Amendments to the Pre-Analysis Plan for "Perceived Relative Income and Revealed Preferences for Clean Air"

Anca Balietti[1], Angelika Budjan[1,2], and Tillmann Eymess[1]

[1]Alfred Weber Institute for Economics, Heidelberg University
[2]Institute of Economics and Law, University of Stuttgart

September 23, 2022

As detailed in the original Pre-Analysis Plan (PAP), the data collection targets a total sample of 2,500 Indian respondents on Amazon Mechanical Turk. The data collection started in May 2022 and is ongoing as of September 2022. Currently, we have reached a total sample of around 800 observations after data cleaning (as per the criteria pre-registered in the PAP). The recruitment rate is much lower than initially anticipated.

We plan to complement the data collection on Amazon Mechanical Turk India with a new sample of 2,500 Indian respondents recruited via the "Dynata" data platform. The data collection with Dynata will start in late September 2022. The research design, survey questionnaire, and analysis plan will stay the same, except for the following changes that apply only to the data collection on Dynata.

Survey experiment structure and content

The following survey items will be *removed* or *changed*:

1. The analysis of data collected so far on Amazon Mechanical Turk shows no statistically significant differences in the main outcome variables between the control (C) and the pure control (PC) groups. Based on this, we will run our survey experiment on Dynata with only four treatment groups, as outlined in Table 1, namely control (C), Income Correction (IC), priming treatment with a low comparison (TL), and priming treatment with a high comparison (TH).

2. The survey items that aim to understand the respondent's baseline adoption of defensive strategies against air pollution are removed from the entry questionnaire (part of step 1 in Table 1).

Table 1: Survey experiment procedure and treatment variation – *updated*

	C	IC	TL	TH
1. Entry questionnaire	X	X	X	X
2. General and personalized air pollution info	X	X	X	X
3. Perception: prior relative own income	X	X		
4. Information on official statistics regarding own relative income		X		
5. Perception: posterior relative own income		X		
6. Perception: relative income of comparison group			X	X
7. Revealed preference elicitation (contribution)	X	X	X	X
8. Stated support for public policies	X	X	X	X
9. Stated adoption of protective measures	X	X	X	X
10. Perception: impact of air pollution on own health	X	X	X	X
11. Preferences for environmental justice	X	X	X	X
12. Preferences for economic equality	X	X	X	X
13. Preferences for redistribution	X	X	X	X
14. Altruism question	X	X	X	X
15. Perception: own happiness	X	X	X	X
16. Political party affiliation	X	X	X	X
17. Feedback	X	X	X	X
18. Research purpose disclaimer	X	X	X	X
19. Relative income disclaimer			X	X

Notes: The table describes the experimental procedure and the treatment variation therein. C = control, IC = income correction, TL = priming treatment with a low (poor) comparison, TH = priming treatment with a high (rich) comparison.

3. The general information on air pollution (see step 2) is shortened and will no longer include information on the adverse effects on the economy. Also, the explanation on how air pollution is measured and the details on the conversion from air pollution concentration to life years lost is removed. Note that respondents are still informed about the fact that air pollution reduces life expectancy. Consequently, respondents are no longer informed about the exact PM2.5 concentration when receiving personalized information on air pollution in their state. Rather, the personalized information only includes by how much the WHO recommended level is exceeded and the average effect of local air pollution on life expectancy.

4. The elicitation of own life years lost (see step 7 in Table 1 from the original PAP and description in section 4.1.3) will be replaced with the following question:

   "*In your opinion, how much do you think your personal health is impacted by air pollution?*".
   Response options: "Not at all", "Only a little", "Moderately", "Quite a lot", "Very much".

Additionally, the item will be asked at a later point (now step 10, see Table 1).

