Beliefs, signal quality, and information sources: Experimental evidence on air quality in Pakistan*

Isra Imtiaz Shotaro Nakamura Sanval Nasim Arman Rezaee

November 15, 2023

Abstract

When governments in developing economies under-provide environmental information, consumers may demand private alternatives depending on their beliefs about the quality of information. We study how information sources shape demand for air quality information via a randomized controlled trial in which we provide day-ahead air pollution forecasts. We make salient one of the information sources: the government vs. a private citizens group. We find that our respondents in Lahore, Pakistan, have a high willingness to pay for the forecast service, yet there is no difference by the assigned source. However, respondents show a significantly higher revealed relative preference for the assigned source, as measured through a donation game. Respondents also believe the government's forecast error is 12 percent higher than the private alternative's. Our findings suggest that respondents have weak priors and malleable preferences for information sources yet expect lower service quality from the government.

JEL: Q53, Q56, D83, H41

Keywords: air quality, beliefs, environmental information, willingness to pay

*Imtiaz: Mahbub ul Haq Research Centre, Lahore University of Management Sciences, Lahore, Pakistan. Nakamura: Federal Trade Commission, Washington DC; email: shotaro.n.nakamura@gmail.com. Nasim: Department of Economics, Colby College. Rezaee: Department of Economics, University of California, Davis. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the Federal Trade Commission, or its Commissioners. We thank Mohammad Malick and Ahsan Tariq for their excellent management of surveys and SMS intervention. This study is funded by grants from the International Growth Center (IGC) and faculty funding from Colby College and UC Davis. We would like to thank the Punjab Environment Protection Department (EPD), the Urban Unit, AirNow, SPRINTARS, and the Pakistan Air Quality Initiative (PAQI) for sharing their data. This study has IRB approval in the US (UC Davis IRB-1336133-1) and Pakistan (LUMS IRB 02242021). The pre-analysis plan is registered with the AEA (AEARCTR-0011489).

1 Introduction

An emerging body of empirical work shows positive 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 measure that could yield considerable public benefit is accessible and reliable air-quality information. A previous experimental study in Lahore, Pakistan, revealed that citizens have a high willingness to pay for information, which helps improve their forecast ability and take on avoidance behaviors (Ahmad et al. 2022). Governments in developing countries, however, often struggle to provide consistent and reliable air quality readings due to resource and capacity constraints or perverse incentives to obscure the true extent of environmental degradation (e.g.Ghanem and Zhang 2014). In response, various stakeholders, including citizens' groups, international and bilateral agencies, and research institutions, have begun providing air quality information.

Alternatives to government services may improve citizens' access to air quality information. However, their efficacy may depend upon citizens' preferences for and beliefs about the accuracy of the information sources. First, sources may differ on when they make readings available and, if so, what the readings are. Second, citizens may have differing preferences for air quality information sources. It is, however, unclear if service quality *per se* drives the demand for information sources or if consumers hold beliefs about the sources that drive the demand for information services.

In this paper, 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 the attributed source: government or a private alternative. We address the following research questions: First, are consumers willing to pay for air quality information, regardless of its attributed sources? Second, do consumers have differential demand for air quality information by the source, even when service quality is held constant? Third, what mechanisms drive the differential willingness to pay or the lack thereof? In other words, how does information affect the relative preferences for sources and underlying beliefs about service quality? Lastly, what do our results imply about the costs and benefits of competing information sources in terms of access to and consumer welfare regarding air quality information?

We address these questions via a randomized information intervention in working-class neighborhoods in Lahore, Pakistan, a context in which there is severe air pollution, limited government services around air quality information, and multiple and often conflicting information sources. In our intervention, we provide identical air quality forecasts to a sample of residents via SMS and randomly vary the attributed source between the government and the private alternative. We develop an ensemble forecast model of day-ahead air pollution using data inputs from multiple sources, including government and private monitors. In one arm, respondents are told that the daily forecasts are constructed using data from the 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 do not have a control arm in which we do not provide air quality forecasts and focus on the differential effects of source attribution instead.

The experimental setup allows us to measure whether and how citizens value air quality information and how they value and trust the sources from which the information comes. We conduct a series of incentivized games in which we measure their willingness to pay for air quality forecasts, elicit their beliefs about air quality levels and the accuracy of our forecast service, and their preferences for the sources in terms of monetary donations to government and private agencies.

We find that residents of working-class neighborhoods in Lahore have high levels of demand for air quality information, yet not differentially so by the salient information source.

The average willingness to pay for two additional months of service after experiencing it for free is PKR 238, equivalent to a month of prepaid mobile and data services. Yet, we do not find significant differences in either their willingness to pay or their forecast ability between the two treatment arms. We hypothesize that the recipients are satisfied with the services they received regardless of treatment arms. We confirm through stated preference measures in the endline survey that recipients are equally satisfied with the reliability and accuracy of the SMS service. Given that actual forecast values are identical between treatments by design, we argue that the information source alone does not lead to differential changes in beliefs about air quality levels.

We find, however, that the recipients' preferences *between* sources within individuals shift significantly as a result of exposure to randomly assigned sources. We measure relative preference for sources via a donation game in which respondents choose to donate a fixed sum between the government and private sources. At baseline, most respondents choose to split their endowment equally between the government and private sources in a donation game. Yet, at the endline, respondents in the government arm favor the government source over the private alternative by 75:25, and vice versa for those in the private arm. Our findings suggest that preferences are relatively malleable in a highly frictional market for information services.

The above findings suggest that respondent update their beliefs about the service quality of the exposed source favorably. Yet, respondents' views on service quality may consist not only of information's accuracy but also of other factors like punctuality and accessibility. We isolate beliefs about information accuracy using outcomes from a set of incentivized forecast games and observe other aspects of service quality with stated preference measures. We find that respondents in the government arm expect a 12% higher error in the SMS forecast than those in the private arm. However, we do not find statistically significant differences in stated preference measures of aspects of service quality. We, therefore, find that those assigned to the government arm believe the information they receive is less accurate than those in the private arm do, while they are equally willing to pay for the information service.

Our results on how consumers value forecast services, their sources, and accuracy suggest several possibilities as to how a social planner could expand access to air quality information. First, a high average willingness to pay for SMS-based forecasts indicates the potential for scaling up such services, regardless of the sources from which the information comes. Second, given that consumers form preferences for an experimentally assigned source against the alternative, sources may affect access to information in the long run once consumers experience and form beliefs about the sources. Third, an equally high willingness to pay for government services as the private alternative, despite the belief that government services are less accurate, suggests that further work is needed to understand the trade-off between accuracy of and the demand for information services.

Our work builds on several strands of the literature to understand how consumers form beliefs and update their preferences in the context of environmental public goods. 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 evaluate the importance of information sources and consumers' beliefs in shaping the demand for such services. Second, we provide relevant insights into the accountability and competition for publicly provided services in developing economies by focusing on environmental services (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 shed light on 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.

The rest of the paper is organized as follows: In section 2, we discuss the information environment for air quality in Lahore, Pakistan. In Section 3, we discuss the experimental design. In Section 4, we discuss the pre-specified outcome variables and the identification strategy in Section 5. In Section 6, we present the pre-specified results and conclude in Section 7.

2 Air quality information in Lahore

Lahore often ranks as one of the worst cities in terms of air quality around the world.[1] Access to accurate and timely information about air quality is critical for assessing the risks of exposure and taking evasive actions. Yet, access to information is limited for the average citizen due to intermittent government services and technological and other frictions in accessing the daily readings. Working-class citizens of Lahore do not access air quality readings, as presented in our baseline survey on Table 1. We find that 9 percent of our respondents state to have accessed air pollution information from a government agency that is responsible for collecting and disseminating environmental information: the Environmental Protection Department (EPD).

The EPD publishes daily reports, including readings from three to four monitor locations, each reporting one of the scheduled pollutants.[2] Data by EPD, however, are often missing in peak smog seasons in late fall and, when available, diverge significantly from those of other

[1] https://www.iqair.com/world-air-quality-ranking

[2] 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 is reported in two different indices: (i) pollutants' concentration and (ii) AQI (Air Quality Index). All of these indices are reported as 24-hour averages.

monitors. Figure 1 shows daily PM2.5 concentrations from EPD and two other sources. The readings are missing for EPD for most of the month of December, usually one of the worst air quality periods, because EPD did not upload readings. Furthermore, daily readings from EPD are hard to access for average citizens, as they are only made public as PDF reports in English on their website and are not well publicized.[3] The EPD 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."

The limited information environment has led citizens' initiatives, like the Pakistan Air Quality Initiative (PAQI), to collect and publish their own data. Started in 2016, PAQI crowd-sources several low-cost air quality monitors (IQAir and PurpleAir) that were 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.[4] PAQI also has a Twitter that disseminates daily readings from Lahore.[5] Yet, the vast majority of the working-class population is unaware of PAQI's data and initiatives, as they may not own smartphones needed to download the AirVisual app or use Twitter.[6] Table 1 also shows that approximately 9% of our sampled households stated to have accessed air quality readings from the AirVisual app at baseline.

Other sources of air quality information exist for Lahore but are less known than the EPD or PAQI or are equally hard to access for the average citizen. The most prominent among such sources is the U.S. Consulate General in Lahore, which has a high-quality monitor within the compound.[7] The Consulate shares their readings on their website and

[3] The daily reports are posted at https://epd.punjab.gov.pk/aqi

[4] e.g., Lahore and Lahore American School

[5] @LahoreSmog

[6] Anecdotally, Twitter is considered to be an upper-middle-class social network in Pakistan, while Facebook, WhatsApp, and voice-based social media services that require less data are popular among working-class populations.

[7] The U.S. Consulate General in Lahore hosts an air quality monitor funded by the U.S. EPA. The

on Twitter, but, to our knowledge, do not actively engage in other forms of dissemination. Another government agency called the Urban Unit owns an air quality monitor but has not been consistently publishing readings for the public's consumption.[8] One could also access forecasts based on satellites and meteorological models.[9] However, we are not aware of any satellite- or other model-based services that actively disseminate air quality information for Lahore or Pakistan.

