Small-Area Projections of COVID-19 Transmission in the United States

Background

The novel coronavirus SARS-CoV-2 has caused a pandemic that has resulted in symptomatic infection in nearly 700,000 individuals with nearly 36,000 deaths in the United States by April 17, 2020. Current models project approximately 200,000-300,000 deaths in the United States due to SARS-CoV-2 in the next few months. These models, however, project for large areas, using a fixed R estimated from national data, do not consider city characteristics (e.g. population density, commuter traffic) that could impact transmissibility, and don't directly measure the effect of social distancing. In addition, these models do not consider time-varying exposures such as temperature and humidity that might affect SARS-CoV-2 transmission. There is an urgent need to project how SARS-CoV-2 transmission will occur in the coming months more selectively across US cities that are very different geographically and with respect to the underlying risks of their populations. Our objective is to project the trajectory of the COVID-19 epidemic in major US cities by assuming that R is not fixed, will vary significantly across the country, and will vary specifically in relationship to temperature, humidity, and social distancing strategies. In conjunction with other national models, these data can provide complementary information to better inform decisions to reopen local communities in the weeks ahead.

Methods

We have selected 260 counties representing 58% of the total US population based on the following characteristics: 1) county containing at least one city with population exceeding 100,000 persons; 2) county includes a state capital; or 3) county is the largest within a less populous state. From these 260 counties we model only those whose SARS-CoV-2 daily case rates exceeded an absolute count of five, thus prioritizing counties within a minimum threshold of disease activity. In early days of a county's outbreak, when the ratio of cases/tests was unstable, kernel smoothing was performed to account for the likelihood that cases from prior days would accumulate when testing capacity increased. Daily case counts of SARS-CoV-2 infection by county were obtained from The New York Times and USAFACTS (usafacts.org). We obtained data about county and population characteristics from the American Community Survey, Behavioral Risk Factor Surveillance System and other surveys of the population. For each county, we obtained data on social distancing defined as the percent change in visits to nonessential businesses in each county. We obtained daily wet-bulb temperatures, which measure the combined effect of temperature and humidity, from the National Oceanic and Atmospheric Administration Local Climatological Data from 2010 to the current date. Estimates for R values in each county were informed by methods from Wallinga & Teunis 2004 and Cori et al. 2013. Distributed lag non-linear mixed effects models (Gasparrini et al 2010) allowed us to examine simultaneously the cumulative exposure-response relationship between daily temperatures over a lag period of 3 to 14 days with R, alongside other local factors, using a random intercept for county. Other factors included social distancing, population density, percentage of population 65 years and older, percent of population under 18 years of age, percent of population over 65 years of age, proportion of uninsured individuals, percent of population below 200 percent of poverty level, prevalence of diabetes, and prevalence of smoking behavior. Interaction effects are identified between temperature and population density, and between social distancing and population density. Post-estimation projections of fitted outbreak curves are elaborated for each county over time.

Preliminary Findings

The trajectory of real-time R has varied greatly across counties, with variance in both the peak and slope of R over time and resultant surge of SARS-CoV-2 cases. Social distancing, population density, and temperature significantly affect the estimations of R over time, while standardizing on population characteristics. The ability of rising temperatures to mitigate transmissibility (or Rs) erodes for counties with higher population density. Estimates have been used to model projected incidence over time as a function of each county's characteristics and in relationship to a five-year average of wet-bulb temperatures across the regions and changing assumptions of social distancing as economies reopen.

Implications
These small area analyses allow real-time modeling of SARS-CoV-2 transmission and provide complementary information to the other national models currently used to project coronavirus transmission. The consideration of county-level characteristics, daily temperatures, and direct measurement of the effect of social distancing directives reveal marked heterogeneity in the magnitude and timing of surges of SARS-CoV-2 cases in the Unites. These results might inform consideration of selective strategies to mitigate SARS-CoV-2 transmission and assess healthcare system capacity constraints across the United States.

Statistical References:

Wallinga, Jacco, and Peter Teunis. "Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures." *American Journal of epidemiology* 160, no. 6 (2004): 509-516.

Cori, Anne, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. "A new framework and software to estimate time-varying reproduction numbers during epidemics." *American journal of epidemiology* 178, no. 9 (2013): 1505-1512.

Gasparrini A1, Armstrong B, Kenward MG. Distributed lag non-linear models. *Statistics in Medicine*. 2010 Sep 20;29(21):2224-34. doi: 10.1002/sim.3940.

We are available to discuss our models at your earliest convenience.
David Rubin, MD, MSCE (rubin@email.chop.edu)
Greg Tasian, MD, MSc, MSCE (tasiang@email.chop.edu)
Jing Huang, PhD (huangj5@email.chop.edu)