Pre-Analysis Plan: Belief formation, signal quality and information sources—Experimental evidence on air quality from Pakistan

Matthew Gibson* Isra Imtiaz† Shotaro Nakamura‡ Sanval Nasim§
Arman Rezaee¶

May 28, 2023

Abstract

An emerging body of empirical work emphasizes the demand for effective mitigation measures against severe seasonal ambient air pollution in developing cities. Yet, governments in developing countries often struggle to provide consistent and reliable air quality information. In such an environment, private alternatives to government services may improve citizens' access to air quality information. However, the efficacy of private alternatives may depend upon citizens' preferences for and beliefs about the accuracy of the information sources. We study how the source of environmental information affects citizens' beliefs about the level of air quality and their protection measures against polluted air. In this field experiment, we vary the salience of information sources—the government vs. a private citizens group—of air pollution forecasts we provide in Lahore, Pakistan. We will measure the changes in people's beliefs about air pollution levels and preferences for government services on air quality information through incentive-compatible elicitation games.

*Department of Economics, Williams College, MA.
†Mahbub ul Haq Research Centre, Lahore University of Management Sciences, Lahore, Pakistan.
‡Department of Economics, University of California, Davis, CA; email: `snnakamura@ucdavis.edu`
§Department of Economics, Colby College, Waterville, ME
¶Department of Economics, University of California, Davis, CA

Contents

1 Introduction

An emerging body of empirical work emphasizes the demand for effective mitigation measures against severe seasonal ambient air pollution in developing cities (e.g. Freeman et al. 2019; Ito and Zhang 2020). One such measure that could yield considerable public benefit is accessible and reliable air-quality information. Our previous experimental study conducted in Lahore, Pakistan, revealed that citizens are willing to pay an average of PKR 91 per month for day-ahead air pollution forecasts (Ahmad et al. 2022). Access to reliable information may allow citizens to form accurate beliefs about air pollution and take mitigation measures. This is evidenced by an increased demand for protective masks in our study (Ahmad et al. 2022).

Governments in developing countries, however, often struggle to provide consistent and reliable air quality information.[1] Consequently, various stakeholders, including citizens' groups, international and bilateral agencies, and research institutions have begun providing air quality information in the absence of reliable local government services.[2] For example, in Lahore, a citizens' group called Pakistan Air Quality Initiative (PAQI) crowd-sources low-cost monitors across the city and publishes their readings on Twitter and a mobile app, at no charge to the users.

Private alternatives to government services may improve citizens' access to air quality information. However, the efficacy of private alternatives may depend upon citizens' preferences for and beliefs about the accuracy of the information sources. First, there are significant variations in the availability of air quality information and readings between sources, as shown in Figure 1. Second, we document heterogeneity in citizens' preferences for air quality information sources in our baseline survey, as shown in Figure 2. It is, however, unclear how much of this variation is driven by their beliefs about the accuracy of the sources or their affinity toward each of the information sources. As such, we study how citizens in a developing city form beliefs about air quality and modify their behavior, as they infer the quality of information from its attributed source. We conduct a randomized information intervention to address the following research questions:

1. Are citizens willing to pay for air quality information, regardless of the sources to which it is attributed?

2. Holding quality constant, are there differential willingness to pay for air quality infor-

[1] There are several reasons behind the unreliable government service provision. First, government agencies may suffer from resource and capacity constraints that prevent them from providing consistent information, as many of the high-quality air quality monitors require regular maintenance. In Lahore, data by the local environmental regulator (Punjab Environment Protection Department, or EPD) are only made public as reports on their website in PDF (URL: `https://epd.punjab.gov.pk`). Second, government agencies may also face perverse incentives to obscure the true extent of environmental degradation by omitting unfavorable information (e.g.Ghanem and Zhang 2014)

[2] One notable example is AirNow by the U.S. Department of State, which has installed monitors at U.S. embassies and consulates and provides readings via their website (URL: `https://www.airnow.gov/international/us-embassies-and-consulates/`) and via Twitter.

mation by the source?

3. What are the mechanisms behind differential willingness to pay?
   3.1. Do citizens adjust their beliefs about air quality levels differentially by the source of information they are exposed to?
   3.2. Do they perceive signal quality differently by the source?
4. Does exposure to information of various sources induce differential policy preferences for environmental services?

In our intervention, we send air quality forecasts via SMS to a sample of Lahore residents in a working class neighborhood. We developed an ensemble forecast model of day-ahead air pollution using data inputs from multiple sources, including government and private monitors. We then experimentally vary the salience of information sources when we give forecasts to the participants, holding constant the actual forecast. In one arm, respondents are told that the forecasts are constructed using data from Punjab Environmental Protection Department (EPD), a government agency responsible for reporting on air quality. In the other arm, respondents are told that the forecasts are constructed using data from a citizens' group called Pakistan Air Quality Initiative (PAQI). We identify treatment effects on measures of their beliefs over air quality and preferences via incentive-compatible elicitation.

We hope to make contributions to several strands of literature including environmental public services in developing economies, public-private competition, and belief formation. First, our work contributes to the emerging body of work on the demand for, and the challenges with the public provision of, environmental services (e.g. Ghanem and Zhang 2014 Freeman et al. 2019; Ito and Zhang 2020). We hope to understand the importance of information source and consumers' beliefs in shaping the demand for such services. Second, we hope our findings will provide relevant insights to policymakers and stakeholders on the supply side of environmental service markets, and contribute the literature on accountability and competition for publicly provided services in developing economies (e.g. Muralidharan and Sundararaman 2015; Das et al. 2016; Jha and Nauze 2022). Third, our work relates to the literature on news media, particularly around mechanisms behind polarization of beliefs and trust in information sources (Gentzkow et al. 2023; Baysan 2022; Chopra et al. 2022). We hope to understand a) the role of beliefs and trust in shaping the demand for environmental information and b) the importance of prior beliefs and conditions under which beliefs about the state of the world and preferences for information services might diverge.

2 Research design

2.1 Sampling

The intervention is conducted in lower-middle-class neighborhoods of National Assembly (NA) constituencies 123 and 124, in northern Lahore. We divide the two constituencies into 200m×200m blocks, and randomly select 100 of them, weighted by population density. Figure 3 shows the selected blocks, plus 20 backups. We then sample 1,010 households from the block centroids by following the left-hand rule: survey every 10 households by spiraling out from the centroid counterclock-wise.

2.2 Randomization

2.2.1 Treatment arms

Figure 4 shows that the sampled households are divided into the following two treatment arms:

- T1: SMS forecasts are attributed to a government agency (EPD)
- T2 SMS forecasts are attributed to a citizens' group (PAQI)

2.2.2 Stratified randomization

We stratify the randomization process into the two treatment groups (T1 vs. T2) on the following variables collected about respondents in the baseline survey:

1. absolute error of incentivized $t+1$ forecast of PM2.5 concentration (i.e., primary outcome 2.)
2. share of donations to government vs. citizens' group (i.e., primary outcome 4.)
3. time spent outdoors (i.e., secondary outcome 5.1.)
4. index: perceived accuracy and approval of government's services on air quality
5. index: perceived accuracy and approval of citizens' groups' services on air quality
6. 1 if comprehended a mock-up of the SMS forecast message without further explanation
7. 1 if reported to have received air pollution information from the EPD
8. 1 if reported to have received air pollution information from AirVisual app (on which PAQI posts air quality readings)

9. Indicators of respondents' main TV news source
10. Asset index: a count of assets (electricity, appliances, vehicles, and number of rooms)

We conducted the blocking and randomization on R using *blockTools*, a package that allows us to block observations on a large number of covariates and can accommodate continuous variables without discretizing them. We use the optimal-greedy algorithm and generate blocks using the Minimum Volume Elipsoid (MVE) estimator. We are primarily concerned about balance on outcome variables at baseline, as well as the "take-up" in terms of exposure and comprehension of our SMS forecast messages. As such, we weigh variables 1. to 6. twice as heavily as the rest of the variables. We follow the advice from Athey and Imbens (2017) that each block contains two units per treatment arm; we specified that a block contains four subjects each (except for the last block, which contains 2). We then assign two subjects per block to T1, and the rest to T2.

