Beliefs, Signal Quality, and Information Sources: Experimental Evidence on Air Quality in Pakistan*

Isra Imtiaz   Shotaro Nakamura   Sanval Nasim   Arman Rezaee

August 7, 2025

Abstract

We study how the perceived source of environmental information—government or non-government—affects consumers' beliefs and demand for air quality forecasts in developing economies. In a randomized experiment in Lahore, Pakistan, we provide identical day-ahead SMS forecasts, varying only the attributed source. Subjects exhibit high willingness-to-pay regardless of source but perceive government forecasts as 12% less accurate. Nonetheless, they ultimately prefer the source assigned to them. Our findings suggest that source exposure—rather than content alone—shapes consumers' beliefs and preferences, with implications for welfare-enhancing access to environmental information in low-capacity settings.
JEL: Q53, Q56, D83, C93
Keywords: air quality, beliefs, environmental information, willingness to pay

*Imtiaz: University of California, Davis. Nakamura: Pennsylvania State University; 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. This study is funded by grants from the International Growth Center (IGC) and faculty funding from Colby College and UC Davis. This study has IRB approval in the US (UC Davis IRB-1336133-1) and Pakistan (LUMS IRB 02242021). This study was pre-registered with AEA (AEARCTR-0011489).

1 Introduction

The economic case for improving information is well-established—for example, environmental quality is a public good, and economic theory predicts that markets will under-supply environmental information (Samuelson 1954). What is not well understood, however, is how citizens—particularly in developing economies—respond to competing sources of information as they emerge. Do developing-country citizens trust and act upon information as it becomes available? Does the perceived source—government versus non-government—shape citizens' beliefs and preferences for this critical public good? Focusing on air quality, we address these questions using a randomized controlled trial in Lahore, Pakistan—a developing city, where severe air pollution, fragmented information, and active competition between government and non-government sources offer an ideal setting to study how developing-country citizens form beliefs and preferences over environmental information.

In developing countries, limited access to air quality information can substantially reduce welfare given significant health and productivity costs of severe ambient air pollution—exacerbated by uncertainty and seasonal variation (WHO 2021; Barwick et al. 2024). According to Greenstone et al. (2024), the average reduction in life expectancy in South Asia due to high concentrations of particulate matter (PM2.5) is 3.5 years, as opposed to just months in the U.S. and Europe.[1] In Lahore—the setting of our study—annual PM2.5 concentrations averaged $124.8\mu g/m^3$ in 2022—16 times the U.S. average ($7.8\mu g/m^3$). Air quality in Lahore also exhibits pronounced seasonal fluctuations based on factors such as crop residue burning (Lan et al. 2022).[2]

Since markets often fail in adequately providing public goods, governments typically assume responsibility in supplying them. However, delivering effective information in developing countries is challenging given weak government capacity, incentives to under-report the extent of environmental degradation, and persistent barriers to access—particularly for low-income households (Ghanem and Zhang 2014; Mu et al. 2024). Consequently, this leads to substantial unmet demand for accurate air quality information and effective pollution mitigation (Ahmad et al. 2022; Freeman et al. 2019; Ito and Zhang 2020). In response, various non-government stakeholders—including citizen-led advocacy groups and international agencies—have begun providing air quality information, potentially increasing awareness and, in some cases, improving local air quality (Jha and La Nauze 2022).

[1]The measure is defined as the average life years gained by a reduction in the concentration of PM2.5 to the 2021 WHO standards ($5\mu g/m^3$).

[2]Average concentration measures are authors' own calculations based on U.S. EPA sources: https://www.airnow.gov/international/us-embassies-and-consulates https://www.epa.gov/air-trends/particulate-matter-pm25-trends.

Despite this growth in information sources, how citizens respond to multiple competing information streams and how their choices ultimately shape the accessibility and quality of public goods remains unclear (Bergstrom et al. 1986; Coury et al. 2024). A robust body of research demonstrates that public-private competition can improve service quality in developing economies (Andrabi et al. 2017; Muralidharan and Sundararaman 2015). Evidence also indicates that credible information about public services can build trust in the government by shifting citizens' beliefs about state capacity (Acemoglu et al. 2020; Khan et al. 2021; Dhinakar Bala et al. 2024). Yet, the mechanisms and attributes governing how citizens form beliefs about the reliability and quality of competing providers requires further investigation. Do consumers view different providers as close substitutes? To what extent does consumer demand reflect perceived service quality rather than innate preferences for particular providers?

In this paper, we study how the identity of an information provider shapes beliefs about—and preferences for—public environmental services in a developing-country setting characterized by competing sources. Our primary innovation is to experimentally isolate the effect of provider identity from that of service quality. Specifically, we randomize the provider attributed as the source of identical air quality forecasts delivered daily to households in Lahore, Pakistan, while holding the actual information and quality constant.

We address several interrelated research questions. First, do urban residents in developing countries exhibit unmet demand for air quality information? Second, does attributing information to a particular sources affect citizens' demand? Third, what mechanisms explain observed differences in demand by source—does source attribution affect beliefs about accuracy, or do innate preferences for particular sources influence demand? Fourth, what are the welfare implications of competition among multiple information sources?

To answer these questions, we implement a randomized controlled trial (RCT) with 1,010 households in lower-middle-income neighborhoods in Lahore. Recognizing that official government air quality information is limited despite severe air pollution, our intervention provides free day-ahead air quality forecasts via mobile SMS messages. We generate these forecasts using an ensemble model incorporating multiple data source (government and non-government).[3] Our experimental design randomly varies only the attributed source of the

[3]We recognize that spatial heterogeneity in air pollution is important—especially in cities such as Lahore where exposure may vary across neighborhoods—but we focus on temporal variation owing to data limitations and the specific focus of our study. First, the lack of an extensive air quality monitoring system in Lahore precludes us from obtaining spatially disaggregated data. Second, we are primarily interested in how source attribution affects consumers' beliefs and preferences, rather than spatially targeted avoidance behavior. Temporal variation suffices for this objective since our intervention and outcomes focus on the salience, perceived quality, and demand for forecasts—not localized exposure.

forecasts—one arm attributes forecasts to the Punjab Environmental Protection Department (EPD, a government agency) while the other arm attributes identical forecasts to the Pakistan Air Quality Initiative (PAQI, a local NGO focused on providing air quality readings using low-cost monitors).[4]

Our intervention differs from existing information services in two key ways: 1) it offers day-ahead forecasts (versus real-time readings); and 2) it improves accessibility for average citizens, who report limited access to air quality information at baseline. We do not include a pure control arm (no information provision) since prior studies already document the benefits of providing air quality information in a similar context (Ahmad et al. 2022). Instead, our design captures the effect of source attribution relative to another provider.

The experimental design allows us to test whether subjects value air quality information and how they perceive and trust different sources of that information. We implement a series of incentivized exercises to: 1) elicit willingness-to-pay (WTP) for SMS-based forecasts; 2) measure beliefs about future air quality levels and the accuracy of the forecasts themselves; and 3) track changes in preferences over information providers. We develop a conceptual framework—in which the demand for environmental information depends not only on subject's beliefs about the state of air quality but also on their beliefs about the accuracy of the source.

We find that residents of working-class neighborhoods in Lahore exhibit high demand for air quality forecasts, regardless of whether the source is a government agency or an NGO. After receiving four months of forecasts for free, subjects' average WTP for continued service (forecasts for two additional months) is 238 Pakistani Rupees (PKR)—roughly equal to the cost of one month of basic mobile and data services—as measured through the Becker-DeGroot-Marshak (BDM) method. The high WTP indicates substantial latent demand for air quality information and satisfaction with our service. We validate these findings through stated preference measures at endline. However, we do not find statistically significant differences in WTP between treatment arms, suggesting that the identify of the attributed source does not influence WTP.

We hypothesize that forecast source may nonetheless influence downstream beliefs and preferences that determine subjects' utility from consuming forecasts. Specifically, we test

[4] We do not experimentally vary the accuracy of the forecasts across treatment arms. All subjects receive the same forecast—we only randomize the attributed source (government vs NGO). This design allows us to isolate the effect of source attribution on consumers' beliefs and preferences while avoiding ethical concerns that may arise from intentionally providing lower-quality information to some respondents. We generate the forecasts using an ensemble model that combines air quality predictions from multiple sources and assigns greater weights to more accurate ones. Thus, all respondents receive the best available forecast, regardless of the attributed source.

whether source attribution affects subjects': 1) beliefs about the state of air quality; 2) beliefs about the accuracy of the forecasts they receive; and 3) relative preference over providers—dimensions of service quality that shape behavior beyond forecast accuracy. We systematically explore evidence on these three channels.

First, we find that attributing forecasts to sources does not affect recipients' beliefs about next-day air quality levels. Using an incentivized prediction task, we elicit respondents' own forecasts of air quality for the next day and observe no significant differences in forecast error between treatment arms. Similarly, we find no evidence that forecast attribution affects avoidance behavior or policy preferences. Since we provided identical forecasts across treatment arms, these results suggest that source attribution alone does not shift expectations about air quality or behaviors related to pollution exposure. We also rule out the possibility that the forecasts lacked useful informational signals—using time-use data, we document correlational evidence that subjects reduce time spent outdoors on more polluted days, consistent with their up-take of and response to forecasts. Overall, subjects process and act on the information provided, but behavioral responses do not vary with the attributed source.

Second, while WTP for forecasts is statistically indistinguishable between arms, subjects perceive government-attributed forecasts as less accurate than those from the NGO. At endline, we ask subjects to predict not only next-day pollution levels but also the forecast they expect to receive from their assigned source—the absolute difference between these incentivized predictions provides a measure of perceived forecast accuracy. We find that subjects in the government arm perceive the same forecasts to be less accurate—reporting a 12% larger expected error than those in the NGO arm—despite receiving identical information. This result suggests that while subjects do not value forecast accuracy at the margin, they may prioritize other attributes—such as accessibility and reliability—when evaluating information services.

Third, we find that revealed and stated preferences for information providers shift substantially as a result of forecast exposure. Using a real-stakes donation task, we elicit relative preferences between the government and NGO. At baseline, most subjects equally divide their endowments between the two. However, by endline, subjects in the government arm allocate their donations to the government by a 75:25 split (the ratio flips for those in the NGO arm). Stated perceptions of accuracy, reliability, and overall approval shift in favor of the assigned source. These results indicate that preferences for sources are highly malleable in a frictional market for information services.

We also estimate heterogeneous treatment effects along pre-specified dimensions of recipients' baseline beliefs about air quality. While we do not find statistically significant heterogeneity in WTP based on these prior beliefs, we do observe differential effects on belief updating. Specifically, among subjects in the government arm (relative to those in the NGO arm), individuals with larger baseline forecast errors—i.e., those whose initial beliefs about air quality were less accurate—exhibit less improvement in endline forecast accuracy. A 100% increase in baseline forecast error is associated with a 26% higher endline error for subjects in the government arm (versus those in the NGO arm). This suggests that even when the information itself is identical across arms, the perceived credibility of the source influences how subjects internalize the forecasts. In particular, subject who who hold less accurate priors appear to update their beliefs more slowly when they receive forecasts from a source they perceive as lower quality.

Our work makes several contributions to the economics literature, addresses concerns over external validity, and provides policy-relevant insights. In particular, we speak to three broad themes: 1) public-private competition in public service provision; 2) belief formation under competing informational signals; and 3) the demand for environmental goods and services.

First, we provide new evidence on how consumers respond to public-private competition in the provision of public services. Our motivation builds on prior work showing how government agencies may face perverse incentives to under-report air pollution levels (Ghanem and Zhang 2014). More broadly, we contribute to the literature on how public-private competition affects the quality of public service delivery (Andrabi et al. 2017; Jha and La Nauze 2022; Muralidharan and Sundararaman 2015) and on trust in the state and perceptions of state capacity (Acemoglu et al. 2020; Khan et al. 2021; Dhinakar Bala et al. 2024). We show that consumers respond to information services regardless of the provider's identity, and that service attributes such as ease of access matter more than accuracy (at least, at the margin)—highlighting policy features that may enhance welfare in information-deficient settings.

Second, we contribute to the literature on belief formation and trust in information sources (Acemoglu et al. 2024; Gentzkow et al. 2023; Baysan 2022; Chopra et al. 2022; Burlig et al. 2024). We provide empirical evidence on how beliefs and trust shape demand for environmental information and on the conditions that shift beliefs about the state of the world and preferences for information services. We find that individuals strongly prefer the source that we randomly assigned to them, suggesting that beliefs are malleable to recent experience. However, this raises concerns about longer-run "lock-in" effects—particularly through platform design or belief polarization (Bowen et al. 2023; Shapiro and Varian 1999). Our findings underscore the importance of understanding how belief formation and trust in information evolve over time and across exposures.