Table 2 shows the summary statistics by the air quality information source. First, we find that compared to AirNow, a U.S. Consulate measure that we define as the ground truth in Section 3.1, both EPD and PAQI measures report lower average PM2.5 concentration levels. Second, even though EPD values are higher than the PAQI values during the pre-intervention period with high pollution, there are 36 fewer observations due to non-reporting. Third, during and after the intervention period, with relatively low pollution, both the EPD and PAQI measures are closer on average to AirNow and are missing on fewer days. In the next section, we discuss how we synthesize multiple sources with varying quality into an

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 U.S. Consulate's compound in Shimla Hills, Lahore. The standards for the monitors installed are provided at `https://www.epa.gov/system/files/documents/2022-12/List_of_FRM_and_FEM.pdf`.

[8]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 members on a daily basis.

[9]One example of such an approach is the Spectral Radiation-Transport Model for Aerosol Species (SPRINTARS), a numerical model that 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 are able to access the hourly forecasts generated by SPRINTARS.

SMS-based forecast service as part of an experimental intervention.

3 Experimental design

Multiple and conflicting information sources create an environment in which consumers have to a) gauge what the true extent of air pollution is and b) decide which sources to access. In our intervention, we provide identical air quality information with experimentally varied information sources. We present the timeline in Figure 3. We randomize a sample of 1,010 households into two treatment arms: government (EPD) and private (PAQI). We optimize power to detect differences between the treatment arms and do not have a pure control arm. The assigned source is made salient when they receive daily air quality forecasts based on an ensemble prediction model. The information provided to the two treatment arms is otherwise identical.

In this section, we first discuss how we define the ground truth under multiple information sources and how we construct the forecast model. We then discuss sampling, randomization, and messaging to the sampled households.

3.1 Defining the ground truth

The existence of multiple and often conflicting information sources creates a conceptual and empirical challenge: defining the ground truth of air quality levels. Because our research questions evolve around the role of information sources on consumers' beliefs about some objective measures of the truth, we choose an independent source of information from either the government or the private alternative. We choose the U.S. Consulate monitor to be the independent source of truth, as it is presumably of the highest quality using the reference method in compliance with the U.S. EPA standards.

The measure of interest is the daily average concentration of PM2.5 (in $\mu g/m^3$). We construct the measure based on hourly readings between 12:00 AM and 4:00 PM from the U.S. Consulate.[10] The time window was selected so that the research team could collect the data, estimate the next-day forecast, and send them via SMS to our sample households between 6:00–8:00 PM. We have chosen this time frame as we learned in Ahmad et al. (2022) that most respondents make plans for the next day in the evening.

We find that air quality readings from EPD, PAQI, and other sources in Lahore are correlated with those from the American Consulate. Yet, there are significant deviations from the American Consulate, and the magnitude of deviation varies by the source. First, Table 3 shows the pairwise correlations between air quality monitoring sources. We find that the American Consulate readings are highly correlated with PAQI's readings ($\rho = 0.82$) but much less so with the EPD ones (0.61). Second, Table 4 shows that the EPD sources have higher deviation on average than PAQI from the American Consulate readings by approximately 20 $\mu g/m^3$.

3.2 Forecast model

Given the ground truth defined in Section 3.1, we construct a model to predict it for the next day ($t + 1$). The objectives of the model are twofold: 1) to provide the most accurate forecast possible given the set of available information, and b) to ensure that information from both EPD and PAQI are used to construct the forecast. The latter is key to ensuring that the messages we convey to our respondents about the use of EPD's and PAQI's data are true. We achieve these objectives by constructing an ensemble forecast from multiple

[10]We rely on other sources when the U.S. Consulate monitor readings are not always available. When only the U.S. Consulate readings are missing, 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 24 May 2023, the U.S. Consulate monitor is missing readings for 16 out of the 97 intervention days. Out of 16 days where the U.S. Consulate is missing data, the Urban Unit is missing data on 4 days.

forecasts, each of which relies on a single information source.

First, we construct four forecast models, all of which predict the ground truth, but the air quality readings we include on the right-hand side are from one monitor. The readings data on the right-hand side are the $t-6$ to t lagged readings of either the U.S. Consulate, EPD, PAQI, or the Urban Unit. Since SPRINTARS already provides predictions based on their model, we simply take their $t+1$ forecast. Each model, except for SPRINTARS, also uses historical meteorological readings and weather forecasts for $t+1$ as inputs. [11] For each of the models, we use an adaptive Lasso model and predict $j+1$ PM2.5 concentration using a model trained on data from Day 1 to Day j, for j going from Day 20 to t. This leads us to have $t-20$ out-of-sample forecasts, the last of which is for Day $t+1$, for each of the models.

We then combine 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.

3.3 Sampling

The intervention is conducted in lower-middle-class neighborhoods of National Assembly (N.A.) 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.

[11] The weather inputs for the model are:

- AccuWeather's $t+1$ forecasts for minimum temperature, maximum temperature, and precipitation in inches, as well as their squared values
- Historical weather data on a 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

Figure 4 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 ten households by spiraling out from the centroid counterclockwise.

3.4 Randomization

Figure 5 shows that the sampled households are divided into two treatment arms. In T1, SMS forecasts are attributed to a government agency (EPD), while in T2, they are attributed to a citizens' group (PAQI). We do not have a pure control group that does not receive SMS forecasts, as the main purpose of this study is to understand the effect of information sources, holding constant the qualities of service and information.

We stratify the randomization process into the two treatment groups on a set of baseline variables that either a) we considered as potential outcome variables, b) proxies of potential outcome variables that we were unable to collect at baseline due to the experimental design, c) some dimensions of heterogeneity that were considered pre-intervention, or d) the household asset index. [12] We use the optimal-greedy algorithm and generate blocks using the Minimum Volume Ellipsoid (MVE) estimator. We are primarily concerned about balance

[12] The final set of stratified variables are as follows:

1. absolute error of incentivized $t + 1$ forecast of PM2.5 concentration (i.e., primary outcome 4.2)
2. share of donations to government vs. citizens' group (i.e., primary outcome 4.4)
3. time spent outdoors (i.e., secondary outcome E.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 the AirVisual app (on which PAQI posts air quality readings)
9. Indicators of respondents' main T.V. news source
10. Asset index: a count of assets (electricity, appliances, vehicles, and number of rooms)

on outcome variables at baseline, as well as the "take-up" in terms of exposure and comprehension of our SMS forecast messages. We follow the advice from Athey and Imbens (2017) that each block contains two units per treatment arm. We then assign subjects to T1 and T2.

3.5 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) is made salient. The daily messages also contain the readings from time *t*. We provide identical t+1 forecasts and daily readings to the two treatment arms, only varying the source made salient.

The subjects also received an introductory message before the start of the daily SMSs and a reminder message every two weeks over the course of the intervention. The daily messages are sent out around 6:00–8:00 PM starting on 18 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).

3.5.1 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]."[13]

3.5.2 Daily forecast messages

The daily messages are sent around 6:00–8:00 PM after collecting the day's data and estimating the forecast for *t+1*. The message on, for instance, 18 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

[13] We use the shorthand "NGO" to refer to organizations of a type, such as PAQI, for the purpose of familiarity with our subjects.

Figure 6 shows screenshots of the daily messages for T1 and T2. Because the text messages are sent from the same number every day, it is easy to compare the forecast values for Day *t* provided on Day *t-1* to the realized value provided on Day *t*.

3.5.3 Fortnightly reminder messages

Starting on Saturday, 4 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."

4 Primary outcome variables

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

4.1 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 respondent's 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.

4.2 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.

4.3 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.

4.4 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 environmental protection purposes: a government institution and PAQI.

5 Identification strategy

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 denote treatment assignment as a vector, whose inputs are equal to 1 if the respondent is assigned to the government arm and 0 if assigned to the citizens' group arm.

5.2 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 4.1)
2. The demand for air quality information is different between the treatment (citizen's group) and control (government) groups (tested on outcome 4.1)
3. Treatment affects beliefs about air quality differentially relative to control (tested on outcome 4.2)
4. Treatment affects the perceived accuracy of air-quality information source relative to control (tested on outcome 4.3)
5. Treatment affects policy preferences for air quality relative to control (tested on outcome 4.4)

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

5.3 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.

5.4 Treatment Effects

5.4.1 Intent to treat

We estimate the treatment effects between subjects as follows;

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

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;

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

We denote Y_0 as the baseline measure of the outcome variable Y. Much of the details about the specification and inference are 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.. We also pre-specified a treatment-on-the-treated identification strategy in the pre-analysis plan. However, we do not find significant first-stage results and therefore put this identification strategy to Appendix Section D.1.

5.5 Heterogeneous effects based on prior beliefs

We also prespecify and test for heterogeneous treatment effects across dimensions that we expect to drive the preferences for air quality information and sources.[14] The dimensions of interests are a) baseline beliefs about, and preferences for, information sources and b) baseline beliefs about air quality levels.

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 aligns with their priors, leading to polarization in preferences and beliefs (e.g., Gentzkow et al. 2023; Chopra et al. 2022).[15] 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 evolves 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.

5.5.1 Measures of the dimensions of heterogeneity

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

[14] We do not, however, adjust for multiple testing in these secondary hypotheses.

[15] 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).

1. donation share of PKR 100 between government's environmental agency vs. citizens' group that tackles air pollution
   - For categorical variables, code as "more to government," "more to citizens' group," and "50-50" or into 10-rupee bins
2. Relative overall approval of government vs. citizen sources: difference in Likert-scale approval measures for the government and citizens' groups for their air quality information services.
   - For a categorical variable, code as "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. Relative beliefs on the accuracy of government vs. citizen sources: difference in Likert-scale measures for the government and citizens' groups for their air quality information's accuracy.
   - For a categorical variable, code as "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

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.

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 4.2: absolute error of incentivized $t+1$ forecast of PM2.5 levels.

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

5.5.2 Estimating equations

The estimating equation to identify the linear ITT effect is as follows:

$$Y_i = \alpha + Z_i\beta + Z_iH_i\theta + H_i\delta + \boldsymbol{X_i}\boldsymbol{\gamma} + \varepsilon_i$$

H_i is the relevant dimension of heterogeneity as a continuous variable and Z_i the treatment assignment variable that is 1 for the Government arm. We interpret the coefficients $\hat{\beta}$ and $\hat{\theta}$ as estimates of average treatment and heterogeneous treatment effects, respectively.