2.3 Forecast model

We designed an ensemble model to forecast PM2.5 concentration for the next day (t+1) by building on Ahmad et al. (2022). In this subsection, we discuss (a) the sources of air quality readings currently available for Lahore, (b) how we define and measure the ground "truth" of air quality levels, and (c) how we construct the ensemble model.

2.3.1 Sources of air quality information

There are several sources of air quality information providing daily readings of PM2.5 concentration. The following are the four major on-the-ground sources that we believe are exhaustive of all publicly available air quality information. We use all four sources, plus a satellite-based measures (SPRINTARS) for our forecast model.

1. **Environment Protection Department (EPD)**: EPD is a department operating under the provincial government of Punjab, Pakistan. We collect their daily readings data from their website. The daily PDF reports include readings from three to four monitor locations, each reporting one of the scheduled pollutants. During our intervention period in the first half of 2023, the reports include readings on carbon monoxide (CO), particulate matter smaller than $2.5\mu m$ ($PM2.5$), and particulate matter smaller than $10\mu m$ ($PM10$). Each of these readings are reported in two different indices: (i) pollutants' concentration; and (ii) AQI (Air Quality Index). All of these indices are reported as 24-hour averages uploaded on the EPD's website everyday around 9-11am. The reports also contain a disclaimer that "[any] other data from any source presenting ambient air quality of any city of Punjab is neither verified nor approved by the EPA Punjab."

2. **Pakistan Air Quality Initiative (PAQI)**: PAQI is a citizens' initiative that crowd-sources collection of air quality readings and provides it via social media and other platforms. Started in 2016, PAQI crowd-sourced several low-cost air quality monitors (IQAir and PurpleAir) that are originally designed for indoor use. PAQI, among other operators, uploads their PM2.5 readings to an online platform named AirVisual. The platform reports both monitor-level and city-level readings at the hourly and daily concentration, going back as far as one month (e.g. Lahore and e.g. Lahore American School). Furthermore, PAQI reports city-level hourly readings from AirVisual on Twitter (such as @LahoreSmog). As such, we take the latter readings posted on Twitter as air quality measures provided by PAQI even though the reading is an average of monitors including ones that do not belong to PAQI.[3]

3. **U.S. Consulate**: The U.S. Consulate General in Lahore hosts an air quality monitor funded by U.S. EPA. The program, called AirNow International, places air quality monitors at U.S. embassies and consulates in mostly developing countries and provides hourly historical readings of $PM_{2.5}$ concentration. The monitor is located within the US Consulate's compound in Shimla Hills, Lahore. The readings can be accessed via the AirNow International's website. The U.S. consular services in Pakistan also share their readings from each consulates via Twitter (e.g. @Lahore_Air).

4. **Urban Unit**: The Urban Unit is a government-owned, yet privately operated entity that addresses urban issues using data in Punjab Province. It was launched as part of a unit in the Planning and Development Department of the provincial government of Punjab in 2005, and was spun off to the private sector with full government ownership in 2012. The unit works on a range of issues pertaining to sustainable urban development, primarily in the realm of environmental services and management. The department owns a high-quality air quality monitor and had previously provided its readings on the banner of their website, but had stopped providing this daily information publicly prior to the beginning of our intervention in early 2023. They have an Environment Dashboard that individuals can sign up for and gain access to historical data on PM2.5 readings but this data is updated at a lag of 10-15 days. We receive hourly average readings of PM2.5 concentration from the Unit's staff member on a daily basis.

5. **SPRINTARS**: Spectral Radiation-Transport Model for Aerosol Species (SPRINTARS) is a numerical model which estimates the effect of aerosols on the climatic system via simulations based on an atmosphere-ocean general circulation model called MIROC. The model and estimates have been developed by the Climate Change Science Section at the Research Institute for Applied Mechanics, Kyushu University (Fukuoka, Japan). SPRINTARS considers both natural and anthropogenic sources of aerosols and categorizes them into suspended particulate matter (SPM), PM2.5, and PM10. Through a collaboration with the model's developers at Kyushu University, we access the hourly forecasts generated by SPRINTARS and construct the $t + 1$ average forecast.

[3] This is because PAQI considers the city-level aggregate measure to be the most comprehensive of air quality information in Lahore and associates itself with it.

2.3.2 Defining the ground "truth"

For both the forecast model and the intervention, we first define the measure (which we refer to as the "truth") that the model predicts and how the readings are collected. The measure of interest is the average concentration of PM2.5 (in $\mu g/m^3$) between 12:00AM and 4:00PM for the day on which it is reported. This is because we send out the daily readings and the $t+1$ forecast between 6:00-8:00PM as we learned in our previous study, Ahmad et al. (2022), that most respondents make plans for the next day in the evening.

The daily readings that we provide as the "truth" and on which the forecast model is trained, are from the U.S. Consulate monitor, as it is presumably of highest quality using the "reference method" in compliance with the U.S. EPA standards.[4]. On days where the U.S. Consulate monitor is missing data, we use the Urban Unit readings, which are also based on a high-quality monitor (BAM-1020 by MET). If both sources are missing, we use readings from PAQI, which are consistently available. As of 24th May, 2023, the U.S. Consulate monitor is missing readings for 16 out of the 97 intervention days. Out of 16 days where U.S. Consulate is missing data, the Urban Unit is missing data on 4 days.

2.3.3 Constructing the ensemble forecast

We use an ensemble model that combines the following $t + 1$ forecasts of the "truth":

1. predictions using data from the U.S. Consulate data
2. predictions using data from the Urban Unit
3. predictions using data from EPD
4. predictions using data that PAQI posts on Twitter
5. *t+1* predictions from the SPRINTARS air pollution model

Combining these prediction models into an ensemble achieves two goals; first it improves our overall predictive ability, and second, it allows us to attribute our predictions to different sources (i.e. EPD or PAQI) when the information is provided to the treatment households.

1. **Constructing individual predictions**
   Each model, except for SPRINTARS uses the following inputs:
   - the lagged readings from a given source on days $t - 6$ to t
   - AccuWeather's *t+1* forecasts for minimum temperature, maximum temperature, and precipitation in inches, as well as their squared values

[4] `https://www.epa.gov/system/files/documents/2022-12/List_of_FRM_and_FEM.pdf`

- Historical weather data on daily average, minimum, and maximum temperature, dew point temperature, wind speed and direction, visibility, and relative humidity from ASOS
- Historical weather data on pressure and precipitation from Weather Underground

Using the Adaptive Lasso model, we predict *j+1* PM2.5 concentration (i.e., the "truth") using a model trained on data from Day 1 to Day j, for j going from Day 20 to t.