In the analysis, we will use this outcome variable, denoted now H_i, for answering the auxiliary research questions AQ-1 and AQ-4 from the PAP. In Section 5.2 of the PAP, the first secondary outcome of interest will be changed from "Perception of own number of life years lost due to air pollution exposure" to "Perception of personal health impacts due to air pollution exposure", where the outcome variable H_i is coded as an integer taking values from 1 to 5, with value 1 corresponding to "Not at all" and value 5 corresponding to "very much", following the Likert scale mentioned above, i.e. $H_i \in \{1, 2, 3, 4, 5\}$. In Section 7.2.1 of the PAP, the outcome variable LYL^p will be replaced by the variable H_i.

5. To determine the personalized information on life years lost on the state level, we use updated data on particulate matter (PM2.5) using the van Donkelaar et al. (2021) data and compute state-level population-weighted averages of PM2.5 concentrations and life years lost. The population data comes from Bondarenko et al. (2020).

6. The priming treatment with a low (poor) comparison household (TL) is changed. In the new TL treatment, we ask participants to place a household living in the same state that has typical characteristics of a very poor household (step 6 in Table 1) in the income distribution. The question will be formulated as follows:

   "*Think of a household in your state where its members:*

   - *have no education*
   - *are unemployed*
   - *cannot afford enough food and clothing*
   - *live in a non-recognized slum with over-crowded rooms*
   - *have no toilet and no access to fresh water.*

   In your opinion, which income group is this household part of?"
   Response options: "Group 1", "Group 2", "Group 3", "Group 4", "Group 5","Group 6", "Group 7", "Group 8", "Group 9", "Group 10".

7. After the elicitation of perceived own income (steps 3 and 5), we include the following question to elicit participants' confidence in their assessment:
   How confident are your with your answer to the previous question?
   Response options: "Not confident at all", "Not very confident","Neutral", "Fairly confident", "Very confident".

8. On the contribution page (step 7), we no longer remind participants of what is the average impact of air pollution on life expectancy in their own state or of their own assessment of how air pollution affects their own life expectancy.

9. We re-structure the questionnaire for the "General preferences" part of the survey experiment described in section 5.3 of the original PAP. The new questionnaire will include the following questions (according to the order in Table 1):

   - Questions on preferences for environmental justice:
     - How much do you agree or disagree with the following statement: "The government should make sure that everyone has equal access to protection measures against air pollution, no matter what their income is."
       [Likert-scale from 1 ("Strongly disagree") to 5 ("Strongly agree")]
     - How much do you agree or disagree with the following statement: "The government should make sure that those with a higher income contribute more to reducing air pollution than those with a lower income."
       [Likert-scale from 1 ("Strongly disagree") to 5 ("Strongly agree")]
   - Question on inequality aversion:
     - How much do you agree or disagree with the following statement: "The gap between the rich and the poor in India is too large."
       [Likert-scale from 1 ("Strongly disagree") to 5 ("Strongly agree")]
   - Question on preference for redistribution:
     - How much do you agree or disagree with the following statement: "It is the responsibility of the government to reduce the gap between the rich and the poor."
       [Likert-scale from 1 ("Strongly disagree") to 5 ("Strongly agree")]
   - Question on altruism:
     - How do you assess your willingness to do good for others without expecting anything in return?
       [Likert scale from 1 ("completely unwilling") to 10 ("very willing")]

   We will no longer standardize the measures on environmental justice, inequality aversion, redistribution, and altruism. The question on political orientation remains unchanged.

10. We remove the survey items on general health and life satisfaction (compare old step 8 to updated step 15 in Table 1).

Consequently, the secondary outcomes 4 (b) and 4 (c) in section 5.2 of the original PAP will no longer be collected.

11. We include a question designed to identify respondents that straight line their responses (*i.e.*, choosing the same response multiple times in a row). The question is formulated as: "Please select the word "Blue" from the options below." Response Options: "Strongly Disagree", "Blue", "Neutral", "Agree", "Strongly Agree". The question is added to step 11 (see Table 1).

12. At the end of the survey, we ask participants to provide feedback. We include the following open-ended questions:

    - Did you find any question unclear or uncomfortable? Did you experience any technical difficulty?
    - Are there other policies to reduce the impact of air pollution that you would think are more appropriate?
    - What did you think about when deciding to contribute or not to an NGO?

