Title: Sample selection appendix for "Cognitive reflection is associated with greater truth discernment for COVID-19 headlines, less trust but greater use of formal information sources, and greater willingness to pay for masks among social media users in Pakistan"
Authors: Ayesha Ali (1), Ihsan Ayyub Qazi (2)
Date: July 12th, 2022
Note: The material contained herein is supplementary to the article named in the title and published in the Harvard Kennedy School (HKS) Misinformation Review.

Appendix E: Sample selection

From September 1, 2020, to September 21, 2020, we conducted a phone survey of 621 social media users who were recruited as part of an earlier field study on misinformation, conducted in the city of Lahore, Pakistan (Ali & Qazi, 2021). The original field study involved a random sample of 750 social media users drawn from low- and middle-income areas of the city. During the baseline survey conducted in May 2019, we obtained contact information and consent to participate in follow-up surveys. We obtained consent and phone numbers of 695 users out of which we were successfully able to contact and complete the phone survey of 621 users (i.e., response rate of 89.3%). Out of the 695 phone numbers, we found that 3.5% were wrong numbers, 5.2% were switched off, and 1.4% did not pick up the phone. Our final dataset included a total of 621 individuals: mean age = 29.3, 50.4% female, mean monthly household expenditure = PKR 35,575 (USD 214.5 in 2020), and a median education between primary school and high school. The phone survey was conducted by our implementation partner, the Survey Wing of the Institute of Development Alternatives (IDEAS).

Field sample recruitment. The original field sample of 750 social media users was randomly drawn from low- and middle-income areas of the city of Lahore. We used inverse population density as a proxy for income. Congestion is common in older and poorer areas of Lahore, due to the non-compact and horizontal pattern of urban development observed commonly in South Asian cities (Ellis & Roberts, 2016; Harriri, 2020). Therefore, we used AsiaPop (2013) satellite data to provide population counts at a spatial resolution of 100 m by 100 m to identify congested low- and middle-income neighborhoods. The areas selected for drawing the sample accounted for nearly 35% of the city's total population and covered seven of the fourteen National Assembly constituencies in the city. The median population density in the selected areas of our study was 109 persons per 100 m by 100 m grid whereas the median density in areas not covered by our sample was 28 persons per 100 m by 100 m grid. Figure D shows the inverse relationship between population count on a 100 m x 100 m grid and the average household's monthly expenditures reported in our survey. A random sample of 200 grids was drawn from the selected areas and five households per grid were surveyed, covering a total of 1000 households in the baseline survey. To initiate the data collection within the chosen grids, a random point (x and y coordinate) was dropped within the grid. The enumerators arrived at the point and used the left-hand rule to survey, within each grid, five households where at least one social media user was present. The definition of social media user for our survey was that the respondent must be at least 18 years of age and be a social media user.

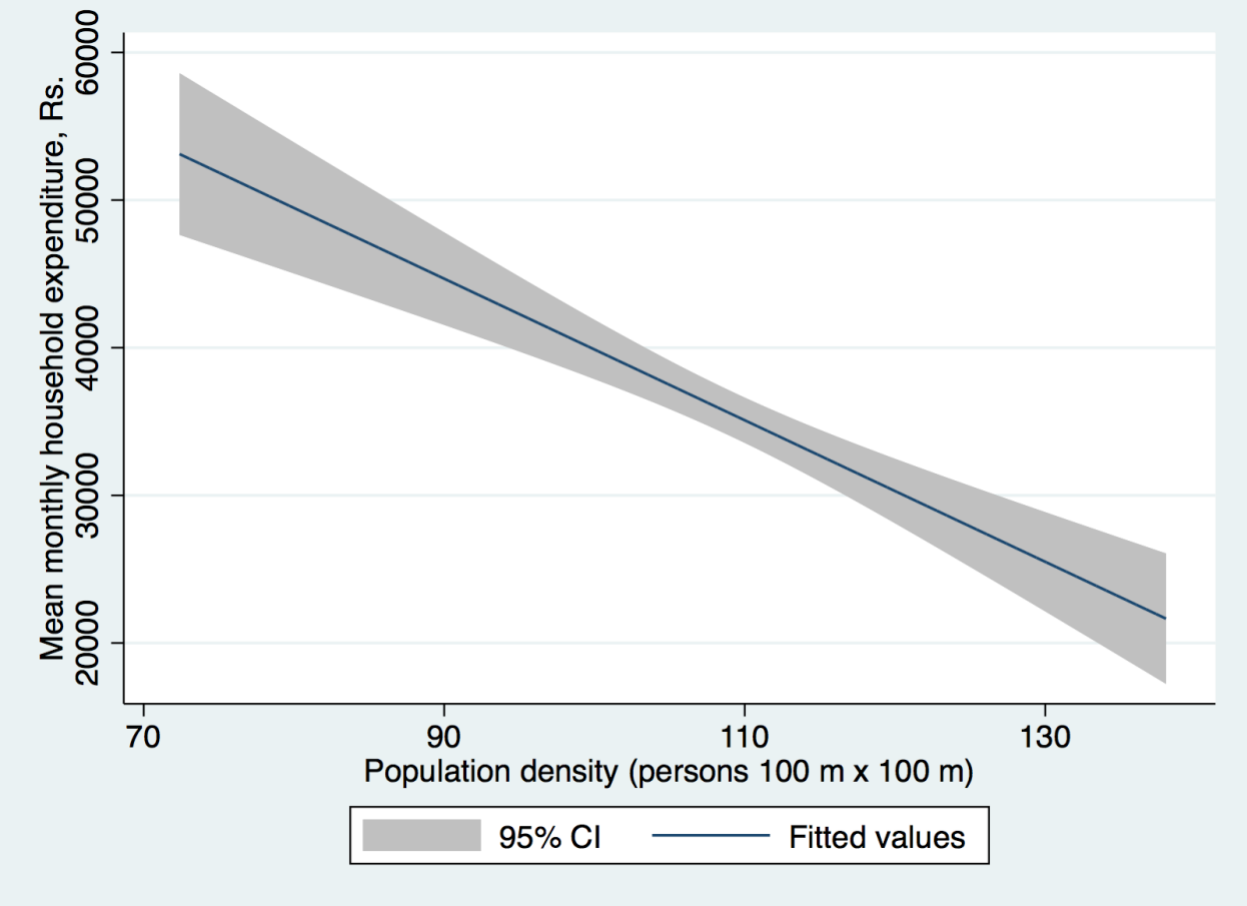

Figure D. Relationship between mean monthly household expenditure and population density (persons per 100m x 100m) in survey data collected in Lahore, Pakistan.