SPRINTARS gives a model-based forecast, so we do not construct our own forecasts.

2. **Combining the forecasts to construct an ensemble model:**
We estimate the root-mean-square error (RMSE) of each model over the period in which we have forecasts. We then weight the forecast based on the sum of RMSE across five models to their own (i.e. $w_i = \frac{\sum_{s \in S} RMSE_s}{RMSE_i * W}$ for a source i in a set of sources S, and W is the sum of all w_i's).

The ensemble forecast is the weighted sum of the individual forecasts.

2.4 Intervention: SMS forecast messaging

The main element of our intervention is the daily provision of the day-ahead (i.e., *t+1*) forecasts of PM 2.5 measures in $\mu g/m^3$ via SMS. In these messages, one of the sources (EPD or PAQI, chosen via the randomization procedure) are made salient. The daily messages also contain the readings from time t. The subjects also received an introductory message before the start of the daily SMS's, and a reminder message every two weeks over the course of the intervention. The daily messages are sent out around 6:00-8:00PM starting on 18th February 2023, and continue through to the end of the endline survey period (currently expected in mid- to late June, 2023). All of these messages are sent out using *OpenCodes*, an API-based system using a short-code service. All messages were in Urdu in the Urdu alphabet (Nastaliq script).

2.4.1 Introductory message

The following messages were sent to the subjects, depending on the assigned treatment arm:

- T1: "Assalam u alaikum! We visited your residence last month and did a survey on Air Pollution in Lahore where you agreed to receive air quality forecast information messages. You will be receiving these messages every day for the next 2 months.
These messages are based on PM 2.5 data which is measured in micrograms per meter cube. The data is collected from the Punjab government's Environmental Protection

Department (EPD) which is tasked with collecting information on Air Pollution. If you have any queries or questions about these messages, please contact the following number [telephone number]."

- T2: "Assalam u alaikum! We visited your residence last month and did a survey on Air Pollution in Lahore where you agreed to receive air quality forecast information messages. You will be receiving these messages every day for the next 2 months. These messages are based on PM 2.5 data which is measured in micrograms per meter cube. The data is collected from a non-governmental organization (NGO) called Pakistan Air Quality Initiative (PAQI [insert phonetic for PAQI in Urdu alphabet]) which collects data on air pollution. If you have any queries or questions about these messages, please contact the following number [telephone number]. "

2.4.2 Daily forecast messages

The daily messages are sent around 6:00-8:00PM, after collecting the day's data and estimating the forecast for *t+1*. We use the shorthand "NGO" to refer to organizations of a type, such as PAQI, for the purpose of familiarity with our subjects. The message on, for instance, 18th February 2023 would look as follows:

- T1: "Actual Air Quality (PM 2.5) on 18-02-23: 179
  Air Quality Forecast (PM 2.5) for 19-02-23 using data From Punjab Government (EPD): 231

- T2: "Actual Air Quality (PM 2.5) on 18-02-23: 179
  Air Quality Forecast (PM 2.5) for 19-02-23 using data From NGO (PAQI [insert phonetic for PAQI in Urdu alphabet]): 231

Figure 5 shows screenshots of the daily messages for T1 and T2. Because the text messages are sent from the same number everyday, it is easy to compare the forecast values for Day *t* provided on Day *t-1* to the realized value provided on Day *t*.

2.4.3 Fortnightly reminder messages

Starting on Saturday, 4th March 2023, reminder messages are sent every two weeks on Saturday about the source and the unit of measurement. The messages by the treatment groups are as follows:

- T1: "The following messages on air pollution (PM 2.5) are based on data from the Punjab Governments Environment Protection Department (EPD). The data is measured in micrograms per meter cube."

- T2: "The following messages on air pollution (PM 2.5) are based on data from a non-government organization (NGO) named Pakistan Air Quality Initiative (PAQI [insert phonetic for PAQI in Urdu alphabet]). The data is measured in micrograms per meter cube."

2.4.4 Eid holiday break

The SMS forecast services were suspended between 21st to 25th April 2023. On Saturday, 15th April 2023, we added the following line after the fortnightly reminder for both treatment arms:

- "We would also like to inform you that this service will be temporarily stopped during the Eid holidays and will be restored immediately after the Eid holidays."

2.5 Power Calculations

We estimate the minimum detectable effect sizes on our primary outcomes at 80% probability, with $\alpha = 0.05$. We assume a 15-percent attrition on our sample of 1,010. We also make conservative adjustments by dividing the α level by the number of tests for which we are identifying minimum treatment effect sizes.

There are two iterations to our power calculations. First, we identified the number of experimental arms and sample size based on the minimum detectable effect sizes during the design phase in June 2022. Out of the five hypotheses we present in this pre-analysis plan, we had only identified two of them during the design phase (and therefore divide α by 2). We then take sample means and standard deviations from survey data used in Ahmad et al. (2022). The outcomes, sample means and standard deviations in parentheses are as follows:

1. Willingness-to-pay (WTP) for SMS-based air quality forecasts: 89.6 (45.2)

2. Absolute error of incentivized $t + 1$ forecast of PM2.5 concentration: 43.4 (43.0)

We find that we are able to detect impacts of 0.27 standard deviations, which equal to PKR 12.3 in the willingness to pay, and 11.7 $\mu g/m^3$ for PM2.5 concentration.

Second, we re-estimate the minimum detectable effect sizes on the five hypotheses that we pre-specify in this document, using new data from the baseline survey when available. The outcomes, hypotheses, sample means, and standard deviations are:

1. Willingness-to-pay (WTP) for SMS-based air quality forecasts is greater than 0 regardless of the source to which the information is attributed: 89.6 (45.2)

2. Willingness-to-pay (WTP) for SMS-based air quality forecasts is differentially affected by treatment: 89.6 (45.2)

3. Absolute error of incentivized $t + 1$ forecast of PM2.5 concentration, divided by the truth, is differentially affected by treatment: 0.72 (0.42)
4. Perceived accuracy of air-quality information source as the absolute error of incentivized guess of the SMS's forecast is differentially affected by treatment: N/A
5. the amount out of PKR 100 donated to a government agency for an environmental cause, as opposed to the citizen's group, is differentially affected by treatment: 50.1 (15.0)

For hypotheses 1. and 2., we use the sample statistics from Ahmad et al. (2022) as we do not collect these outcomes in the baseline of this study. We do not have relevant statistics available from either the baseline or from Ahmad et al. (2022) for hypothesis 3., but we expect the outcome variable for it to have a similar distribution to the one for hypothesis 3..

We find that we are able to detect impacts of 0.43 standard deviations, which equals to PKR 19.4 in the willingness to pay (for hypothesis 2.), 0.18 for hypothesis 3., and 6.4 for hypothesis 5.. For the test of means for hypothesis 1., we find that we are powered to detect that willingness to pay is greater than PKR 3.6.

Although the minimum detectable impact is fairly large in terms of standard deviations, the treatment effect sizes are relatively small in the outcomes' units. Furthermore, there are several reasons why our assumptions may not hold or statistical precision could be improved. First, we plan to improve precision by including controls selected via a double-post-selection method using LASSO. Assuming a 30-percent reduction in standard errors, the minimum detectable effects would be 0.30 standard deviations. Second, the willingness-to-pay statistic from Ahmad et al. (2022) may be outdated after two years of high inflation.