We also estimate a model where the dimension of heterogeneity is categorical. The estimation equation is as follows:

$$Y_i = \alpha + Z_i\beta + \sum_{j\in J} Z_iH_i\theta_j + \sum_{j\in J} H_i\delta_j + \boldsymbol{X_i}\boldsymbol{\gamma} + \varepsilon_i$$

H_i is the relevant dimension of heterogeneity as a categorical variable, and each category is denoted as j. We interpret the coefficients $\hat{\beta}$ and $\hat{\theta}_j$ as estimates of the average treatment effect and heterogeneous treatment effect for a group $H_i = j$, respectively.

6 Results

6.1 Checks on balance

We test the balance of variables used for blocking between the two treatment arms as well as other additional variables. 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. Table 1 shows the balance on the variables used in the blocking procedure. We do not find statistically significant differences in any of the primary outcomes for which we have baseline measures or other variables over which we stratified our randomization.

6.2 Prespecified outcomes

Table 5 shows the coefficients and their standard errors of the intend-to-treat estimates for the five prespecified primary hypotheses using post-double-selection LASSO. Here, by "treatment," we mean being assigned to the government arm, as opposed to the citizen's group arm. Table 6 shows the p- and q-values of the corresponding columns. In the following subsections, we center our analysis on the four prespecified primary outcomes and five hypotheses presented in Table 5. We then complement the findings with non-primary outcomes and analyses.

6.3 Willingness to pay for air quality information

We find that the respondents have a high willingness to pay, but not differentially between treatment arms. Column 1 in Table 5 shows that the respondents are willing to pay PKR 238 for two months of air quality forecast services. This amount is roughly equivalent to a month of popular prepaid mobile and data services, often referred to as the "social" bundle by major carriers in Pakistan. For example, the Social Plus plan by Jazz includes 10Gb of data, 300 minutes of calls in-network, 50 minutes out-of-network, and 1,000 SMS messages and is priced at PKR 260 as of August 2023. [16] Figure 7 also shows the distribution of the willingness-to-pay for air quality forecasts as demand functions, indicating considerable heterogeneity.

We find, however, that there are no statistically and economically significant differences between the treatment arms in their willingness to pay for the forecasts. Column 2 in Table 5 shows that those assigned to the Government arm are willing to pay only PKR 0.33 more on average, and the difference is not statistically significant from zero. The small coefficient and standard error also exclude any economically meaningful difference between the two treatment arms. These results corroborate our secondary analysis in Appendix Section E on avoidance behaviors and policy preferences, where we do not find statistically significant differences by the treatment arms.

What would explain the high demand for information, yet no distinction in their differential willingness to pay by the information source? We hypothesize that the recipients are satisfied with the services they received regardless of treatment arms. We confirm through stated preference measures at the endline on the recipients' overall satisfaction with the service, their belief about punctuality, and their beliefs about the accuracy of our service. The results shown in Table 7 indicate recipients' satisfaction with the SMS forecast services in stated measures. The Likert scale means of l satisfaction, reliability, and accuracy are around 4 out of 5. The table also shows that the satisfaction with the SMS services is not statistically different between treatment arms.

[16] https://jazz.com.pk/prepaid/monthly-social-plus

6.4 Beliefs about air quality

We also find that different information sources do not lead to differential beliefs about air quality levels (i.e., "state of the world"). We measure the respondents' beliefs about air quality via incentivized t+1 forecasts. We would expect differences in the forecast error by treatment arm if information sources affect the magnitude by which recipients update their beliefs about the state of air quality toward our SMS forecast. Column 3 on Table 5 shows that those assigned to the government arm have, on average, a five percentage-point higher forecast error than in those in the private arm, although the difference is not statistically significant. The magnitude is also relatively small relative to the private-arm average error of 73% of the actual reading.

Other measures of beliefs about air pollution levels confirm our findings in Column 3, Table 5. First, we do not find statistically significant treatment effects on other definitions of air quality forecast, such as in level and absolute differences, as shown in Appendix Table A.4. Second, we also do not see significant differences by treatment in stated measures of concern about air quality. Appendix Table A.5 shows statistically insignificant results on the Likert-scale measure of concern about air quality and on the number of days in the last week that the respondents believed to have had good air quality.

Overall, we do not find that exposure to an information source alone leads to differential changes in beliefs about the state of air quality. One possible takeaway would be that consumers do not care about information sources. We cannot rule out this possibility, at least in terms of the demand for information and their beliefs about air quality. Another possibility is that individual consumers develop differential preferences for and beliefs about the quality of sources as they gain exposure through the SMS intervention. Such differential beliefs and preferences are possible in an environment with scarce information and limited access to air quality readings outside of our SMS forecast services.

6.5 Preferences for sources

We show in Section 6.3 that recipients of our SMS forecast service do not value the goods differentially treatment assignment. These are results on the differential demand for an identical good between individuals in two treatment arms. We find, however, that the recipients' preferences *between* sources within individuals shift significantly as a result of exposure to randomly assigned sources. We identify such effect based on our primary measure on preferences between sources.

For our primary measure of respondents' preference between sources, we conduct donation games with financial stakes. In the baseline and endline surveys, we ask respondents to allocate PKR 100 between government and private air quality monitoring sources, which the survey team donates to respective agencies. We argue that the relative allocations, as well as changes to them between baseline and endline, identify respondents' preferences for information sources.

We find that the respondents shift a larger fraction of their donations to the experimentally assigned sources at endline. Figure 2 shows the baseline distribution for both treatment groups and Figure 10 at the endline by treatment group. The figures show most respondents split the donations 50:50 at baseline, but their preferences diverge significantly by treatment arm at the endline. More than 90 percent of respondents who are assigned to the Government arm donate more to the government at the endline, as opposed to the private alternative. On the other hand, more than 90 percent of respondents assigned to the Private arm donate more to the private alternative at the endline. The average ratio between the assigned source and the other is approximately 75:25. Column 5 in Table 5 confirms that those assigned to the Government arm donate PKR 54 more to the government, on average, relative to the respondents in the Private arm.

Furthermore, we find evidence of a higher willingness to pay for information from the experimentally assigned source when we look *within* individuals. After the endline elicitation of willingness to pay for the SMS service, they are asked *hypothetically* how much they would be willing to pay if the forecast were to come from the other source (i.e., from the private group for those assigned to government, and vice versa). Column 4 in Table 8 shows that the respondents are, on average, willing to pay PKR 16 less for the alternative source than for the experimentally assigned one. Although the hypothetical WTP measure is not a revealed preference measure, it is in line with our findings from other measures of preferences for sources. Overall, our findings highlight the possibility that preferences are relatively malleable in a highly frictional market for information services.

6.6 Beliefs about service quality

Results discussed in Section 6.5 suggest that respondent update their beliefs about service quality of the exposed source favorably. Yet, service quality and its perception consist not only of how accurate the forecasts are, but also of other factors like how easy to access the information and whether it is consistently provided without delay.

As such, we isolate the respondents' beliefs about the accuracy of the SMS forecasts using outcomes from two incentivized forecast games in the endline survey. We conduct two types of incentivized elicitation regarding air quality forecasts: 1) respondents' belief about the actual air quality level tomorrow and 2) their guess of the SMS forecast. We argue that the absolute difference between the two measures is the respondents' belief about the quality of SMS forecasts. We then attribute the differences between treatment arms to respondents' beliefs about the sources' service quality.

We note that respondents indicate their general satisfaction with the service—including accuracy—in unincentivized elicitations as discussed in Table 7. Yet, it is possible that unincentivized and discretized measures may contain noise and fail to capture respondents' beliefs about nuanced parameters with precision.

We find that the respondents in the Government arm believe in larger SMS forecast errors than in the Private alternative. Column 4 on Table 5 shows the difference to be 2.8 points in terms of the concentration measure ($\mu g/m^3$). The effect size is 12% of the Private arm's mean (22.7). The effect is statistically significant at the 5% level and survives adjustments to multiple hypothesis testing, as shown in Table 6.

6.7 Heterogeneous treatment effects

We conduct a pre-specified analysis on heterogeneous treatment effects, as described in Section 5.5. We do not find strong evidence that the consumers respond differentially based on their prior beliefs about the information sources' service quality. At the same time, we find some evidence that consumers with higher baseline forecast errors have higher endline forecast errors if they are assigned to the government arm.

First, with respect to the consumers' preferences for sources, we do not find strong evidence of heterogeneous treatment effects except on endline donations to the government. Tables 9 to 11 show the linear heterogeneous treatment effect estimates and their categorical equivalents in Appendix Tables A.6 to A.8. Coefficients on interaction terms from 9 to 10 are not generally statistically significant. One exception is the negative interaction terms for the endline donation outcome (Column 5), which is likely because the outcome measure has a ceiling at PKR 100. In other words, those who report to prefer the government in baseline would donate more to the government and would not be able to increase donations to the government beyond PKR 100. One exception is the marginally significant interaction term in Column 3, Table 11, but this result is not corroborated with a categorical specification in Table A.8.

Second, with respect to the consumers' baseline forecast error, we find evidence of an adverse heterogeneous treatment effect of the government assignment on forecast error and on respondents' beliefs about the SMS's error. Table 12 shows the linear estimates, and Appendix Table A.9 the categorical equivalent. We find that for those assigned to the government arm relative to the citizens' group alternative, having a 100% larger baseline forecast error is associated with having 26% higher endline forecast errors. In other words, those with higher baseline errors update their priors less about air pollution levels relative to similar individuals if they are assigned to the government arm v.s. the citizens' group. Similar causal effects also exist on the respondents' beliefs about the SMS's errors but are less precisely estimated.

Two takeaways emerge as a result of the pre-specified analysis. First, it seems likely that there are no strong heterogeneous effects on the consumers' demand for air quality information based on consumers' priors about the sources. This may indicate that the consumers have relatively weak priors about information sources, and their beliefs are relatively malleable. Second, even when attributions to information sources do not meaningfully affect the demand for the ultimate service (air quality information), consumers with less accurate beliefs about air quality may update their beliefs slowly when they are assigned to the government source, presumably the lower quality of the two.

