# "Press 1 for Roads": Bridging Communication Gaps in Political Representation

Miriam Golden[*]

Saad Gulzar[†]

Luke Sonnet[‡]

November 21, 2024

**Abstract:** To investigate whether lack of information is one reason politicians may be unresponsive to voter preferences, we conduct a randomized control trial with senior politicians in Pakistan that collects preferences using Interactive Voice Response (IVR). IVR allows politicians to script questions for voters and voters to respond using cell phones. There is strong interest by politicians in soliciting opinion via IVR; additionally, relatively high numbers of voters respond. Nonetheless, politicians fail to use the information provided to modify their on-the-ground engagement with voters or types of service delivery. We also do not observe improvements in voter evaluations of politicians or in electoral support for them. A forecasting exercise shows that experts find these outcomes unexpected. One implication is that information may not be the primary determinant of the gap between what voters want and what politicians do.

**Key words:** political responsiveness, political representation, legislators, information and communications technology, randomized control trial, service delivery, Pakistan

---

[*]Corresponding author. Stanford University (`golden@ucla.edu`)

[†]Princeton University (`gulzar@princeton.edu`)

[‡]Independent researcher (`luke.sonnet@gmail.com`)

# 1 Introduction

Studies of political representation of citizens by elected politicians in democratic settings draw on at least two alternative theories: some concentrate on deliberate malfeasance by "bad" politicians and others on informational deficits on the part of either politicians or voters. In the first framework, representation is distorted because rent-seeking types hold elected office but have low intrinsic interest in good representation in the first place (Besley, 2006; Gulzar, 2021). In the second view, politicians are poorly informed about voters' preferences (Bafumi and Herron, 2010; Broockman and Skovron, 2018; Kalla and Porter, 2021; Pereira, 2021; Walgrave et al., 2023) and inadequately constrained because voters are not able to fully observe political performance (Dunning et al., 2019). In this framework, all politicians potentially rent-seek but nonetheless accurately represent voters' preferences under properly functioning informational and electoral constraints (Ferejohn, 1986).

We build on the latter stream of research, which naturally generates the following question: If elected politicians knew what voters wanted, would they deliver more representative policies and government goods and services? To investigate this, we designed a field experiment to directly address politicians' informational deficits regarding citizen preferences. Our experiment partnered with 40 Pakistani provincial-level elected politicians, each representing about 300,000 citizens. We offered 20 of the politicians opportunities to script and record questions for citizens. We disseminated the questions via Interactive Voice Response (IVR) technology as robocalls to cell phones, allowed a representative sample of households to respond, and then presented aggregate-level responses to representatives. Politicians were encouraged to script and record follow-up calls that acknowledged and responded. The intervention offered politicians a novel opportunity to gain vastly more information about the preferences of voting-age adults in their constituencies during the run-up to an election than they were able to do on their own. It offered citizens a novel channel to communicate directly with their elected representative. Our expectations were that this

would result in more favorable evaluations by citizens of their representative, higher levels of political participation by respondents, and an improved ability to evaluate politicians on the basis of performance. We also expected politicians to take advantage of a new and more inclusive channel of communication with constituents by leveraging it with electoral ends in mind.

We find high enthusiasm by politicians for involvement in the exercise: more than twice as many volunteered to participate as we were able to accommodate. This demonstrates that politicians are willing to engage in costly action to receive information about citizen preferences. We also find that using IVR with this citizen population results in much higher uptake rates than reported by other studies that have sought to induce communication by citizens to politicians (see below, Table 2). On these important measures of communication, our intervention was successful. However, the intervention did not change the frequency or type of on-the-ground engagement by politicians with the population or the evaluations of politicians by citizens. On the pre-registered respondent-level outcomes of political support for and evaluation of politicians, the intervention did not produce significant changes.

Our main contribution is to document that although politicians may be eager to receive more information about public opinion, improving information delivery without also improving other aspects of governance may dead-end the process. Contrary to expectations — our expectations as investigators, the expectations of the politicians who volunteered to participate in the intervention, and the expectations of a large panel of experts we consulted (discussed below) — even apparently well-intentioned politicians encounter roadblocks to responsiveness. These previously hidden obstacles emerge when the pre-existing low information environment undergoes improvement: at that point, politicians confront demands they cannot fulfill. Clearing informational bottlenecks should be partnered with enlargements of state capacity.