2.6 Project timelines

The project timelines are as follows:

1. Design Phase: -January 2023
2. Pilot baseline survey: December 2022 - January 2023
3. Baseline survey: January 2023 - February 2023
4. SMS intervention: February 2023 to May 2023
5. Endline survey: May 2023 to June 2023

3 Data

3.1 Survey data

3.1.1 Survey frequency

We conduct the following surveys:

- Baseline survey (11th to 31st January 2023)
- Endline survey (29th May to mid/late June 2023)

3.1.2 Survey modules

In the baseline survey, we ask for demographics, some of the outcome measures (i.e., outcomes that are not contingent on the subjects' having experienced the forecast service), and dimensions of heterogeneity. Detailed survey instruments are included in the appendix. We provide detailed descriptions on outcome and other variable definitions in Section 4.

The baseline survey modules are as follows:

- Identification of a decision maker in the household as the respondent and consent
- Household roster and their demographics
- Awareness about air pollution in Lahore and access to information
- Donation game between EPD and PAQI, and stated preferences for the sources
- Stated beliefs in their trust of government services
- Incentivized forecast of air pollution (PM 2.5) concentration tomorrow
- Attitudes and behaviors regarding air pollution
- Time use survey and outdoor activities
- Participation in local community and civil society
- Access to news sources and preferred channels
- Household assets

The endline survey modules are as follows:

- Identification of the same respondent as in the baseline and consent

- Incentivized forecast of air pollution (PM 2.5) levels tomorrow and incentivized guess of the SMS's forecast
- Value elicitation of the SMS forecast service through a bidding game using the BDM method
- Access to information about air pollution, and stated satisfaction about the SMS forecast service
- Donation game between EPD and PAQI, and stated preferences for the sources
- Preferences for air quality-related policies via hypothetical scenarios
- Attitudes and behaviors regarding air pollution
- Time use survey and outdoor activities
- Stated mask usage
- Interest in filing complaints about air pollution to government authorities

3.2 Air quality data

We collect air quality readings data from five different sources for the forecast model and for the intervention. We provide further detail on each of the data sources in Section 2.3.1.

3.3 Weather Data

We also collect weather data as inputs for the forecast model, as described in further detail in Section 2.3.3.

- **AccuWeather:** We scrape daily forecasts on maximum and minimum temperatures and precipitation probability from AccuWeathern for Lahore at `https://www.accuweather.com/en/pk/lahore/260622/daily-weather-forecast/260622`. AccuWeather uses NOAA's (National Oceanic and Atmospheric Administration) data and construct its own forecasts.
- **ASOS**: We also collect detailed meteorological data collected by weather stations at airports. The data sources are called Automated Surface/Weather Observing Systems (ASOS/AWOS), or more generically METeorological Aerodome Reports (METARs). We use a web repository of these data sets hosted by Iowa State University's Iowa Environmental Mesonet and collect data for a station named "[OPLA] LAHORE(CIV/MIL)" via the following link: `https://mesonet.agron.iastate.edu/request/download.phtml?network=PK__ASOS`.

- **Weather Underground**: We also collect data on average and minimum atmospheric pressure and daily total precipitation from Weather Underground (URL: `https://www.wunderground.com/weather/pk/lahore`).

4 Outcome variables

4.1 Primary outcomes

Following the hypotheses listed in Section 1, we identify primary outcomes of interest. There are four outcomes, over which we test five primary hypotheses. The primary outcomes are constructed from incentivized games in the endline survey. They are defined as follows:

1. Demand for air quality information as the willingness-to-pay (WTP) for SMS-based air quality forecasts

    - The outcome is defined as the amount respondents' are willing to pay in PKR. We elicit respondents' willingness to pay for the SMS forecast using the Becker-DeGroot-Marshak (BDM) method (Becker et al. 1964). In the endline survey, we ask for the respondents' willingness to pay for the SMS-based air quality forecast messages. They have been receiving these messages for the past three months and we ask for their willingness to pay for an additional two months. In the prompt, we make the experimentally assigned source salient by reminding them that the forecast is built using data from the said source. The bid's ceiling is set at PKR 400.

2. Beliefs about air quality levels as the absolute error of incentivized $t + 1$ forecast of PM2.5 concentration

    - The outcome is defined as the absolute difference between the actual PM2.5 concentration and the respondent's forecast, divided by the actual PM2.5 concentration. In both baseline and endline surveys, we ask respondents to make an incentivized guess of the air pollution level on day $t + 1$. In the baseline survey, we show respondents a table containing the average, minimum, and maximum of the average daily PM2.5 concentration over the last calendar week. We then ask them to forecast tomorrow's average PM2.5 concentration. Respondents receive PKR 250 if their guess falls within 5% of the actual levels, PKR 150 if within 10%, and PKR 50 if within 20%. In the endline, we first elicit the forecast without the table containing the information from the previous calendar week. We then allow the respondents to revise their forecast after showing them the table.

3. Perceived accuracy of air-quality information source as the absolute error of incentivized guess of the SMS's forecast

- The outcome is defined as the absolute difference between the respondent's guess of the PM2.5 forecast generated by our model and their own forecast for $t+1$. In the endline survey, we not only ask respondents to forecast the actual PM2.5 concentration for tomorrow, but also the value of our SMS forecast. The guess is financially incentivized, as in the guess for the actual PM2.5 concentration for tomorrow.

4. Preference for information source as the share of donations to government vs. citizens' group

    - The outcome is defined as the share of PKR 100 donated to a government agency for an environmental cause, as opposed to the citizen's group. We offer an opportunity to donate PKR 100 between two sources for environment protection purposes: a government institution and PAQI.

4.2 Secondary outcomes

We present other variables that are of interest, but for which we do not correct for multiple testing. We first present outcomes that are alternative definitions of, or otherwise related to, the primary outcomes. We then list other complementary outcomes.

1. Demand for air quality information (related to Primary Outcome 1.)

    1.1. Stated satisfaction of the SMS service

    - The outcome is defined as the Likert scale, with 5 the most favorable. We ask respondents to rank their overall satisfaction with the SMS forecast service in the past three months.

    1.2. Stated belief in the reliability of SMS forecast service

    - The outcome is defined as the Likert scale, with 5 being "strongly agree." We ask respondents if they agree with the statement that the SMS forecasts have been provided frequently and on time.

    1.3. Approval of government and citizens' groups air quality information service

    - The outcome is defined as the Likert scale, with 5 being "strongly agree." We ask respondents if they agree with the statement that they approve of the job EPD or PAQI, respectively, does to address air quality in Lahore.

    1.4. Stated belief in the reliability of government and citizens' groups air quality information

    - The outcome is defined as the Likert scale, with 5 being "strongly agree." We ask respondents if they agree with the statement that EPD or PAQI, respectively, provide air quality measurements frequently and on time.

1.5. Access to other forms of air quality information
    - The outcome is defined as the number of air quality information sources the respondents have accessed in the past.

2. Policy preferences and collective action for air quality (related to Primary Outcome 4.)

    2.1. Prefers the local government to invest in air quality vs. other policies
    - The outcome is defined as 1 if they prefer the government invest in air quality v.s. other policy goals. We ask a hypothetical scenario in which the local government has PKR 100 million to allocate either towards improving air quality or towards investing in one of three other goals (education, health, and waste management, in three separate scenarios).

    2.2. Takes a document on how to file a complaint to the local government
    - The outcome is defined as 1 if the respondent takes a pamphlet. At the end of the endline survey, we prompt the respondent that EPD is a government agency responsible for addressing air quality issues in Lahore. We tell the respondents that we have a document that shows them how to file a complaint to the EPD, and ask if they would like a copy.

