

## Modeling and the Covid-19 Pandemic: A Local Area Perspective

David A. Swanson, Ph.D.\*  
([dswanson@ucr.edu](mailto:dswanson@ucr.edu))

Modeling even simple aspects of the covid-19 pandemic is more than challenging in the United States because of the sparsity of data. There is no comprehensive testing and little is known about the efficacy of testing. Adding to the uncertainty, we not only lack information on the total numbers infected, but we do not know if those who have been infected can be re-infected, the proportion of those who are infected who are asymptomatic, and when and how long they may infect others. In short, you have a forecaster's nightmare. The "sparse data" issue is compounded because most, if not all, epidemiological models are complex. They are designed to provide a lot of information needed in the fact of a pandemic (see, e.g., Magel and Webb, 2020). Making this nightmare even worse is the fact that while at least some data may be available to use at the national level, applying a typical epidemiological model to a subnational area such as a county is virtually impossible without having to endure a very heavy "assumption burden." All of these issues leave local officials and residents literally in the dark when it comes to trying to get a picture of what may be coming and how to prepare for it. And it is in these small, local areas that the battles are being fought.

However, all is not lost for local areas. I have been producing county level forecasts that employing simple methods and concepts that need no more than the sparse data available in the U.S. (Swanson (2020a, 2020b, 2020c) With a simple geometric model using only the cumulative daily count of confirmed cases, and the "impact analysis" framework (Swanson et al., 2009), I have developed forecasts for Whatcom County, Washington (population  $\approx 225,000$ , with one university, Western Washington University, located in the county seat, Bellingham), which is north of Seattle and on the southern border of British Columbia, Canada. The "impact analysis" framework looks at how an event might unfold if it was left to run its course relative to interventions designed to alter that course. It is not perfect in that it is not a controlled experiment. However, to paraphrase George P. Box (Box and Draper, 1987: 426), while it may be an approximation, it is useful (Swanson et al. 2009).

The initial forecast was a "baseline" that launched from March 28<sup>th</sup>, showing what the county could expect in the absence of an "intervention," which in this case was containment measures (Swanson, 2020a). As of the date of the expected peak of the initial surge, April 25<sup>th</sup>, the baseline showed that 6,151 confirmed cases were expected.

About a week after the baseline forecast was released, I followed with the first update (Swanson, 2020b). The update used data that reflected the initial effects of containment measures put in place by the Governor of Washington, Jay Inslee, on March 25<sup>th</sup>. Like the baseline, it was based on a simple geometric model. As of the date of the expected peak of the initial surge, April 25<sup>th</sup>, this first update showed that 2,696 confirmed cases were expected. By reducing the initial rate of growth by 2.47 percent in less than a week, the containment measures led to 56 percent reduction in the cumulative number of expected cases by April 25<sup>th</sup>. These results showed the local officials and the

general public that the sacrifices made by the many people who strove to adhere to the containment measures, which included foregoing work and income, were paying a dividend in cases averted and lives saved.

The second update (Swanson, 2020c) continued with more good news in terms of the reduction of cases being brought about from the containment measures, an 82 percent reduction in the total number of confirmed cases relative to the baseline as of April 25<sup>th</sup>.

In addition to providing the 2<sup>nd</sup> update in terms of the simple geometric model I had been using, I turned to a more complex model, exponential in nature, to generate a forecast because there were now 17 days of data available. To implement this model, I used the exponential model function found under the “curve fitting” choices in the NCSS statistical software (<https://www.ncss.com/software/ncss/>). In addition to using all of the information available, I employed the exponential model to assess the adequacy of the geometric model and its results. This is a move that I had stated would be done once data were sufficient to support more complex models (Swanson, 2020a).

The exponential model yielded a forecast of 961 total confirmed covid-19 cases as of April 25<sup>th</sup>, which was only 157 cases fewer than that forecasted by the simple geometric model, a relative difference of -16.3 percent). Looking at the 95 percent prediction interval accompanying the exponential model, I found that the results of the simple geometric model. (1,118) fell well within the interval, which had 715 cases as the lower limit and 1,207 cases as the upper limit. All in all, these results suggested that the simple geometric model had done a reasonable job with the sparse data available.