We make three additional specific contributions in this paper. First, we provide new

*descriptive* data relevant to the underpinnings of democratic accountability in the global south (Auerbach, 2019; Siddiqui, 2022; Allie, 2023; Cheema et al., 2023). It is commonly believed that accountability fails in less developed settings because interactions between politicians and citizens are clientelistic and elite-skewed (Wantchekon, 2003; Stokes, 2005; Larreguy, Marshall and Querubin, 2016; Cruz, Labonne and Querubin, 2017); this parallels research in the developed world showing that politicians are more responsive to the preferences of citizens with higher socio-economic status (Costa, 2017; Broockman and Skovron, 2018; Persson and Sundell, 2024). Using a variety of data collection instruments to evaluate pre-existing citizen-politician interactions, we document that politics is neither as clientelistic nor as elite-skewed as believed. Our data document that although there is substantial demand for individual benefits such as employment, most citizen-initiated requests are for community improvements. In addition these requests come from a representative sample of households, though substantial deficiencies like the frequency of communication and level of political engagement by women remain. We suggest that future research would benefit from more descriptive data about the status quo representative process.

The second specific contribution that we make is what we might call an *operational* one. The accountability literature has deployed numerous experimental interventions to study political communication initiated by voters (Grossman, Humphreys and Sacramone-Lutz, 2014; Chong et al., 2015; Kruks-Wisner, 2018; Arias et al., 2019; Buntaine, Nielson and Skaggs, 2019; Dunning et al., 2019; Bussell, 2019; Grossman, Humphreys and Sacramone-Lutz, 2020). These have often experienced such low take-up that it has been impossible to study politician responsiveness in the aftermath. Instead, we approach the problem from the politicians' side by testing uptake of political communication when initiated by politicians. We document large improvements relative to this literature, allowing us to investigate what, if anything, politicians do with the new information they receive.

Finally, our research has a *technological* component (Bussell, 2012). It shows that

in a low-information setting, new communications technologies, such as IVR, can vastly enlarge the scope and speed with which public authorities receive information from citizens. In the pre-existing low-information status quo setting, citizens are forced to approach politicians in person and the communication process essentially proceeds on a piecemeal basis. Aggregate phone-based communication improves the geographic scope and efficiency of interactions between politicians and the public, accelerating and widening interactions. Since our experiment, the use of IVR has become quite common in contemporary politics in South Asia. We expect that the lessons from our work will travel well to the next generation of studies of new communication technologies, such as smartphone based communication and social media. Assessing how these technologies shape political communication remains an important avenue of research.

Our paper proceeds as follows. We first provide qualitative descriptive information about the context we study and how the literature characterizes the pre-existing representative process. We then detail the experimental design and implementation we used. We then describe how IVR compares along various dimensions with status quo political communication and interaction channels. A next section presents downstream experimental results of our intervention and also results of forecasts to document whether our main results were expected. A final section interprets and concludes.

# 2   Context and Status Quo Political Engagement

Our study is set in Pakistan, a large but understudied lower-middle income democracy. Our intervention augments existing face-to-face interactions between politicians and voters by introducing technologically-based communication through an experiment featuring Interactive Voice Technology. IVR allows politicians to script and record questions for voters; we disseminate the questions as robocalls to voters' cell phones. Voters answer the questions

using the number keys on their phones.[1] We then aggregate responses and present them to the politician, who can follow up with an additional call that acknowledges and responds to what he learns about voter preferences. The follow-up call closes the communication loop by informing voters that they have been heard and their preferences recorded. We conduct the experiment with the expectation that voters will respond to new, unsolicited, high-frequency and two-way contact with greater political support for their representative and also by moving towards performance-based voting criteria (Soo, Weinberg and Dommett, 2020). We expect politicians to respond to the opportunity to communicate directly with large numbers of citizens and to receive aggregated information on their policy preferences by targeting improved service delivery and other visible activities that might improve reelection chances. This expectation draws on theories of distributive politics and political ambition (Mayhew, 1974; Ferejohn, 1974; Golden and Min, 2013).

