Social Distancing Metrics and Estimates of SARS-CoV-2 Transmission Rates: Associations between mobile telephone data tracking and R

Christopher P. Morley, PhD[1,2]

Kathryn B. Anderson, MD PhD MSPH[3,4]

Jana Shaw, MD MPH,[5,1]

Telisa Stewart, DrPH[1,6]

Stephen J. Thomas, MD[7,4,10]

Dongliang Wang, PhD[1]

[1]Department of Public Health & Preventive Medicine, SUNY Upstate Medical University, Syracuse, NY

[2]Department of Family Medicine & Department of Psychiatry & Behavioral Sciences, SUNY Upstate Medical University, Syracuse, NY

[3]Department of Internal Medicine, Division of Hospital Medicine, SUNY Upstate Medical University, Syracuse, NY

[4]Department of Microbiology and Immunology, SUNY Upstate Medical University, Syracuse, NY

[5]Department of Pediatrics, Division of Pediatric Infectious Disease, SUNY Upstate Medical University, Syracuse, NY

[6]Department of Urology & Department of Geriatrics, SUNY Upstate Medical University, Syracuse, NY

[7]Department of Medicine, SUNY Upstate Medical University, Syracuse, NY

[8]Institute for Global Health & Translational Science, SUNY Upstate Medical University, Syracuse, NY

Corresponding Author:

Christopher P. Morley, PhD

Chair, Department of Public Health & Preventive Medicine

SUNY Upstate Medical University

Syracuse, NY

morleycp@upstate.edu

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 eight counties surrounding the city of Syracuse, NY. 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 (β=1.002, p=.001), Visitation (β=.887, p=.012), and Encounters (β=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.

Introduction

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 over 230,000 deaths.[3]

Once infected, an individual appears capable of transmitting the virus whether they are asymptomatic, pre-symptomatic, or symptomatic making non-pharmacologic public health interventions challenging.[4] Variance in global testing capacity make 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. Further, 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 US implemented social distancing by rolling out stay home orders and closing non-essential 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 over 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 U.S., down to the level of the county.

In the face of the pandemic, many local departments of health, as well as healthcare organizations, have been conducting local epidemic modeling and surveillance operations. New York State has become a center for the epidemic in the U.S., requiring significant planning and preparation on the part of hospitals and healthcare systems.[13] Our own region, located in the middle of New York State (Central New York, or CNY), has a metropolitan center in Syracuse, NY, located in Onondaga County. The county serves as a healthcare 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 were 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 which requires regular calculation at regular time intervals. The R 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 one 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. Further, as communities contemplate the phased re-opening 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.

Methods

In order 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, NY, were assessed as predictors of weekly average rate of reproduction (Rt) value, from time of first case (generally early-mid March) to April 15th, 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 County; and seven 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 days and 4 days, respectively.

Unacast Data

Unacast[12] utilizes mobile telephone tracking data to calculate four 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 70-80% reduction in movements. The averages for each day are compared to the corresponding days (i.e 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: A: > 70% decrease; B: 55-70% decrease; C: 40-55% decrease; D: 25-40% decrease; F: <25%

- *Daily Visitation Difference* (Visitation) evaluates the change in the non-essential visits. Essential venues include such places as food stores, pet stores, and pharmacies. Non-essential travel comprises of places like retail groups that have been determined to be non-grocery stores.

- *Daily Encounters* (Encounters) evaluates the absolute value of the number of encounters, compared to 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 two devices being in the same place at the same time regardless of prior human behavior. The encounter is defined by the space between two devices (50 meters or less) and time (60 minutes or less). A national average encounter density score is calculated by the baseline measurement before the COVID-19 outbreak (February 10-March 8, 2020). The scoring range includes: A:>94%; B: 82-94%, C: 74-82%; D 40-74%; F: <40%

Each of these three 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 that a worse grade (less distancing).

In addition to the scale variables, Unacast represents county-level performance as ordinal A through F grades, were >70% reduction equals an 'A'. The numerical, ordinal equivalents are:

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), were a higher grade should hypothetically lead to a lower R(t).

Analysis

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

MeanR(t) = the mean reproduction rate for a week in a county

SDv = Each of the four social distancing variables, aligned with the first day of each weekly average

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 heteroskedastic variance, and for endogenous county characteristics.

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

MeanR(t) = SDv + ὲ

MeanR(t) = SDv +RuralPct + ὲ

Each case represented one day in one 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 20th were matched with the mean R(t) for the week of March 20th – March 26th 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.

Additionally, we calculated Pearson correlation coefficients for Distance, Visitation, and Encounters with Mean R(T), in order 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.

