A New Database to Measure Changes in the Association Between Income, Race, and Life Expectancy Over Time

Raj Chetty, John N. Friedman, Nathan Hendren, and Michael Stepner

Abstract: We propose to update the income-mortality gradient statistics from Chetty et al. (JAMA, 2016, funded by the NBER SSA RRC) to the present day, as the start of a project that will aim to release these statistics recurrently on an annual basis. This project will result in annual statistics on the income-mortality gradient at the race-by-gender-by-county level in the U.S., which will augment our understanding of more general patterns of income inequality and support other researchers in their own work.

To illustrate the value of such annual statistics, we will use them to analyze how the COVID-19 pandemic has affected the income-mortality gradient - to what extent has the large increase in mortality from this disease fallen on the poor, and how does that vary by race, gender, and location? In order to include all deaths resulting from COVID (or prevented by the various non-pharmaceutical inventions), we will calculate the income gradient of excess mortality by comparing deaths in 2020 and 2021 to those in previous years. Unlike the previous JAMA paper, which relied entirely on tax data, we will combine datasets housed within the U.S. Census Bureau, including the SSA NUMIDENT File (deaths), IRS Tax Data (income and location, since 1989), and the Decennial Censuses from 2000 and 2010 (race, gender, and ethnicity). The panel structure of these data (which can be linked over time at the individual level within the secure Census computing environment) will allow us to measure income not only at a point in time but over more than 30 years. We can thus distinguish between the effects of transitory income shocks and permanent income, as well as measure income during earlier working years for individuals who have retired; this is especially important given higher mortality rates among the elderly. We will further disaggregate this income-mortality gradient to explore how it varies by race and ethnicity, across space, and potentially by other individual and household variables (e.g., disability status, education, wealth, health insurance coverage).

These results will speak directly to the distributional impacts of several SSA programs on the populations that they seek to serve. More generally, our analysis of the variation in the income-mortality gradient will inform our understanding of the structural factors that contribute to disparities by race and ethnicity.

Funding: We request ½ of a 1/9th for each of the four co-authors, plus funding for half of the time of an RA who will be based at the NBER.