    2.3. Plans to file a complaint to the local government about air quality
    - The outcome is defined as 1 if a respondent intends to file a complaint to the EPD about air quality.

3. Beliefs about air quality levels (related to Primary Outcome 2.)

    3.1. Unincentivized guesses of air quality in comparison (endline only)
    - 1 if correctly guesses that yesterday's air quality is better than the day before yesterday.
    - 1 if correctly guesses that today's air quality is better than yesterday.

    3.2. Number of days with satisfactory air quality
    - The outcome is defined as the number of days in the last week with satisfactory air quality. What would constitute "satisfactory" air quality is subjective and is left to the respondents' interpretation.

    3.3. Concern about air quality
    - The outcome is defined as the Likert scale, with 5 being "strongly agree". We ask respondents if they agree with the statement that they are "concerned about air quality in general" in the last week.

4. Perceived accuracy of air-quality information source (related to Primary Outcome 3.)

    4.1. Weight put on a government reading in a hypothetical scenario

- The outcome is a continuous value between 0 and 1, indicating the weight the respondents put on an EPD reading as opposed to a PAQI one. We present a hypothetical scenario in which there are readings of the PM2.5 concentration from two sources: government (EPD) and citizens' group (PAQI). One of the sources (chosen at random) is $50\mu g/m^3$, and the other is $100\mu g/m^3$. We then ask the respondent what they think the true concentration level is, between 50 and 100. We then construct $(\frac{|V_g - V_r|}{50})$, where V_g is the value assigned to EPD and V_r is the respondent's guess on the truth. This is a hypothetical scenario and the game is not incentivized.

4.2. Stated belief in the accuracy of the SMS forecasts

- The outcome is defined as the Likert scale, with 5 being "strongly agree". We ask respondents if they agree with the statement that the SMS forecasts we have provided in the past three months are accurate. We make the experimentally assigned source salient by reminding them that the forecast is built using data from the said source.

4.3. Stated belief in the accuracy of government and citizens' group's air quality information

- The outcome is defined as the Likert scale, with 5 being "strongly agree". Aside from their beliefs in the accuracy of the SMS forecasts, we ask respondents if they agree with the statement that air quality measurements published by EPD or PAQI, respectively, are accurate.

5. Avoidance behaviors

5.1. Outdoor time use

- The outcome is defined as the number of hours spent outside. We ask respondents the type of activity (sleep, paid work, home work, leisure, travel, and other) they conducted for each hour of the previous day, and whether it was indoors or outdoors. We aggregate the number of hours the respondent engaged in any outdoor activity.
- We also plan to estimate impact by the type of outdoor activities (sleep, paid work, home work, and leisure), as well as the share of each type of activity spent outdoors. We conduct analysis on these outcomes to identify mechanisms, but do not adjust for multiple testing.

5.2. Access to high-quality masks

- The outcome is 1 if the respondent shows a high-quality mask to the enumerator. We ask if the respondents have been given or purchased any masks for air pollution, and if so, to show one to the enumerator. We identify respondents who show a N90/95 mask. We also collect information on what other types of masks (e.g. surgical masks, cloth) the respondents show.

5.3. Adjust their time use because of air pollution

- The outcome is defined as the Likert scale, with 5 being "strongly agree." We ask respondents if they agree with the statement that they reduced the number of hours worked significantly in response to poor air quality.
- We also ask respondents how many hours they would have spent outdoors if the pollution level was, hypothetically, 150 on average. This is asked after we measure how many hours they usually spend outdoors on a typical day and is meant to capture behavior changes, if any, due to poor pollution levels.

5 Empirical Analysis

5.1 Exogenous variable

Our main exogenous variable is treatment assignment between the arm where the government (EPD) was made salient as the source, as opposed to the citizens' group (PAQI). We refer to being in the citizens' group arm as being in the "treatment," and the government arm as being in the "control" for the rest of this document. Let Z_g denote treatment assignment as a vector, whose inputs equal to 1 if the respondent is assigned to the government arm and 0 if assigned to the citizens' group arm.

5.2 Checks on balance

We test balance between the two treatment arms denoted by Z_g. There are two groups of baseline measures: covariates W, which were used for blocking, and others X which were not. We report a standard balance table for each of the covariate groups. The statistics we present include means for the two treatment arms, differences between the two treatment arms, and t-tests of the null hypothesis of zero difference. If we find that one of the covariates in X is not balanced across treatment, we include it as a control in all regressions for analysis. We also test for balance across the alignment variable A_a on the same sets of covariates, but exclude those who reported to prefer EPD and PAQI equally at baseline.

In addition, we run a regression of the following form, and estimate the F statistic and the p-value with heterogeneity-robust standard errors.

$$Z_{g_i} = \boldsymbol{X_i'}\boldsymbol{\eta_1} + \boldsymbol{W_i'}\boldsymbol{\eta_2} + \epsilon_i$$

Table 1 shows balance on the variables used in the blocking procedure.

In the working paper, we will also evaluate balance on attrition by assessing if attritors and non-attritors differ on observables, when interacted with treatment assignment. First, we report attrition rates by the two experimental arms. We will then compare attritors and non-attritors on observables as follows, where D is an attrition dummy;

$$D_i = Z_{g_i}\kappa_1 + Z_{g_i}\boldsymbol{X_i'}\boldsymbol{\kappa_2} + Z_{g_i}\boldsymbol{W_i'}\boldsymbol{\kappa_3} + \boldsymbol{X_i'}\boldsymbol{\kappa_4} + \boldsymbol{W_i'}\boldsymbol{\kappa_5} + \omega_i$$

Again, we estimate the F statistic and the p-value obtained with heterogeneity-robust standard errors.

5.3 Pre-specified hypotheses

The following are the five hypotheses that we test and for which we correct for multiple testing.

1. The demand for air quality information is greater than zero regardless of the treatment assignment group (tested on outcome 1.)
2. The demand for air quality information is different between the treatment (citizen's group) and control (government) groups (tested on outcome 1.)
3. Treatment affects beliefs about air quality differentially relative to control (tested on outcome 2.)
4. Treatment affects the perceived accuracy of air-quality information source relative to control (tested on outcome 3.)
5. Treatment affects policy preferences for air quality relative to control (tested on outcome 4.)

The above hypotheses correspond, in order, to the research questions specified in Section 1.

5.4 Test of positive willingness to pay for air quality information

To test for hypothesis 1., we simply use a t-test to see if the willingness to pay for the SMS forecasts is higher than 0. We pool the two treatment arms and conduct a one-tail test.

5.5 Treatment Effects

5.5.1 Intent to treat

We estimate the treatment effects between subjects as follows;

$$Y_i = \alpha + {Z_{g_i}}'\beta + \boldsymbol{X_i'}\boldsymbol{\gamma} + \varepsilon_i$$

The matrix $\boldsymbol{X}$ includes control variables selected through a double-post-selection method using LASSO, as in Belloni et al. (2014). Given that we are agnostic as to which information source is more likely to shift beliefs, preferences, and beliefs related to air quality, our hypothesis tests are two-tailed: $\beta \neq 0$.

With the above estimating equation, we test hypotheses 2. and 4..

