medRxiv preprint doi: https://doi.org/10.1101/2020.05.29.20117044; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Predicting Growth of COVID-19 Confirmed Cases in Each U.S. County with a Population of 50,000 or More

Leon S. Robertson

Professor Adjunct

Yale University School of Public Health

Email: leon.robertson@yale.edu

ABSTRACT

Models of local spread of COVID-19 are needed to assist local governments and health care providers to prepare for surges of clinical cases in their communities. National and state based models are inadequate because the virus is introduced and spreads at different rates in local areas. Models based on motor vehicle traffic measured by tracking cell phones movement in relation to daily numbers of confirmed cases are being developed but this study finds that indicator less effective in predicting future cases than time since the first case in a county. Each county has its own function of time since the first case that can be used to predict increases in reported cases two weeks in advance for each of 914 counties in the U.S. with populations of 50,000 or more inhabitants.

INTRODUCTION

Projected COVID-19 cases and deaths from various mathematical models using different methods initially produced quite different results in the U.S. because of the variance in assumptions about how the virus and people would behave. As more data became available the projections converged somewhat but still varied substantially [1]. The early projections did serve the purpose of motivating politicians, however belatedly, to adopt policies to slow the spread of the virus. In the U.S. the timing of stay-at-home (shutdown) orders varied among the states. The business operations and other gatherings prohibited during the shutdown also varied somewhat among the states but all left plenty of leeway for the virus to continue to spread. Several state governors did not issue such orders and many announced partial or complete termination of the orders in late April and early May, 2020. In Wisconsin and Oregon the order was voided by judges. Some issued standards for physical distancing in businesses and other organization as well as wearing masks in public places after the shutdown but enforcement will be problematic as it was in states with stay-at-home orders. If local medical care facilities are to be prepared for an influx of cases, it is important to be able to predict the rise in numbers in local areas as early as possible. Large swaths of counties in the U.S. have no hospital beds, and if they have them, no intensive care units. More than half of U.S. counties have no intensive care beds [2].

While I was working on a model to predict local upsurges in cases using motor vehicle traffic as a predictor, the Policy Lab of Children's Hospital of Philadelphia announced a short-term projection model using numerous variables, including daily motor vehicle traffic, to estimate transmission rates of the virus and alert local areas of increases [3]. When I contacted the developers asking for information on their model, they provided no information saying that it is still under development.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.29.20117044; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Estimating the transmission rate of the COVID19 virus is problematic because of its behavior. Many people who are infected do not experience significant or any symptoms but shed virus that infects others in physical proximity or in contact with surfaces where it dwells for a time. For example, tests of a sample of the population of Westchester County New York, the first "hot spot" in that state, indicated that 16.7 percent of people had antibodies to COVID-19 [4]. If the sample is representative of the population, 161,673 people in Westchester County had been exposed to enough of the virus to produce antibodies (.167 x 967,506 people in the population) but there were only 31,294 "confirmed cases" reported by the County Health Department as of May 10, 2020. Some 81 percent of those who may be positive for antibodies but not reported as cases either experienced mild enough symptoms to avoid seeking help or no symptoms at all. In New York City, 21 percent of the sample had antibodies but in Bronx County only about 13.8 percent of those had turned up in the "confirmed cases" count by May 10, about 86 percent did not. The difference between Westchester and Bronx counties could be a result of fluctuations in sampling or they could represent differences in help-seeking among people with relatively high (Westchester) or low (Bronx) incomes or other differences among the populations. Whether people with antibodies are immune to future infections, and if so for how long, is unknown. These characteristics of virus spread alone make the application of traditional epidemic modeling difficult. These models require information on infection and recovery rates as well as immunity [5] which involves testing vast numbers of people.

Human behavior is obviously a major factor in the spread of a human transported infectious disease. Human responses to the news that the COVID-19 virus was spread by breathing, coughing, sneezing, talking and singing as well as touching surfaces by those infected varied from substantial risk avoidance behavior (e.g., reducing physical proximity to other people, frequent hand washing, wearing face masks) to mockery of those who did so and protests against requirements to do so [6]. Photographs and videos of street protesters against shutdowns showed many people in close proximity to one another with no face masks [7].

