stay-at-home" orders to further limit human contacts. Helpfully, platforms collecting and aggregating human movement information by tracking mobile phone data and global positioning system loggers became widely available at no cost and have been in use for more than a decade.[10,11] One such company, Unacast,[12] created an online platform that utilizes mobile phone data tracking to generate a score for gauging social distancing effectiveness in the United States, down to the level of the county.

In the face of the pandemic, many local departments of health, as well as health care organizations, have been conducting local epidemic modeling and surveillance operations. New York State has become a center for the epidemic in the United States, requiring significant planning and preparation on the part of hospitals and health care systems.[13] Our own region, located in the middle of New York State (Central New York, or CNY), has a metropolitan center in Syracuse, New York, located in Onondaga County. The county serves as a health care and commerce hub for a number of less densely populated counties surrounding it. Syracuse serves as the home of the region's only level 3 trauma hospital and academic medical center. Monitoring the course of the epidemic was therefore crucial to both population management and facility planning, in addition to general health messaging. As a part of this process, a team of public health scientists was creating epidemic models and generating a daily R value to estimate viral transmission. The R value refers to the reproduction number that describes an average number of new cases generated by an infected individual.[14] This is a moving number that requires regular calculation at regular time intervals. The R at any given time point is $R(t)$. An $R(t)$ value below 1 is an estimate that each infected individual will, on average, infect less than 1 new person. R values therefore offer an indication of whether an epidemic is growing or declining. It is also a crucial parameter in the estimate of SEIR epidemic models.[15]

Social distancing may be flattening the epidemic curve, but it is also blamed for severe economic consequences. It is therefore essential to demonstrate whether the costs of social distancing are having the desired effect. Furthermore, as communities contemplate the phased reopening of aspects of their economy, they will require real-time measures that correspond with risk of viral transmission. In this brief report, we present one such tool for tracking community contact rates and thus transmission potential. The analysis was conducted to monitor the local impact of social distancing measures.

Methods

To assess the impact of social distancing in CNY, variables representing publicly available mobile telephone movement data, tracked and graded by Unacast across 8 counties surrounding the city of Syracuse, New York, were assessed as predictors of weekly average rate of reproduction ($R(t)$) value, from time of first case (generally early mid-March; March 6 was the date of the first case in the region, in Herkimer County) to April 15, 2020, in each county. See Table 1 for notable COVID-19 milestone dates in CNY and first case presentations per county.

Counties analyzed represent the main urban center of the region (Syracuse), situated in Onondaga

TABLE 1
Notable Dates Relative to COVID-19 in Central New York

Events	Date
First case identified, per county	
Cayuga County	Mar 17, 2020
Cortland County	Mar 16, 2020
Herkimer County	Mar 6, 2020
Madison County	Mar 16, 2020
Oneida County	Mar 13, 2020
Onondaga County	Mar 10, 2020
Oswego County	Mar 17, 2020
Tompkins County	Mar 8, 2020
Onondaga County cancels St Patrick's parade/gatherings	Mar 12, 2020
School closings	First wave—Mar 16, 2020 Second wave—Mar 19, 2020
First COVID-positive case	Mar 16, 2020
Drive-up testing begins	
Restaurants close	
Stay-at-home order	Mar 22, 2020
Universal masking—Upstate	Mar 27, 2020
Universal masking—Business	Apr 15, 2020
Universal masking public	Apr 17, 2020

County, and 7 neighboring counties that feed patient flow to the Syracuse area: Cayuga, Cortland, Herkimer, Madison, Oneida, Oswego, and Tompkins counties.

Calculation of R(t)

We applied the method proposed by Cori et al[14] to estimate the time-varying $R(t)$ over 7-day window, using our daily incidence data and the mean and standard deviation of serial interval distribution, estimated by Du et al,[16] of 5 and 4 days, respectively.