There were 99 directly-elected, all male Members of the Provincial Assembly (MPAs) in the province of Khyber Pakhtunkhwa (KP) in 2017–18, each representing a single-member district. MPAs assist voters with personal and community problems and control access to the state. In our baseline survey, 55 percent of respondents report that the MPA can get roads fixed in communities and 49 precent say that an MPA can help family members get a job. Thus, voters have confidence in the ability of their provincial representative to provide community and individual assistance, giving them a reason to want to communicate their policy preferences to members of this level of government.

Of the 40 provincial representatives we worked with, 27 had been first elected in 2013.

---

[1]More than 80 percent of Pakistanis have cellphone access, making it feasible to use them for widespread political communication (Pakistan Telecom Authority, 2021). Also, because marketing robocalls are infrequent in Pakistan, households are not already saturated with unsolicited cell phone calls. This makes it likely that households will answer the phone when it rings. Finally, households have no communication channels to reach politicians other than face-to-face interactions, giving them an incentive to use a technologically-based medium. These factors suggest that our setting is suitable for an intervention based on cell-phone communication.

In that election, the reelection rate of incumbent KP MPAs was only 11 percent. Thus, the politicians we partner with were likely to be highly insecure in their offices. This would have given them incentives to improve whatever dimensions of representation they believed could enhance reelection probabilities.[2]

Our respondents were a random sample of (male) heads-of-households; selection and enrollment is described below (see Section 3). Eighty-six percent of them report having voted in the prior general election. Although this figure is surely inflated by recall and social desirability biases, we use the terms "citizens," "voters," "households," and "respondents" interchangeably in what follows.[3] In Pakistan, there is evidence that voters have low party identification (Zaman and Mushtaq, 2022), making it difficult for politicians to distinguish supporters from other adult citizens.

In the status quo representative process, delegations of residents approach their provincial representative and make face-to-face requests for assistance. Precisely for this reason, politicians return at least weekly to their constituency offices where, as one ethnographer described it, "crowds of applicants wait outside to see the politician or a personal assistant in order to get the all-important 'chit' of paper ..." (Wilder, 1999, p. 199). Semi-structured interviews by one of us with more than three dozen MPAs and Members of the National Assembly (MNAs) elicited complaints about the constant need to attend constituency "weddings and funerals" in order to maintain visibility among and trust of electors. When in their home towns, politicians interact with anywhere from 20 to more than a hundred supplicants over the course of a day (Wilder, 1999, p. 199).[4] This is the main mode of

---

[2]Of the 40 politicians we worked with, all but seven ran again in 2018. This corroborates that most of them wished to retain elected office.

[3]Turnout in the prior 2013 elections is reported to be 53 percent (Gallup Pakistan, 2013).

[4]Although the anthropological research that studies on-the-ground politics has generally been conducted in the province of Punjab and our research was conducted instead in Khyber Pahktunkhta, we have no reason to believe there exist substantial differences in village-level interactions between politicians and voters that would make this description inaccurate for KP.

communication between politicians and the people they represent. In the setting where we work, elected officials do not themselves run public opinion polls nor do they have dependable on-the-ground political party operatives or machines feeding them information. Instead, they gather information about citizen opinions via face-to-face interactions.

Given the size of the constituencies (averaging 300,000 persons), MPAs are personally acquainted with only a fraction of their electors. They have no means to communicate directly with large numbers of voters except via face-to-face interactions that require them to travel to the numerous settlements in their home constituencies. For the MPAs, enrolling in the IVR experiment offered a way to gain potentially valuable information about public opinion that they did not possess. It also offered a way to communicate at a much larger scale with constituents, and in particular to communicate with constituents who did not approach them in person.

The literature reports that direct contact between Pakistani voters and politicians are skewed towards higher-income male co-partisans (Martin, 2014; Liaqat, Cheema and Mohmand, 2020). In this political culture, "leaders redirect public resources to benefit kin, friends and clients" (Martin, 2016, p. 67). In a setting that anthropologists characterize as one of asymmetric factional hierarchies based on status and power (Barth, 1965; Lyon, 2004), politicians seek to assemble large clienteles but, as our semi-structured interviews in KP repeatedly revealed, limited resources mean politicians are unable to provide assistance to most who request it. Since politicians thus make discretionary allocative decisions, other scholars report that "those who are richer and part of the village elite have greater bargaining power vis-à-vis leaders than poorer, non-elite members ..." (Mohmand, 2019, p. 24). The standard assumption is that access to politicians is dependent on income, gender, and partisanship.