Travel data based on tracing cell phone movements indicate that travel decreased prior to the adoption of stay-at-home orders in many metropolitan areas but increased in time later [8] suggesting that risk avoidance began before the orders and deteriorated thereafter. Since manifestation of symptoms lags infection by the virus for some two weeks, data projecting numbers two weeks in advance could give local areas at least a warning of acceleration in cases assuming that the data were available promptly. A more desirable model would project cases based on daily available data on the past growth of cases in the county. It turns out that a model based on that idea is a better predictor of future cases than cell phone movement.

The purpose of this paper is to report successful prediction of accumulated confirmed cases two weeks before the cases are confirmed and to provide equations to local officials and health care providers who may use them to prepare for surges in cases.

METHODS

medRxiv preprint doi: https://doi.org/10.1101/2020.05.29.20117044; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

I used ordinary least squares regression of prior changes in vehicle traffic and separately time from the first case in a county to predict the number of confirmed cases two weeks in the future. The study was limited to counties with 50,000 or more population to avoid random variation in small numbers. The two predictor variables could not be combined into one model because they are highly correlated as noted in the results section.

The hypothesized predictive equation using travel data is: $\log(\text{accumulated cases}_{t+14})=a +b[\log(\text{travel}_t/\text{travel}_{t-14}) \times \log(\text{accumulated cases}_t)]$ where t is time in days and a is the number of cases prevalent at the first observation and b indicates the slope of the curve in time for a given county. A separate a and b was estimated for each county by ordinary least squares regression of the cases each day for 45 days through May 23, 2020. Goodness of fit was indicated by the correlation coefficient squared. The 45 day period from March 17 was chosen as the days to estimate the regression parameters because of large changes in traffic volume prior to that date that occurred before there were enough cases for predictive power.

The hypothesized predictive equation for time from the first reported case is $\log(\text{accumulated cases}_{t+14})=a +b[\text{days since first case at time t})$. Again, a and b were estimated for each county by ordinary least squares regression.

Daily numbers of accumulated confirmed cases in each county through May 23, 2020 were downloaded from usafacts.org [9]. Estimated daily vehicle travel in each county from March 1 through May 12 was provided by Streetlight Data, Inc. that relies on tracking smart phone movement [10]. Estimated 2019 population for each county was obtained from the U.S. Department of Agriculture website [11] based on U.S. Census Bureau estimates.

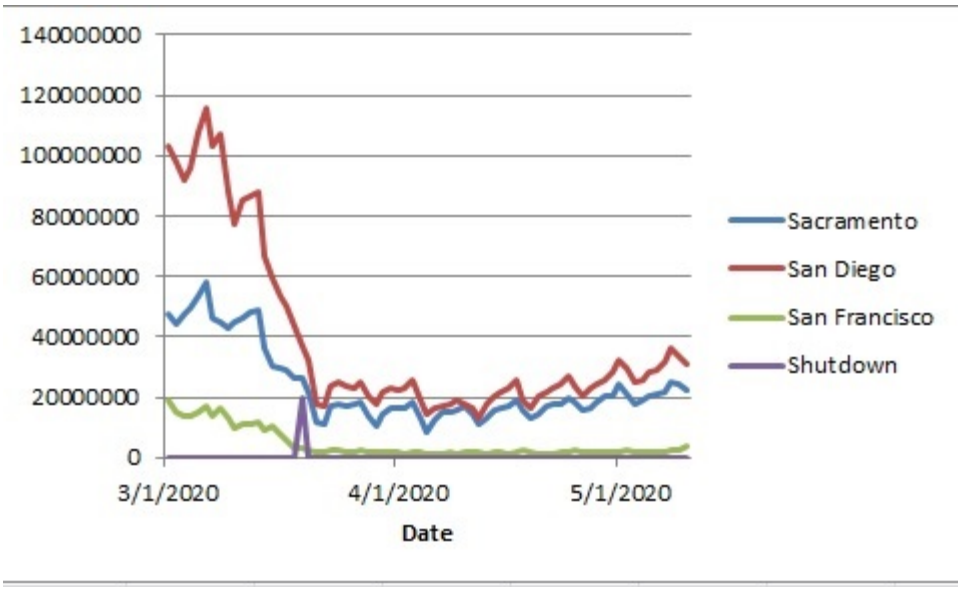
RESULTS

Data on all variables were available for 991 counties. The total population of these counties (288,662,000) was about 88 percent of the estimated 2019 U.S. population.