Unacast data

Unacast[12] utilizes mobile telephone tracking data to calculate 4 variables representing different aspects of social distancing, down to the level of the county:

- *Daily distance difference* (Distance) evaluates the change in the overall average distance traveled, comparing pre-COVID (defined as before March 8, 2020) travel to the day of evaluation. Grades were assigned using the region demonstrating the strongest distancing (Italy) as a benchmark; they demonstrated a 70% to 80% reduction in movements. The averages for each day are compared with the corresponding days (ie, Friday pre-March 8, 2020, vs Friday post-March 8, 2020). A percent change is calculated and translated into a letter grade. The letter grade includes the following: A, more than 70% decrease; B, 55% to 70% decrease; C, 40% to 55% decrease; D, 25% to 40% decrease; F, less than 25%. (*Note:* Unacast numerical ranges are reported with overlapping values; these values were not pertinent to the analysis, as we utilized the continuous variables available for each measure.)
- *Daily visitation difference* (Visitation) evaluates the change in the nonessential visits. Essential venues include such places as food stores, pet stores, and pharmacies. Nonessential travel comprises places such as retail groups that have been determined to be nongrocery stores.
- *Daily encounters* (Encounters) evaluate the absolute value of the number of encounters compared with a national baseline. The variable represents a summation of encounters per square kilometer of land area for a given county. A potential human encounter is generated by 2 devices being in the same place at the same time regardless of prior human behavior. The encounter is defined by the space between 2 devices (≤50 m) and time (≤60 minutes). A national average encounter density score is calculated by the baseline measurement before the COVID-19 outbreak (February 10 to March 8, 2020). The scoring range includes the following: A, more than 94%; B, 82% to 94%, C, 74% to 82%; D, 40% to 74%; F, less than 40%.

Each of these 3 variables is represented as a negative scale, with a lower (more negative) number representing a larger reduction from the baseline. A positive relationship between each variable with $R(t)$ values would therefore represent a worse grade (less distancing).

In addition to the scale variables, Unacast represents county-level performance as ordinal A through F grades, where more than 70% reduction equals an "A." The numerical, ordinal equivalents are as follows:

5.0 = A
4.7 = A−
4.3 = B+
4.0 = B
3.7 = B−
3.3 = C+
3.0 = C
2.7 = C−
2.3 = D+
2.0 = D
1.7 = D−
1.3 or lower = F

Unlike the negative linear scale variables, the overall average variable moves inversely to $R(t)$, where a higher grade should hypothetically lead to a lower $R(t)$.

Analysis

Each of the 4 variables were modeled as simple predictors of weekly $R(t)$ using the AREG procedure in SPSS, v.26. AREG accounts for autocorrelation, and the Cochrane-Orcutt estimation was implemented with an AR1 covariance structure. The models were constructed where:

Mean $R(t)$ = The mean reproduction rate for a week in a county.

SDv = Each of the 4 social distancing variables aligned with the first day of each weekly average. The variables are expressed as change from baseline (so most are negative). A less negative (ie, larger) value indicates poorer performance on each measure for the county. Relationships between each SDv and $R(t)$ would therefore be larger and positive if a county's performance was poor, and $R(t)$ consequently rises.

RuralPct = Percentage of each county's population that qualifies as rural; this variable simultaneously controlled for county as an instrumental variable to control for subunit of heteroscedastic variance and for endogenous county characteristics.

We calculated both simple unadjusted and county covariate-adjusted models, represented by the following:

$$\text{Mean } R(t) = \text{SDv} + \hat{\varepsilon}$$

$$\text{Mean } R(t) = \text{SDv} + \text{RuralPct} + \hat{\varepsilon}$$

Each case represented 1 day in 1 county, with the social distancing variables for each day being matched with the mean $R(t)$ for the week that followed. So, for example, the social distancing variables for March 20 were matched with the mean $R(t)$ for the week of March 20 to March 26 for Onondaga county. This data structure allowed for the hypothesized temporal precedence of distancing leading to changes in $R(t)$ to be built into the models. Because the estimates of $R(t)$ in the first few days of each county's outbreak tended to be inflated, due to testing and case identification backlogs, only cases where mean $R(t)$ was less than 3 were included in the linear regression modeling procedures.