Third, our paper adds to the growing literature on the determinants of demand for environmental goods and services in developing economies. Existing work across China and South Asia—as well as some developed economies—document unmet demand for pollution avoidance strategies (Ahmad et al. 2022; Barwick et al. 2024; Freeman et al. 2019; Ito and Zhang 2020; Metcalfe and Roth 2025). In contexts where avoidance demand appears low Greenstone et al. (2021), emerging evidence points to the role of (mis)beliefs and limited awareness (Chowdhury et al. 2025). Our work complements these studies by focusing on the demand for more accessible information channels—such as mobile-based forecasts—rather than government monitors deployed to meet regulatory requirements. Recent studies from the US show that demand for private low-cost monitors is concentrated in wealthier and less polluted areas, raising inequity concerns in access to environmental public goods (Coury et al. 2024; Zivin et al. 2024). Taken together, our findings highlight how consumer beliefs—and not just infrastructure—can help mitigate informational market failures, particularly in underserved communities.

The rest of the paper flows as follows: Section 2 describes the landscape of air quality information in Lahore. Section 3 provides the experimental design. Section 4 defines outcome measures while Section 5 outlines the identification strategy. Section 6 presents our main findings. We then develop a conceptual framework to interpret the results and explore potential mechanisms. Section 8 concludes with policy implications.

2 Context

2.1 Air quality information sources

Punjab Province's Environmental Protection Department (EPD) is the primary regulatory body in Lahore with a mandate to protect the environment and meet the national environmental quality standard, including providing environmental information (The Parliament of Pakistan 1997). EPD is expected to publish daily reports of scheduled pollutants.

Daily readings by EPD, however, are hard to access and often unavailable for reasons that are unclear. EPD only makes public the daily readings in English on their website and does not, to our knowledge, publicize their data in a consistent and timely manner in an accessible form such as social media.[5] Those interested in the readings would have to access

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

the website and download a daily PDF, an example of which is shown in Appendix Figure A.1. This is a significant hurdle for an average citizen with limited access to the internet and the lack of proficiency in English.[6] Furthermore, EPD's PM2.5 readings are missing at a much higher frequency than other sources. Figure 1 shows that the readings were missing for most of December 2022, usually one of the worst air quality periods due to the seasonal smog.[7] As a result, we find that only around 9 percent of the working-class citizens in Lahore report having accessed air quality readings from EPD, as shown in Appendix Table A.1.

Limited information from EPD has led citizens' initiatives, like the Pakistan Air Quality Initiative (PAQI), to collect and publish their own data to fill the void left by government services. Established in 2016, PAQI crowd-sources several low-cost air quality monitors (IQAir and PurpleAir) installed at private homes, businesses, and educational institutions. 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 levels, going back as far as one month.[8] PAQI also has a Twitter account that disseminates daily readings from Lahore.[9]

Yet, the vast majority of the working-class population is unaware of PAQI's or other sources' initiatives, as they may be constrained in their smartphone's data capacity to access specialized apps for air quality or Twitter.[10] Appendix Table A.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 challenging for the average citizen to access. The most prominent among such sources is called AirNow, a high-quality monitoring system operated by the U.S. Consulate General in Lahore.[11] The Consulate shares their readings on their website and on Twitter, but, to our knowledge, does not actively

[6] According to GSMA (2024), 53% of men and 33% of women have access to mobile internet. However, much smaller shares (26% for men and 11% for women) access mobile internet on a daily basis for multiple use cases.

[7] EPD does not usually disclose reasons behind data unavailability, though we know from anecdotal experience in maintaining a ground monitor in Lahore that monitors need to occasionally go through maintenance and recalibration. However, we do not have information on the exact reason behind the lack of data in December 2022.

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

[9] @LahoreSmog

[10] 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.

[11] The U.S. Consulate General in Lahore hosts an air quality monitor funded by the U.S. EPA. The AirNow International program places air quality monitors at U.S. embassies and consulates in mainly 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`.

engage in other forms of dissemination.

2.2 Multiple, conflicting information sources

Information friction and uncertainty about service quality may lead to a world in which air quality readings differ significantly between information sources. We find that the service quality of EPD and PAQI differ significantly in terms of data availability and correlation with an independent U.S. Consulate measure. Appendix Table A.2 shows the summary statistics of daily PM2.5 readings by the air quality information source before and during our intervention. Even though EPD values are higher than the PAQI values during the pre-intervention period with high pollution, EPD has 36 fewer observations due to non-reporting. Appendix Tables A.3 and A.4 also show that the EPD readings are less correlated with, and have higher deviations from, the U.S. Consulate measures than the PAQI ones do.

As a result, consumers must decide whether, and from which source, they want to consume air quality information. To do so, they need to gauge the true extent of air pollution and assess the veracity, reliability, and other relevant attributes of a given information source. Inferring service quality, however, may be difficult in a context where there is political polarization and concerns for misinformation in social media (Davies 2023; Hirshleifer et al. 2023). In fact, the Provincial government has signaled its desire to push back by suggesting possible legal action against "fake air quality data" (Raza 2021). Furthermore, EPD reports now 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" as shown in Appendix Figure A.1.

3 Experimental design

To address how attribution to a source affects consumers' beliefs and demand for information, we conduct an intervention in which we hold the signal quality constant and vary the information sources that we make salient. We randomly sample 1,010 households into two treatment arms: government (EPD) and NGO (PAQI). We optimize power to detect differences between the treatment arms and do not have a pure control arm. We provide the same day-ahead forecast and the day-of readings to the two treatment arms. In each message, we make salient the information source to which recipients are assigned. We capture changes in recipients' beliefs and preferences about air quality and information sources through lab-in-the-field games. The timing of our intervention and surveys is shown in Appendix Figure

A.2.

3.1 Defining the measure of 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 revolve around the role of information sources in consumers' beliefs about some objective measures of true air quality levels, we must choose an independent source that is neither the EPD nor PAQI to be our ground truth. We choose the U.S. Consulate monitor as such a measure because its readings are likely accurate measures of the ground truth at the monitoring station. The U.S. Department of State requires consular monitors and data quality to be of the same technical standard as ones used by the U.S. EPA domestically (White 2018).

The measure for which we construct a forecast model is the daily average concentration of PM2.5 (in $\mu g/m^3$) at the U.S. Consulate. We construct a single daily measure and abstract away from spatial variation in air quality, as our respondents reside in a relatively confined, single subsection of the city.[12] The daily based on hourly readings between 12:00 AM and 4:00 PM.[13] The time window is selected so that the research team can collect the day's readings, estimate the next-day forecast, and send the SMS with the day's readings and the next-day forecast to our sample households between 6:00–8:00 PM. We identified via our pilot that this timing allows SMS recipients to make plans for the next day based on the SMS forecast.

3.2 Forecast model

We construct an ensemble prediction model of the U.S. consulate readings for the next day ($t+1$). The model's objectives are twofold: 1) to provide the most accurate forecast possible given the available information, and b) to ensure that information from both EPD and PAQI are used to construct the forecast. Ensemble modeling allows us to selectively attribute to a source (EPD or PAQI) while holding the actual forecast values constant.

[12]Previous studies have shown that exposure to air pollution vary by socioeconomic status within cities (e.g., Hsiang et al. 2019). However, this spatial variation is not a concern for our randomization and intervention, as the accuracy of ground truths or forecasts does not depend on treatment arms.

[13]We rely on other sources when the U.S. Consulate monitor readings are unavailable. When 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 the Urban Unit is also missing, we use PAQI, whose readings are always available. As of 24 May 2023, the U.S. Consulate monitor is missing readings for 16 of the 97 intervention days. Out of 16 days where the U.S. Consulate is missing data, the Urban Unit is missing on 4 days.

First, we construct four forecast models, all of which predict the ground truth, as inputs for the ensemble forecast model. Each model contains readings from only one monitor as inputs. 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 its $t+1$ forecast. Each model, except for SPRINTARS, also uses historical meteorological readings and weather forecasts for $t+1$ as inputs. For each of the input models, we use an adaptive Lasso framework 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. We have $t-20$ out-of-sample forecasts, the last of which is for Day $t+1$, for each model.

We then combine the input 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. The ensemble forecast is the weighted sum of the individual forecasts.[14] We find that the ensemble-model forecasts are significantly more accurate than the incentivized forecasts by those in the experimental sample at baseline, as shown in Appendix Table A.6.

3.3 Sampling and demographics

We conduct our intervention in lower-middle-class neighborhoods of National Assembly Constituencies 123 and 124 in northern Lahore. We divide the two constituencies into 200m×200m grids and randomly select 100 blocks, weighted by population density. We then sample 1,010 households from the selected block centroids following the left-hand rule: survey every ten households by spiraling out from the centroid counterclockwise.

The sample consists of overwhelmingly male, middle-aged household decision makers who earn a living outside through non-salaried employment. Appendix Table A.7 shows summary statistics and balance on main demographic variables. Because the inclusion criteria for our intervention is that they have access to a cellphone and can make decisions based on the SMS forecasts they receive, the sample are over 95% male, mostly married with children and other family members to support, and are around 39 years old on average. They report to have worked 11 hours on average the day prior to the survey, of which 5 are outside. Only about

[14] Inclusion of both EPD and PAQI sources in the ensemble model does not worsen forecast accuracy, relative to picking either source. Appendix Table A.5 shows that forecast errors of EPD and PAQI-based LASSO models are equivalent. In fact, forecasts from Sprintars's climate model (and not a LASSO model based on its air quality estimates) has significantly higher errors, to the overall worse performance of the implemented ensemble model than any single-source LASSO models or an ensemble model of only EPD and PAQI inputs.

27% of the respondents report their main source of income as either salaried employment or pension, meaning that the rest make their living through other forms of non-salaried private enterprise or as day laborers. As such, our intervention sample consists of male household heads in working-class neighborhoods of a developing city.

3.4 Randomization

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 the NGO (PAQI). We do not have a pure control group that does not receive SMS forecasts, as the purpose of this study is to understand the effect of the source, holding constant service quality and forecast values.

We stratify the randomization process into the two treatment groups on a set of baseline variables that 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, and d) the household asset index.We use the optimal-greedy algorithm and generate blocks using the Minimum Volume Ellipsoid (MVE) estimator. We are primarily concerned about balance on outcome variables at baseline and 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 split the subjects into 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. The daily messages are sent from the beginning of the intervention on 18 February, 2023, and continues through the endline survey. The daily messages as part of the intervention ends on 20 June, 2023. All messages were in Urdu. The English translation of the messages on, for instance, 20 May 2023 is as follows:

- T1: "Actual Air Quality (PM 2.5) on 20-05-23: 106
  Air Quality Forecast (PM 2.5) for 21-05-23 using data From Punjab Government (EPD): 120."

- T2: "Actual Air Quality (PM 2.5) on 20-05-23: 106
  Air Quality Forecast (PM 2.5) for 21-05-23 using data From NGO (PAQI): 120."

Figure 2 also shows screenshots of the daily messages for T1 and T2. The messages are sent around 6:00–8:00 PM daily, after collecting the day's data and estimating the forecast for *t+1*. The daily messages also contain the readings from time *t*. 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*. The subjects also receive an introductory message before the start of the daily SMSs and a reminder message every two weeks over the course of the intervention, as discussed in further detail in Appendix Section B.1.

4 Primary outcome variables

We identify four primary outcomes, constructed using incentivized games, with which we test five primary hypotheses.

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

In the endline survey, we ask for the respondent's willingness to pay for the SMS-based air quality forecast messages for two additional months. 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 prompt, we make the experimentally assigned source salient by reminding them that the forecast is built using data from the said source. The bid ceiling is PKR 400.

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

In baseline and endline surveys, we ask respondents guess the air pollution level on the next day. We show respondents a table containing the average, minimum, and maximum of the average daily PM2.5 concentration over the last calendar week and 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%. 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.

4.3 Perceived accuracy of the source as the absolute difference between own and SMS forecasts

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. 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 $t + 1$.

4.4 Preference for information source as the share of donations to government vs. NGO

In baseline and endline surveys, we offer an opportunity to donate PKR 100 between two sources for environmental protection: a government institution and PAQI. The outcome is defined as the share of PKR 100 donated to a government agency for an environmental cause, as opposed to the NGO.

5 Identification strategy

5.1 Exogenous variable

Our exogenous variable is treatment assignment between the arm where the government (EPD) was made salient as the source, as opposed to the NGO (PAQI). For expository purposes, we refer to being in the government arm as being in the "treatment," and the NGO arm as the "control." 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 NGO arm.

5.2 Test of positive willingness to pay for air quality information

First, we conduct a t-test to see if the willingness to pay for the SMS forecasts is higher than 0. We pool the two treatment arms.

