County-level Projections of COVID-19 Transmission in the United States

Background
The novel coronavirus SARS-CoV-2 has caused a pandemic that has already infected 6.1 million people with more than 185,000 deaths in the United States as of September 2, 2020. As local areas are responding to emerging risk of outbreaks, most models that are predicting the case burden over the next few months project for large areas (state or national), using a fixed reproduction number (R) estimated from national data, do not consider city characteristics (e.g., population density, demographics) 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 remains an urgent need to project how SARS-CoV-2 transmission will occur week to week more selectively across U.S. 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 local U.S. communities 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 guide communities as they respond to emerging risk in the weeks ahead.

Methods
We have expanded our projections to include 819 counties with active outbreaks (up from 211 in our original analyses), representing 82% of the total U.S. population, which have been curated in several ways. The principal inclusion criteria since the project's inception has been 1) counties with at least 100,000 residents or counties with the state capital city; 2) counties with more than 3 days of daily case counts exceeding 5, as of June 7, 2020; and 3) counties with average daily case counts exceeding 5 (minimum threshold activity), between the date of disease outbreak in the county and July 27, 2020. We have now also added counties with a population exceeding 40,000 residents and a minimum population density of 250 people per square mile, and have had at least 5 daily cases between June 23 and July 6, 2020. Additionally, we now include any county with an average of 20 or more daily cases between June 23 and July 6, 2020, regardless of other inclusion criteria.

Daily case counts of SARS-CoV-2 infection by county are obtained from The New York Times and USAFACTS (usafacts.org). We obtain 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 obtain data on social distancing, defined as the percent change in visits to non-essential businesses in each county (Unacast). We obtain daily wet-bulb temperatures (an index that combines temperature and humidity) from the National Oceanic and Atmospheric Administration Local Climatological Data from 2010 to the current date. R estimates in each county are informed by methods from Wallinga & Teunis 2004 and Cori et al. 2013. Distributed lag non-linear mixed effects models (Gasparrini et al. 2010) allow us to examine simultaneously the cumulative exposure-response relationship between daily wet-bulb temperatures over a lag period of 4 to 14 days with R. In addition to the coefficients from the distributed lag non-linear models, the predictive models also include, at the county level, a rolling average of lagged social distancing between 4-14 days, population density, diabetes prevalence, percent of residents over 65 years of age, percent of population below 200% of poverty level, household crowding, absolute testing positivity (rolling 7-day average), change of testing capacity (rolling 7-day average), and random effects for county, metropolitan area and climate zone (Baechler et al., 2015). Models predicting Rs and daily cases over the next four weeks are trained using data from between April 1 and July 3, 2020, on a subset of 571 counties. The final model was selected

from six models based on predictive accuracy of observed data in the last two weeks of May. For predictions, we use historical averages of daily temperature and humidity; testing positivity rate and social distancing are held constant at their current 7-day averages. As more counties continue to be added to the dataset, county random effects are being assigned based on the county's nearest neighbor. All states are now represented in the model.

Preliminary Findings

The trajectory of real-time R varies 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/humidity significantly affect the estimations of R over time, while standardizing on population characteristics. Higher temperatures and humidity have not reduced transmission risk beyond spring-time temperatures; if anything, the gathering effects of summer have conferred increased risk since springtime. To the degree that even springtime temperatures mitigate transmissibility (or Rs) is poor for most counties with higher population density and among communities that reopen too quickly, compared to their peak social distancing in April.

Implications

These county-level 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 United States. These results can inform consideration of selective strategies to mitigate SARS-CoV-2 transmission and assess health care system capacity constraints across the country. They can also serve as an early warning indicator for communities and their residents to reassess and modify behavior in response to forecasted resurgence risk.

Statistical References:

Baechler, Michael C., et al. "Building America Best Practices Series: Volume 7.3: Guide to Determining Climate Regions by County." Pacific Northwest National Laboratory (2015): 1-50.

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.

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.

For more information on our models, contact:

David Rubin, MD, MSCE (rubin@email.chop.edu)
Greg Tasian, MD MSCE (tasiang@email.chop.edu)
Jing Huang, PhD (huangj5@email.chop.edu)

Updated September 2, 2020