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 (β=1.002, p=.001), Visitation (β=.887, p=.012), and Encounters (β=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 (β=1.702, p<.001) having the largest apparent effect when adjusted for rurality. **Table 2** contains additional information. Additionally, the overall grade was also associated with Mean R(t) in both the unadjusted (β=-297, p<.001) and adjusted (β=-.298, p<.001) calculations.

All three 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.

The overall grade for the day was also associated with Mean R(t), in both the full (N=225) and outlier-restricted (n=166) datasets. 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 **Figure 1a-b** for more detail.

Discussion

Social distancing has helped lower the transmission rate of SARS-CoV-2, and to flatten the COVID-19 epidemic curve in CNY. Furthermore, the Unacast measures appears 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. Additionally, 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. Additionally, with more data, the relative importance of the different measures may become more apparent. For example, number of encounters was the mostly highly correlated of the three 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. It is possible that moderate measures may be effective in slowing transmission, while balancing a slow and cautious re-opening 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.

Acknowledgements

We are grateful to Unacast (https://www.unacast.com/) for making COVID-19 resources freely available. Roles: CP Morley led the analysis; D Wang calculated R(t) daily and acquired covariate data; KB Anderson rechecked & verified analyses, and conducted supplementary analyses; TM Stewart facilitated access to Unacast data; J Shaw contributed background research, and was a key clinical consultant; S Thomas was a key interpreter of the data. All authors participated in interpretation of the data and writing of the manuscript.

References

1. Fauci AS, Lane HC, Redfield RR. Covid-19 - Navigating the uncharted. N. Engl. J. Med. 2020;382(13):1268–9.
2. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020 [Internet]. [cited 2020 May 4];Available from: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020
3. COVID-19 United States Cases by County - Johns Hopkins Coronavirus Resource Center [Internet]. [cited 2020 May 4];Available from: https://coronavirus.jhu.edu/us-map
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).

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):451–7.
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).
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).
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 México 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 [Internet]. [cited 2020 May 4];Available from: https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html
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):1023–32.
11. Tatem AJ, Huang Z, Das A, Qi Q, Roth J, Qiu Y. Air travel and vector-borne disease movement. Parasitology 2012;139(14):1816–30.
12. Schema for Covid-19 Social Distancing Scoreboard [Internet]. [cited 2020 Apr 26];Available from: https://www.unacast.com/covid19/docs/schema-for-covid-19-social-distancing-scoreboard
13. Singer AJ, Morley EJ, Henry MC. Staying Ahead of the Wave. N Engl J Med 2020;
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):1505–12.

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(1).

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).

Table 1 - Notable dates relative to COVID-19 in CNY

Events	Date
First Case identified, Per County	
Cayuga County	17-Mar-20
Cortland County	16-Mar-20
Herkimer County	6-Mar-20
Madison County	16-Mar-20
Oneida County	13-Mar-20
Onondaga County	10-Mar-20
Oswego County	17-Mar-20
Tompkins County	8-Mar-20
Onondaga County Cancels St. Patrick's parade/gatherings	12-March-20
School closings	1st wave - 16-March-20 2nd wave - 19-March-20
First COVID pos case	
Drive-up testing begins	16-March-20
Restaurants close	
Stay-at-home order	22-March-20
Universal masking-Upstate	27-March-20
Universal masking -Business	15-April-2020
Universal Masking Public	17-April-20

Table 2 - AR1 Linear regression models for effect of each Social Distancing variable upon Mean R(t) per week

	Unadjusted	Adjusted*
Overall Grade	-.297 ($p<.001$; $R2=.096$)	-.298 ($p<.001$; $R2=.096$)
Distance	1.002 ($p=.012$; $R2 =.039$)	1.007 ($p=.011$; $R2=.040$)
Visitation	.887 ($p=.017$; $R2=.035$)	.930 ($p=.014$; $R2=.038$)
Encounters	1.070 ($p=.001$; $R2=.069$)	1.702 ($p<.001$; $R2=.102$)

*Adjusted for percent rural per county

Table 3 - Correlation coefficient by Central New York county (sorted by county population density)

County	Encounters	Distance	Visitation	Density (per km2)
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

Numbers of essential visits correlated better with R in higher-density areas

All comparisons were significant with all $p<0.01$

Figure 1 a-b: Visualization of Mean R(t) by ordinal grade. Figure 1a includes all measurements; Figure 1b includes measurements of $R(t)<3$, to eliminate early outlier estimates. Differences in both trends are significant at $p=.001$.