Is the Relationship between Income Inequality and Population Health Robust? Evidence from a Multilevel Analysis

(Extended Abstract)

Yaqiang Qi

California Center for Population Research

and

Department of Sociology

University of California, Los Angeles

In recent decades, studies of social and contextual influences on population health have attracted a lot of attention from social scientists. There is convincing evidence that both individual's socioeconomic status and characteristics of the community where they live are strongly associated with their health. In particular, there have been increasing concerns on the relationship between income inequality and population health both within and across countries. It is found that even among the wealthy countries, there is considerable variation in life expectancy. Although one of the richest countries in the world, life expectancy in U.S. is five years shorter than in Japan, and three years shorter than in Sweden. More generally, ecological studies show that, provided other things equal, population health is negatively associated with unequal distribution of national income. The similar correlation between income inequality and population health is also reported across 50 states of the United States.

Various pathways have been proposed through which income inequality harms population health: Income inequality can affect individual health through individual's self-assessment of relative position, and relative deprivation is an important source of chronic stress, which has been shown to be a risk factor for many common diseases; societal hierarchy disrupts social cohesion and social trust, and increases crime and violence; social disruption, crime and violence has spillover effects that may affect social wellbeing including health. On the contrary, it is suggested that the observed correlation between income inequality and population health at the aggregate level is spurious and just reflects a nonlinear effect of individual income on health; in other words, income inequality per se does not harm population health directly and the association between income inequality and population health disappears when individual income is properly controlled. Apparently, a multilevel framework that includes both aggregate

level and individual level data is necessary for detecting the "true" effect of income inequality on population health. This paper is an attempt to fulfilling such an analysis.

To do this, I will need health and income measures at the individual level and income inequality estimates at the country level for a large number of societies. The World Values Surveys (WVS) collect self-reported general health status and other individual characteristics from 118,519 respondents of 81 countries, which provides a good source of individual level data for my analysis. The United Nations publish income inequality estimates and other country level indicators for the world, which will be my main source for the aggregate level data.

A general form of multilevel analysis for the problem in question can be expressed in two equations:

$$H_{ij} = \beta_{0j} + \beta_1 Age_{ij} + \beta_2 Sex_{ij} + \beta_3 Edu_{ij} + \beta_4 Inc_{ij} + \beta_5 Inc_{ij}^2 \quad (1)$$

$$\beta_{0j} = \eta_{00} + \eta_{01} Gini_j + \alpha_{0j} \quad (2)$$

where H_{ij} is the logit of reporting poor/fair health status for the i th person of the j th country. Individual characteristics (e.g., age, sex, education, and household income) and country-level characteristics (e.g., a country's Gini coefficient) are included simultaneously. In this framework, the genuine effect of income inequality on population health can be tested by whether $\hat{\eta}_{01} \cong 0$.

However, this approach may provide misleading results in that it is well known that self-rated health status is not a perfect measure of individual's true health status and is subject to substantial reporting errors, which probably vary by social and cultural contexts. Moreover, this approach, although not inherent, does not use the full information of self-rated health by collapsing the ordinal categories into a dummy variable. I will adapt a different modeling strategy that will reduce reporting biases of self-rated health: A residualization approach.

Here is a brief description of this approach:

First, estimate an ordered probit model of self-rated health for each country. The equation is as follows:

$$\Pr(h_i^{SR} \le j | X_i) = \Phi(\alpha_j + X_i' \beta) \quad (3)$$

where X_i' indicates controlled covariates such as age, education, marital status, and individual income.

Second, using the estimated coefficients, predict the fitted probability of reporting each of the possible categories of self-rated health for each respondent. For example, in a standard five-category question on self-rated health, it will give five probabilities for each respondent. Let p_{1i} to p_{5i} be the corresponding fitted probabilities of reporting "very good", "good", "fair", "poor" and "very poor" health for the *i* th respondent. Then we have $\sum_{k=1}^{5} p_{ki} = 1$.

Third, for each individual, compute the cumulative probability that the respondent is in a better category than the one actually reported, plus half of the probability of being in the actually reported category. For instance, if the i th respondent reports that his health status is "fair", his cumulative probability is equal to $p_{1i} + p_{2i} + 0.5p_{3i}$; if the i th respondent reports that his health status is "poor", his cumulative probability is equal to $p_{1i} + p_{2i} + p_{3i} + 0.5p_{4i}$. This cumulative probability stands for a measure of "relative illness" of the individual.

Fourth, construct a score of relative illness based on the cumulative probabilities. A normalizing transformation (inverse normal function) can be used for computing a continuous score based on these cumulative probabilities. And a mean relative illness score will be constructed for each country.

Finally, the country means of relative illness score will be estimated as a function of income inequality and other country level covariates.

This approach will reduce within-country reporting biases in self-rated health status by residualization. However, please note that, if a country has an overall tendency to optimistically report better health than the true status, it is not explicitly adjusted by this approach. For instance, it is possible that self-rated health is measuring the truth health status with three potential sources of errors:

$$H_{ij}^{SR} = H_{ij} + \varepsilon_{ij} + \delta_{ij} + \mu_j \quad (4)$$

where H_{ij} is the true health status for the i th person of the j th country, and ε_{ij} is an error term that is random across all observations, δ_{ij} indicates individual-specific nonrandom errors that interact with individual characteristics such as age, sex, education, and income, and finally μ_j is a country-specific error term. If μ_j is not random and is determined by unmeasured country features such as culture, the approach fails to control for it. However, as long as μ_j is not correlated with income inequality, it will still provide an unbiased estimate for the association between income inequality and population health.