7 Conclusion

We study how residents living under uncertainty about the state of air pollution and information quality provided by multiple sources form beliefs and demand for information services. We conduct a randomized control trial in which we randomly attribute air quality forecast services to one of two sources: government and private. We first address whether lower middle-income citizens in developing cities are willing to pay for air quality information.

We then investigate if the random attribution leads to a differential demand for information or beliefs about air quality levels. We then study whether respondents hold varying beliefs about the information's accuracy or exhibit preferences *between* information sources.

We find that consumers in working-class neighborhoods of Lahore have a high willingness to pay for air quality information, yet it does not seem to be differentially driven by the associated source. We also find a strong preference toward the information sources to which the consumers are exposed. Furthermore, we find those assigned to the government arm believe the information they receive is less accurate than those in the private arm do, while they are equally willing to pay for the information service. Our results suggest that expanding access to air quality information may improve social welfare and provide insights into how a policymaker or a social planner may approach such expansion.

First, residents of Lahore value air pollution information, corroborating existing insights from (Ahmad et al., 2022). We find that our respondents—residents of a working-class neighborhood of Lahore—are willing to pay PKR 238 on average to continue receiving air pollution forecasts for another two months. This amount roughly translates to the cost of monthly prepaid mobile and data services. Thus, scaling the service across the city—with close to 14 million residents—will lead to large public benefits.

Second, the source of information does not affect residents' demand for air quality information or their beliefs about the state of air pollution. We do not find evidence that telling respondents that the forecasts they receive stem from a government or private source leads to differences in willingness to pay for the SMS forecasts. It may be that service quality is the dominant factor on demand over information sources, as most respondents would not have easily accessible and reliable air quality readings outside of our intervention. Yet, given that consumers form preferences for an experimentally assigned source against the alternative, sources may affect access to information in the long run once consumers experience and form beliefs about the sources.

Third, an equally high willingness to pay for government services as the private alternative, despite the belief that government services are less accurate, suggests that further work is needed to understand the trade-off between accuracy and the demand for information services. Several theories could explain why we see differential beliefs about the accuracy of information, yet no differential willingness to pay, by sources. One possibility is quality targeting, in that consumers are satisfied with the forecast quality of certain error levels, below which their willingness to pay is unchanged. This explanation may make sense in the world where the interpretation of air quality readings is "lumpy," i.e., air quality readings are often classified into non-linear tiers such as red, yellow, green for hazardous, unsafe, and satisfactory. Another theoretical explanation is that preferences for sources matter more than the information's quality. Such a theoretical explanation would have opposite policy implication to that of quality targeting, especially if our intervention could lead to polarized beliefs about the sources' service quality. Further work is needed to understand the trade-off between accuracy of and the demand for information services.

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. The demand for news: Accuracy concerns versus belief confirmation motives. 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`.

8 Figures

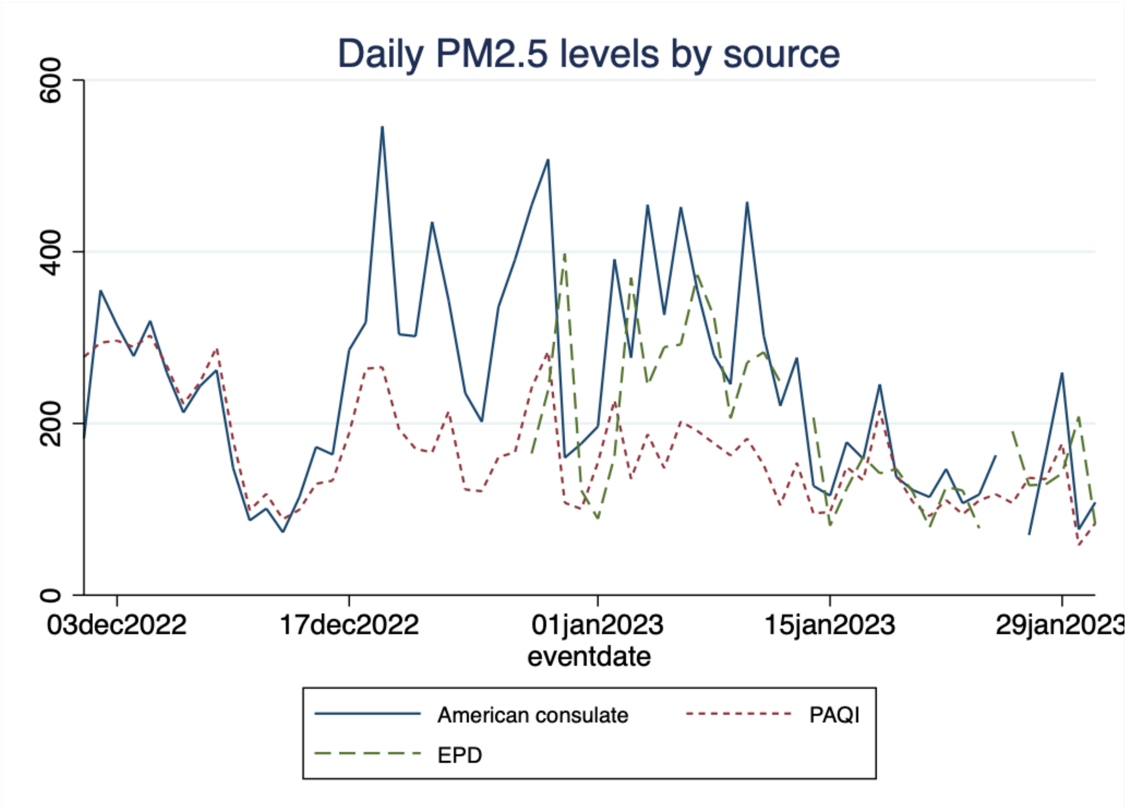


Figure 1: This figure shows the daily average PM2.5 concentration (in $\mu g/m^3$) levels by sources. "American consulate" refers to readings from the air quality monitor at the American consulate in Lahore. We treat this reading as the ground truth. "PAQI" refers to readings from the average of lower-cost 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.

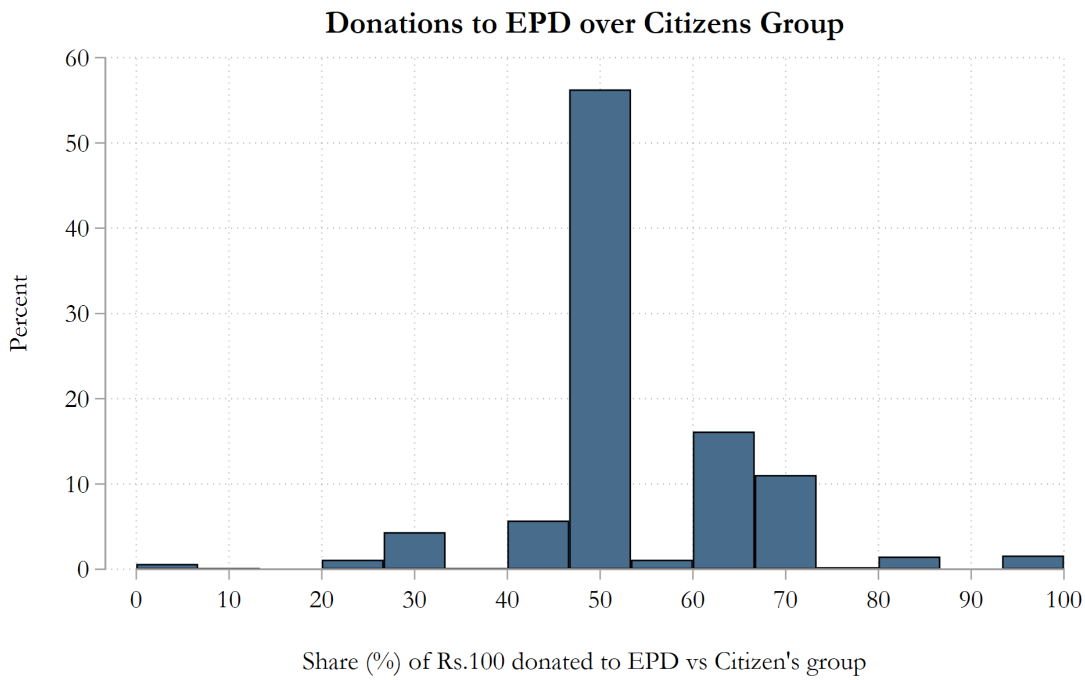


Figure 2: This figure shows the result from the donation game in our baseline survey, in which we asked respondents to split PKR 100 between government (EPD) and private (PAQI) sources.

