Dashboard 3: The Survival Convolution Model for Forecasting the COVID-19 Pandemic

Professor Yuanjia Wang

A survival convolution model (SurvCon; *Wang et al., 2020*) was developed to help the *COVID-19 Forecast Hub* and US *Center of Disease Control* (CDC) to forecast COVID-19 cases (Figure 1) and deaths in the near future.

The model was created by Professor Yuanjia Wang, PhD student Qinxia Wang, post-doctoral fellow Shanghong Xie in the Department of Biostatistics at Columbia University, and Professor Donglin Zeng at UNC-Chapel Hill. It is now part of the COVID-19 Forecast Hub which predicts the course of the COVID-19 pandemic in the US and provides weekly forecasts to the CDC. The methodology has been reported in Frontiers in Public Health. The open source software is available at the Github web-site *COVID19BIOSTAT*.

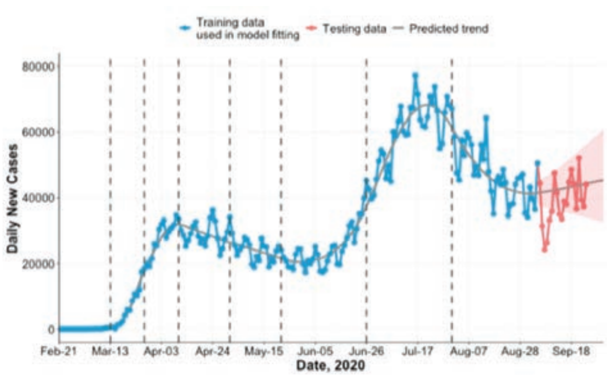

Figure 1. Reported Daily Incident COVID-19 cases and forecasted trend. Model trained using data until 9/4/2020 (blue). Testing data till 9/24 (red). Dashed lines are knots. Model suggests an uptick in COVID new cases since September.

SurvCon uniquely combines nonparametric statistical curve fitting with underlying mechanisms of infectious disease transmissions known from research in epidemiology. It is inspired by the Susceptible-Exposed-Infectious-Recovered (SEIR) models, although it does not involve all of the components of the SEIR. This reduces the parameters and assumptions that SurvCon requires to achieve robustness. Nevertheless, the model does take into account the main features that distinguish SARS-Cov-2 from other coronaviruses, i.e., its substantial rate of pre-symptomatic transmissions. This rate is modeled by piece-wise linear functions placed at meaningful event dates, a procedure which provides flexibility while retaining parsimony. The model is calibrated using official reports of COVID-19 incident cases and incident deaths. It accurately predicted the national-level apex of the first surge of COVID-19 cases in early April. It currently predicts about 225,700 total observed COVID-19 deaths in the US by November 1st, with a 10% chance of fewer than 222,100, and a 10% chance of more than 230,000.

Another unique feature of SurvCon is that it leverages natural experiment designs to study the effects of government non-pharmaceutical interventions (NPIs) such as social distancing, stay-at-home orders, and mask mandates. Such quasi-experiments are often used to estimate the effects of public health interventions or health policies when randomized controlled trials are not feasible. Intervention effects are estimated by comparing transmission rates before and after implementation of an NPI under assumptions of local randomization and continuity (i.e., the trends observed before an NPI would continue had the NPI not been implemented). This approach shows a large effect of NPIs implemented immediately after the declaration of a national public health emergency in the US on March 13th.

Since May, states around the United States have entered a 'reopening' phase. Government responses to the pandemic vary considerably. Since July, our team has been working on refined analyses of state-level data, and has compared the effectiveness of different reopening strategies, their effects on the second surge of COVID cases and deaths, and more recently, the effects of school reopening. New features being developed include leveraging community mobility data provided by Google and Apple to capture changes in population movements during the pandemic, incorporating census and state-level health data to adjust for confounding, and developing causal inference methods to evaluate NPI effects.

Such refined state-level analysis is expected to reveal the heterogeneity of NPI effects across key indicators of COVID-19 (e.g., timing of NPI implementation, urban vs suburban areas, race/ethnicity, poverty levels) on transmission rates. These insights will help guide the implementation of precision public health interventions.

Shanghong Xie

In this project I conducted a literature search, produced figures and tables, maintained our GitHub website, and submitted our death and case predictions weekly to the COVID-19 Forecast Hub for the CDC COVID-19 Ensemble Forecasts.

It is challenging to accurately predict the course of the epidemic, especially its peak and end. Most models of COVID-19 are deterministic, sensitive to initial values, and accurate only for one-week predictions. Our model can capture the trend(s) in the epidemic over time and predict its peak and end dates well, while being simple and robust. It requires only (officially reported) confirmed cases and deaths, without any individual-level information.

I have learnt several lessons from the project. First, we need scientific knowledge of COVID-19 (e.g., incubation distribution, reporting delay distribution) in order to accurately predict the course of the epidemic. Second, it is not easy to have good long-term predictions without incorporating some specific assumptions, given that the public health interventions (mitigation strategies, reopening) are changing over time. Third, the trajectory of the epidemic differs across states, given that they implement different policies to contain it.

Although there are some widely used traditional models of epidemic diseases, biostatisticians can still contribute to this field by modeling disease transmission and by designing study plans for treatment and vaccine trials. I am grateful to have been involved in the project, since our predictions are being used by the CDC for policy making. It has also been interesting to attend the weekly CDC COVID-19 forecasts calls to learn the approaches developed by other teams. In the future, I am enthusiastic about the possibility of conducting other COVID-19 related research, such as developing state and county-level predictions and investigating the effects of school reopening and wearing masks.

Qinxia Wang

I have been working on implementing the SurvCon model using Tensorflow, and on maintaining code to update the model with new data to provide real time forecasts of COVID-19 cases and deaths in the US on a weekly basis. We initially used grid search to provide the best-fitting estimates of the model parameters. However, this became very time-consuming as the pandemic continued and the volume of observed data increased. This encouraged us to look for a more efficient algorithm.

Tensorflow is an open-source library that can automatically compute gradients for any well-specified models and provide fast algorithms. Since I did not have much prior experience with Tensorflow, it took me some time to become familiar with it and implement it in SurvCon. However, my new implementation has significantly improved our computing capacity and has allowed us to provide state-level forecasts with permutations to compute confidence intervals.

One important lesson I have learnt in this process is that I should not be afraid to step out of my comfort zone and try new, unfamiliar tools. It may take me a while to learn them, but they can be beneficial for my research going forward. I also learned that it is important to be highly organized in both coding and project management. Since it is almost certain that we will revisit our code and make changes adaptively, building documented functions and code will save considerable time down the road.

Overall, I really appreciate the opportunity to have worked on this project. I have learned a great deal, and have been privileged to be a part of the research community that brings valuable insights to the pandemic and has a positive impact on it.