# 3    Experimental Design and Implementation

**Sources of data**   Our study uses mixed methods, analyzing information that was collected via nine separate instruments, enumerated in Table 1. The table also reports the source of data for all the figures and tables in the paper. In early 2017 we undertook a (a) *pilot* survey with a single MPA to demonstrate the project's operational feasibility; the take-up rate for the pilot was extremely high, with 31 percent of respondents answering at least one IVR question, although results were statistically insignificant for changes in respondent opinions (Golden, Gulzar and Sonnet, 2017). In late 2017, we distributed a (b) *baseline* survey to enroll 14,400 voters in the study. We undertook the IVR experiment itself in 2018 and it generated (c) *intervention* data on content, take-up, and response rates. In 2019, we conducted an (d) *endline* survey to collect information on outcomes. In 2020, we conducted a (e) *descriptive* telephone survey of 3,600 respondents, subsampled from the initial 14,400, from whom we collected information on preexisting (status quo) political communication.[5] During the intervention, we surveyed 240 (f) *key informants* across the full 20 treated constituencies to collect information about politician behavior. We also merged household-level data with 2019 polling station level (g) *election* data that we assembled. In 2019, we also conducted lengthy face-to-face semi-structured (h) *interviews* with approximately three dozen MPAs and MNAs from KP. Finally, in 2019 we surveyed 400 students, academics, and policy implementors in advanced countries and in Pakistan in a (i) *forecasting* exercise that asked respondents to tell us whether they thought the intervention would be successful, using data from the 2017 pilot.

Our data permit two separate levels of analysis. Given the large sample size of individual voters, we are able to conduct quantitative analyses of responses to questions for voters both in treatment and in control both before and after the intervention. With only 20

---

[5]We had hoped to observe face-to-face interactions between MPAs and citizens but the Covid-19 pandemic made that infeasible except for a single day-long pilot observation. We used that to inform the questions in the descriptive survey instrument.

Table 1: Datasets collected and used for each table and figure in the paper

| Dataset | Sample size | Notes | Tables and figures |
|---|---|---|---|
| **Respondent level** | | | |
| *Pilot* survey (2017) | 1,218 HHs | Conducted with one MPA. | Data used in forecasting exercise presented in Figure 7 |
| *Baseline* survey (2017–2018) | 14,399 HHs | Every respondent baselined (except for one duplicate phone number). Some tables use only a subset (e.g. {H2}) when appropriate. Often IVR compliance statistics (e.g. answering the IVR question) come from this data as it is available regardless of whether we endline. | Table I.2; Figures 2 and 6 |
| *Intervention* data (2018) | 1,247 HH | Respondents who answered the phone and answered a question. | Table D.1 |
| *Endline* survey (2018) | 13,988 HHs | Every respondent we could reach for the endline. Note that some tables that use this data use only a subset (e.g. {H2}) when appropriate and when noted in the table/table notes. | Tables 3, G.1, G.2, G.3, I.1, I.3, and I.4; Figure H.1 |
| *Descriptive* survey (2020) | 2,863 HHs | Those we could reach via phone out of 3,600 HHs randomly subsampled from the experimental group assigned to {H2}. | Figures 3, 4, 5, and 6. |
| **Polling station area level** | | | |
| *Key informant* survey (2018) | 240 PS areas | PS level data largely from mean responses of two key informants in each of the 6 matched pairs of treatment and control polling station areas per MPA. | Table J.1 |
| *Electoral* data (2019) | 300 PS areas | The max set of polling stations at which electoral data available (larger than key informant data due to cost of interviewing key informants). | Table K.1 |
| **Politician/constituency level** | | | |
| *Interviews* (2019) | 35 KP MPAs and MNAs | Hour-long face-to-face semi-structured open-ended interviews. | Paper text |
| **Other** | | | |
| Expert *forecasts* (2019) | 400 respondents | Solicited in-person and over email. Includes undergraduate students, research/development practitioners, and academics around the world. Information provided for forecasts drew on results of pilot study. | Figure 7 |