On April 17<sup>th</sup>, I published the third update (Swanson 2020d). This one used a three-parameter logistic model, which was selected from the curve fitting choices offered by the NCSS Statistical Software System (<https://www.ncss.com/software/ncss/>). This model was selected because it was clear the initial explosive growth indicated by the baseline forecast had been brought sufficiently under control by the containment measures that the surge was near its peak and on the verge of plateauing. In addition, there was a sufficient set of observations to support this model. Like the earlier updates, this third update brought welcome news, particularly in light of the sacrifices made by the many people who strove to adhere to the containment measures, which included foregoing work and income. These sacrifices paid a huge dividend in cases averted and lives saved. Per the third update, by April 25<sup>th</sup>, these sacrifices were expected to bring about a 95 percent reduction in the initial expected number of confirmed cases as shown in the baseline forecast. I noted that it was a tremendous achievement that had done a lot to reduce the risk to the first responders, healthcare, grocery, and other workers who had put themselves at higher levels of risk by staying at essential jobs.

What is my take-away from this experience to-date? Namely, that when employed within an “impact analysis” framework, simple models selected with reasonable judgment (Use a geometric model in the initial stages of a surge, not a linear model) can provide a reasonable view of the future with only sparse data to support

them. The primary force driving this idea is that in the absence of the information provided by such models framed by the impact analysis perspective, people living in counties and small towns will have some idea of what they might be facing, rather than remaining in the dark. As more data become available, these simple models can be replaced with more complex ones, including those designed to provide more information such as R0 (the rate of infection) Case Fatality Rates, and hospitalization rates.

Before closing, I note that Green and Armstrong (2015) found no evidence that complex models are more accurate than simple ones. Their argument is not inconsistent with the view underlying Beirman's "Two Cultures," which is that problem solving is more important to those identifying with Beirman's "culture" than the adherence to traditional methods and protocols found in the other "culture." (Raper, 2020). In closing, I note that in the case of complex models, they clearly supply more information than does the simple geometric model, but this is of little use if the models lack sufficient data to implement and are "assumption heavy" (Jewell, Lewnard, and Jewell, 2020).

\*Professor Emeritus, University of California Riverside, and affiliated Faculty, Center for Studies in Demography and Ecology, University of Washington.

#### References

- Box, G. E. P., and N. Draper (1987). *Empirical Model-Building and Response Surfaces, 1<sup>st</sup> Edition*. John Wiley & Sons. Hoboken, NJ.
- Green, K. and J. Armstrong (2015). "Simple versus complex forecasting: The evidence." *Journal of Business Research* 68: 1678-1683.
- Jewell, N., J. Lewnard, and B. Jewell. (2020). Caution Warranted: Using the Institute for Health Metrics and Evaluation Model for Predicting the Course of the COVID-19 Pandemic. *Annals of Internal Medicine*. (<https://annals.org/aim/fullarticle/2764774/caution-warranted-using-institute-health-metrics-evaluation-model-predicting-course>).
- Magel, P., and G. Webb. (2020). "Predicting the number of reported and unreported cases for the COVID-19 epidemic in South Korea, Italy, France and Germany." *MedRxiv* (<https://www.medrxiv.org/content/10.1101/2020.03.21.20040154v1>).
- Raper, S. (2020.) "Leo Breiman's Two Cultures." *Significance* 17 (1): 34-37.
- Swanson, D. (2020a). "Forecasting the COVID-19 Surge Peak in Whatcom County." *The Northwest Citizen* (<https://nwcitizen.com/>), posted March 30<sup>th</sup>, 2020.
- Swanson, D. (2020b). "Monitoring the COVID-19 Surge Peak in Whatcom County - Update I." *The Northwest Citizen* (<https://nwcitizen.com/>), posted April 3<sup>rd</sup>, 2020.
- Swanson, D. (2020c). "Monitoring the COVID-19 Surge Peak in Whatcom County - Update 2." *The Northwest Citizen* (<https://nwcitizen.com/>), posted April 10<sup>th</sup>, 2020.
- Swanson, D. (2020d). "Monitoring the COVID-19 Surge Peak in Whatcom County - Update 3." *The Northwest Citizen* (<https://nwcitizen.com/>), posted April 17<sup>th</sup>, 2020.
- Swanson, D., R. Forgette, J. McKibben, M. Van Boening, and L. Wombold (2009). "The Socio-Demographic and Environmental Effects of Katrina: An Impact Analysis Perspective". *The Open Demography Journal*. 2 (11): 36-46