5.3 Treatment Effects

We estimate the treatment effects between subjects using one of the following equations:

$$Y_i = Z_i^{'}\beta + \gamma Y_{0i} + \boldsymbol{X_i^{'}\delta} + \varepsilon_i \tag{1}$$

$$Y_i = \alpha + Z_i{}^{'}\beta + \boldsymbol{X_i^{'}\gamma} + \varepsilon_i \tag{2}$$

We use Equation 1 for outcomes that have baseline measures and Equation 2 for those that do not. 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: $\beta \neq 0$.[15]

6 Results

6.1 Checks on balance

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

6.2 Tests of information spillovers

One potential concern for internal validity is informational spillovers. For instance, individuals in the Government arm may learn that their neighbors receive similar SMS-based forecast services from the NGO source. We minimize this concern via design by ensuring that sampled households are sufficently spaced out at every 10 doors. To identify the extent of informational spillovers, however, we ask the respondents at the endline survey whether they have heard about, or have independently accessed air quality readings from, EPD and

[15]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 exclude this identification strategy from our anlaysis.

PAQI.

Appendix Table A.8 shows the results. We find that approximately 50% of respondents report to know about the source to which they are assigned at the endline. However, we also find that 2.2% of respondents assigned to the EPD arm report to know about PAQI and 0.4% of them vice versa. Furthermore, only very small fractions of respondents report to have independently accessed either EPD or PAQI readings at the endline, and these levels are not statistically significantly different between treatment arms.

6.3 Pre-specified outcomes

Table 1 shows the coefficients and their standard errors of the intend-to-treat estimates for the five pre-specified primary hypotheses using post-double-selection LASSO. Here, by "treatment," we mean being assigned to the government arm instead of the NGO arm. Table 1 also shows the p- and q-values of the corresponding columns. In the following subsections, we focus our analysis on the four pre-specified primary outcomes and five hypotheses. We then complement the findings with non-primary outcomes and analyses.

6.4 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 1 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 3 also shows the distribution of the willingness-to-pay for air quality forecasts as demand functions, indicating considerable heterogeneity.

However, there are no statistically and economically significant differences between the treatment arms in their willingness to pay for the forecasts. Column 2 in Table 1 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 significantly different from zero. The small coefficient and standard error exclude any economically meaningful difference between the two treatment arms. We also do not find evidence that the distributions of the willingness to

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

pay are significantly different between treatment arms. Appendix Table A.9 shows that the densities of the 50-rupee bins are not statistically significantly different in 7 out of 8 bins. Appendix Table A.10 also shows that we fail to reject the null in the Kolmogorov-Smirnov and Wilcoxon rank-sum tests.

What would explain the high demand for information, yet no distinction in their differential willingness to pay by the information source? One possibility is that our experiment failed to make the information source salient or that information sources do not matter in forming or shaping recipients' beliefs and preferences. We rule out this possibility in the following sections. Another possibility is that the source affects recipients' beliefs and preferences for air quality information that factor into their utility from consuming the air quality forecast. But such effects may not be picked up in the aggregate willingness-to-pay measure if a) the effects are symmetric between treatment arms or b) the magnitude of the effects is small at the margin.

As such, we highlight three attributes that we hypothesize are key to the recipients' willingness to pay for air quality information and test how the treatment assignment affects them (see Equation 9 in Appendix E). The first attribute is the recipients' beliefs about the state of air quality; if the treatment affects recipients' beliefs about air quality levels and their ability to forecast it, then that may affect their willingness to pay for future air quality information. The second is the recipients' beliefs about the accuracy of information. The third is the recipients' relative preference between the two sources. We examine these three attributes systematically in the following subsections.

6.5 Beliefs about the state, i.e., air quality levels

We find that different information sources do not lead to differential beliefs about air quality levels nor improve forecast ability in one treatment group relative to another. We measure the respondents' beliefs about air quality via incentivized 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 1 shows that those assigned to the government arm have, on average, a five percentage-point higher forecast error than in those in the NGO arm, although the difference is not statistically significant. The magnitude is also small relative to the NGO-arm average error of 73% of the actual reading.

Other measures of beliefs about air pollution levels confirm this finding. 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.13. Second, we also do not find significant differences between treatment arms in stated measures of concern about air quality. Appendix Table A.14 shows statistically insignificant results on the Likert-scale measure of concern about air quality and the number of days in the last week that the respondents believed to have had good air quality.

Furthermore, the information sources do not differentially affect avoidance behaviors and policy preferences. These results are shown in Appendix Tables A.15 to A.19. We do not generally find statistically significant differences by treatment arms on time spent outdoors, policy preferences for air quality over other issues, and their willingness to file a complaint to the local government. Appendix Table A.16 shows, however, that the recipients in the Government arm are 3.5 percentage points less likely to report to have a mask and 4.3 percentage points less likely to have used a mask in the last week. We refrain from over-interpreting these results due to the concerns about multiple hypothesis testing and the lack of effects on similar outcomes such as time use.

One may be concerned that recipients do not update their beliefs about air quality levels and, therefore, do not engage in any avoidance behavior. This is unlikely, as demonstrated via the following correlational relationships between air quality and outdoor time use. Appendix Table A.17 shows the correlations between PM2.5 information and respondents' outdoor time use at baseline and endline. The table shows that a $10\mu g/m^3$ increase in the PM2.5 concentration reduces respondents' stated outdoor time use by approximately 3 minutes per day at endline. This negative correlation is absent at baseline, suggesting that respondents adjust their time use on worse air quality days.

Overall, we do not find strong evidence that exposure to an information source alone leads to differential changes in beliefs about the state of air quality or in subsequent avoidance behaviors. Beliefs about air quality levels and the demand for avoidance behavior *per se* may affect the willingness to pay for air quality information. However, our treatments did not have differential effects in these beliefs and are unlikely to have driven differential willingness to pay for air quality information between treatment arms.

6.6 Beliefs about SMS forecast's accuracy

Next, we focus on the recipients' beliefs about the accuracy of information. Note that recipients in the two treatment arms receive identical readings, forecasts, and messages other than the source to which the information is attributed. As such, differential beliefs about the accuracy of the SMS forecasts should be formed through the source to which the information is attributed.

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. The absolute difference between the two measures captures the respondents' belief about the quality of SMS forecasts conditional on their own beliefs about air quality the next day.

We find that the respondents in the Government arm believe in larger SMS forecast errors than in the NGO arm. Column 4 on Table 1 shows the difference to be 2.8 points in the concentration measure ($\mu g/m^3$). The effect size is 12% of the NGO arm's mean (22.7). The effect is statistically significant at the 5% level and survives adjustments to multiple hypothesis testing.

Those assigned to the Government arm believe, on average, that the SMS forecasts are less accurate than those in the NGO arm despite having statistically indistinguishable willingness to pay for the forecast services. This result suggests that recipients either have a relatively low willingness to pay for forecast precision on the margin or care about other attributes such as punctuality and ease of access. As such, we examine how recipients' beliefs about service quality associated with a particular source are shifted and, more generally, how their preferences between sources shift due to our intervention.

6.7 Preferences for sources

We conduct donation games with financial stakes for our primary measure of respondents' preference between sources. In the baseline and endline surveys, respondents PKR 100 between government and NGO sources, which the survey team donates to respective agencies. The relative allocations, as well as changes to them between baseline and endline, identify respondents' preferences for information sources with real financial stakes.

We find that the respondents shift a larger fraction of their donations to the experimentally assigned sources at endline. Figure 4 shows the distributions for both treatment groups at baseline and endline. 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 assigned to the Government arm donate more to the government at the endline, as opposed to the NGO one. On the other hand, more than 90 percent of respondents assigned to the NGO arm donate more to the NGO at the endline. The average ratio between the assigned source and the other is approximately 75:25. Column 5 in Table 1 confirms that those assigned to the Government arm donate PKR 54 more to the government, on average, relative to the respondents in the NGO 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 BDM, they are asked hypothetically how much they would be willing to pay if the forecast were to come from the other source (i.e., NGO for those assigned to government, and vice versa). Column 4 in Table 2 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 based on revealed preference, this finding aligns with other findings from other measures of preferences for sources.

We do not have other revealed-preferences measures that would help us identify exactly what components, besides forecast accuracy, of the information sources' services recipients value. As such, we collect a set of stated-preference measures of the recipients' approval of the sources' reliability (i.e., punctuality of their forecasts), accuracy, and overall service quality. We collect these measures for both Government and NGO service providers for each recipient.

Table 3 shows that recipients assigned to the Government arm have significantly higher approval for it regarding reliability, accuracy, and overall satisfaction than those assigned to the NGO arm. Similarly, Table 4 shows the symmetrical results for those assigned to the NGO arm. We find that the recipients are satisfied with the services they received regardless of treatment arms, and they associate their satisfaction with the source made salient by our intervention. Overall, we find evidence that preferences are relatively malleable in a highly frictional market for information services.

6.8 Heterogeneity by baseline beliefs

6.8.1 Pre-specified dimensions of heterogeneity

So far, we have identified average intent-to-treat effects of information sources on beliefs and preferences. However, one may be concerned that the evolution of beliefs and preferences may depend on consumers' priors. Average effects may mask significant heterogeneous treatment effects, leading to divergence in beliefs and preferences based on consumers' baseline characteristics.

To address such concerns, we pre-specify four dimensions to test potential heterogeneity, as described in detail in Appendix Section D.1. They are: a) a measure of relative preferences between the sources using the baseline donation game, b) a relative measure overall approval between sources, c) a relative measure of beliefs about the sources' accuracy, and d) beliefs about air quality levels. All of these measures are collected at baseline.

We focus the beliefs and preferences between information sources at baseline based on the emerging body of work on media bias, trust for information sources, and polarization. Previous work has shown 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).[17] 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.

We focus on baseline beliefs about air quality to see if the extent of belief-updating depends on the accuracy of baseline beliefs and the recipients' beliefs about signal quality. Those 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. Critically, the extent to which such individuals update their beliefs would also depend on their beliefs about the strength of the signal, which they may glean from the information source.

6.8.2 Heterogeneous treatment effects

We estimate heterogeneous treatment effects based on the pre-specified dimensions.[18] The empirical specifications are described in Appendix Section D.2. We do not find strong evidence that the consumers respond differentially based on their prior beliefs about the information sources' service quality. We also find evidence that consumers with higher baseline forecast errors have higher endline forecast errors if they are assigned to the government arm.

First, we do not find strong evidence of heterogeneous treatment effects based on consumers' preferences for sources, except for endline donations to the government. Appendix Tables A.20 to A.22 show the linear heterogeneous treatment effect estimates and their cate-

[17]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).

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

gorical equivalents in Appendix Tables A.24 to A.26. Coefficients on interaction terms from Appendix Tables A.20 to A.21 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, Appendix Table A.22, but this result is not corroborated with a categorical specification in Appendix Table A.26.

Second, we find an adverse heterogeneous treatment effect based on the consumers' baseline forecast error on forecast error and respondents' beliefs about the SMS's error Appendix Table A.23 shows the linear estimates, and Appendix Table A.27 the categorical equivalent. For those assigned to the government arm relative to the NGO one, 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 than similar individuals if assigned to the government arm v.s. the NGO. Similar causal effects also exist on the respondents' beliefs about the SMS's errors but are less precisely estimated.

Two takeaways emerge as from the pre-specified analysis. First, there are no strong heterogeneous effects based on consumers' priors about the sources on the demand for air quality information. This confirms our results from Section 6.7 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 update their beliefs more slowly when they are assigned to the government source, which they believe to have lower quality.

7 Mapping results to a conceptual framework

Our empirical results show that, although the source does not differentially affect the recipients' demand for air quality information, it affects several underlying beliefs and preferences. To put further structure to our findings, we map our empirical results to a conceptual framework that specifies the utility of consuming an air quality forecast service in terms of its attributes. We establish links between pre-specified hypotheses and the conceptual framework. We then identify which attributes of the utility function shift in response to an information source that is exogenously made salient. We present the highlights in this section and provide further details in Appendix Section E.

We specify a consumer's utility function consisting of attributes underpinned by their beliefs. First, consumers hold beliefs about the state variable, i.e., the air quality level. They value the forecast information because it provides additional signals about the state and helps them take better mitigation measures against air pollution. Second, they also hold beliefs about signal quality (i.e., the accuracy of the SMS forecast) for a given source. Third, they may also hold beliefs and preferences about a source that is not tied to signal quality (e.g., consistent availability) as well as innate preferences for a given source, all of which we bundle into an attribute. Such beliefs and preferences factor into the utility function as attributes. In our experiment, we elicit consumers' utility gains from a signal—the SMS air quality forecast—whose source we exogenously vary. We denote the utility gained from accessing an air quality forecast from source $s \in \{G, P\}$ of a consumer i assigned to treatment arm $a \in \{G, P\}$ (G(overnment) or P(rivate), i.e., NGO) on day t as $u^s_{i,a,t}$. We measure $u^G_{i,G,t}$ and $u^P_{i,P,t}$ through BDM, and $u^P_{i,G,t}$ and $u^G_{i,P,t}$ through hypothetical willingness-to-pay measures.