MPAs enrolled in the IVR experiment, we conduct qualitative analyses of their involvement in and reactions to the intervention. Information at the level of the MPAs draws on data gathered from home constituencies during the intervention and also on post-intervention in-depth interviews.
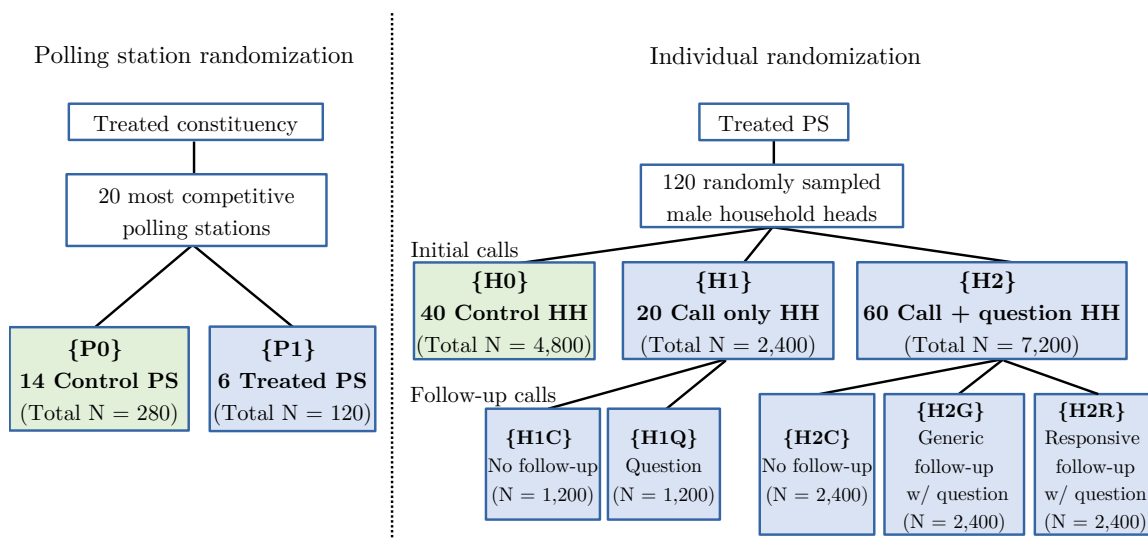
**Content of IVR**  IVR calls are made in two stages. Stage one contains two sections. The MPA first introduces himself and credit-claims for recent activities and, second, he asks a question seeking feedback. A sample script is available in Appendix D. The intervention bundles political advertising and credit-claiming with preference elicitation. This bundling parallels pre-existing face-to-face interactions in the setting where we work.

The research team collated responses to the MPA's question and provided him aggregate demographics from the baseline survey as well as aggregate IVR responses. This provides the MPA with information about how respondents evaluate him and his political party, among other things. In a low-information setting without political polling, this information is new for the MPA.

In the second stage of IVR calls, the MPA records a follow-up call where he reports what he learned in the first stage and how he intends to proceed. He also records a second question, to which voters respond. Details on the timeline appear in Appendix C.

**Sampling and randomization**  Of the 99 directly-elected MPAs in the KP assembly, our implementing partner identified an initial 47 who expressed interest in working with IVR to interact with citizens. We randomly select 40 of these; the 40 MPAs generally come from slightly more competitive and urban constituencies than other MPAs. Blocking on political party, we randomly select 20 of the 40 MPAs to enroll into treatment and thus give access to the IVR technology through our research team. The other 20 are put into control. Figure 1 provides a diagram of the randomization process at both the polling station and household levels.

Figure 1: **Experimental design at the polling station and household levels**



The MPAs in our sample are elected from constituencies with between 50 and 294 precincts, which in Pakistan are commonly referred to as polling stations (PS). The median number of registered voters per KP polling station is $1,333$; as we have noted, each MPA thus represents an average of 300,000 people. As a comparative benchmark, we note that this is 25 percent smaller than the average size of a district that elects a representative to the California state legislature, which is about 400,000 persons.

Within each treated constituency, we select the 20 polling stations with the smallest absolute margin of victory of the incumbent MPA. We randomize six of the 20 into treatment {P1} and 14 into control {P0} in a two-step process. We first randomly sample 12 polling stations out of the 20 and then, for those 12 polling stations, create matched pairs using a Mahalanobis distance score that incorporates the total number of registered voters and raw vote totals for large parties.[6] We assign one of each pair to treatment and the other to control.