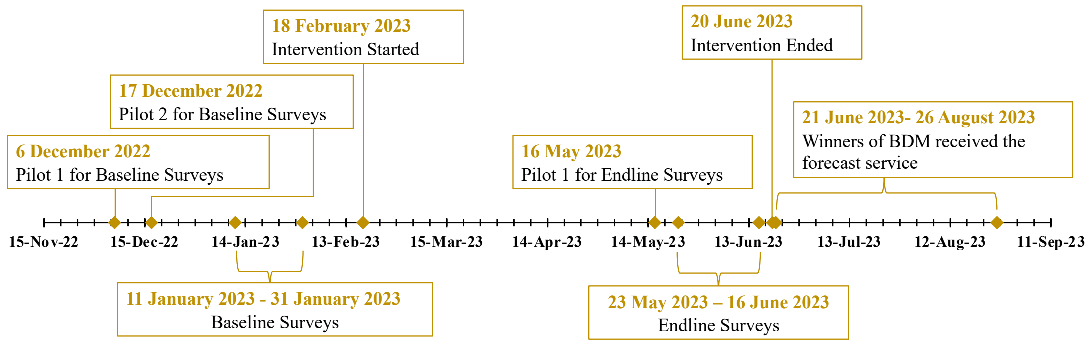


Figure 3: Timeline of intervention and surveys

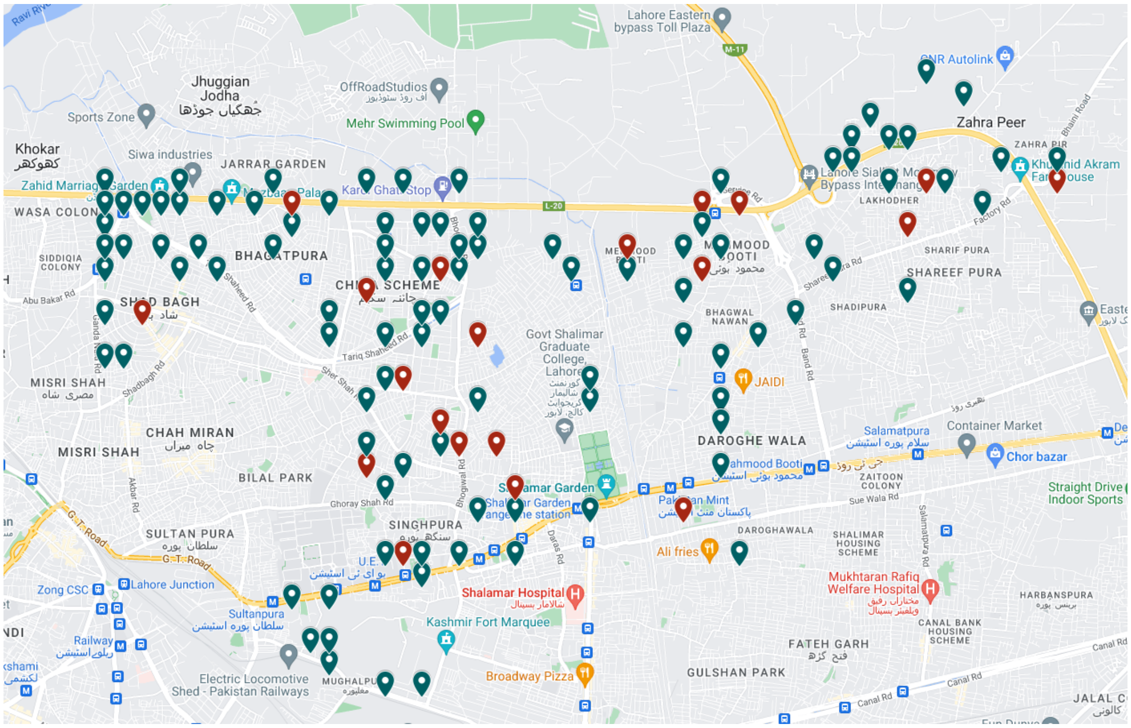


Figure 4: Sampling coordinates in NA-123 and NA-124 constituencies in Lahore, Pakistan

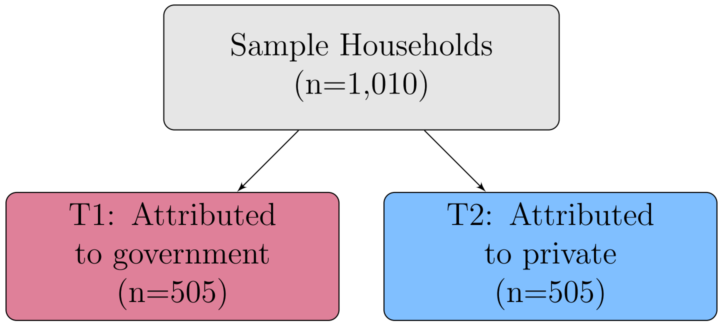
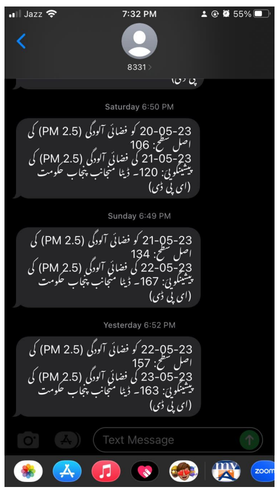


Figure 5: Treatment Groups

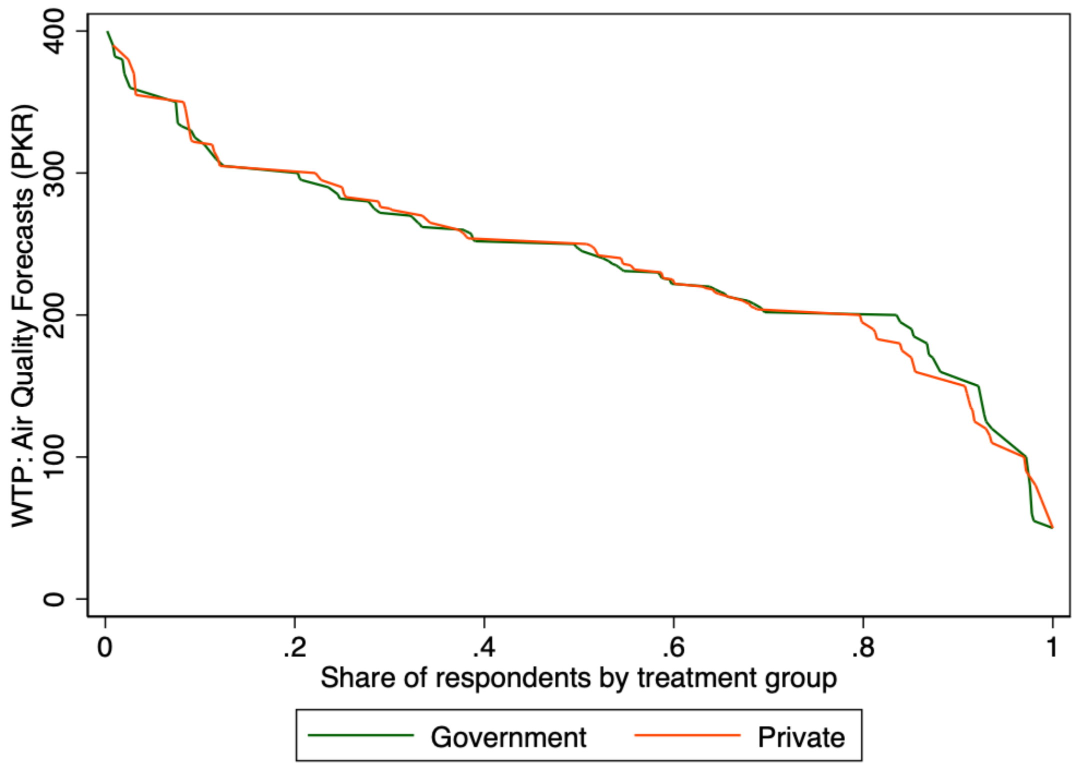


(a) T1: Daily messages

(b) T2: Daily messages

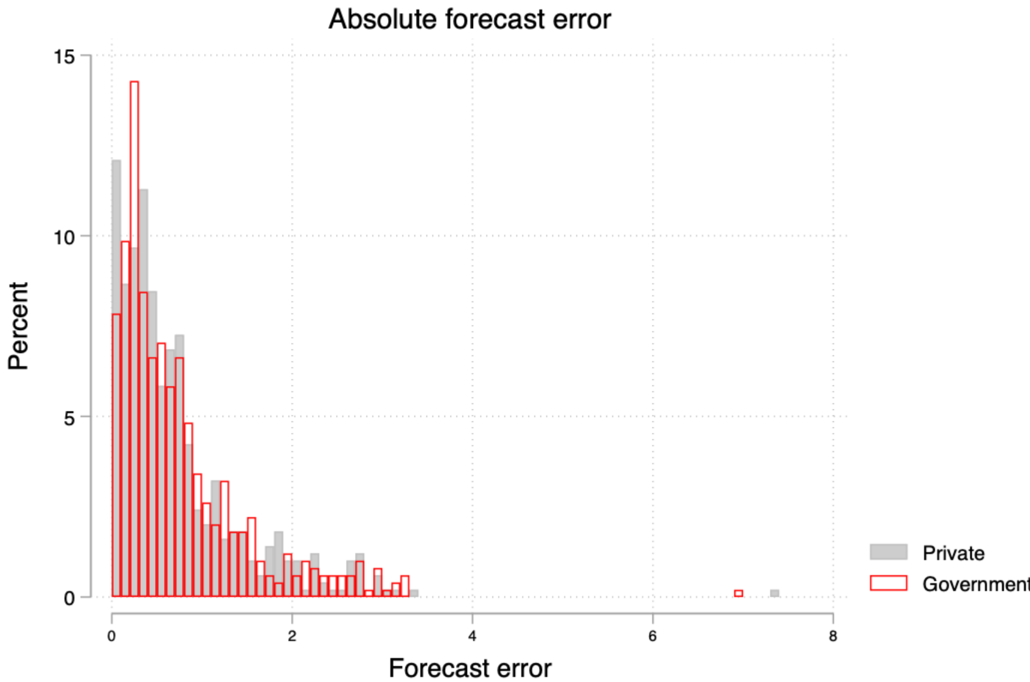
Figure 6: Sample messages to respondents

Figure 7: Demand curves for air pollution forecast by treatment



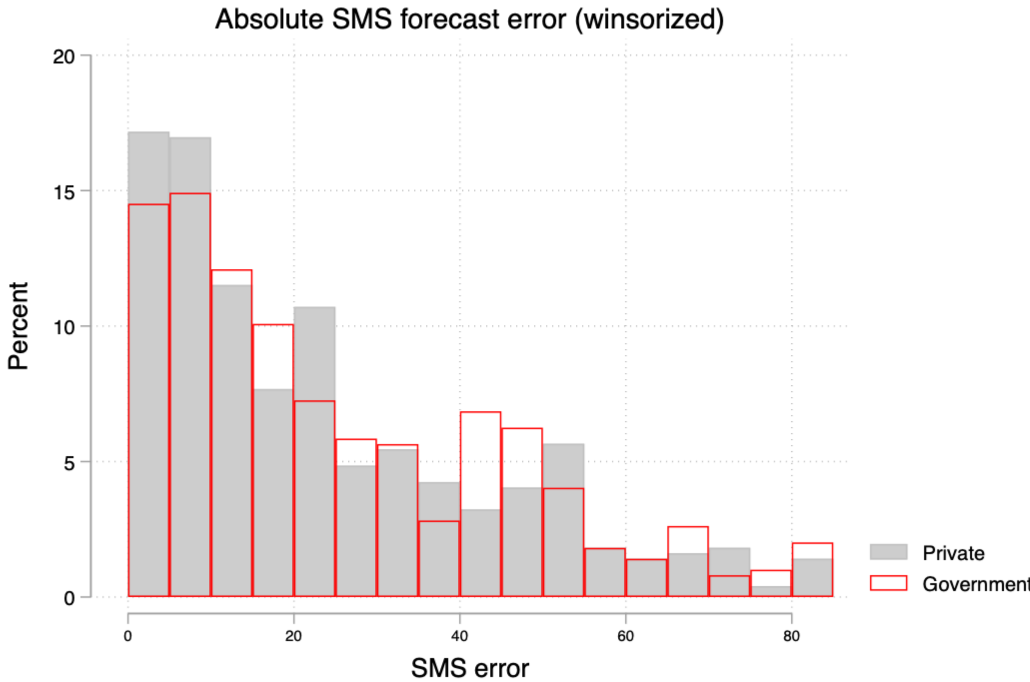
This figure shows the distributions of respondents' bids for two months of air pollution forecast service from the endline survey. "Government" corresponds to the arm in which the EPD source is made salient, and "Private" the PAQI source.