We estimate the treatment effects within subjects as follows;

$$Y_i = Z_{g_i}^{'}\beta + \gamma Y_{0i} + \boldsymbol{X_i'\delta} + \varepsilon_i$$

We denote Y_0 as the baseline measure of the outcome variable Y. Much of the details about the specification and inference is the same as in the between-subject model; we select the vector of controls $\boldsymbol{X}$ via a double-post-selection method with LASSO, and estimate p-values using randomization inference. Our hypothesis tests are also two-sided, i.e., $\beta \neq 0$.

With the above estimating equation, we test hypotheses 3. and 5..

5.5.2 Treatment on the treated

We define take-up of our intervention as looking at our forecasts via the SMS, which we do not observe. Instead, we construct a proxy of this measure from the endline survey, where we ask, "[during] the service period, how many days out of the week did you read the message?" This question is asked to everyone in the sample, as we send SMS forecasts to both treatment arms (i.e. no pure control group). We denote the number of days a subject i reports to have seen the SMS as R_i. We code "not sure" and "refused to respond" as $R_i = 0$. A subject's takeup is $P_i = \frac{R_i}{7}$, i.e. the fraction of forecasts respondents report to have seen. We acknowledge that R_i is likely measured with error, and that the reported value may depends on salience of the SMS forecasts and other factors that may be influenced by treatment. As such, we interpret R_i as a measure of attention to the SMS forecasts, which we exogenously vary.

The treatment-on-the-treated (TOT) effects is estimated using 2SLS, with Z_g or $\boldsymbol{A}$ instrumenting for P. We present the following first and second-stage specifications for a within-subject model with Z_g as an instrument.

$$P_{Ti} = \eta_T + Z_g^{'}\phi_T + \nu_T Y_{0i} + \boldsymbol{X_i'\theta_T} + \upsilon_{Ti}$$

$$Y_i = \alpha + \widehat{P'}\beta + \gamma Y_{0i} + \boldsymbol{X_i'\delta} + \varepsilon_i$$

$\widehat{P}$ is the instrumented "takeup." Much of the rest of the specification and testing remain the same as in the ITT; we include the same set of controls in the first- and second-stage regressions and carry out two-sided tests on the same set of outcomes. The between-subject models are analogous to the equations above, except for the latter in which we omit $\nu_T Y_{0i}$ and γY_{0i}.

5.6 Heterogeneous effects

We consider dimensions of heterogeneity that we expect to drive the preferences for air quality information sources. We are primarily interested in (a) baseline beliefs about, and preferences for, information sources, and (b) baseline beliefs about air quality levels and its deviation from the truth.

The first dimension is informed by an emerging body of work on media bias, trust for information sources, and polarization. Theoretical and empirical work in this literature shows that agents may place heavier weights on information from a source that align with their priors, leading to polarization in preferences and beliefs (e.g. Gentzkow et al. 2023; Chopra et al. 2022).[5] If, on the other hand, agents do not exhibit belief confirmation or do not hold strong priors about the sources' quality, they may shift their priors more strongly to information from a source that they are less exposed to at baseline. As such, it is *a priori* unclear how the demand for the sources evolve based on their baseline preferences and beliefs. The second dimension is of more standard Bayesian concern, in that individuals who are less well-informed about air quality levels may hold priors with more deviations from the truth. Those individuals may therefore update their beliefs more strongly toward the truth based on the signals they receive and value the SMS forecasts more.

5.6.1 Dimensions of heterogeneity

For the dimension of heterogeneity on baseline preferences for, and beliefs about, the sources of air quality information, we use the following proxies:

1. donation share of PKR 100 between government's environmental agency vs. citizens' group that tackles air pollution

2. categorical variable of overall approval: "Government-leaning" if the respondents' Likert-scale approval measure for the government is greater than that for the citizens' group, "Citizens' group-leaning" if vice versa, and "neutral" if they equally approve the two sources

3. categorical variable of accuracy: same as above, except the Likert-scale measure captures respondents' beliefs about accuracy of the sources' air quality information services.

For robustness, we also consider other definitions of baseline preferences and beliefs, such as the original Likert scales used to construct the proxies above, as well as the respondents' primary news sources' political leanings.

[5] This may be driven by "belief confirmation," i.e. they prefer sources that distort information toward their prior beliefs (Mullainathan and Shleifer 2005), or driven by uncertainty about accuracy of information sources, inducing an individual to put heavier weights on their preferred source (Gentzkow and Shapiro 2006).

For the dimension of heterogeneity on baseline beliefs about air quality and its deviation from the truth, we use the following proxy:

- baseline outcome variable 2.: absolute error of incentivized $t + 1$ forecast of PM2.5 levels.

We also use several other definitions of baseline beliefs to test, for instance, for asymmetry based on the direction of the error.

5.6.2 Estimating equations

For brevity, we present the specification for within-subject analysis of the ITT effects. The between-subject analysis and TOT effects follows a similar structure. The estimating equation is as follows:

$$Y_i = \alpha + Z_{g_i}\boldsymbol{H_i'}\boldsymbol{\beta} + \boldsymbol{X_i}\boldsymbol{\gamma} + \varepsilon_i$$

H_i is the relevant dimension of heterogeneity, coded as a matrix consisting of vectors of dummies that may represent bins of an underlying continuous or categorical variable. We include all bins so that we separately estimate treatment effects treatments for each bin (i.e. $\boldsymbol{\beta}$ is a vector) and Z does not enter the equation separately.

We also estimate specifications where the underlying dimension of heterogeneity is a continuous variable (such as the donation share and the absolute forecast error). The estimation equation in such case would be as follows:

$$Y_i = \alpha + Z_{g_i}{H_i}'\beta + H_i\beta_h + \boldsymbol{X_i}\boldsymbol{\gamma} + \varepsilon_i$$

In this specification, H also enters separately to control for baseline level of the dimension of heterogeneity.

References

Husnain F Ahmad, Matthew Gibson, Fatiq Nadeem, Sanval Nasim, and Arman Rezaee. Forecasts: Consumption, Production, and Behavioral Responses. 2022.

S. Athey and G.W. Imbens. The Econometrics of Randomized Experiments. In *Handbook of Economic Field Experiments*, volume 1, pages 73–140. Elsevier, 2017. ISBN 978-0-444-63324-8. doi: 10.1016/bs.hefe.2016.10.003. URL `https://linkinghub.elsevier.com/retrieve/pii/S2214658X16300174`.

Ceren Baysan. Persistent Polarizing Effects of Persuasion: Experimental Evidence from Turkey. *American Economic Review*, 112(11):3528–3546, 2022. ISSN 0002-8282. doi: 10.1257/aer.20201892. URL `https://www.aeaweb.org/articles?id=10.1257/aer.20201892`.

Gordon M. Becker, Morris H. Degroot, and Jacob Marschak. Measuring utility by a single-response sequential method. *Behavioral Science*, 9 (3):226–232, 1964. ISSN 1099-1743. doi: 10.1002/bs.3830090304. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/bs.3830090304`. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bs.3830090304.

A. Belloni, V. Chernozhukov, and C. Hansen. Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies*, 81(2):608–650, April 2014. ISSN 0034-6527, 1467-937X. doi: 10.1093/restud/rdt044. URL `https://academic.oup.com/restud/article-lookup/doi/10.1093/restud/rdt044`.

Felix Chopra, Ingar Haaland, and Christopher Roth. CESifo Working Paper no. 9673. 2022.