The right panel of Figure 1 shows household-level randomization. A random walk sampled 120 households within each treated PS area, starting at a central location in the

---

[6]We define a large party as any party that received more than 100 votes in any of the 12 PSs or that received an average of 20 votes across all 12 PSs in the prior (2013) elections.

PS catchment area. We enroll male heads-of-household only, owing to obstacles collecting phone numbers of women.[7] Our total sample is 14,400 male heads of household. Because the last national census conducted by the Government of Pakistan that was available when we undertook our intervention dated to 1998, we have no data frame that would allow us to know how representative our sample is. However, the descriptive statistics that characterize it (see Figure 6) seem reasonable given what we know about the province.

We survey enrolled participants once before (baseline) and once after (endline) the intervention; in order to deliver the intervention, we collect phone numbers during the baseline. Blocking on co-partisanship with the MPA, 40 of the 120 PS-level households are placed into a control condition and receive no contact other than the surveys {H0}, 20 receive an IVR call with a credit-claiming message but no question(s) {H1}, and 60 receive a credit-claiming message as well as an IVR question(s) {H2}. Total sample sizes are reflected by the $N$ values reported in each box.

A second stage splits those who receive first-stage calls to either receive no follow-up call or receive a follow-up call containing specific components. Respondents in {H1C} and {H2C} receive no follow-up call, while all other respondents receive a follow-up call in which the MPA asks a new question via IVR {H1Q, H2G, H2R}. Respondents in {H1Q} receive only the new IVR question, while respondents in {H2G} and {H2R} also receive a *generic* or *responsive* message, respectively, where the MPA acknowledges the first stage IVR question. The *responsive* message details how the MPA will act based on the information collected in the initial robocall in which he asked for constituent input while the *generic* message simply thanks respondents for their input. The main goal of these randomizations is to estimate the total effect of the most interactive and deepest IVR communication we could generate {H2R} as well as marginal effects of call components.

---

[7]Using female enumerators, ninety percent of the women we asked refused to provide their phone numbers or did not have regular access to a phone. The exclusion of women may raise ethical issues. These are discussed in Appendix B.

# 4    Communication Transformations Offered by IVR

We first report descriptive results, comparing patterns of status quo communication with those elicited by IVR technology. In the status quo, as we have noted, citizens meet face-to-face with their provincial representatives, often traveling in small delegations to do so. Thus, we comparing these interactions with the new communication channel that we established using phone calls.

**Uptake** Under IVR, 73 percent of respondents answer the phone. The proportion of respondents who answered at least one question is 17.3 percent. Compared to other ICT interactions and compared to our *descriptive* data on rates of face-to-face interactions between politicians and voters, this rate is high, as documented in Table 2.

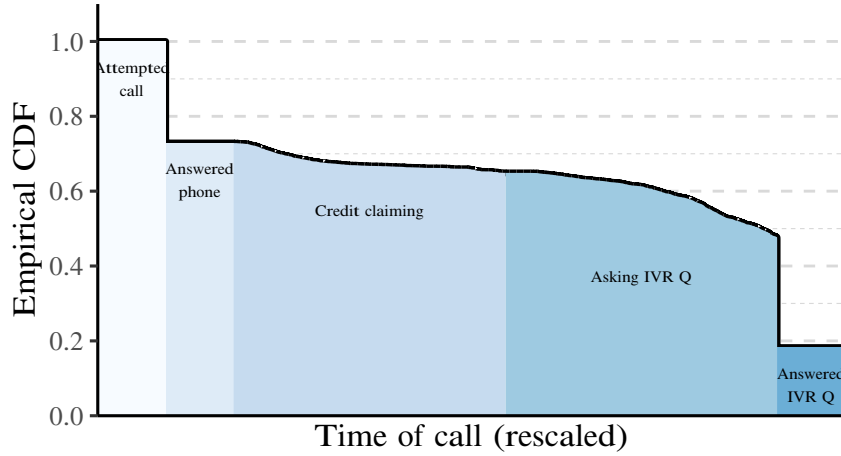Table 2: Take-up Rates and Modes of Communication Across Studies