In addition, we calculated Pearson correlation coefficients for Distance, Visitation, and Encounters with mean $R(t)$ to further assess individual county effects. We also projected the relationship between the ordinal overall daily grade (A through F) and daily $R(t)$ value, with significance of differences in means assessed via analysis of variance. All procedures were conducted in SPSS, v.26, and checked in R. As all data were publicly available and aggregated, this study does not meet the criteria for human subject research.

TABLE 2

AR1 Linear Regression Models for Effect of Each Social Distancing Variable Upon Mean $R(t)$ per Week

	Unadjusted	Adjusted[a]
Overall grade	−0.297 ($P < .001$; $R^2 = 0.096$)	−0.298 ($P < .001$; $R^2 = 0.096$)
Distance	1.002 ($P = .012$; $R^2 = 0.039$)	1.007 ($P = .011$; $R^2 = 0.040$)
Visitation	0.887 ($P = .017$; $R^2 = 0.035$)	0.930 ($P = .014$; $R^2 = 0.038$)
Encounters	1.070 ($P = .001$; $R^2 = 0.069$)	1.702 ($P < .001$; $R^2 = 0.102$)

[a] *Adjusted for percent rural per county.*

Results

A total of 225 observations were available across the 8 counties, with 166 meeting the mean $R(t) < 3$ outlier criterion for the regression models (counties had variable number of days in the study period, from first case on March 6 to April 15 for last social distance measurement used in the analysis). Measurements for Distance ($\beta = 1.002$, $P = .012$), Visitation ($\beta = .887$, $P = .017$), and Encounters ($\beta = 1.070$, $P = .001$) were each predictors of $R(t)$ for the following week. These trends were robust to adjustment for the percentage of rural occupancy in each county, with Encounters ($\beta = 1.702$, $P < .001$) having the largest apparent effect when adjusted for rurality. Table 2 contains additional information. In addition, the overall grade was associated with mean $R(t)$ in both the unadjusted ($\beta = -.297$, $P < .001$) and adjusted ($\beta = -.298$, $P < .001$) calculations.

All 3 scale variables were correlated with mean $R(t)$ in all 8 counties. Visitation (essential visits) correlated more strongly with $R(t)$ in higher-density populations. See Table 3 for county-by-county Pearson correlation coefficients, ordered by county population density.

TABLE 3

Correlation Coefficient by Central New York County (Sorted by County Population Density)[a]

County	Encounters	Distance	Visitation	Density (per km^2)
Cayuga	0.953	0.566	0.74	45
Herkimer	0.674	0.608	0.56	46
Oswego	0.833	0.765	0.671	49
Madison	0.763	0.681	0.798	63
Oneida	0.885	0.594	0.806	75
Tompkins	0.744	0.612	0.836	80
Cortland	0.788	0.556	0.84	99
Onondaga	0.942	0.87	0.884	200

[a] *Numbers of essential visits correlated better with R in higher-density areas. All comparisons were significant with all Ps < .01.*