Based on the utility and other belief measures, our empirical analysis tests whether treatment assignment differentially affects the demand for air quality information and its attributes, i.e., whether being assigned G to affects $u^G_{i,G,t}$ and its attributes differently from being assigned to P affecting $u^P_{i,P,t}$. We also analyze whether treated individuals prefer to receive forecasts from the source to which they are experimentally assigned against the alternative, i.e., if $u^G_{i,G,t} > u^P_{i,G,t}$ and $u^P_{i,P,t} > u^G_{i,P,t}$ and changes in beliefs about which attributes explain the difference.

Our empirical results map to the following descriptions of the model's terms. First, we fail to reject $u^G_{i,G,t} \neq u^P_{i,P,t}$ but find that $u^G_{i,G,t} > u^P_{i,G,t}$ and $u^P_{i,P,t} > u^G_{i,P,t}$, i.e., the willingness to pay is not differentially affected between treatment arms but recipients prefer the source to which they are assigned. Second, we do not find that their air quality forecast error is different between treatment arms, i.e., their beliefs about the state variable are not differentially affected. Third, recipients assigned to the government arm believe that the SMS forecasts have higher errors than those in the NGO arm.

Our empirical results suggest that consumers put a relatively small weight on the signal quality attribute and/or that they value attributes specific to the assigned source other than signal quality. On these potential mechanisms, we provide correlational evidence in Appendix Section E. First, we show that $u^G_{i,G,t}$ is not correlated with the SMS forecast error, but $u^P_{i,P,t}$ is, suggesting that consumers receiving government forecasts do not value signal quality on the margin. Second, we evaluate what aspects of the service quality, as measured through Likert-scale statements, correlate with $(u^G_{i,G,t} - u^P_{i,G,t})$ and $(u^P_{i,P,t} - u^G_{i,P,t})$. We find that it is their approval of the assigned source in terms of reliability. These correlational results suggest that consumers do not value the forecast's precision at the margin but rather value the reliability of the service that they have experienced.

8 Conclusion

We study how consumers form beliefs and preferences for a public good, an environmental information service, when there is uncertainty about the state of environmental quality as well as the service quality of suppliers. We conduct a randomized control trial in which we randomly attribute air quality forecast services to one of two sources: the government and an NGO. We evaluate if the random attribution leads to a differential demand for air quality forecasts and beliefs about air quality levels and information sources. We also investigate 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 high demand for air quality information, yet not differentially so by the associated source. Yet, we find that those assigned to the government arm believe the information they receive is less accurate than those in the NGO arm, while they are equally willing to pay for the information service. Consumers shift their preference toward the source to which the they are exposed. Our findings provide insights into a market for a public good with competing suppliers and high levels of friction to access them; consumers have limited access to air quality information at baseline yet have a high demand for it as measured through an incentivized elicitation. They prefer a source to which they are exposed, yet may not value the accuracy of the information at the margin, at least as long as they generally approve of the service quality.

Our results have policy implications for governments, multilateral organizations, and civil society on improving access to environmental information.

First, our findings suggest sizable potential welfare gains from increasing access to environmental information where there are damages from environmental degradation. Residents of a working-class neighborhood of Lahore are willing to pay PKR 119 (≈USD 0.50 as of August 2022) per month for air pollution forecasts after the conclusion of the free information intervention. This amount roughly equates to 50% of the cost of monthly prepaid mobile and data services. Scaling the service across the city—with close to 14 million residents—will likely lead to a large benefit based only on individual consumers' willingness to pay for information. In contrast, we find in the baseline survey that only around 9 percent of the working-class citizens in Lahore report having accessed air quality readings from a given source, underlying the market failure in the provision of public goods.

Second, our findings show limited sensitivity of the overall willingness to pay to the source and to their believed accuracy, suggesting that the welfare gains would come from access to information regardless of the source. A social planner, therefore, should increase citizens' access to air quality information regardless of the source. Note, however, that such takeaway may be conditional on a certain signal and service quality we maintained as part of the experiment. Potential concerns remain as to whether, outside the experimental setting, there are trade-offs between accuracy, reliability, and operational costs between information sources that have different hardware and other technical capacities. Government agencies tend to have higher quality monitors that can monitor multiple scheduled pollutants and meet technical standards, while crowd-funded NGO need to rely on low-cost monitors of uncertain quality (US EPA 2024). Consumers may be harmed if, for instance, NGO air quality readings are significantly less accurate than the government ones to the point where consumers would care about the difference in accuracy.

We provide two pieces of evidence that, at least in Lahore's context, such trade-offs are unlikely to be of concern. First, we find that non-government sources perform better across several metrics of accuracy and reliability than government ones. Appendix Tables A.2 to A.4 show that the government (EPD) air quality readings have higher noise than the NGO (PAQI) ones based on low-cost monitors when we define the truth to be the third-party U.S. Consulate measures. Second, when it comes to predictive modeling, forecasts with inputs from government readings perform as well as those using non-government inputs, as shown in Appendix Table A.28 in terms of forecast errors. In other words, there is limited scope for improving air quality forecasts for the city of Lahore by introducing additional devices or sources.

Third, our findings show shifts in consumers' preferences for the experimentally assigned source against alternatives from a baseline of seemingly weak priors. However, our experimental evidence is unable to speak to the long-term effects on consumers' preferences for a source. For instance, consumers' preferences for an information source may become less malleable as they experience services from it. Such patterns could lead to polarized beliefs about the sources' accuracy and service quality in the long run. Shifting consumers to receive information from one source to another may have a negative welfare impact, as we find that consumers have a lower willingness to pay for information from a less familiar source. Further work is needed to understand consumers' longer-term belief updating process and preferences.

References

Acemoglu, D., C. G. Aksoy, C. Baysan, C. Molina, and G. Zeki (2024). Misperceptions and demand for democracy under authoritarianism. Technical report, National Bureau of Economic Research.

Acemoglu, D., A. Cheema, A. I. Khwaja, and J. A. Robinson (2020). Trust in state and nonstate actors: Evidence from dispute resolution in pakistan. *Journal of Political Economy 128*(8), 3090–3147.

Ahmad, H. F., M. Gibson, F. Nadeem, S. Nasim, and A. Rezaee (2022). Forecasts: Consumption, Production, and Behavioral Responses.

Andrabi, T., J. Das, and A. I. Khwaja (2017). Report cards: The impact of providing school and child test scores on educational markets. *American Economic Review 107*(6), 1535–1563.

Athey, S. and G. Imbens (2017). The Econometrics of Randomized Experiments. In *Handbook of Economic Field Experiments*, Volume 1, pp. 73–140. Elsevier.

Barwick, P. J., S. Li, L. Lin, and E. Y. Zou (2024). From fog to smog: The value of pollution information. *American Economic Review 114*(5), 1338–1381.

Baysan, C. (2022). Persistent Polarizing Effects of Persuasion: Experimental Evidence from Turkey. *American Economic Review 112*(11), 3528–3546.

Becker, G. M., M. H. Degroot, and J. Marschak (1964). Measuring utility by a single-response sequential method. *Behavioral Science 9*(3), 226–232.

Belloni, A., V. Chernozhukov, and C. Hansen (2014, April). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies 81*(2), 608–650.

Bergstrom, T., L. Blume, and H. Varian (1986). On the private provision of public goods. *Journal of Public Economics 29*(1), 25–49.

Bowen, T. R., D. Dmitriev, and S. Galperti (2023). Learning from shared news: When abundant information leads to belief polarization. *The Quarterly Journal of Economics 138*(2), 955–1000.

Burlig, F., A. Jina, E. M. Kelley, G. V. Lane, and H. Sahai (2024). Long-range forecasts as climate adaptation: Experimental evidence from developing-country agriculture. Technical report, National Bureau of Economic Research.

Chopra, F., I. Haaland, and C. Roth (2022). The demand for news: Accuracy concerns versus belief confirmation motives.

Chowdhury, A., T. Garg, M. Jagnani, and M. Mattsson (2025). Misbeliefs, experience, and technology adoption: Evidence from air purifiers in bangladesh. *Available at SSRN 5284948*.

Coury, M., L. Falconer, and A. La Nauze (2024). Wildfire smoke and private provision of public air-quality monitoring.

Davies, C. (2023, 4 May). Polarised politics are tearing Pakistan apart. *BBC*.

Dhinakar Bala, A., M. Mattsson, and S. Vyas (2024). Why do governments make promises they cannot keep? evidence from india's national clean air programme. Technical report, SSRN.

Freeman, R., W. Liang, R. Song, and C. Timmins (2019, March). Willingness to pay for clean air in China. *Journal of Environmental Economics and Management 94*, 188–216.

Gentzkow, M. and J. Shapiro (2006). Media Bias and Reputation. *Journal of Political Economy 114*(2), 280–316. Publisher: The University of Chicago Press.

Gentzkow, M., M. B. Wong, and A. T. Zhang (2023). Ideological Bias and Trust in Information Sources.

Ghanem, D. and J. Zhang (2014, September). 'Effortless Perfection:' Do Chinese cities manipulate air pollution data? *Journal of Environmental Economics and Management 68*(2), 203–225.

Greenstone, M., T. Ganguly, C. Hasenkopf, N. Sharma, and G. Hrishikesh (2024). Annual update. *Air Quality Life Index*.

Greenstone, M., K. Lee, and H. Sahai (2021). Indoor air quality, information, and socioeconomic status: Evidence from delhi. In *AEA Papers and Proceedings*, Volume 111, pp. 420–424. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

GSMA (2024). The Mobile Gender Gap Report.

Hirshleifer, S., M. Naseem, A. A. Raza, and A. Rezaee (2023). The spread of (mis) information: A social media experiment in pakistan.

Hsiang, S., P. Oliva, and R. Walker (2019). The distribution of environmental damages. *Review of Environmental Economics and Policy*.

Ito, K. and S. Zhang (2020, May). Willingness to Pay for Clean Air: Evidence from Air Purifier Markets in China. *Journal of Political Economy 128*(5), 1627–1672. Publisher: The University of Chicago Press.

Jha, A. and A. La Nauze (2022, November). US Embassy air-quality tweets led to global health benefits. *Proceedings of the National Academy of Sciences 119*(44), e2201092119. Publisher: Proceedings of the National Academy of Sciences.

Khan, A., S. Nasim, M. Shaukat, and A. Stegmann (2021). Building trust in the state with information: Evidence from urban punjab. *Journal of Public Economics 202*, 104494.

Lan, R., S. D. Eastham, T. Liu, L. K. Norford, and S. R. Barrett (2022). Air quality impacts of crop residue burning in india and mitigation alternatives. *Nature communications 13*(1), 6537.

Metcalfe, R. D. and S. Roth (2025). Making the invisible visible: the impact of revealing indoor air pollution on behavior and welfare. Technical report, National Bureau of Economic Research.

Mu, Y., E. Rubin, and E. Y. Zou (2024). What's missing in environmental self-monitoring: Evidence from strategic shutdowns of pollution monitors. *Review of Economics and Statistics*, 1–45.

Mullainathan, S. and A. Shleifer (2005, September). The Market for News. *American Economic Review 95*(4), 1031–1053.

Muralidharan, K. and V. Sundararaman (2015, August). The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India. *The Quarterly Journal of Economics 130*(3), 1011–1066.

Raza, A. (2021, 17 Nov). Punjab minister wants FIA action against fake air quality data. *Financial Times*.

Samuelson, P. A. (1954). The pure theory of public expenditure. *The Review of Economics and Statistics 36*(4), 387–389.

Shapiro, C. and H. Varian (1999). *Information rules: A strategic guide to the network economy.* Harvard Business School Press.

The Parliament of Pakistan (1997). Pakistan Environmental Protection Act.

US EPA (2024). Low–Cost Air Pollution Monitors and Indoor Air Quality. Technical report.