The travel variables show a remarkable pattern. Motor vehicle travel decreased substantially in many counties during early March, 2020 as warnings of the potential spread of COVID-19 were widely publicized. For example, Figure 1 shows the dramatic decline in traffic in Sacramento and San Diego Counties in California but less so in San Francisco County. It was this pattern that led to the decision to base the regression analysis on only 45 days from April 4, 2020 which is 14 days after the trend in traffic volume reversed through May 23, 2020 when the analysis began. To account for 14 days of change in vehicle traffic in the model testing, the series would have been extended further back into March when the miles were falling dramatically. The data for most counties fit the travel model fairly well. The average R^2 is 0.80 and only 9 percent were below 0.70.

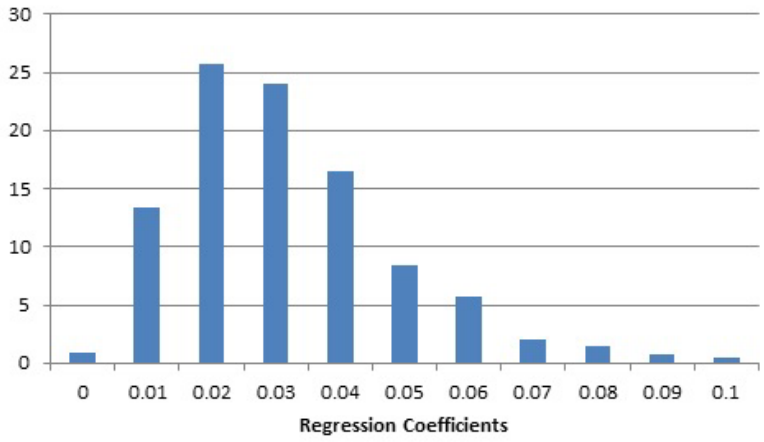
medRxiv preprint doi: https://doi.org/10.1101/2020.05.29.20117044; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 1. Estimated Daily Vehicle Miles Traveled in Selected California Counties



The second model based on days since the first confirmed case in a given county was even more effective at predicting total accumulated cases 14 days in advance. The average R^2 was 0.91 with only 3 percent less than 0.70.

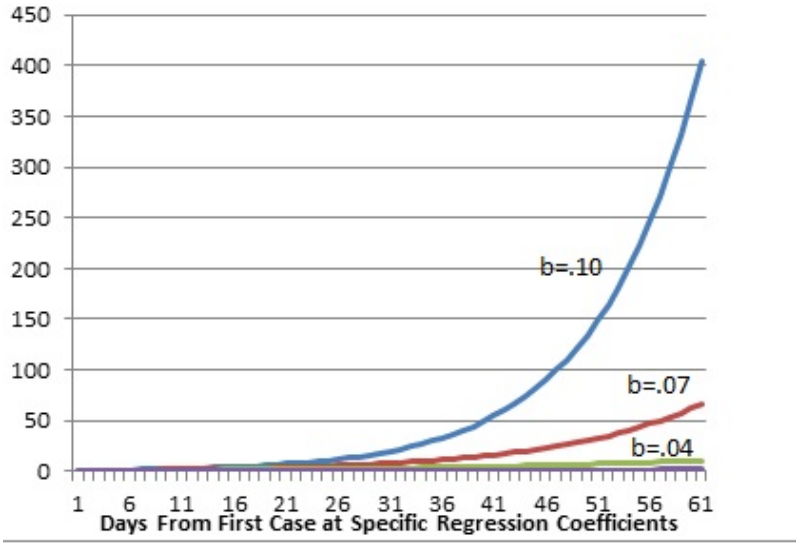
Figure 2. Percent of Regression Coefficients Among U.S. Counties from the First Case Model



medRxiv preprint doi: https://doi.org/10.1101/2020.05.29.20117044; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The distribution of regression coefficients among counties is displayed in Figure 2. Almost two thirds were 0.03 or less but the higher ones indicate substantial increases in number of cases per day despite the shutdown. Higher coefficients are indicative of a curve bending upward, far from the flattening needed to indicate containment of the virus. Figure 3 shows the effect of coefficients within 60 days of the first case. Notice that the upturn in cases occurs in a shorter period of time when coefficients are relatively higher.

