Big Data Science and AI to Flatten the COVID-19 Pandemic: Merits and Pitfalls

*Asaph Young Chun, Paul Choi, Eunjeong Jung

*Director-General, Statistics Research Institute | Statistics Korea

Keywords: Big Data Science, AI, COVID-19, Pandemic, Prediction Science

Flattening the COVID-19 epidemic curve, like every struggle against a pandemic, is a race against time. Predicting the course of COVID-19 pandemic in the 21st century has been as challenging as forecasting the Spanish flu that took away over 55 million lives across the world in the early 20th century. As the COVID-19 pandemic enters its second year, its variants have been identified in the United Kingdom and South Africa and threaten to overrun the promise of vaccines that have been developed in record time.

The purpose of this paper is to show genesis and evolution of modeling that has nudged government decisions in Korea to flatten the pandemic curve in the first year of COVID-19. Modeling COVID-19 pandemic is a prerequisite to informing evidence-based policymaking. Governments across the globe have varied measures of non-pharmaceutical interventions (NPI), such as masking, social distancing, and business closures, often based on models that capture one's own nation's characteristics and available data.

In the case of Korea, science-and-data based modeling has been a cornerstone of policymaking since COVID-19 infected the Korean population in February, 2019. We have initially used the model of Incidence Decay with Exponential Adjustment (IDEA) based on a well-established model that classifies population into the three groups: Susceptibles, Infected and Recovered (SIR). We have turned to integrating multiple models grounded on mathematics, statistics, and data science in a transdisciplinary manner. The data used in modeling included official organic data and commercial big data, both of which enhanced the rigor of scientific modeling. The outcomes of big data science-based modeling in Korea include as follow: 1) nearcasting models quantifying the impact of non-pharmaceutical interventions, such as wearing masks, social distancing, and school/business closures; 2) prediction of infection cases, patients in ICU and mortality, leveraging artificial intelligence, mathematics, statistics, and computer science in a transdisciplinary manner; and 3) cross-cutting models and metrics showing tradeoffs among health merits, economic effects, and post-COVID19 change in social life. The paper points out issues for future research and conditions for further innovation.