Data collection

As previously mentioned, the data collection will be conducted by the panel provider "Dynata". The following aspects of the data collection will be changed:

- The base incentive is no longer US$1. Rather, all respondents receive the customary compensation for a 10min survey with their respective primary panel platform.
- The bonus payment will be in Indian rupees (INR), *i.e.,* respondents receive 120 INR ($\approx$ US$1.50).
- The full sample ($N = 2,500$) will be representative of the national population with respect to gender and age based on information from the 2011 census.[1]
- The steps to ensure data quality as outlined in section 3.2 of the original PAP will be changed. In the updated design, we will exclude the following cases from the analysis:
    - Respondents that answer "No" to the question: "In your honest opinion, should you be allowed to continue with the survey or should we disconnect you because you did not devote your full attention to the questions so far?".

[1]The representativeness with respect to age is subject to uncertainty due to the ability to reach respondents aged 45 and above.

- Respondents that need more than three attempts in the last question during the 'reading and comprehension exercise'.
- Respondents that give answers during the end of survey feedback that are unambiguously automated (utilizing a bot program to fill out surveys) or otherwise entirely nonsensical.
- Respondents that have impossibly short completion times of less than 240 seconds for the full survey experiment (steps 1 to 19 in Table 1) or less than 90 seconds in the exit questionnaire (steps 7 to 16 in Table 1).
- Respondents with response patterns that can be classified as straight-lining (i.e., choosing the same response option multiple times in a row). We classify the following as straight-lining:
  1. Respondents that cannot answer the question correctly that is designed to identify straight-lining participants (see the list on "Survey experiment structure and content" above).
  2. All of the 13 questions between steps 8. and 13. in Table 1 have five response options displayed horizontally. Respondents that use the same response option for all these questions (even if they correctly respond to the question designed to "catch" straight-lining).

Analysis

For the analysis, we will largely proceed as described in the original PAP. Differences are the following:

- We change the equations used to test for a treatment effect on primary and secondary outcomes in the IC treatment. We will estimate

$$D_i = \alpha_s + \beta_1 IC_i + \beta_2 \text{Neg. bias}_i + \beta_3 IC_i \times \text{Neg. bias}_i + X_i'\Gamma + \epsilon_i. \quad (1)$$

  instead of Eq. (5) of the original PAP. Here, "Neg. bias" denotes those respondents with a negative misperception (those that perceive to be poorer than what is indicated by their reported income level), *i.e.*, when $I_i^p < I_i$ with I_i as the reported income level and I_i^p as the perceived income (the prior) of respondent i. Similarly, Eq. (10) in the original PAP is replace with the following:

$$y_i = \alpha_s + \beta_1 IC_i + \beta_2 \text{Neg. bias}_i + \beta_3 IC_i \times \text{Neg. bias}_i + X_i'\Gamma + \epsilon_i. \quad (2)$$

- In Section 8 of the PAP, in the sub-section "Potential Channels", we will check how the different responses to general preferences (environmental justice, preferences for redistribution, preferences for inequality/inequality aversion, altruism) are impacted by the treatments. For each measure, we will estimate reduced form models, controlling for treatment dummies, personal characteristics, and state fixed effects. Standard errors will be clustered at the state level. This analysis replaces the approach in Eq. (13) of the original PAP.

References

Bondarenko, M., Kerr, D., Sorichetta, A., and Tatem, A. (2020). Census/projection-disaggregated gridded population datasets for 189 countries in 2020 using built-settlement growth model (bsgm) outputs.

van Donkelaar, A., Hammer, M. S., Bindle, L., Brauer, M., Brook, J. R., Garay, M. J., Hsu, N. C., Kalashnikova, O. V., Kahn, R. A., Lee, C., et al. (2021). Monthly global estimates of fine particulate matter and their uncertainty. *Environmental Science & Technology*, 55(22):15287–15300.