Figure 3. Slope of Accumulation of COVID-19 Cases When Regression Coefficients Are at Specified Levels



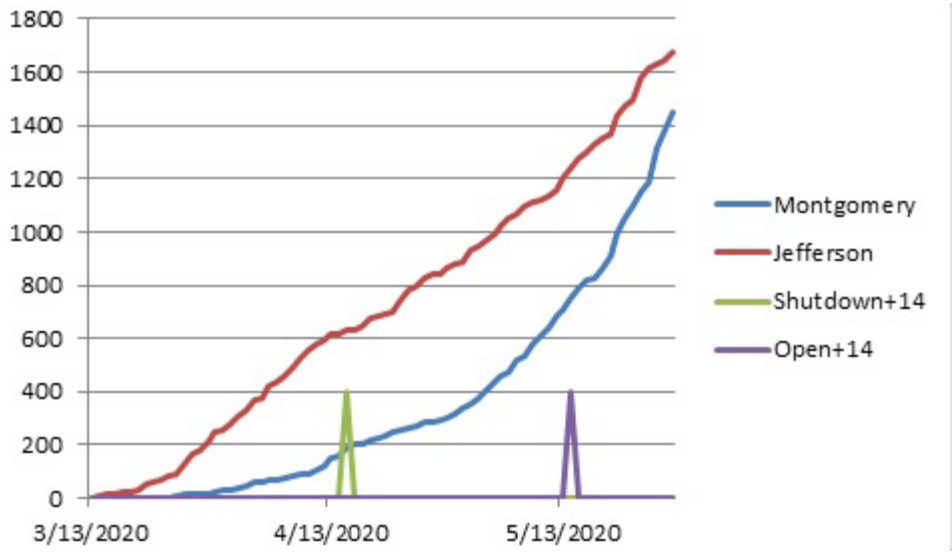
An attempt to combine the two predictors (travel and cases since the first day) into one equation produced problematic results. As seen in selected cities in Figure 1, travel was increasing in time during subsequent days after the low around March 22, 2020. This means that travel and time from the first case are highly correlated in most counties, a condition that distorts regression coefficients.

A scientific calculator can be used to calculate the number of cases in a county expected 14 days in the future from any given day. Multiply the regression coefficient (b) for that county times the number of days since the first case in the county, add the log of cases at origin (a) and use the ex function on the result to get the expected number of accumulated cases. Subtract the number of accumulated cases today from the expected number to get the number of new cases per day expected in 14 days.

For example, on May 23, 2020, Montgomery County, Alabama was 59 days from its first case and had accumulated 1,147 confirmed cases. It's expected accumulated cases by June 6 is: e [(59 x .052)+5.79]=7,030, some 5,883 more than on May 23. The county's intensive care beds were full by May 23 and COVID-19 patients were being sent to Birmingham in Jefferson County [13] which had its own fast-rising case load at the time (Figure 3). The spike points in Figure 3 on 4/14/2020 and 5/14/2020 indicate the beginning and end of Alabama's stay-at-home order respectively, lagged 14 days to account for the delay in manifestation of the symptoms among those infected. The order is associated with a moderation in the increase in cases but it appears that defiance of the order began about two weeks after its issuance in Montgomery County and earlier in Jefferson County.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.29.20117044; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 3. Growth in COVID-19 Cases in Montgomery and Jefferson County (Birmingham), Alabama



DISCUSSION

These data further illustrate that the spread of COVID-19 is very different in local areas mainly dependent on the number of days from the first reported case to whatever day one looks at the accumulated cases. The data do not predict what will happen in the longer term but it can provide state and county officials and medical care facilities with estimated case counts 14 days in advance within the months until a vaccine is widely available or there is enough testing and quarantine of infected carriers to reduce interpersonal transmission. The parameters for the equation for each county can be downloaded as an Excel file at www.nanlee.net. The 914 counties with an R^2 of .70 or more are included in the file online.

Without the shutdown, the COVID-19 virus would have caused more than enough severe illnesses to overwhelm the medical care system sooner in many cities in the U.S. [12]. As for rural America, low density reduces the individual risk but the medical system in many counties is without hospitals and those that have them are often without intensive care beds. Whether the virus will continue to surge in larger cities and invade the less dense areas of the country depends on several contingencies. There is insufficient data to judge whether there will be a seasonal decline during the summer as there has been with other manifestations of corona viruses but projection based on assumption of seasonal variation suggest that the virus will continue to spread even with some hiatus during warm weather [13]. The surge in cases in countries with warm climates such as Brazil and India suggest that warm weather will not have a major effect.