Figure 8: Absolute forecast error by treatment



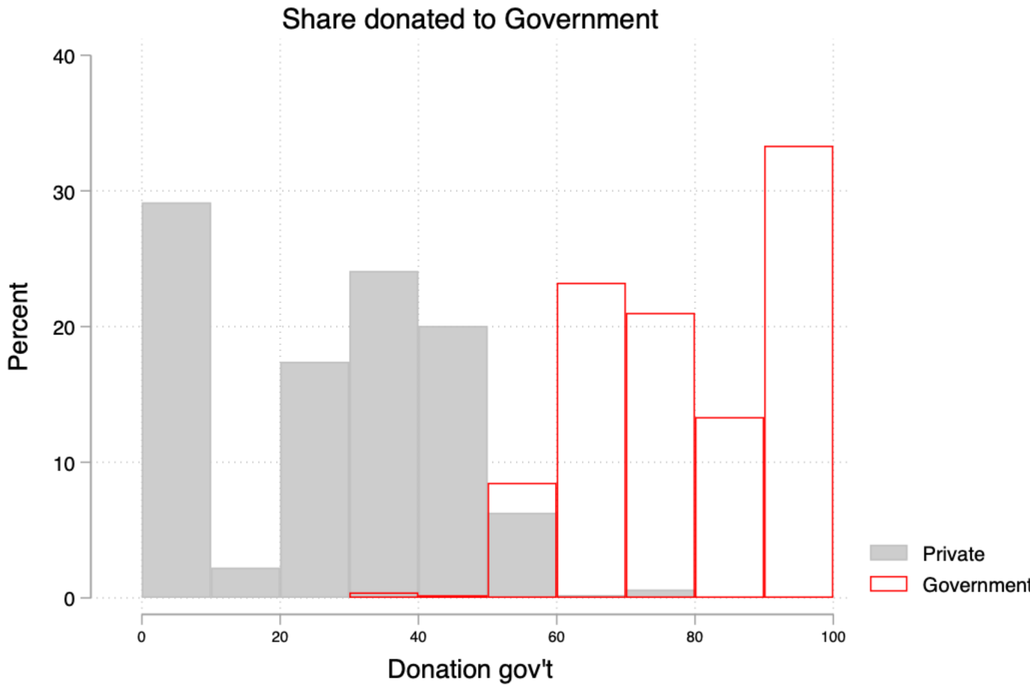
This figure shows the distributions of respondents' absolute forecast error in the endline survey. The measure is defined as the absolute difference between their forecast air pollution level on the next day and the actual reading, divided by the actual reading. "Government" corresponds to the arm in which the EPD source is made salient, and "Private" the PAQI source.

Figure 9: Estimate of SMS forecast error by treatment



This figure shows the distributions of respondents' beliefs about the absolute error of SMS forecasts, measured at the endline survey. The measure is defined as the absolute difference between their forecast air pollution level and their guess of the SMS forecast on the next day. "Government" corresponds to the arm in which the EPD source is made salient, and "Private" the PAQI source.

Figure 10: Donation to government sources vs private



This figure shows the distributions of respondents' donations to a government agency vs. a non-government entity for environmental protection, measured at the endline survey. The measure is defined as the amount out of PKR 100 donated to the government source. "Government" corresponds to the arm in which the EPD source is made salient, and "Private" the PAQI source.

9 Tables

Table 1: Balance table of outcomes and key demographic variables at baseline

Variable	(1) Private Mean/(SE)	(2) Government Mean/(SE)	(1)-(2) Pairwise t-test Mean difference
Baseline forecast error	0.725 (0.019)	0.714 (0.019)	0.011
Baseline donation gov't	50.139 (0.682)	50.119 (0.654)	0.020
Baseline: hours spent outside	7.403 (0.204)	7.440 (0.198)	-0.037
Stated preference for citizens group	0.013 (0.042)	-0.011 (0.043)	0.024
Stated preference for government	-0.009 (0.043)	-0.010 (0.043)	0.001
Comprehended the text message without explanation	0.768 (0.019)	0.766 (0.019)	0.002
Received air pollution info from: EPD	0.087 (0.013)	0.083 (0.012)	0.004
Received air pollution info from: AirVisual App	0.097 (0.013)	0.089 (0.013)	0.008
Index: Sentiment on air quality	-0.019 (0.032)	0.010 (0.032)	-0.029
Asset index	0.020 (0.046)	-0.026 (0.043)	0.046
F-test of joint significance (F-stat)			0.210
Number of observations	504	504	1008

Notes: This table presents sample means and standard deviations by treatment arms, mean differences and their t-tests, and the two-tailed significance. All measures come from the baseline survey. "Baseline forecast error": baseline measure of the pre-specified forecast-error outcome. "Baseline donation gov't": baseline measure of the preference for the government source vs the citizen's group. "Baseline: hours spent outside": time spent outdoors, as calculated from a time-use log. "Stated preference for citizens group": indexed measure of respondents' stated beliefs that a) air quality readings from the citizens' initiative are accurate, and that b) they approve of the job that the citizens' igroup s doing to address air quality. "Stated prefernce for government": indexed measure of respondents' stated beliefs that a) air quality readings from the government are accurate, and that b) they approve of the job that the government is doing to address air quality. "Comprehended the text message without explanation": When the respondent was shown a mock-up of a text message they will receive, they understood it without further explanation. "Received air pollution info from: EPD": self-reported to have accessed air quality readings from EPD."Received air pollution info from: Air Visual App": self-reported to have accessed air quality readings from the AirVisual App, on which PAQI disseminates air quality information. "Index: Sentiment on air quality": indexed measure that a) respondents care about air quality in places they live, b) they have been concerned about air quality in general in the last week, c) their quality of life is significantly affected at home, their performance at work or school is significantly affected, d) their sleep is affected, they reduced the number of hours worked, and e) the number of days in the last week with unsatisfactory air quality. "Asset index": indexed measure of the household's ownership of electronic appliances. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table 2: Summary statistics: PM2.5 readings by source

	(1)		
	mean	sd	count
Pre-intervention			
AirNow	209.0	114.9	107
EPD	171.4	86.7	80
PAQI	161.7	64.3	116
Urban	210.5	108.3	107
Sprintars	69.2	19.5	116
During/post-intervention			
AirNow	63.5	39.4	175
EPD	55.7	36.0	173
PAQI	58.7	31.0	190
Urban	92.4	70.9	134
Sprintars	59.1	14.5	186
Total			
AirNow	118.7	104.6	282
EPD	92.3	78.4	253
PAQI	97.7	68.3	306
Urban	144.9	106.9	241
Sprintars	63.0	17.3	302

Notes: "Pre-intervention": time period prior to our intervention (Feb 18), i.e., the period with high levels of PM2.5 concentrations. "During/post-intervention": Period since February 18, when there are relatively low PM2.5 concentrations. "Total": readings from November 1, 2022 to August 26, 2023. "AirNow": U.S. Consulate readings. "Urban": The Urban Unit (Provincial Government of Punjab). "PAQI": Pakistan Air Quality Initiative. "EPD": Environment Protection Department (Provincial Government of Punjab). "Sprintars": Satellite-based measure.

Table 3: Correlations between readings

	(1) AirNow	EPD	PAQI	Urban	Sprintars
AirNow	1				
EPD	0.61***	1			
PAQI	0.82***	0.70***	1		
Urban	0.76***	0.58***	0.73***	1	
Sprintars	0.20***	0.28***	0.27***	0.23***	1

Notes: Pairwise correlation measures of air quality readings by source. "AirNow": U.S. Consulate readings. "Urban": The Urban Unit (Provincial Government of Punjab). "PAQI": Pakistan Air Quality Initiative. "EPD": Environment Protection Department (Provincial Government of Punjab). "Sprintars": Satellite-based measure. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table 4: Deviations of monitor readings from the American Consulate readings

	RMSE:All	RMSE:pre	RMSE:during/post	MAD:All	MAD:pre	MAD:during/post
EPD	75.6	121.4	41.2	46.4	87.3	28.1
PAQI	63.7	100.8	23.8	34.6	69.5	14.6
Urban	67.6	62.1	71.5	38.7	41.6	36.5
Sprintars	113.4	177.4	44.2	73.2	143.6	32.1

Notes: Deviation from the American Consulate readings by source. RMSE: Root mean squared error. MAD: mean absolute difference. "All": readings from November 1, 2022 to August 26, 2023. "pre": time period prior to our intervention (Feb 18), i.e., the period with high levels of PM2.5 concentrations. "during/post": Period since February 18, when there are relatively low PM2.5 concentrations. "AirNow": U.S. Consulate readings. "Urban": The Urban Unit (Provincial Government of Punjab). "PAQI": Pakistan Air Quality Initiative. "EPD": Environment Protection Department (Provincial Government of Punjab). "Sprintars": Satellite-based measure.