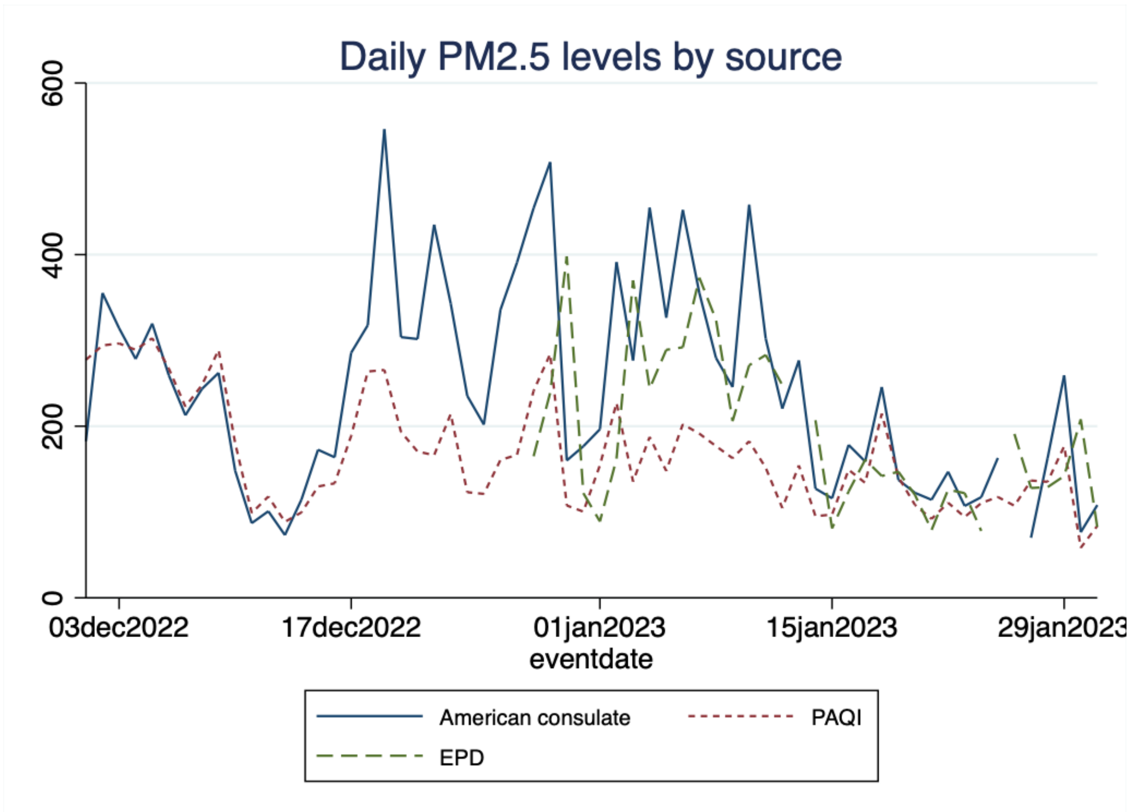
White, J. E. (2018). AirNow Global – Platform for International Collaboration.

WHO (2021). *WHO global air quality guidelines: particulate matter (PM2. 5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide.* World Health Organization.

Zivin, J. S. G., B. Krebs, and M. J. Neidell (2024). Private adoption of public good technologies: The case of purpleair. Technical report, National Bureau of Economic Research.

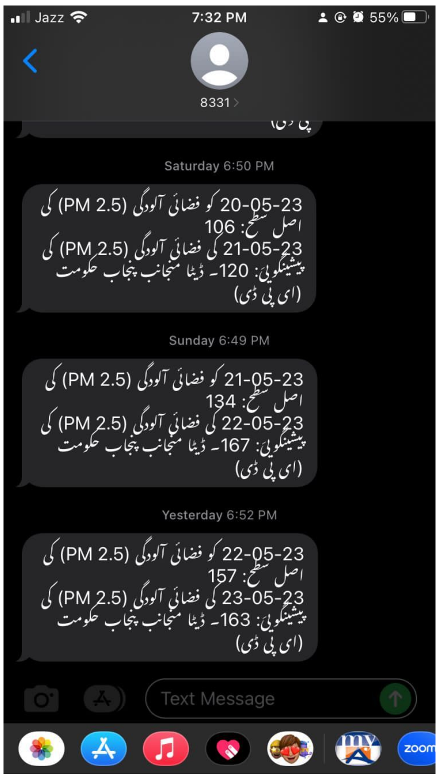
9 Figures

Figure 1: Air quality readings from three main sources



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), an NGO based 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: Sample messages to respondents

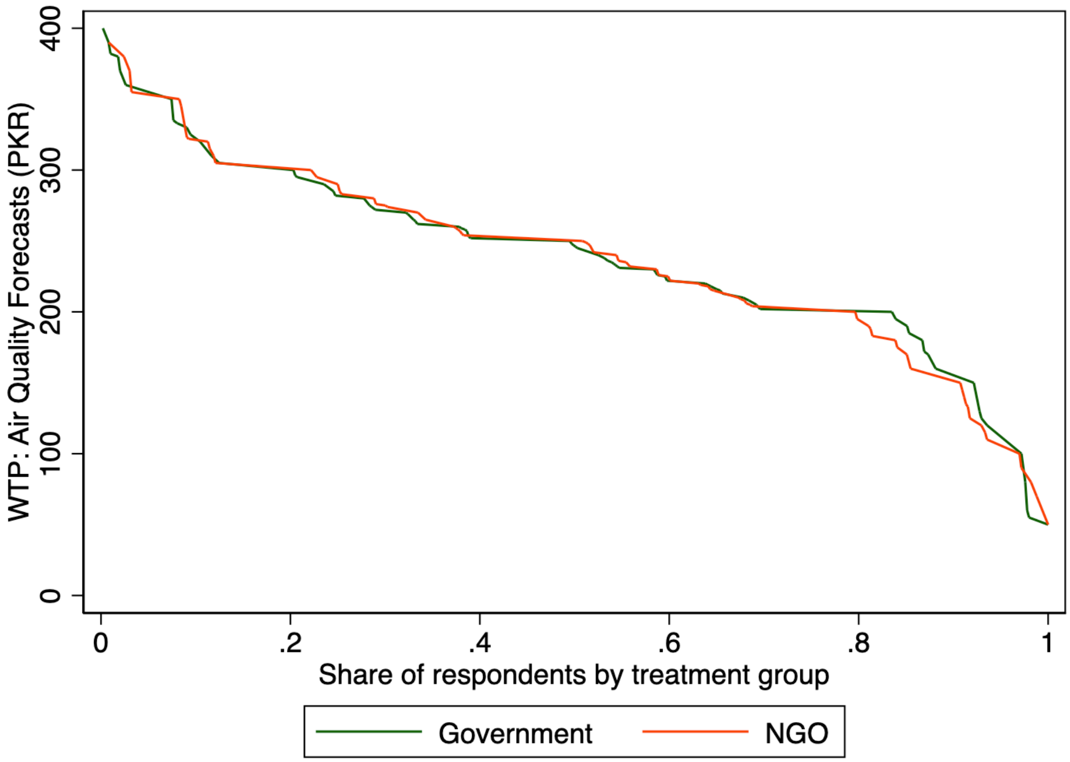


(a) T1: Government (EPD)

(b) T2: NGO (PAQI)

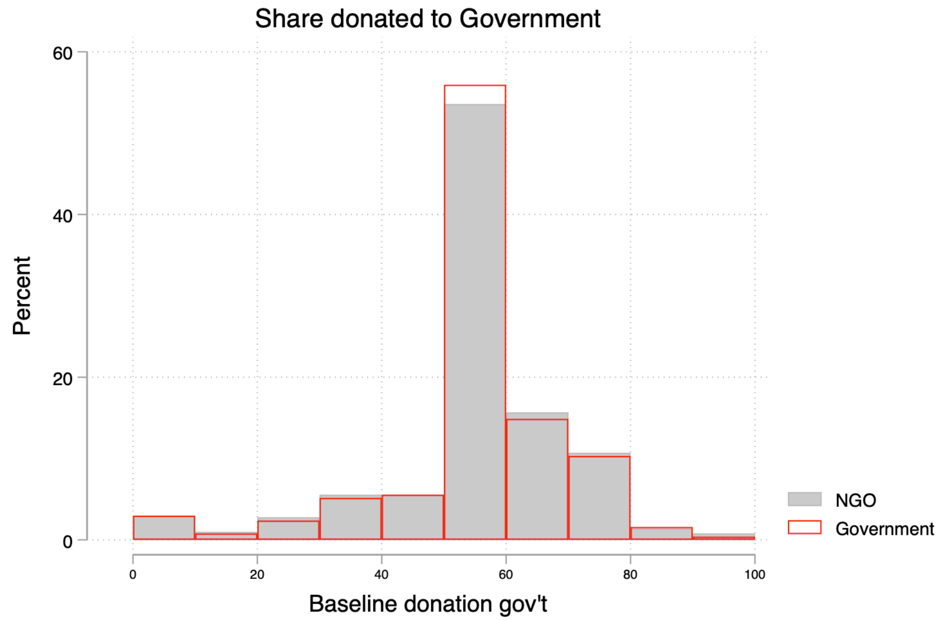
The figures above are screenshots from research managers' cellphones showing daily messages from T1 (EPD) and T2 (PAQI). Daily messages are delivered from the same short code (8331) so that recipients can compare daily readings and forecasts. For 20 May, 2023, the daily messages read as follows:
T1: Actual Air Quality (PM 2.5) on 20-05-23: 106
Air Quality Forecast (PM 2.5) for 21-05-23 using data From Punjab Government (EPD): 120
T2: Actual Air Quality (PM 2.5) on 20-05-23: 106
Air Quality Forecast (PM 2.5) for 21-05-23 using data From NGO (PAQI): 120

Figure 3: Demand curves for air pollution forecast by treatment

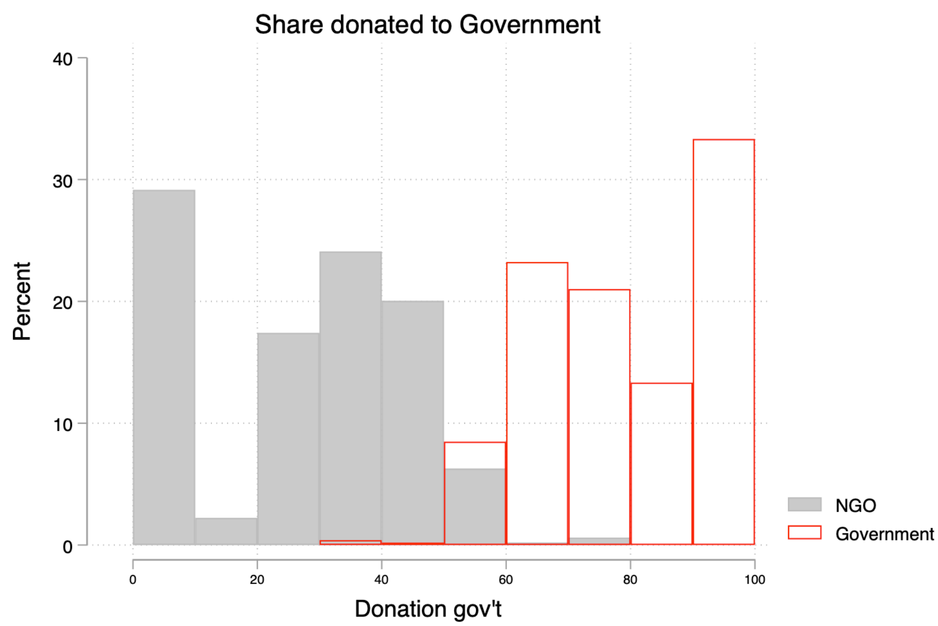


This figure shows the distributions of respondents' bids for two months of air pollution forecast service from the endline survey. "Government" corresponds to T1, the arm in which the EPD source is made salient. "NGO" corresponds to T2, the arm in which the PAQI source is made salient.

Figure 4: Baseline and endline donation to government sources vs NGO



(a) Baseline



(b) Endline

This figure shows the distributions of respondents' donations to a government agency vs. a non-government entity for environmental protection, measured at the baseline and endline surveys. The measure is defined as the amount out of PKR 100 donated to the government source. "Government" corresponds to T1, the arm in which the EPD source is made salient. "NGO" corresponds to T2, the arm in which the PAQI source is made salient.

10 Tables

Table 1: Pre-specified 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)
P value	0	.927	.208	.029	0
Q value	.001	.351	.116	.03	.001
Observations	993	993	993	991	989
Endline mean of NGO		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***. We also show the critical values for the "Constant" and "Gov't arm" coefficients. "P value:" Unadjusted p-values. "Q value": Benjamini Krieger Yekutieli (2006) sharpened q-values.