Although travel monitored by cell phone movement does not produce the best prediction model, it is close enough to suggest a worrisome scenario. Previous research indicates that motor vehicle travel is related to temperature, increasing as temperature increases [14]. If warmer temperatures fail to curb the spread of the COVID-19 virus cases, an acceleration of infections in summer as people travel to vacation and other locations is likely. We do not know the degree to which those who survived infection

medRxiv preprint doi: https://doi.org/10.1101/2020.05.29.20117044; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

will have immunity and for how long. We do not know how many people will be gravely ill and survive or die before the curve is flattened by changes in behavior or tracing and quarantine. The development of a successful vaccine in time to curb the exponential slope in cases and deaths in many counties is unlikely.

CONFLICT OF INTEREST

The author has no financial or other interest in Streetlight Data. The company graciously supplied the data promptly upon request.

REFERENCES

1. Bui, Q., et al. Coronavirus models are nearing consensus, but reopening could throw them off again. The New York Times, May 12, 2020. https://www.nytimes.com/interactive/2020/05/12/upshot/coronavirus-models.html, accessed May 12, 2020.

2. Schulte, F., et al. Millions of older Americans live in counties with no ICU beds as pandemic intensifies. Kaiser Health News, March 20,2020 online. https://khn.org/news/as-coronavirus-spreads-widely-millions-of-older-americans-live-in-counties-with-no-icu-beds/#lookup, accessed May 12, 2020.

3. Rubin, D., et al. Small-area projections of COVID-19 transmission in the United States. Children's Hospital of Philadelphia. http://policylab.chop.edu/sites/default/files/PolicyLab-Small-Area-Analysis-COVID-19-Transmission.pdf, accessed May 24, 2020.

4. Saplakoglu, Y. 1 in 5 people tested in New York City had antibodies for the coronavirus. Livescience.com. https://www.livescience.com/covid-antibody-test-results-new-york-test.html, accessed May 18, 2020.

5. Perez, L. and Dragicevic, S. An agent-based approach for modeling dynamics of contagious disease spread. International Journal of Health Geographics, August 5, 2009. https://ij-healthgeographics.biomedcentral.com/articles/10.1186/1476-072X-8-50#citeas, accessed May 10, 2020.

6. Jones, T. Trump supporters harassed Arizona reporters for wearing masks. Poynter. May 5, 2020. https://www.poynter.org/reporting-editing/2020/trump-supporters-harassed-arizona-reporters-for-wearing-masks/, accessed May 18, 2020.

7. Fernandez, M. In photos: Protesters storm Michigan Capitol over coronavirus restrictions. Axios. April 30, 2020. https://www.axios.com/michigan-protestors-state-capitol-coronavirus-ebcb34b6-5a48-47fb-b305-360c38baa16a.html, accessed 5/18,2020.

8. Ghader, S. et al. Observed mobility behavior data reveal "social distancing" inertia. Arvix.org, undated. https://arxiv.org/ftp/arxiv/papers/2004/2004.14748.pdf, accessed May 18, 2020.

9. USAfacts.org. COVID-19 Map by County and State Usafacts.org, Coronavirus Locations. https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/, accessed May 12, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.29.20117044; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

10. Streetlight data, Inc. https://www.streetlightdata.com/?utm_source=Google-Adwords&utm_medium=Paid-Search&utm_campaign=StreetLight-Brand&utm_term=%2Bstreetlight%20%2Bdata&streetlightdata.com&gclid=CjwKCAjwqpP2BRBTEiwAfpiD-5x7_51pAwGwsQb_tUIs-VXnf1fhwfcPBQaBPfrLlO-qyFYnFwNpShoCzmMQAvD_BwE, accessed May 19,2020.

11. U.S. Department of Agriculture. https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/, accessed May 13, 2020.

12. Kissler, S.M., et al., Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science 368:860-868, 2020. Online at: https://science.sciencemag.org/content/368/6493/860, accessed May 25, 2020.

13. Lemos, G. and Waldrop, T. With just one ICU bed available, Montgomery, Alabama, is sending sick patients to Birmingham. CNN, https://www.cnn.com/2020/05/21/us/montgomery-alabama-icu-bed-shortage/index.html, accessed May 27, 2020.

14 Wei, L. and Wehby, GL. Comparison of estimated rates of coronavirus disease 2019 (COVID-19) in border counties in Iowa without a stay-at-home order and border counties in Illinois with a stay-at-home order. JAMA Network Open, May 15,2020. https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2766229, accessed May 18,2020.

15. Robertson, L.S. Reversal of the road death trend in the U.S. in 2015–2016: An examination of the climate and economic hypotheses. Journal of Transport & Health, 9:161-168, 2018.