Table 5: Prespecified hypotheses: ITT

	(1) WTP	(2) WTP	(3) Forecast error	(4) SMS error	(5) Donation gov't
Constant	237.5***				
	(2.19)				
Gov't arm		0.33	0.051	2.82**	53.8***
		(3.68)	(0.040)	(1.29)	(1.04)
Observations	993	993	993	991	989
Endline mean of PVT		237.2	0.73	22.7	22.9

Notes: Model: PDSLASSO. "WTP": Willingness to pay for two months of SMS air quality forecasts. "Forecast error": the absolute difference between their forecast air pollution level on the next day and the actual reading, divided by the actual reading. "SMS error": the absolute difference between their forecast air pollution level and their guess of the SMS forecast on the next day. "Donation gov't": amount out of PKR 100 donated to the government source. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table 6: Adjustments for multiple hypothesis testing on prespecified hypotheses

	(1) WTP	(2) WTP	(3) Forecast error	(4) SMS error	(5) Donation gov't
P value	0	.927	.208	.029	0
Q value	.001	.351	.116	.03	.001

Notes: We show the critical values for the "Constant" and "Gov't arm" coefficients in the corresponding columns of Table 5. "P value:" Unadjusted p-values. "Q value": Benjamini Krieger Yekutieli (2006) sharpened q-values.

Table 7: ITT: Stated preference measure on satisfaction with the SMS service

	(1) SMS: Satisfaciton index	(2) SMS: Satisfied	(3) SMS: Reliable	(4) SMS: Accurate
Gov't arm	0.056	0.055	0.052	0.020
	(0.040)	(0.038)	(0.038)	(0.040)
Observations	990	990	990	988
Endline mean of PVT	-0.053	4.07	4.05	3.85

Notes: We present estimates of effects on the stated-preference measures on the respondents' satisfaction with the SMS service. We ask if they are overall satisfied with the service (Column 2), if they think the service is reliable and on time (Column 3), and if they believe the forecasts are accurate (Column 4), in the Likert scale where positive values indicate approval. Column 1 shows the unweighted and standardized sum of the three measures in Columns 2 and 4. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table 8: ITT: Alternative definitions of the WTP outcome

	(1) WTP	(2) WTP (other)	(3) diff(WTP)	(4) diff(WTP)
Gov't arm	0.33	0.55	-0.21	
	(3.68)	(3.66)	(0.54)	
Constant				15.9***
				(0.60)
Observations	993	993	993	993
Endline mean of PVT	237.2	221.2	16.0	

Notes: "WTP": The prespecified outcome measuring the willingness to pay for two months of SMS air quality forecasts, where the assigned source is made salient. "WTP if other source": hypothetical WTP if the forecast were to come from the other source not assigned to them. "diff(WTP sources)": the difference between the willingness to pay for the assigned vs. the other sources. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table 9: Heterogeneous effects: Baseline donation to government

	(1) WTP	(2) WTP	(3) Forecast error	(4) SMS error	(5) Donation gov't
Baseline donation gov't	0.43**	0.83	-0.011**	-0.048	0.50***
	(0.17)	(0.66)	(0.0053)	(0.11)	(0.13)
Gov't arm		4.39	-0.052	-0.25	95.2***
		(16.6)	(0.22)	(2.81)	(3.02)
Gov't arm × Baseline donation gov't		-0.074	0.0020	0.052	-0.83***
		(0.30)	(0.0040)	(0.056)	(0.056)
Observations	993	993	993	991	989
Endline mean of PVT		237.2	0.73	22.7	22.9

Notes: Heterogeneous treatment effects by donation to government at baseline. "WTP": Willingness to pay for two months of SMS air quality forecasts. "Forecast error": the absolute difference between their forecast air pollution level on the next day and the actual reading, divided by the actual reading. "SMS error": the absolute difference between their forecast air pollution level and their guess of the SMS forecast on the next day. "Donation gov't": amount out of PKR 100 donated to the government source. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table 10: Heterogeneous effects: Baseline overall approval of government v.s. citizens' group's source

	(1) WTP	(2) WTP	(3) Forecast error	(4) SMS error	(5) Donation gov't
Relative stated pref for govt: Approval	8.50***	0	0	0	0
	(2.20)	(.)	(.)	(.)	(.)
Gov't arm		0.32	0.050	2.73**	53.8***
		(3.68)	(0.040)	(1.29)	(0.94)
Gov't arm × Relative stated pref for govt: Approval		1.89	-0.020	0.35	-14.5***
		(4.17)	(0.062)	(1.05)	(0.79)
Observations	990	990	990	988	986
Endline mean of PVT		237.0	0.73	22.7	23.0

Notes: Heterogeneous treatment effects by a relative measure of overall approval for the government source to the citizens' group's. The measure "Relative stated prf for govt: Approval" is a standardized difference of Likert-scale questions in which the respondents evaluated their overall approval of the government's and citizen's group's job in addressing air quality in Lahore. "WTP": Willingness to pay for two months of SMS air quality forecasts. "Forecast error": the absolute difference between their forecast air pollution level on the next day and the actual reading, divided by the actual reading. "SMS error": the absolute difference between their forecast air pollution level and their guess of the SMS forecast on the next day. "Donation gov't": amount out of PKR 100 donated to the government source. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table 11: Heterogeneous effects: Baseline belief on information accuracy

	(1) WTP	(2) WTP	(3) Forecast error	(4) SMS error	(5) Donation gov't
Relative stated pref for govt: Accuracy	8.03***	0	0	0	0
	(2.17)	(.)	(.)	(.)	(.)
Gov't arm		-0.40	0.069*	3.14**	51.3***
		(3.60)	(0.039)	(1.51)	(0.96)
Gov't arm × Relative stated pref for govt: Accuracy		-1.04	-0.070*	-0.55	-13.7***
		(3.76)	(0.039)	(1.29)	(0.95)
Observations	948	948	948	947	945
Endline mean of PVT		236.4	0.71	23.4	23.8

Notes: Heterogeneous treatment effects by a relative measure of beliefs about the accuracy of the government source's and the citizens' group's air quality readings. The measure "Relative stated prf for govt: Accuracy" is a standardized difference of Likert-scale questions in which the respondents evaluated their how accurate the government's and citizen's group's air quality readings are. "WTP": Willingness to pay for two months of SMS air quality forecasts. "Forecast error": the absolute difference between their forecast air pollution level on the next day and the actual reading, divided by the actual reading. "SMS error": the absolute difference between their forecast air pollution level and their guess of the SMS forecast on the next day. "Donation gov't": amount out of PKR 100 donated to the government source. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table 12: Heterogeneous effects: Baseline forecast error

	(1) WTP	(2) WTP	(3) Forecast error	(4) SMS error	(5) Donation gov't
Baseline forecast error	-1.90	32.0	2.14***	7.49	31.4***
	(5.15)	(26.7)	(0.30)	(9.11)	(7.32)
Gov't arm		-6.57	-0.14*	-1.69	63.8***
		(8.12)	(0.084)	(2.38)	(2.10)
Gov't arm × Baseline forecast error		10.6	0.26**	6.66*	-13.8***
		(8.94)	(0.11)	(3.77)	(2.39)
Observations	993	993	993	991	989
Endline mean of PVT		237.2	0.73	22.7	22.9

Notes: Heterogeneous treatment effects by the baseline forecast error. "Baseline forecast error" the baseline outcome measure of respondents' forecast error. "WTP": Willingness to pay for two months of SMS air quality forecasts. "Forecast error": the absolute difference between their forecast air pollution level on the next day and the actual reading, divided by the actual reading. "SMS error": the absolute difference between their forecast air pollution level and their guess of the SMS forecast on the next day. "Donation gov't": amount out of PKR 100 donated to the government source. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

A Appendix tables

Table A.1: Accuracy of individual respondents' and SMS' forecasts

Variable	N	(1) Private Mean/(SE)	N	(2) Government Mean/(SE)	N	(1)-(2) Pairwise t-test Mean difference
Baseline: abs(Ind. forecast - truth)/truth	504	0.725	504	0.714	1008	0.011
		(0.019)		(0.019)		
Endline: abs(Ind. forecast - truth)/truth	496	0.720	497	0.762	993	-0.042
		(0.031)		(0.032)		
Endline: abs(SMS forecast - truth)/truth	496	0.694	497	0.710	993	-0.016
		(0.030)		(0.029)		
Endline: abs(Ind. forecast - truth)/abs(SMS forecast - truth)	496	2.043	497	1.987	993	0.056
		(0.101)		(0.096)		

Notes: Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table A.2: Prespecified hypotheses: ITT (winsorized)

	(1) WTP	(2) WTP	(3) Forecast error	(4) SMS error	(5) Donation gov't
Constant	237.4***				
	(2.18)				
Gov't arm		0.31	0.051	1.89*	53.8***
		(3.67)	(0.037)	(1.01)	(1.04)
Observations	993	993	993	991	989
Endline mean of PVT		237.2	0.73	22.7	22.9

Notes: We winsorize the outcome variables at the 1st and 99th percentiles. "WTP": Willingness to pay for two months of SMS air quality forecasts. "Forecast error": the absolute difference between their forecast air pollution level on the next day and the actual reading, divided by the actual reading. "SMS error": the absolute difference between their forecast air pollution level and their guess of the SMS forecast on the next day. "Donation gov't": amount out of PKR 100 donated to the government source. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table A.3: Adjustments for multiple hypothesis testing on prespecified hypotheses (winsorized)

	(1) WTP	(2) WTP	(3) Forecast error	(4) SMS error	(5) Donation gov't
P value	0	.932	.177	.061	0
Q value	.001	.284	.113	.065	.001

Notes: We show the critical values for the "Constant" and "Gov't arm" coefficients in the corresponding columns of Table A.2. "P value:" Unadjusted p-values. "Q value": Benjamini Krieger Yekutieli (2006) sharpened q-values.