Table 2: 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 NGO	237.2	221.2	16.0	

Notes: "WTP": The pre-specified 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 3: ITT: Stated preference measure on satisfaction with the Government service

	(1) Gov't: Approval index	(2) Gov't: Reliable	(3) Gov't: Accurate	(4) Gov't: Approve
Gov't arm	2.25***	1.41***	1.34***	1.33***
	(0.054)	(0.037)	(0.039)	(0.037)
Observations	989	980	950	989
Endline mean of NGO	-2.27	3.07	3.05	3.10

Notes: We present estimates of effects on the stated-preference measures on the respondents' satisfaction with the Government's service. We ask if they are overall satisfied with the service (Column 3), if they think the service is reliable and on time (Column 4), and if they believe the forecasts are accurate (Column 2), in the Likert scale where positive values indicate approval. The measure for Column 1 is a standardized sum of measures in Columns 2 and 4. The measures for Columns 2 through 4 are in the 5-point Likert scale. 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 4: ITT: Stated preference measure on satisfaction with the NGO service

	(1) NGO: Approval index	(2) NGO: Reliable	(3) NGO: Accurate	(4) NGO: Approve
Gov't arm	-1.69***	-1.34***	-1.49***	-1.37***
	(0.039)	(0.035)	(0.037)	(0.033)
Observations	982	972	958	986
Endline mean of NGO	1.70	4.46	4.47	4.42

Notes: We present estimates of effects on the stated-preference measures on the respondents' satisfaction with the NGO source's service. We ask if they are overall satisfied with the service (Column 3), if they think the service is reliable and on time (Column 4), and if they believe the forecasts are accurate (Column 2), in the Likert scale where positive values indicate approval. The measure for Column 1 is a standardized sum of measures in Columns 2 and 4. The measures for Columns 2 through 4 are in the 5-point Likert scale. 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 and figures

Table A.1: Balance table of outcome variables at baseline

Variable	(1) NGO 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 NGO. "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 NGO are accurate, and that b) they approve of the job that the NGO is doing to address air quality. "Stated preference 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": An Indexed measure of the household's ownership of electronic appliances. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table A.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 A.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 A.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 A.5: Errors of LASSO forecasts from individual sources and ensembles

	RMSE:All	RMSE:pre	RMSE:during/post	MAD:All	MAD:pre	MAD:during/post
EPD	68.2	103	39.4	47.3	78.6	31
PAQI	67.1	101.7	39.1	46	76.9	30.3
AirNow	66.3	99.4	39.9	46	74.3	31.6
Urban	67.5	101.7	40.1	46.3	76.7	30.8
Sprintars	115	178.6	48.5	75.8	145.7	35
Ensemble (model in SMS)	77.8	111.2	52.7	55.1	83.1	40.5
Ensemble (only using EPD & PAQI)	67.6	102.2	38.9	46.5	77.7	30.3

Notes: The table presents errors of LASSO forecst models using individual reading sources and of ensemble models. RMSE: Root mean squared error. MAD: mean absolute difference. The row labeled "Ensemble (model in SMS)" shows errors of the model we deployed for the intervention. The row labeled "Ensemble (only using EPD & PAQI)" shows errors of an alternative model using only EPD and PAQI LASSO forecasts as inputs in the ensemble. "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. "All": readings from November 1, 2022 to August 26, 2023.

Table A.6: Respondent forecasts, SMS forecasts, and their errors at baseline

	(1)		
	mean	sd	count
Baseline forecast	238.0	59.4	1008
Baseline truth (t + 1)	146.7	47.7	1008
Baseline: SMS forecast (replicated)	161.5	38.8	1008
Abs(baseline forecast - truth)	95.0	48.1	1008
Abs(truth - SMS forecast(replicated))	44.1	27.5	1008

Notes: This table shows t+1 baseline respondent forecasts, SMS forecasts, and their errors against actual readings the truth (American Consulate readings). "Baseline forecast": incentivized forecast by respondents at baseline. "Baseline truth (t+1)" American Consulate readings the next day of the surveyed date. "Baseline: SMS forecast (replicated)": the SMS model's forecast for the corresponding day. "Abs(baseline forecast - truth)": the absolute difference between respondents' forecast and American Consulate readings. "Abs(truth - SMS forecast(replicated))": the absolute difference between the SMS model's forecasts and American Consulate readings.

Table A.7: Balance table of key demographic variables at baseline

Variable	N	(1) NGO Mean/(SE)	N	(2) Government Mean/(SE)	N	(1)-(2) Pairwise t-test Mean difference
Registered own cellphone for SMS forecasts	503	0.988 (0.005)	504	0.984 (0.006)	1007	0.004
Male	504	0.958 (0.009)	504	0.974 (0.007)	1008	-0.016
Age	504	39.431 (0.508)	504	39.091 (0.540)	1008	0.339
Married	504	0.859 (0.016)	504	0.851 (0.016)	1008	0.008
Someone in family has respiratory issues	504	0.038 (0.008)	504	0.018 (0.006)	1008	0.020*
Baseline: hours spent outside	504	7.403 (0.204)	504	7.440 (0.198)	1008	-0.037
Hours Spent on Work	504	11.479 (0.125)	504	11.354 (0.120)	1008	0.125
Baseline: hours spent working outside	504	5.319 (0.198)	504	5.302 (0.198)	1008	0.018
Total members of this household	504	6.633 (0.128)	504	6.742 (0.136)	1008	-0.109
N. elderly	504	0.339 (0.026)	504	0.308 (0.025)	1008	0.032
N. children	504	1.726 (0.079)	504	1.845 (0.083)	1008	-0.119
Years of Formal Education	504	8.552 (0.218)	504	8.302 (0.210)	1008	0.250
Owned their own house	504	0.720 (0.020)	504	0.754 (0.019)	1008	-0.034
Has electricity at home	504	0.990 (0.004)	504	0.992 (0.004)	1008	-0.002
Total Number of Air Conditioners	504	0.133 (0.016)	504	0.161 (0.028)	1008	-0.028
Total Number of Fans	504	3.365 (0.052)	504	3.399 (0.049)	1008	-0.034
Asset index	504	0.020 (0.046)	504	-0.026 (0.043)	1008	0.046
Relative socioeconomic status	504	3.010 (0.029)	503	3.014 (0.027)	1007	-0.004
Main source of income is salaried employment or pension	504	0.266 (0.020)	503	0.280 (0.020)	1007	-0.014
F-test of joint significance (F-stat)						1.059
F-test, number of observations						1005

Notes: This table presents numbers of observations, 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. The following are variable definitions that may require additional explanations. "Registered own cellphone for SMS foreacasts": When signing up for the SMS forecast intervention, the respondent gave their own phone number (having a cellphone number they can receive SMS on was a requirement for inclusion)."Someone in family has respiratory issues": The respondent reported that at least one person in the family has respiratory issues. "Baseline: hours spent outside": The number of hours respondents reported to have spent outside the previous day. "Hours spent on work": The number of hours respondents reported to have spent on paid work the previous day. "Baseline: hours spent working outside": The number of hours respondents reported to have spent on paid work outside the previous day. "Asset index": An Indexed measure of the household's ownership of electronic appliances. "Relative socioeconomic status": A likert-scale question on where the respondent ranks relative to other households in this neighborhood (1 being a lot worse off and 5 a lot better off). "Main source of income is salaried employment or pension": The respondent reports that their main source of income is either salaried private employment, government employment, or government or private pension. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table A.8: Knowledge about, and independent access to, EPD and PAQI readings

Variable	N	(1) NGO Mean/(SE)	N	(2) Government Mean/(SE)	N	(1)-(2) Pairwise t-test Mean difference
Knows EPD	496	0.022 (0.007)	497	0.515 (0.022)	993	-0.493***
Knows PAQI	496	0.496 (0.022)	497	0.004 (0.003)	993	0.492***
Accesses EPD	496	0.014 (0.005)	497	0.024 (0.007)	993	-0.010
Accesses PAQI	496	0.002 (0.002)	497	0.000 (0.000)	993	0.002
F-test of joint significance (F-stat)						2057.740***
F-test, number of observations						993

Notes: This table shows summary statistics and tests of differences between the treatment groups of their knowledge about, and independent access to, EPD and PAQI readings. The measures are collected at the endline survey. Two-tailed significance: p<0.1*; p<0.05**; p<0.01***.

Table A.9: Treatment effects on the distribution of willingness-to-pay in 50-PKR bins

	(1) 1-50	(2) 51-100	(3) 101-150	(4) 151-200	(5) 201-250	(6) 251-300	(7) 301-350	(8) 351-400
Gov't arm	0.0083	-0.0082	-0.026**	0.019	-0.00021	0.0053	0.0072	-0.0060
	(0.0068)	(0.011)	(0.013)	(0.020)	(0.024)	(0.024)	(0.015)	(0.0090)
Observations	1008	1008	1008	1008	1008	1008	1008	1008
Endline mean of NGO	0.012	0.052	0.079	0.16	0.30	0.26	0.087	0.032

Notes: Model: PDSLASSO. The outcomes are dummy variables that equals 1 if the endline willingness-to-pay (WTP) falls in the bin, and 0 otherwise. 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.10: Tests of distribution of the willingness-to-pay—Kolmogorov-Smirnov and Wilcoxon rank-sum tests

	(1) Kolmogorov–Smirnov statistic	(2) p-value
K-S NGO is smaller	.041	.44
K-S Government is smaller	-.021	.81
K-S Combined	.041	.807
Wilcoxon rank-sum test	.	.919

Notes: The table reports the test statistics and asymptotic p-values from the two-sided Kolmogorov–Smirnov and Wilcoxon rank-sum tests. The two groups are the Government and NGO arms. The Kolmogorov–Smirnov statistic is the supremum of the differences between the two groups. In a row labeled "K-S NGO is smaller," we test the hypothesis that the distribution of the willingness to pay is lower for the NGO arm. In a row labeled "K-S Government is smaller," we test the hypothesis that the distribution of the willingness to pay is lower for the Government arm. "K-S Combined" is the combined test statistic. Wilcoxon rank-sum test (also known as the Mann-Whitney two-sample statistic) is a test where the null hypothesis is that the willingness-to-pay measure for Government and NGO arms is drawn from the same distribution.

Table A.11: Pre-specified 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 NGO		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.12: Adjustments for multiple hypothesis testing on pre-specified 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.11. "P value:" Unadjusted p-values. "Q value": Benjamini Krieger Yekutieli (2006) sharpened q-values.

Table A.13: 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 NGO	0.73	0.47	40.0	6.23

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.14: 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 NGO	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.15: 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 NGO	5.14	3.89	3.65

Notes: Standard errors are reported in parentheses. For the "[endline]: hours spent outside" variable, 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. For the "[endline]: Hrs (stated)" variable, we ask respondents to state how many hours they spent outside the previous day, as opposed ot aggregating the hours using the time-use module. For the "[endline]: Hrs (if bad day)" variable, we ask respondents how many hours they would spend on a bad air quality day. 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.16: Secondary outcomes: Mask use

	(1) Has mask	(2) Shows mask	(3) Uses mask
Gov't arm	-0.035*	-0.011	-0.043**
	(0.021)	(0.015)	(0.021)
Observations	993	993	993
Endline mean of NGO	0.20	0.099	0.19

Notes: Column 1 "Has mask" refers to a binary outcome in which the surveyed individual responded to have purchased or been given a mask. Column 2 indicates whether the individual showed a mask to the enumerator. Column 3 indicates whether the individual reported to have used a mask in the last week. 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.17: Correlations: Readings, forecasts, and outdoor time use

	(1) Baseline: hours spent outside	(2) Endline: hours spent outside	(3) Endline: hours spent outside	(4) Endline: hours spent outside
PM2.5 reading	-0.00060			
	(0.0010)			
PM2.5 reading		-0.0054**		-0.0049**
		(0.0023)		(0.0023)
PM2.5 forecast (SMS)			-0.0055	-0.0045
			(0.0037)	(0.0037)
Observations	1007	992	992	992

Notes: Standard errors are reported in parentheses. We regress baseline and endline outdoor time use measures on PM2.5 readings and forecasts on the relevant days. 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.18: 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 NGO	0.050	0.077	0.17

Notes: Standard errors are reported in parentheses. 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 for Columns 1, 2, and 3, respectively). 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.19: 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 NGO	0.85	0.12

Notes: Standard errors are reported in parentheses. "Takes info:" 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. The outcome is defined as 1 if the respondent takes a pamphlet. "Plans to complain:" The outcome is defined as 1 if a respondent intends to file a complaint to the EPD about air quality. 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.20: 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 NGO		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.21: Heterogeneous effects: Baseline overall approval of government v.s. NGO 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 NGO		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 NGO'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 NGO'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 A.22: 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 NGO		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 NGO'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 NGO'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 A.23: 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)	(7.31)
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 NGO		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***.

Table A.24: Heterogeneous effects: Baseline donation to government (as categories)

	(1) WTP	(2) WTP	(3) Forecast error	(4) SMS error	(5) Donation gov't
More to NGO	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 NGO × 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***
		(14.0)	(0.15)	(3.01)	(2.87)
Observations	993	993	993	991	989
Endline mean of NGO		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.25: Heterogeneous effects: Baseline overall approval of information sources (as categories)

	(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 NGO		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 NGO's. The measure is based on Likert-scale questions in which the respondents evaluated their overall approval of the government's and NGO's job in addressing air quality in Lahore. "Approval: Neutral": approves of government as much as the NGO. "Approval: Govt": approves of the government more than the NGO. "Approval: Citizen": approves of the NGO 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.26: Heterogeneous effects: Baseline belief on information accuracy (as categories)

	(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 NGO		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 NGO's air quality readings. The measure is based on Likert-scale questions in which the respondents evaluated their how accurate the government's and NGO's air quality readings are. "Accuracy: Neutral": believes government is as accurate as the NGO. "Accuracy: Govt": believes that the government is more accurate than the NGO. "Accuracy: Citizen": believes that the NGO 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.27: Heterogeneous effects: Baseline forecast error (as categories)

	(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 NGO		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***.