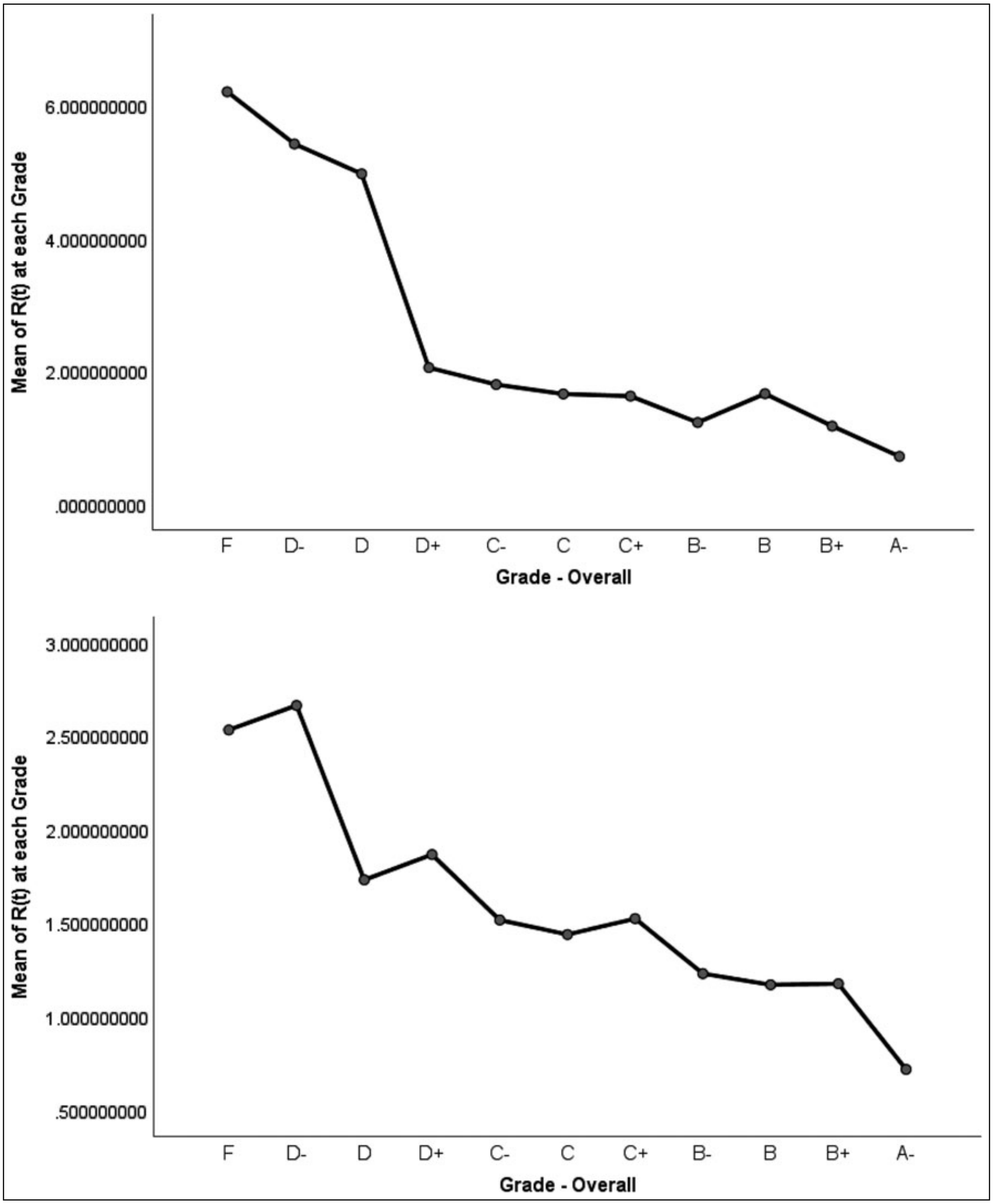

FIGURE Visualization of Mean *R*(*t*) by Ordinal Grade[a]: (A) Includes All Measurements and (B) Excludes Values of *R*(*t*) ≥ 3, to Eliminate Early Outlier Estimates
[a]Differences in both trends are significant at $P = .001$.

The overall grade for the day was also associated with mean $R(t)$, in both the full ($N = 225$) and outlier-restricted ($n = 166$) data sets. A distinct drop-off in mean $R(t)$ occurs when overall distancing grades move from D+ to C− and continues to drop as overall grades are higher. It is important to note that no county achieved an "A" rating (>70% reduction in overall social distancing) over the time period of our analysis. These trends were significant ($P < .001$ for each). See the Figure (parts A and B) for more detail.