Jishnu Das, Alaka Holla, Aakash Mohpal, and Karthik Muralidharan. Quality and Accountability in Health Care Delivery: Audit-Study Evidence from Primary Care in India. *American Economic Review*, 106(12):3765–3799, December 2016. ISSN 0002-8282. doi: 10.1257/aer.20151138. URL `https://www.aeaweb.org/articles?id=10.1257/aer.20151138`.

Richard Freeman, Wenquan Liang, Ran Song, and Christopher Timmins. Willingness to pay for clean air in China. *Journal of Environmental Economics and Management*, 94: 188–216, March 2019. ISSN 0095-0696. doi: 10.1016/j.jeem.2019.01.005. URL `https://www.sciencedirect.com/science/article/pii/S0095069617308902`.

Matthew Gentzkow and Jesse M. Shapiro. Media Bias and Reputation. *Journal of Political Economy*, 114(2):280–316, 2006. ISSN 0022-3808. doi: 10.1086/499414. URL `https://www.jstor.org/stable/10.1086/499414`. Publisher: The University of Chicago Press.

Matthew Gentzkow, Michael B Wong, and Allen T Zhang. Ideological Bias and Trust in Information Sources. 2023.

Dalia Ghanem and Junjie Zhang. 'Effortless Perfection:' Do Chinese cities manipulate air pollution data? *Journal of Environmental Economics and Management*, 68(2):203–225, September 2014. ISSN 0095-0696. doi: 10.1016/j.jeem.2014.05.003. URL `https://www.sciencedirect.com/science/article/pii/S0095069614000400`.

Koichiro Ito and Shuang Zhang. Willingness to Pay for Clean Air: Evidence from Air Purifier Markets in China. *Journal of Political Economy*, 128(5):1627–1672, May 2020. ISSN 0022-3808. doi: 10.1086/705554. URL `https://www.journals.uchicago.edu/doi/full/10.1086/705554`. Publisher: The University of Chicago Press.

Akshaya Jha and Andrea La Nauze. US Embassy air-quality tweets led to global health benefits. *Proceedings of the National Academy of Sciences*, 119(44):e2201092119, November 2022. doi: 10.1073/pnas.2201092119. URL `https://www.pnas.org/doi/10.1073/pnas.2201092119`. Publisher: Proceedings of the National Academy of Sciences.

Sendhil Mullainathan and Andrei Shleifer. The Market for News. *American Economic Review*, 95(4):1031–1053, September 2005. ISSN 0002-8282. doi: 10.1257/0002828054825619. URL `https://www.aeaweb.org/articles?id=10.1257/0002828054825619`.

Karthik Muralidharan and Venkatesh Sundararaman. The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India. *The Quarterly Journal of Economics*, 130(3):1011–1066, August 2015. ISSN 0033-5533. doi: 10.1093/qje/qjv013. URL `https://doi.org/10.1093/qje/qjv013`.

6 Figures

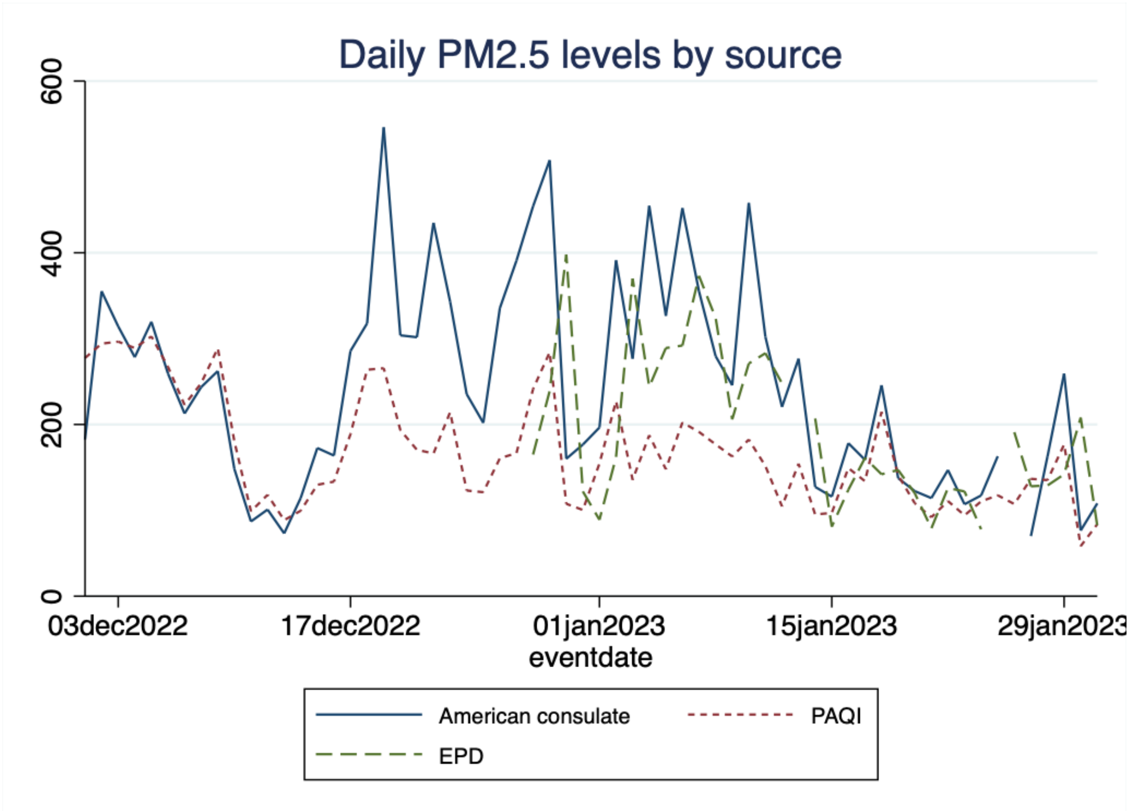


Figure 1: This figure shows the daily average PM2.5 concentration (in $\mu g/m^3$) levels by sources. “American consulate” refers to readings from the air quality monitor at the American consulate in Lahore. We treat this reading as the ground truth. “PAQI” refers to readings from the average of lower-cost air quality monitors managed by Pakistan Air Quality Initiative (PAQI) in Lahore. “EPD” refers to readings from air quality monitors managed by the Environmental Protection Department (EPD) of the Government of Punjab Province.