Table A.28: Forecast errors of the individual and ensemble models

	RMSE:All	RMSE:pre	RMSE:during/post	MAD:All	MAD:pre	MAD:during/post
EPD	68.2	103	39.4	47.3	78.6	31
PAQI	67.1	101.7	39.1	46	76.9	30.3
AirNow	66.3	99.4	39.9	46	74.3	31.6
Urban	67.5	101.7	40.1	46.3	76.7	30.8
Sprintars	115	178.6	48.5	75.8	145.7	35
Ensemble (model in SMS)	77.8	111.2	52.7	55.1	83.1	40.5
Ensemble (only using EPD & PAQI)	67.6	102.2	38.9	46.5	77.7	30.3

Notes: Forecast errors against 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. "Ensemble (model in SMS)" is constructed based on a weighted average of all sources listed. "Ensemble (only using EPD & PAQI)" is constructed based on a weighted average of EPD and PAQI forecasts.

Figure A.1: Example of the daily EPD report (December 17, 2022)

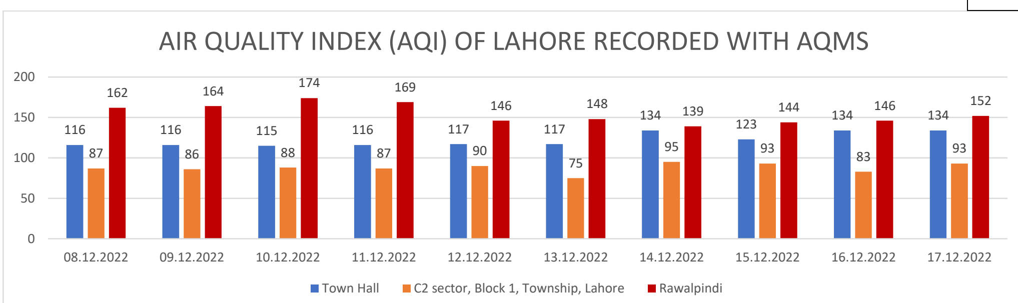
AIR QUALITY INDEX (AQI) OF LAHORE RECORDED WITH AQMS ON 17.12.2022 (BASED ON PREVIOUS 24 HOURS DATA)

Station name	Parameter used to calculate AQI	PEQS value µg/m³	Conc. (µg/m³) used to calculate AQI	AQI
Town Hall, Lahore (Based on 8 hours maximum average value recorded in last 24 hours)	Carbon Monoxide (CO)	**5 (mg/m3)**	6.72	134
C2 sector, Block 1, Township, Lahore	Carbon Monoxide (CO)	**5 (mg/m3)**	4.68	93
Shallay Valley Range Road, Rawalpindi	Particulate matter (PM_{10})	**150**	253	152

AQI limits of EPA, Punjab

0-50	Good
51-100	Satisfactory
101-150	Moderate
151-200	Unhealthy for sensitive group
201-300	Unhealthy
301-400	Very unhealthy
401-500	Hazardous

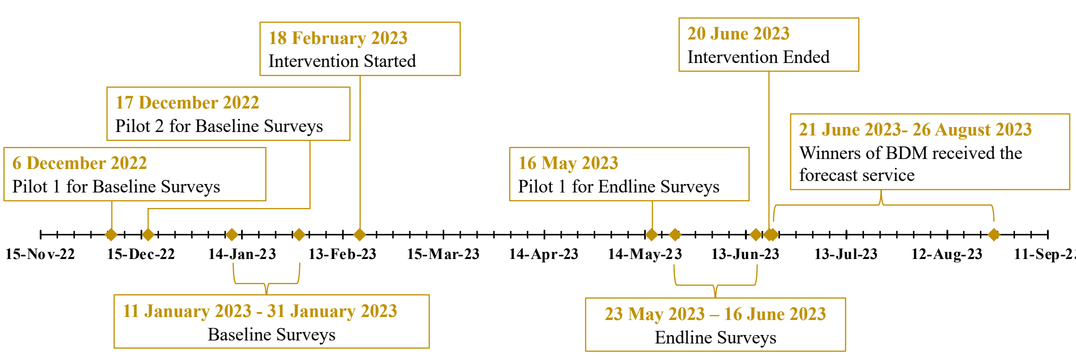
Max.Temp	**7°C - 09°C**
Max.RH	**80% - 90%**
Wind	**NW/05-10 KPH**



Note: i. The report prepared based on instruments generated data and AQI of Lahore calculated based on revised breakpoints of EPA Punjab
ii. The source of meteorological conditions is https://rmcpunjab.pmd.gov.pk/WWW/AirQualityUpdate.php.

Disclaimer: Any other data from any source presenting ambient air quality of any city of Punjab is neither verified nor approved by the EPA Punjab

Figure A.2: Timeline of intervention and surveys



B Additional implementation details

B.1 Additional SMS texts

B.1.1 Introductory message

The following messages were sent to the subjects shortly before the beginning of the SMS intervention, which started on 18 February, 2023.

- 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) which collects data on air pollution. If you have any queries or questions about these messages, please contact the following number [telephone number]."[19]

B.1.2 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."

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

- 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). The data is measured in micrograms per meter cube."

B.2 Survey data

B.2.1 Survey frequency

We conduct the following surveys:

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

B.2.2 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.3 Air quality data

Other than the ones discussed in Section 2.1, we collect air quality reading data from the following sources for the forecast model and the intervention.

A government agency called the Urban Unit owns an air quality monitor but has not been consistently publishing readings for the public's consumption. It 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.

One could also access forecasts based on satellites and meteorological models. 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. However, we are not aware of any satellite- or other model-based services that actively disseminate air quality information for Lahore or Pakistan.

B.4 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 NGO, 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 Specification: heterogeneous treatment effects

D.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:

1. donation share of PKR 100 between government's environmental agency vs. NGO that tackles air pollution
   - For categorical variables, code as "more to government," "more to NGO," 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 NGOs 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 NGO, "NGO-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 NGOs 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 NGO, "NGO-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.

D.2 Estimating equations

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

$$Y_i = \alpha + Z_i\beta + Z_iH_i\theta + H_i\delta + \boldsymbol{X_i\gamma} + \varepsilon_i \tag{3}$$

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:

$$Y_i = \alpha + Z_i\beta + \sum_{j \in J} Z_iH_i\theta_j + \sum_{j \in J} H_i\delta_j + \boldsymbol{X_i\gamma} + \varepsilon_i \tag{4}$$

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.

E Conceptual framework

We specify a consumer's utility function and how information from a given source alters their beliefs and utility. In our model, consumers hold beliefs about the state variable, i.e., the air quality level. They value the forecast information, with which they can take better mitigation measures against air pollution. We express their value of accessing the forecast information from a given source as a utility function.[20] Consumers also hold beliefs about signal quality (i.e., the accuracy of the SMS forecast) about a source, as well as beliefs and preferences about a source that is not tied to the signal itself. Such beliefs and preferences factor into the utility function as attributes. We exogenously vary the source to which we attribute the signal in our experiment.

E.1 Set-up

The state variable over which consumers form beliefs and receive signals is the air quality for the next day, denoted as q_{t+1}. There are two sources of signals for air quality $s \in \{G, P\}$, government and NGO, respectively. The sources send out SMS forecasts, i.e., signals of air quality for day $t+1$ on day t, denoted as ${f_{s,t}}^{t+1}$. We model the consumer's willingness to pay for information from source s when they have received signals up to day t from $a \in \{G, P\}$. a is the information source to which the consumer is already exposed on day t and need not equal s.

On day t, consumer i holds beliefs about the air quality level tomorrow $(E_{i,t}(q_{t+1}))$. Each day, they receive a signal on air quality levels tomorrow, ${f_{s,t}}^{t+1}$, and update their beliefs about the air quality for the next day. Before they receive the signal, they also have beliefs about the signal quality of the SMS forecast for tomorrow $(E_{i,t}({f_{s,t}}^{t+1}))$. Consumers' beliefs may also be updated based on the information they have received up to day t (ι) from source a, which we denote as $z(\iota, a)$.

E.2 Utility function

A consumer gains utility by accessing an air quality forecast for day $t + 1$ from source $s \in \{G, P\}$ on day t. Consumer i's utility, $u^s_{i,a,t}$, is defined as follows:

$$u^s_{i,a,t} = \alpha_i + \beta_t + \delta g(E_{i,t}[f_{s,t+1}], E_{i,t}[q_{t+1}]; z(\iota, a)) + \theta b^s_{i,t}(z(\iota, a)) + \epsilon_{i,s,t} \tag{5}$$

[20]For simplicity, we implicitly assume that the cost of accessing any forecast services is sufficiently high that they would only consume information from one source.

In the utility function, the constant individual term is expressed as α_i. The time-varying term β_t captures the value of information that varies by the day and observable conditions they face. The function $g()$ expresses the consumer's beliefs about signal quality $E_{i,t}[f_{s,t+1}]$ conditional on their belief about air quality $E_{i,t}[q_{t+1}]$ and its accuracy. Furthermore, consumers may have a preference for a source s that is not tied to the signal itself, which we express as $b^s_{i,t}(z(\iota, a))$. Lastly, we include an i.i.d. error term $\epsilon_{i,s,t}$.

E.3 Beliefs about signal quality and updating

In Equation 5, $g()$ is an unspecified function. We introduce an additional structure about $g()$ and the belief-updating process to approximate $g()$ as a linear function of the consumer's own forecast accuracy and their belief about the SMS forecast's signal quality. We elicit both of these measures in incentivized games from the surveys, allowing us to map our conceptual framework to the empirical tests.

We first note that the value of the SMS forecast to a consumer should depend on the accuracy of their own beliefs about air quality without access to the forecast. We define this measure to be $E[|q_{t+1} - E_{i,t}(q_{t+1}; z(\iota, a))|]$. For instance, the value of a signal may be greater for individuals with noisier beliefs without access to the signal. In such a case, we should expect:

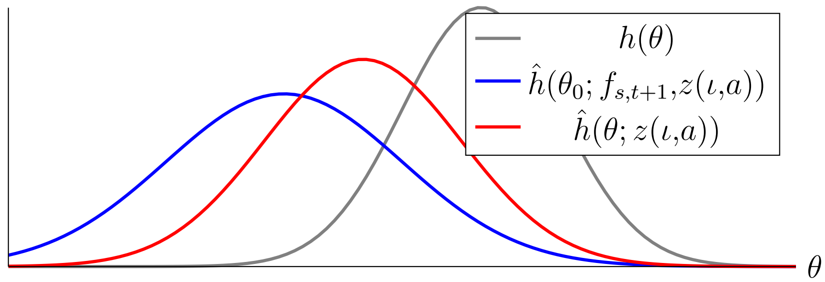
$$\frac{dg}{dE[|q_{t+1} - E_{i,t}(q_{t+1}; z(\iota, a))|]} > 0 \tag{6}$$

We also assume that the value of the SMS forecast to a consumer should depend on the additional signal that they believe the forecast provides, conditional on their own beliefs about air quality without the forecast. We define such a measure to be the following:

$$\begin{aligned} E[|q_{t+1} - E_{i,t}(q_{t+1}; f_{s,t+1}, z(\iota, a))|] - E[|q_{t+1} - E_{i,t}(q_{t+1}; , z(\iota, a))|] \\ = |E_{i,t}(q_{t+1}; f_{s,t+1}, z(\iota, a)) - E_{i,t}(q_{t+1}; z(\iota, a))| \end{aligned} \tag{7}$$

The consumer engages in a Bayesian belief-updating process about the expected value of q_{t+1} based on the signal, $f_{s,t+1}$. Appendix Figure E.1 describes such a process. Suppose that the SMS forecast signal θ is drawn from a Normal distribution of the true air quality level, whose probability density function is $h()$. The priors and posterior beliefs of the distribution are denoted as $\hat{h}(\theta; z(\iota, a))$ and $\hat{h}(\theta; f_{s,t+1}, z(\iota, a))$, respectively. Then the extent to which $\hat{h}()$ shifts toward $h()$ depends on the deviation of $f_{s,t+1}$ from the mean of the prior distribution, as well as the consumer's perception about the signal quality.