Table A.4: ITT: Alternative definitions of the forecast outcome

	(1) abs(own - truth)/truth	(2) (own - truth)/truth	(3) abs(own - truth)	(4) (own - truth)
Gov't arm	0.051	0.060	-0.48	3.39
	(0.040)	(0.049)	(2.17)	(2.85)
Observations	993	993	993	993
Endline mean of PVT	0.73	0.47	40.0	91.9

Notes: We present estimates of effects on forecast outcomes with different definitions, where "own" stands for the respondent's own forecast of the air quality level the next day, and "truth" the actual readings on the corresponding day. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table A.5: ITT: Concerns about air quality

	(1) Care about AQ	(2) N. days good air
Gov't arm	0.0088	0.025
	(0.046)	(0.054)
Observations	992	961
Endline mean of PVT	2.59	3.16

Notes: We present estimates of effects on measures of concern about air quality. "Care about AQ": a Likert-scale measure of how much the respondent cares about air quality in the places they live and work. "N. Days good air": Number of days in the last week with acceptable air quality. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table A.6: Heterogeneous effects: Baseline donation to government

	(1) WTP	(2) WTP	(3) Forecast error	(4) SMS error	(5) Donation gov't
More to Pvt	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
50-50	3.23	-1.85	-0.0078	0.71	10.5***
	(7.57)	(13.3)	(0.11)	(2.89)	(2.94)
More to Govt	13.8*	-2.18	-0.13	-0.42	25.1***
	(7.54)	(16.5)	(0.14)	(3.88)	(3.52)
Gov't arm		-10.2	0.015	1.13	72.4***
		(12.6)	(0.13)	(1.95)	(2.52)
More to Pvt × Gov't arm		0	0	0	0
		(.)	(.)	(.)	(.)
50-50 × Gov't arm		12.9	0.042	1.59	-15.3***
		(13.6)	(0.15)	(2.71)	(3.03)
More to Govt × Gov't arm		12.1	0.0027	1.39	-37.6***
		(13.9)	(0.15)	(3.01)	(2.87)
Observations	993	993	993	991	989
Endline mean of PVT		237.2	0.73	22.7	22.9

Notes: Heterogeneous treatment effects by donation to government at baseline. "WTP": Willingness to pay for two months of SMS air quality forecasts. "Forecast error": the absolute difference between their forecast air pollution level on the next day and the actual reading, divided by the actual reading. "SMS error": the absolute difference between their forecast air pollution level and their guess of the SMS forecast on the next day. "Donation gov't": amount out of PKR 100 donated to the government source. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table A.7: Heterogeneous effects: Baseline overall approval of information sources

	(1) WTP	(2) WTP	(3) Forecast error	(4) SMS error	(5) Donation gov't
Approval: Neutral	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
Approval: Govt	23.2***	2.72	0.056	6.04*	21.7***
	(4.60)	(10.4)	(0.100)	(3.12)	(2.56)
Approval: Citizen	11.8	22.7*	0.049	-4.09	5.78*
	(7.68)	(13.0)	(0.13)	(3.32)	(3.33)
Gov't arm		1.76	0.16**	3.91*	73.1***
		(7.17)	(0.070)	(2.36)	(1.63)
Approval: Neutral × Gov't arm		0	0	0	0
		(.)	(.)	(.)	(.)
Approval: Govt × Gov't arm		1.02	-0.16*	-1.94	-39.6***
		(8.54)	(0.091)	(3.08)	(2.00)
Approval: Citizen × Gov't arm		-13.0	-0.25	-3.67	-13.7***
		(14.0)	(0.16)	(3.44)	(3.28)
Observations	993	993	993	991	989
Endline mean of PVT		237.2	0.73	22.7	22.9

Notes: Heterogeneous treatment effects by a relative measure of overall approval for the government source to the citizens' group's. The measure is based on Likert-scale questions in which the respondents evaluated their overall approval of the government's and citizen's group's job in addressing air quality in Lahore. "Approval: Neutral": approves of government as much as the citizens' group. "Approval: Govt": approves of the government more than the citizens' group. "Approval: Citizen": approves of the citizens' group more than the government. "WTP": Willingness to pay for two months of SMS air quality forecasts. "Forecast error": the absolute difference between their forecast air pollution level on the next day and the actual reading, divided by the actual reading. "SMS error": the absolute difference between their forecast air pollution level and their guess of the SMS forecast on the next day. "Donation gov't": amount out of PKR 100 donated to the government source. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table A.8: Heterogeneous effects: Baseline belief on information accuracy

	(1) WTP	(2) WTP	(3) Forecast error	(4) SMS error	(5) Donation gov't
Accuracy: Neutral	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
Accuracy: Govt	24.0***	8.88	-0.29**	2.14	23.1***
	(5.08)	(10.8)	(0.13)	(3.04)	(2.82)
Accuracy: Citizen	17.5**	30.2**	0.059	6.26**	11.9***
	(7.78)	(12.5)	(0.15)	(3.16)	(3.41)
Gov't arm		12.0	0.10	5.98**	77.4***
		(7.76)	(0.078)	(2.77)	(1.70)
Accuracy: Neutral × Gov't arm		0	0	0	0
		(.)	(.)	(.)	(.)
Accuracy: Govt × Gov't arm		-17.1*	-0.091	-4.95	-40.7***
		(9.09)	(0.098)	(3.39)	(2.06)
Accuracy: Citizen × Gov't arm		-13.5	-0.049	-7.50*	-21.0***
		(14.0)	(0.14)	(3.86)	(3.19)
Observations	993	993	993	991	989
Endline mean of PVT		237.2	0.73	22.7	22.9

Notes: Heterogeneous treatment effects by a relative measure of beliefs about the accuracy of the government source's and the citizens' group's air quality readings. The measure is based on Likert-scale questions in which the respondents evaluated their how accurate the government's and citizen's group's air quality readings are. "Accuracy: Neutral": believes government is as accurate as the citizens' group. "Accuracy: Govt": believes that the government is more accurate than the citizens' group. "Accuracy: Citizen": believes that the citizens' group is more accurate than the government. "WTP": Willingness to pay for two months of SMS air quality forecasts. "Forecast error": the absolute difference between their forecast air pollution level on the next day and the actual reading, divided by the actual reading. "SMS error": the absolute difference between their forecast air pollution level and their guess of the SMS forecast on the next day. "Donation gov't": amount out of PKR 100 donated to the government source. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table A.9: Heterogeneous effects: Baseline forecast error

	(1) WTP	(2) WTP	(3) Forecast error	(4) SMS error	(5) Donation gov't
Baseline error below median	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
Baseline error at or above median	-1.12	11.4	-0.18*	-9.29***	7.21***
	(4.39)	(9.36)	(0.096)	(2.41)	(2.51)
Gov't arm		-1.29	-0.062	0.11	60.0***
		(6.09)	(0.061)	(1.44)	(1.56)
Baseline error below median × Gov't arm		0	0	0	0
		(.)	(.)	(.)	(.)
Baseline error at or above median × Gov't arm		2.20	0.22**	5.58*	-12.2***
		(7.93)	(0.088)	(2.88)	(2.19)
Observations	993	993	993	991	989
Endline mean of PVT		237.2	0.73	22.7	22.9

Notes: Heterogeneous treatment effects by baseline forecast error. "Baseline error below median": their baseline error is lower than the median. "Baseline error at or above median": their baseline error is at or higher than the median. "WTP": Willingness to pay for two months of SMS air quality forecasts. "Forecast error": the absolute difference between their forecast air pollution level on the next day and the actual reading, divided by the actual reading. "SMS error": the absolute difference between their forecast air pollution level and their guess of the SMS forecast on the next day. "Donation gov't": amount out of PKR 100 donated to the government source. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: $p<0.1$*; $p<0.05$**; $p<0.01$***.

B Data

B.1 Survey data

B.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)

B.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 of outcomes 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 in 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 the 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 with 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

B.2 Air quality data

We collect air quality reading 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.

B.3 Weather Data

We also collect weather data as inputs for the forecast model, as described in further detail in Section 3.2.

- **AccuWeather:** We scrape daily forecasts on maximum and minimum temperatures and precipitation probability from AccuWeather 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 constructs 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`).

C Power Calculations

We estimate the minimum detectable effect sizes on our primary outcomes at 80% probability, with $\alpha = 0.05$. We assume 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 is 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 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.

D Alternative identification strategies

D.1 Treatment on the treated

We define takeup 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?" 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 depend on the 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 or $\boldsymbol{A}$ instrumenting for P. We present the following first and second-stage specifications for a within-subject model with Z as an instrument.

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

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

$\widehat{P}$ 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}.

E Secondary outcomes

We present other variables that are of interest but for which we do not correct for multiple testing.

E.1 Avoidance behaviors

- Outdoor time use
  - The outcome is defined as the number of hours spent outside. We ask respondents the type of activity (sleep, paid work, homemaking, 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.
- 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 an N90/95 mask. We also collect information on what other types of masks (e.g., surgical masks, cloth) the respondents show.

Table E.10: Secondary outcomes: Time use

	(1) Endline: hours spent outside	(2) Endline: Hrs (stated)	(3) Endline: Hrs (if bad day)
Gov't arm	-0.036	0.010	0.0063
	(0.15)	(0.12)	(0.11)
Observations	993	993	993
Endline mean of PVT	5.14	3.89	3.65

Notes: Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table E.11: Secondary outcomes: Mask use

	(1) Has mask	(2) Shows mask	(3) Shows mask
Gov't arm	-0.035*	-0.011	-0.011
	(0.021)	(0.015)	(0.015)
Observations	993	993	993
Endline mean of PVT	0.20	0.099	0.099

Notes: Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

E.2 Policy preferences and collective action for air quality

- 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).
- 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.
- 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.

Table E.12: Secondary outcomes: Preference for air quality policies over other domains

	(1) AQ over Educ	(2) AQ over health	(3) AQ over waste
Gov't arm	0.0038	0.0046	0.021
	(0.012)	(0.014)	(0.020)
Observations	992	992	993
Endline mean of PVT	0.050	0.077	0.17

Notes: Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table E.13: Secondary outcomes: Demand for filing complaints about air quality

	(1) Takes info	(2) Plans to complain
Gov't arm	-0.016	-0.0024
	(0.018)	(0.017)
Observations	993	993
Endline mean of PVT	0.85	0.12

Notes: Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.