Discussion

Social distancing has helped lower the transmission rate of SARS-CoV-2 and flatten the COVID-19 epidemic curve in CNY. Furthermore, the Unacast measures appear to be reasonable approximations for the extent of social distancing. While a rating of A− or higher may be necessary to reduce $R(t)$ below 1 (and hence stop viral transmission), moderate levels of social distancing, corresponding to Unacast grades of C− or higher, appear to have dropped $R(t)$ below 1.5.

There are several limitations to our study. The first is that a comparison with $R(t)$ daily measurement is not a comparison with the identification of new cases. Unfortunately, with a variety of tests in use throughout our region, with accompanying variation in lag times between symptom emergence, testing, and test result reporting, daily case counts are erratic. However, comparisons between the SEIR models we have generated and real-time surveillance data suggest that our calculations of $R(t)$ are reasonable approximations of epidemic trends in our region and have been consistent over time. In addition, we employed a *de facto* lag to examine the effect of Unacast scores on the average $R(t)$ value in the following week. There may be different lag periods that are more precise. Owing to the pressing nature of decision making around social distancing, however, we opted to quickly decide upon a lag period for the purposes of this report. A future study, informed by more data, should examine a wider range of lag periods. In addition, with more data, the relative importance of the different measures may become more apparent. For example, number of encounters was the most highly correlated of the 3 measures with $R(t)$. Other measures (distance and numbers of visits) are also correlated but limited by some lack of resolution. For example, delivery drivers are deemed to be "essential" workers but would appear in tracking as making repeated and multiple home visits and would not be discernible from casual visits between friends, for example.

In conclusion, our findings support the continued use of social distancing measures to reduce transmission of SARS-CoV-2. This conclusion is consistent with the rapidly emerging preprint literature, which includes a number of other studies attempting to measure the effect of social distancing and stay-at-home orders, and all of which find significant effects of distancing upon SARS-CoV-2 transmission through a variety of methods.[17-21] It is possible that moderate measures may be effective in slowing transmission, while balancing a slow and cautious reopening of some business and commerce activities with the protection of the health of the public. However, reopening businesses, although important for financial health, risks eroding the already fatigued public's resilience for continued social distancing. We would strongly urge caution in doing so, and employing social distance monitoring may be one tool local officials can use to determine the speed, extent, and, potentially, the need to reverse reopening initiatives. The monitoring of social distancing also is useful in interpreting epidemiological models and to inform the assumptions underlying those models. Finally, Unacast grading or similar distancing measures are potentially effective public communication tools to reinforce social distancing.

Implications for Policy & Practice

Why we undertook this analysis:

- Social distancing and the associated economic and social impacts of business and school closures, isolation, and deferred services are politically difficult to defend if they are not impactful on the spread of the SARS-CoV-2 virus.
- We analyzed the impact of social distancing orders on the weekly reproduction rate, or *R*(t), of the virus that causes COVID-19 in an 8-county region of New York State surrounding the primary academic health center in the area.
- Measurements of social distancing became publicly available via Unacast, via aggregated mobile phone data, making assessment of the impact of these policies on viral transmission possible.

What we found:

- Our local calculations of *R*(t) were associated with Unacast measures of social distancing.
- The Unacast grade on a given day was associated with the 7-day average *R*(t) for the week that followed.
- The better the distancing, the lower the *R*(t).

Implications

- In advance of more knowledge about the effectiveness of specific interventions (such as universal masking, hygiene, and physical distancing), broad social distancing appears to have been very useful in limiting the spread of SARS-CoV-2 at the beginning of the outbreak.
- Public health professionals may use this information to defend the initial measures, as well as future shelter-in-place or social distancing actions should the needs arise.

References

1. Fauci AS, Lane HC, Redfield RR. Covid-19—navigating the uncharted. *N Engl J Med*. 2020;382(13):1268-1269.
2. WHO Director-General's opening remarks at the media briefing on COVID-19—11 March 2020. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020. Accessed May 4, 2020.

