COVID Mortality Prediction Pre and Post Vaccine Availability

Yoosoon Chang (Indiana University), Sangmyung Ha (Indiana University), Joon Y. Park (Indiana University), Kosali Simon (Indiana University)

Abstract:

This study is motivated by the following three observations. First, the US experienced over 1.2 million COVID deaths during the pandemic era. Many of these deaths occurred in the hospital. Among COVID patients admitted to hospital, a substantial fraction died. Second, there are several known risk factors for covid mortality, conditional on covid hospitalization. Examples are pre-existing conditions such as diabetes; however, there may be other factors that will be uncovered by a prediction exercise. Finally, there was a drastic change in mortality when vaccines were available, and there also was a change in mortality when the dominant strains of COVID-19 changed. It could be that there are a few different regimes in effect during the pandemic—pre and post vaccine availability, for example. It could be that the risk factors for covid mortality in the hospital changed after vaccine availability.

We utilize various machine learning (ML) tools to examine the predictors of COVID mortality in the hospital and changes in betas in pre vs post vaccine era in 2020 and 2021, respectively, from the large dimensional nation-wide hospitalization data obtained from the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID) and National Inpatient Sample (NIS), which provide more than 6 million observations each year.

Specifically, we aim to answer the following specific questions using the rich database of COVID era hospitalizations and effectively leveraging machine learning tools. What are the predictors of COVID mortality in the hospital during the initial phase of the pandemic? Do the predictors from a machine learning exercise confirm findings published in the medical literature at the time (obtained using small-sample observational studies), such as to the shapes of nonlinear relationships between COVID risk factors and age? What changes occur in the predictors of COVID mortality (and the intensity of effect reflected in the coefficients, i.e., the betas) after vaccine availability compared to before? How do different phases of vaccine uptake and COVID variants affect the predictions, and does it vary by the vaccine take-up rate across different demographic groups?

To answer these questions, we conduct our analysis with the region-level percent vaccinated as an explanatory variable or by interacting it with a time period dummy (indicating whether it is early or late in the COVID era after vaccines), as well as with indicators of chronic conditions (such as diabetes) that capture risk factors. We expect these results to enhance understanding of which population groups (both in demographic factors as well as in terms of pre-existing chronic health conditions) benefited the most from the different phases of vaccine availability in reducing COVID related death rates during hospitalization. This research also illustrates how using machine learning tools on a large-scale database uncovers relationships that were not well understood in observational studies that were the only possible option at the time. Thus, it is helpful for understanding the benefits to enabling more real-time data access for future public health emergencies.