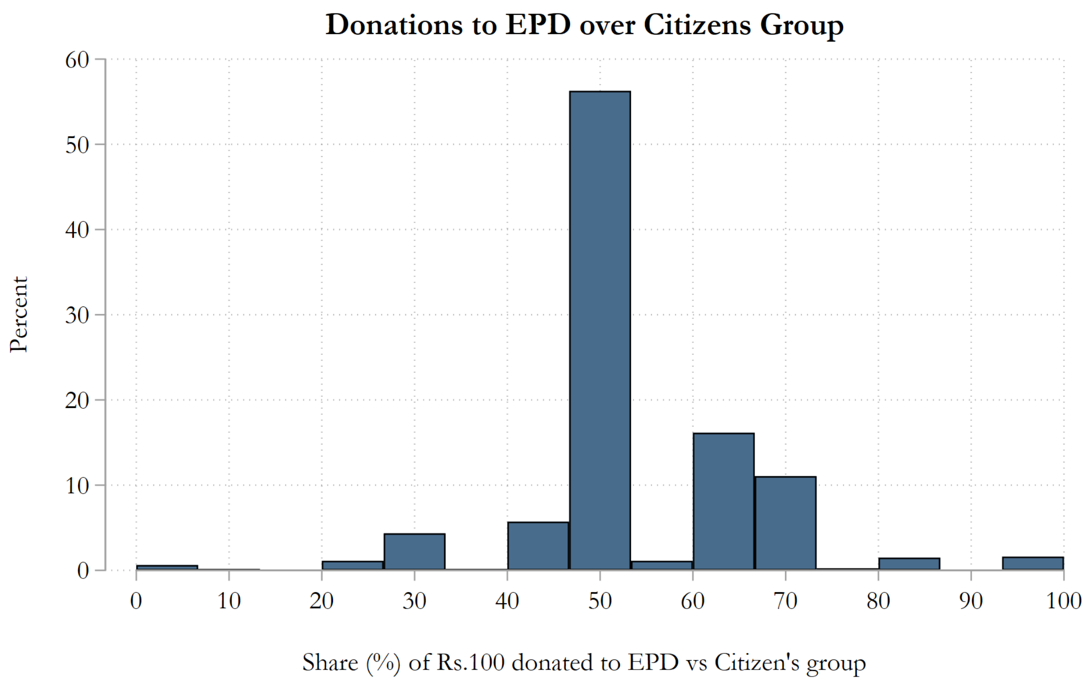


Figure 2: This figure shows the result from the donation game in our baseline survey, in which we asked respondents to split PKR 100 between the government (EPD) and private (PAQI) sources.

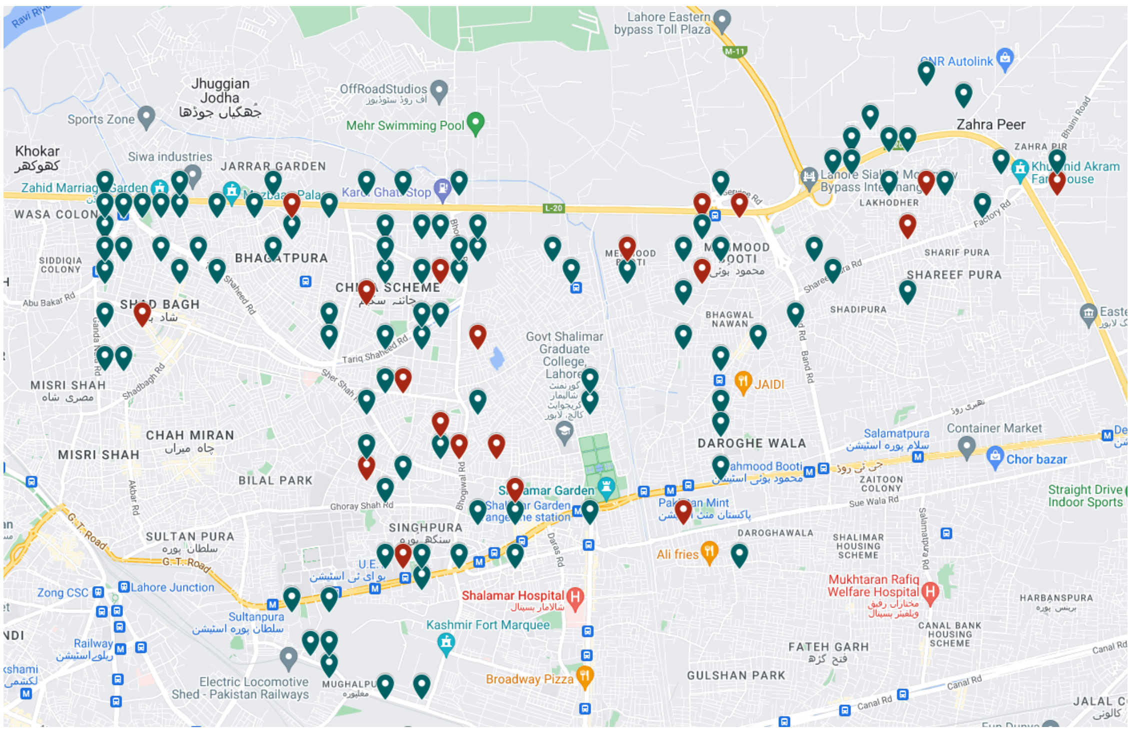


Figure 3: Sampling coordinates in NA-123 and NA-124 constituencies in Lahore, Pakistan

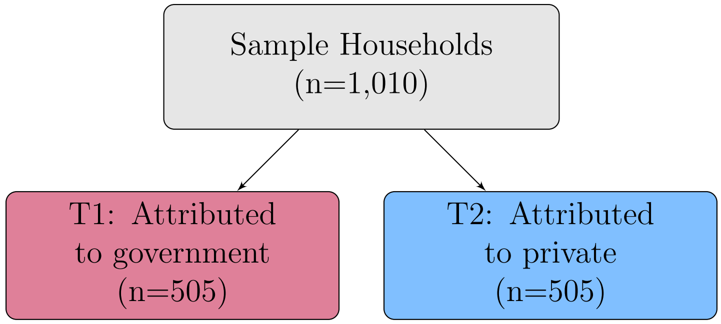
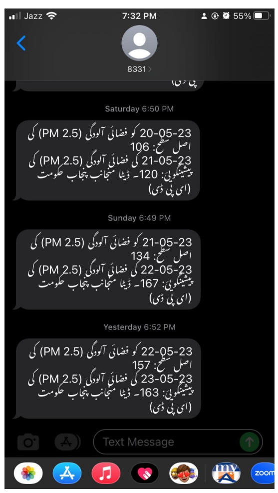


Figure 4: Treatment Groups



(a) T1: Daily messages

(b) T2: Daily messages

Figure 5: Sample messages to respondents

7 Tables

Table 1: Balance Table

Variable	(1) No Mean/(SE)	(2) Yes Mean/(SE)	(1)-(2) Pairwise t-test Mean difference
Absolute Difference of PM 2.5 Truth and Forecast	96.014	93.812	2.202
	(2.092)	(2.183)	
Ratio of Absolute Difference and Truth for PM 2.5	0.724	0.713	0.011
	(0.019)	(0.019)	
Share (%) of Rs. 100 donated to EPD	50.139	50.059	0.079
	(0.681)	(0.655)	
Hours Spent Outdoors	7.414	7.446	-0.032
	(0.204)	(0.197)	
Stated preference for citizens group	0.009	-0.008	0.017
	(0.042)	(0.043)	
Stated preference for government	-0.013	-0.008	-0.005
	(0.043)	(0.043)	
Comprehended the text message without explanation	0.768	0.766	0.002
	(0.019)	(0.019)	
Received air pollution info from: EPD	0.087	0.083	0.004
	(0.013)	(0.012)	
Received air pollution info from: AirVisual App	0.097	0.089	0.008
	(0.013)	(0.013)	
Index: Sentiment on air quality	-0.022	0.008	-0.029
	(0.032)	(0.032)	
Main TV channel: Geo News	0.428	0.457	-0.030
	(0.022)	(0.022)	
Main TV channel: ARY News	0.156	0.139	0.018
	(0.016)	(0.015)	
Main TV channel: City 42	0.081	0.095	-0.014
	(0.012)	(0.013)	
Main TV channel: Express News	0.081	0.081	0.000
	(0.012)	(0.012)	
Index: Household asset ownership	0.022	-0.022	0.043
	(0.046)	(0.043)	
F-test of joint significance (F-stat)			0.359
Number of observations	505	505	1010

Significance: ***=.01, **=.05, *=.1. Errors are robust.