| Study Authors | Country | Mode of Communication | Take-up Rate (%) |
|---|---|---|---|
| Grossman, Humphreys and Sacramone-Lutz (2014) | Uganda | SMS | 5.8 |
| Leo et al. (April 2015) | Various | IVR | 4.3 |
| Ferrali et al. (2016) | Uganda | SMS | 2–7 |
| Grossman, Michelitch and Santamaria (2017) | Uganda | SMS | 4.8 |
| Erlich et al. (2018) | South Africa | SMS+ | 2–14 |
| Blair, Littman and Paluck (2019) | Nigeria | SMS | 0.1 |
| Buntaine, Hunnicutt and Komakech (2020) | Uganda | SMS | 10 |
| Grossman, Humphreys and Sacramone-Lutz (2020) | Uganda | SMS | <0.1 |
| Shaul-Cohen and Lev-On (2020) | Israel | SMS | 4–18 |

*Notes:* SMS refers to Short Messaging Service. SMS+ refers to a 5-channel study.

**Dropoff** While 73 percent of respondents answer the phone, about half are still on the call when the MPA asks a question. The biggest drop off occurs after the question is asked but before respondents answer, as is shown in Figure 2.

**Scope** In Figure 3, we depict the distribution of in-person interactions and IVR calls across polling stations (left) and constituencies (right). In substantial numbers of polling stations, no one has met his MPA in person in the last year. This number goes to almost zero
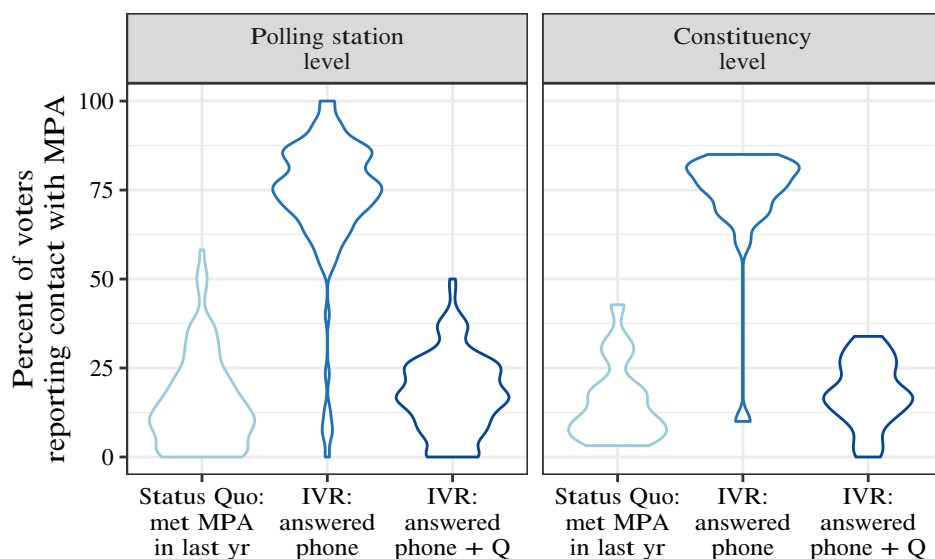
13

Figure 2: Dropoff during IVR call



*Notes:* Proportion of respondents who remain on the IVR call at various stages of the call. The x-axis is rescaled for illustrative purposes to account for differing call lengths across MPAs.

