On the Correlation Between Immigration and Viral Spread

Jason Richwine

April 28, 2020

Last week the Cato Institute's Alex Nowrasteh and Andrew Forrester published a self-defeating analysis of how immigration influences the spread of COVID-19 in the United States. After controlling for population density and state-level variation, they found no relationship at the county level between the foreign-born share of the population and the rate of viral infections and deaths. Their model shows that population density, not the foreign-born share, seems to have a strong association with the virus. Therefore, the authors conclude that immigration has not accelerated viral spread in the United States.

Their analysis is an own-goal (to borrow a soccer term), for the simple reason that immigration increases population density. Unless immigration somehow causes a major decrease in childbearing or life expectancy among natives, it's true almost by definition that more immigration means more residents per square mile. (Immigration has added 43 million people to the United States since 1990, and none have created any new land!) Therefore, if we take the Nowrasteh-Forrester model seriously, then the virus would have spread less quickly in the United States if we had had lower levels of immigration. Similarly, accepting more immigrants will cause the virus to spread faster than it otherwise would.

The self-defeating nature of their analysis is obscured somewhat by the inaccurate language the authors use to introduce their findings. The headline says "immigration is *not correlated* with COVID-19." The second paragraph starts with, "The rate of COVID-19 cases and deaths by county in the United States is *not correlated* with the share of the population that is foreign-born." These claims are false. In fact, if my own calculations are correct, the foreign-born share correlates with case rates at 0.23 and with death rates at 0.15.

What Nowrasteh and Forrester really mean is that the *partial correlations* are zero after controlling for density and state-level variation. They acknowledge using control variables later in the piece, of course, but that does not make their initial claims any less wrong. It would be nonsense for someone to say, "The correlation between X and Y is zero, and by that I mean the partial correlation between X and Y after controlling for Z is zero." Those are two separate claims requiring two separate calculations. Had the authors honored the distinction, their own-goal would have been more obvious.