3. Johns Hopkins Coronavirus Resource Center. COVID-19 United States cases by county. https://coronavirus.jhu.edu/us-map. Accessed May 4, 2020.
4. Furukawa NW, Brooks JT, Sobel J. Evidence supporting transmission of severe acute respiratory syndrome coronavirus 2 while presymptomatic or asymptomatic. *Emerg Infect Dis*. 2020;26(7). doi:10.3201/eid2607.201595.
5. Lasry A, Kidder D, Hast M, et al. Timing of community mitigation and changes in reported COVID-19 and community mobility—four U.S. metropolitan Areas, February 26-April 1, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(15):451-457.
6. Nasrullah M, Breiding MJ, Smith W, et al. Response to 2009 pandemic influenza A H1N1 among public schools of Georgia, United States—fall 2009. *Int J Infect Dis*. 2012;16(5):e382-e390.
7. Chowell G, Echevarría-Zuno S, Viboud C, et al. Characterizing the epidemiology of the 2009 influenza A/H1N1 pandemic in Mexico. *PLoS Med*. 2011;8(5):e1000436.
8. Herrera-Valdez MA, Cruz-Aponte M, Castillo-Chavez C. Multiple outbreaks for the same pandemic: local transportation and social distancing explain the different "waves" of A–H1N1pdm cases observed in Mexico during 2009. *Math Biosci Eng*. 2011;8(1):21-48.
9. See which states and cities have told residents to stay at home. *The New York Times*. https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html. Accessed May 4, 2020.
10. Shirley MDF, Rushton SP. Where diseases and networks collide: lessons to be learnt from a study of the 2001 foot-and-mouth disease epidemic. *Epidemiol Infect*. 2005;133(6):1023-1032.
11. Tatem AJ, Huang Z, Das A, Qi Q, Roth J, Qiu Y. Air travel and vector-borne disease movement. *Parasitology*. 2012;139(14):1816-1830.
12. Unacast. Schema for Covid-19 social distancing scoreboard. https://www.unacast.com/covid19/docs/schema-for-covid-19-social-distancing-scoreboard. Accessed April 26, 2020.
13. Singer AJ, Morley EJ, Henry MC. Staying ahead of the wave. *N Engl J Med*. 2020;382(18):e44.
14. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am J Epidemiol*. 2013;178(9):1505-1512.
15. Wang H, Wang Z, Dong Y, et al. Phase-adjusted estimation of the number of coronavirus disease 2019 cases in Wuhan, China. *Cell Discov*. 2020;6:10.
16. Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Meyers LA. Serial interval of COVID-19 among publicly reported confirmed cases. *Emerg Infect Dis*. 2020;26(6):1341-1343.
17. Fowler JH, Hill SJ, Levin R, Obradovich N. The effect of stay-at-home orders on COVID-19 cases and fatalities in the United States. *medRxiv*. 2020. doi:10.1101/2020.04.13.20063628.
18. Friedson A, McNichols D, Sabia J, Dave D. *Did California's Shelter-in-Place Order Work? Early Coronavirus-Related Public Health Effects*. Cambridge, MA: National Bureau of Economic Research; 2020. NBER Working Paper No. 26992.
19. Chen MK, Zhuo Y, de la Fuente M, Rohla R, Long EF. Causal estimation of stay-at-home orders on SARS-CoV-2 transmission. *ArXiv.org*. May 2020. https://arxiv.org/abs/2005.05469
20. Abouk R, Heydari B. The immediate effect of COVID-19 policies on social distancing behavior in the United States. *MedRxiv*. April 2020. doi: https://doi.org/10.1101/2020.04.07.20057356
21. Thakkar N, Burstein R, Hu H, Selvaraj P, Klein D. Social distancing and mobility reductions have reduced COVID-19 transmission in King County, WA. https://covid.idmod.org/data/Social_distancing_mobility_reductions_reduced_COVID_Seattle.pdf. Published March 2020. Accessed June 25, 2020.