Figure E.1: Prior and posterior beliefs of h based on signal $f_{s,t+1}$



As such, we conjecture that the extent to which the beliefs shift in response to the signal is proportional to the difference between the prior belief about the expected air quality and the SMS signal. In other words, the following is true to an approximation.

$$|E_{i,t}(q_{t+1}; f_{s,t+1}, z(\iota, a)) - E_{i,t}(q_{t+1}; z(\iota, a))| \propto |E_{i,t}(f_{s,t+1}; z(\iota, a)) - E_{i,t}(q_{t+1}; z(\iota, a))| \quad (8)$$

Based on the additional structure introduced above, we can rewrite $g()$ as a linear Taylor approximation in Equation 9. In our field experiment, we measure each of the components of the equation by exogenously varying a.

$$\begin{aligned} u^s_{i,a,t} &= \alpha_i + \beta_t + \gamma * |E_{i,t}(f_{s,t+1}; z(\iota, a)) - E_{i,t}(q_{t+1}; z(\iota, a))| \\ &\quad + \omega * E[|q_{t+1} - E_{i,t}(q_{t+1}; z(\iota, a))|] + \theta b^s_{i,t}(z(\iota, a)) + \epsilon_{i,s,t} \end{aligned} \quad (9)$$

The utility function allows us to set up a framework for the empirical exercise we conduct. We map the utility function to the hypotheses in the next subsection.

E.4 Pre-specified hypotheses and their links to the utility function

We define five pre-specified hypotheses based on outcomes 4.1 through 4.4, as defined in Section 4. These hypotheses address whether the willingness to pay measure is greater than zero and whether the four primary outcome variables have different levels between the two treatment arms. We organize pre-specified hypotheses and map them to components of the utility function specified as Equation 9.

The left-hand side variable ($u^G_{i,G,t}$ or $u^P_{i,P,t}$) is estimated via the bidding game using the Becker-DeGroot-Marschak (BDM) method. The bid $v^a_{i,s,t}$ in this game would maximize expected utility if $v^a_{i,s,t} = u^s_{i,a,t}$. As such, we observe distributions of $u^G_{i,G,t}$ and $u^P_{i,P,t}$. We also observe $u^P_{i,G,t}$ and $u^G_{i,P,t}$ through a hypothetical willingness-to-pay survey question in which we asked about the respondent's demand for information coming from a source to which they are not assigned.

We also observe the right-hand-side components of the utility function through survey responses and incentivized elicitation. The consumer's belief about signal quality $E_{i,t}(f_{s,t+1}; z(\iota, a))$ is elicited through the forecast game of SMS forecast and $E_{i,t}(q_{t+1}; z(\iota, a))$ through the forecast game of the air quality level the next day. We observe the realized air quality level of the next day q_{t+1} through air quality readings. We also observe proxies of $b^s_{i,t}(z(\iota, a))$ through stated preferences and beliefs measures about the service quality of each of the sources.[21]

The following are the five pre-specified null hypotheses with links to the primary outcome variables and components of the utility function shown as Equation 9:

1. The demand for air quality information is equal to zero regardless of the treatment assignment group. This hypothesis is tested on outcome 4.1 and equates to $u^G_{i,G,t} = 0$, $u^P_{i,P,t} = 0$ in the conceptual framework.

2. The demand for air quality information is not different between the treatment (NGO) and control (government) groups. This hypothesis is tested on outcome 4.1 and equates to $u^G_{i,G,t} = u^P_{i,P,t}$ in the conceptual framework.

3. Treatment does not differentially affect beliefs about air quality relative to control. This hypothesis is tested on outcome 4.2 and equates to $E[|q_{t+1} - E_{i,t}(q_{t+1}; z(\iota, G))|] = E[|q_{t+1} - E_{i,t}(q_{t+1}; z(\iota, P))|]$ in the conceptual framework.

4. Treatment does not affect the perceived accuracy of the air-quality information source relative to control. This hypothesis is tested on outcome 4.3 and equates to $|E_{i,t}(f_{s,t+1}; z(\iota, G)) - E_{i,t}(q_{t+1}; z(\iota, G))| = |E_{i,t}(f_{s,t+1}; z(\iota, P)) - E_{i,t}(q_{t+1}; z(\iota, P))|$ in the framework.

5. Treatment does not affect relative preferences between information sources. This hypothesis is tested on outcome 4.4 and equates to $b^G_{i,t}(z(\iota, G)) = b^P_{i,t}(z(\iota, G))$ and $b^P_{i,t}(z(\iota, P)) = b^G_{i,t}(z(\iota, P))$ in the conceptual framework.

[21] We do not observe individual component α_i in the data. As such, we average it out between treatment groups, relying on balance in individual characteristics from randomization. We also control for β_t based on the date of the endline survey.

E.5 Mapping the empirical results to the conceptual framework

Our empirical results show that, although the source does not differentially affect the recipients' demand for air quality information, it affects several underlying beliefs and preferences. To put further structure to our findings, we map our empirical results to Equation 9. We follow the links between pre-specified hypotheses and the conceptual framework highlighted in Section E.4 and identify which attributes of the utility function shift in response to making salient the information source.

We start with measures of the demand for air quality information from a given source. Empirical results in Section 6.4 show that there is a positive demand for air quality information, i.e., $u^G_{i,G,t} > 0$, $u^P_{i,P,t} > 0$. However, we fail to reject the null that there is no differential willingness to pay between treatment groups, i.e., $u^G_{i,G,t} = u^P_{i,P,t}$. Yet, it is unclear if and how the treatment affects each attribute of the utility function and to what extent the changes in beliefs matter to the overall utility. In other words, after averaging out individual fixed effects via randomization and controlling for time fixed effects, failing to reject the null of $u^G_{i,G,t} = u^P_{i,P,t}$ implies the following equation:

$$\begin{aligned}&\gamma * |E_{i,t}(f_{G,t+1}; z(\iota, G)) - E_{i,t}(q_{t+1}; z(\iota, G))| + \omega * E[|q_{t+1} - E_{i,t}(q_{t+1}; z(\iota, G))|] + \theta b^G_{i,t}(z(\iota, G)) \\ &= \gamma * |E_{i,t}(f_{P,t+1}; z(\iota, P)) - E_{i,t}(q_{t+1}; z(\iota, P))| + \omega * E[|q_{t+1} - E_{i,t}(q_{t+1}; z(\iota, P))|] + \theta b^P_{i,t}(z(\iota, P))\end{aligned} \tag{10}$$

Equation 10 suggests that, even if the right and left-hand sides are equal, individual components of the equations could be differentially affected by treatment. To highlight this point we estimate correlational relationships between the willingness to pay and other prespecified outcomes by treatment arm, with results shown on Appendix Table E.1. For the NGO arm, we find a negative correlational relationship between the willingness to pay and the SMS error measure, i.e., beliefs about signal quality $|E_{i,t}(f_{P,t+1}; z(\iota, P)) - E_{i,t}(q_{t+1}; z(\iota, P))|$. This relationship, however, is absent for the Government arm. These results are consistent with other empirical results that consumers in the Government arm believe lower signal quality, but do not have lower willingness to pay.

Next, we identify which belief measures in Equation 10 are differentially affected by the information source. Results in Section 6.5 show that there are no differential beliefs about air quality levels between government and NGO sources, i.e., $E[|q_{t+1} - E_{i,t}(q_{t+1}; z(\iota, G))|] = E[|q_{t+1} - E_{i,t}(q_{t+1}; z(\iota, P))|]$. However, results from Section 6.6 show that those assigned to the government arm believe in worse SMS signal quality than those in the NGO arm, i.e.,

$|E_{i,t}(f_{G,t+1}; z(\iota, G)) - E_{i,t}(q_{t+1}; z(\iota, G))| > |E_{i,t}(f_{P,t+1}; z(\iota, P)) - E_{i,t}(q_{t+1}; z(\iota, P))|$.

The results above highlight two possibilities. One is that the effect of differential beliefs about signal quality between two sources is offset by other beliefs, i.e., $\theta b^G_{i,t}(z(\iota, G)) > \theta b^P_{i,t}(z(\iota, P))$, assuming $\gamma < 0$. We do not find evidence of such an offsetting mechanism. Based on stated-preference measures in Tables 3 and 4, we find that the respondents improve their approval of the assigned source in the equal magnitude between treatment arms. Thus, we conclude that another possibility is more likely: respondents do not value the precision of SMS forecasts at the current margin of error.

Lastly, we find that the exogenous attribution to a source increases demand for it relative to other alternative sources driven by improved beliefs about signal quality and other aspects of the information source. Table 2 shows that when we compare within individuals between their assigned source and a hypothetical counterpart, recipients have significantly higher demand for the assigned one, i.e., $u^G_{i,G,t} > u^P_{i,G,t}$ and $u^P_{i,P,t} > u^G_{i,P,t}$. These results imply the following condition for those assigned to the government arm (and symmetrically for those in the NGO one):

$$\begin{aligned} &\gamma * |E_{i,t}(f_{G,t+1}; z(\iota, G)) - E_{i,t}(q_{t+1}; z(\iota, G))| + \theta b^G_{i,t}(z(\iota, G)) \\ &> \gamma * |E_{i,t}(f_{P,t+1}; z(\iota, G)) - E_{i,t}(q_{t+1}; z(\iota, G))| + \theta b^P_{i,t}(z(\iota, G)) \end{aligned} \tag{11}$$

Condition 11 shows that the within-person difference in the valuation of government v.s. NGO sources may come from their belief in signal quality or in other unrelated factors. Unfortunately, we cannot empirically observe $E_{i,t}(f_{P,t+1}; z(\iota, G))$. However, Condition 11 is consistent with the results from Tables 3 and 4, which show that respondents increase their approval of their assigned source in terms of its signal quality, reliability (e.g., promptness of the daily forecast SMS messages), and on their overall satisfaction in similar magnitudes.

Furthermore, we correlationally demonstrate which aspects of Condition 11 lead the inequality, particularly their valuations of signal quality vis-à-vis reliability and other attributes. For respondents assigned to the Government arm, we observe $u^G_{i,G,t}$ and $u^P_{i,G,t}$ based on incentivized and hypothetical willingness-to-pay measures. $u^G_{i,G,t} - u^P_{i,G,t} > 0$ underpins the inequality in Condition 11. We regress $u^G_{i,G,t} - u^P_{i,G,t}$ on the differences in Likert-scale approval of government source to the NGO one in terms of signal quality, reliability, and overall service quality. We also have equivalent measures for those in the NGO arm, i.e., $u^P_{i,P,t}$, $u^G_{i,P,t}$, and the corresponding Likert-scale measures.

Appendix Table E.2 shows the results. We find that, for both Government and NGO arms, the Likert-based measure of reliability is more strongly correlated with the difference in willingness to pay than accuracy or overall satisfaction. These correlational results indicate that consumers update their beliefs about the reliability and timeliness of the assigned source more strongly than about other aspects of the service, leading to an increased demand for the said source.

Table E.1: Relationships between prespecified outcome variables

	(1) WTP: Government	(2) WTP: NGO
Forecast error	6.93*	5.89
	(4.18)	(5.16)
SMS error	0.080	-0.39**
	(0.084)	(0.18)
Donation gov't	-0.24	-0.45
	(0.26)	(0.29)
Observations	494	494

Notes: Model: PDSLASSO. The table shows results of regressing the endline willingness to pay for two months of SMS air quality forecasts on forecast error, SMS error, and share of donations to the government. Column 1 is restricted to those in the Government arm, and Column 2 those in the NGO arm. "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 E.2: Relationships between differences between the sources in willingness to pay and beliefs about service quality

	(1) diff(WTP): Government	(2) dif(WTP): NGO
Diff(Govt - NGO Reliable)	1.97**	-4.91**
	(0.78)	(2.06)
Diff(Govt - NGO Accurate)	1.56	-0.64
	(1.09)	(0.93)
Diff(Govt - NGO Approve)	-1.76	1.62
	(1.07)	(1.28)
Observations	494	492

Notes: Model: PDSLASSO. The table shows results of regressing the differences between sources on the willingness to pay and the differences between sources on beliefs about service quality. Column 1 is restricted to those in the Government arm, and Column 2 those in the NGO arm. "diff(WTP): Government": differences between the elicited willingness to pay for the forecast service that is experimentally associated with the Government and a hypothetical willingness to pay if the information came from the NGO. "diff(WTP): NGO": differences between the elicited willingness to pay for the forecast service that is experimentally associated with the NGO and a hypothetical willingness to pay if the information came from the government. "Diff(Govt - NGO Reliable)": differences in stated beliefs (measured in the Likert scale) that the Government source is reliable and on time, relative to the NGO one. "Diff(Govt - NGO Accurate)": differences in stated beliefs (measured in the Likert scale) that the Government source is accurate, relative to the NGO one. "Diff(Govt - NGO Approve)": differences in overall approval (measured in the Likert scale) of the Government source, relative to the NGO one. Standard errors are reported in parentheses. All regressions include randomization-strata fixed effects, and heteroskedasticity-robust standard errors are used. Two-tailed significance: $p<0.1$*; $p<0.05$**; $p<0.01$***.