for IVR contact, defined as answering the phone; IVR thus provides complete geographic penetration, including to remote areas, where residents are typically denied direct political access to their MPA. Results are similar if we restrict attention to only IVR respondents who answer a question or if we consider constituency-level differences instead of polling station-level differences. Finally, we also document that citizens who live further from the MPA's constituency office are disadvantaged by the need to travel to face-to-face meetings (see Table E.1). We do not observe the same in the case of IVR contact.
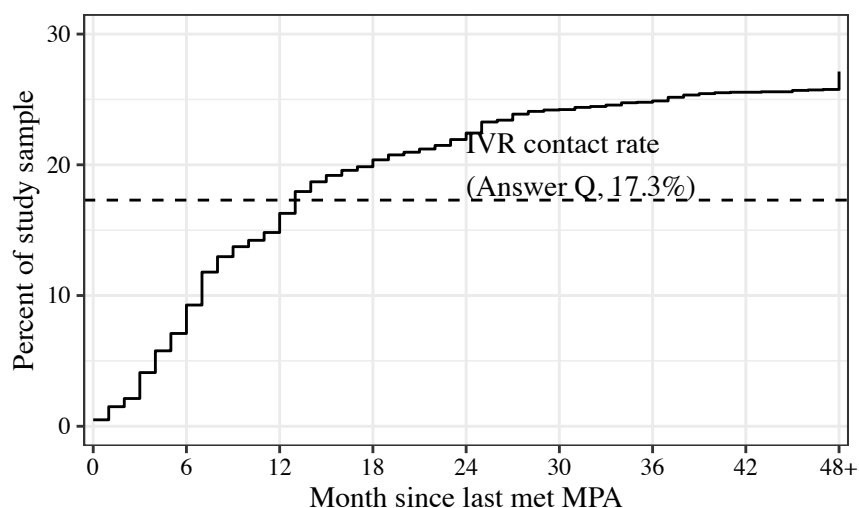
**Scale**   Next, we analyze the total number of constituents who interact with their elected representative in any given period. Only about a third of all households report *ever* having met their MPA in person; we, therefore, take this as the upper-limit on face-to-face interactions over *any* period of time. IVR is not only able to more than double the rate of interaction in just a single round of calls, it also establishes as much two-way interaction in one round of calls as occurs face-to-face over a full year, as we show in Figure 4.

Figure 3: Distribution of contact rates across polling stations (left) and constituencies (right)



*Notes:* Distribution of contact between politicians and voters. For example, in the left panel, polling stations with no reported contact lie at zero and those where everyone reports contact with their MPA lie at 100.
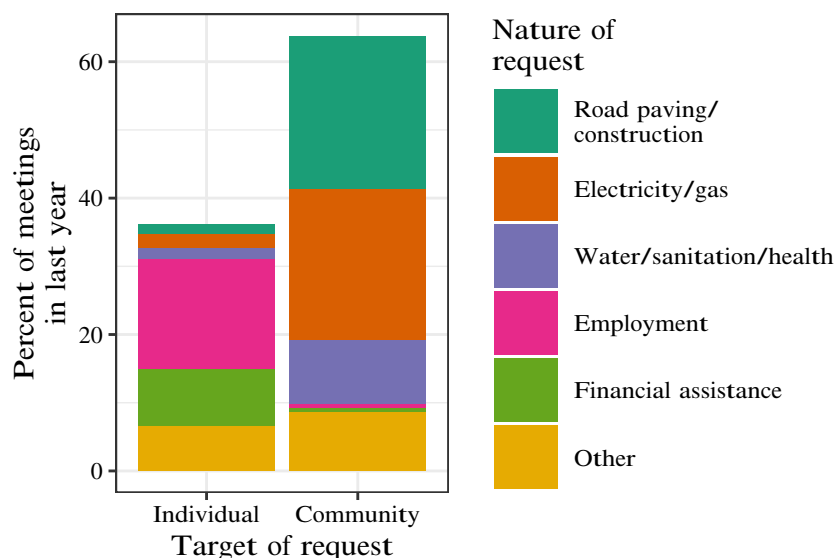
Figure 4: Descriptive evidence on scale of communication transformations under IVR



*Notes:* The solid (increasing) line plots the cumulative distribution of time since survey respondents last met their MPA in person. The dashed (flat) line is the average rate of contact under IVR (where contact is defined as having answered the question).

**Content**   Figure 5 examines what kinds of requests individuals make when meeting their MPA face-to-face in the last year. We bin requests into two "targets:" requests that relate to the individual or their household (labeled "individual"), and those that relate to the broader community. Two-thirds of requests seek community improvements. The data thus show that even in face-to-face interactions, voters more frequently request basic public infrastructure than individual clientelistic goods. However, face-to-face interactions almost never involve discussions of public policy and legislation. This is in contrast to IVR; Table D.1 shows the distribution of answers of those responding to the first-stage IVR question. A quarter of first-stage IVR respondents suggest their representative prioritize legislation. Three-quarters convey preferences for development goods (roads, electricity, piped water) or for improvements in health and education; the latter is especially important for respondents residing in rural areas, where basic facilities are often absent or inadequate (e.g. schools without lavatories).

Figure 5: Target and nature of requests made to MPA in face-to-face meetings



*Notes:* Data are subset to respondents who report having met their MPA in the last year. Targets of requests are coded by whether they pertain to the individual and his household (labeled "individual") or whether they pertain to the community.
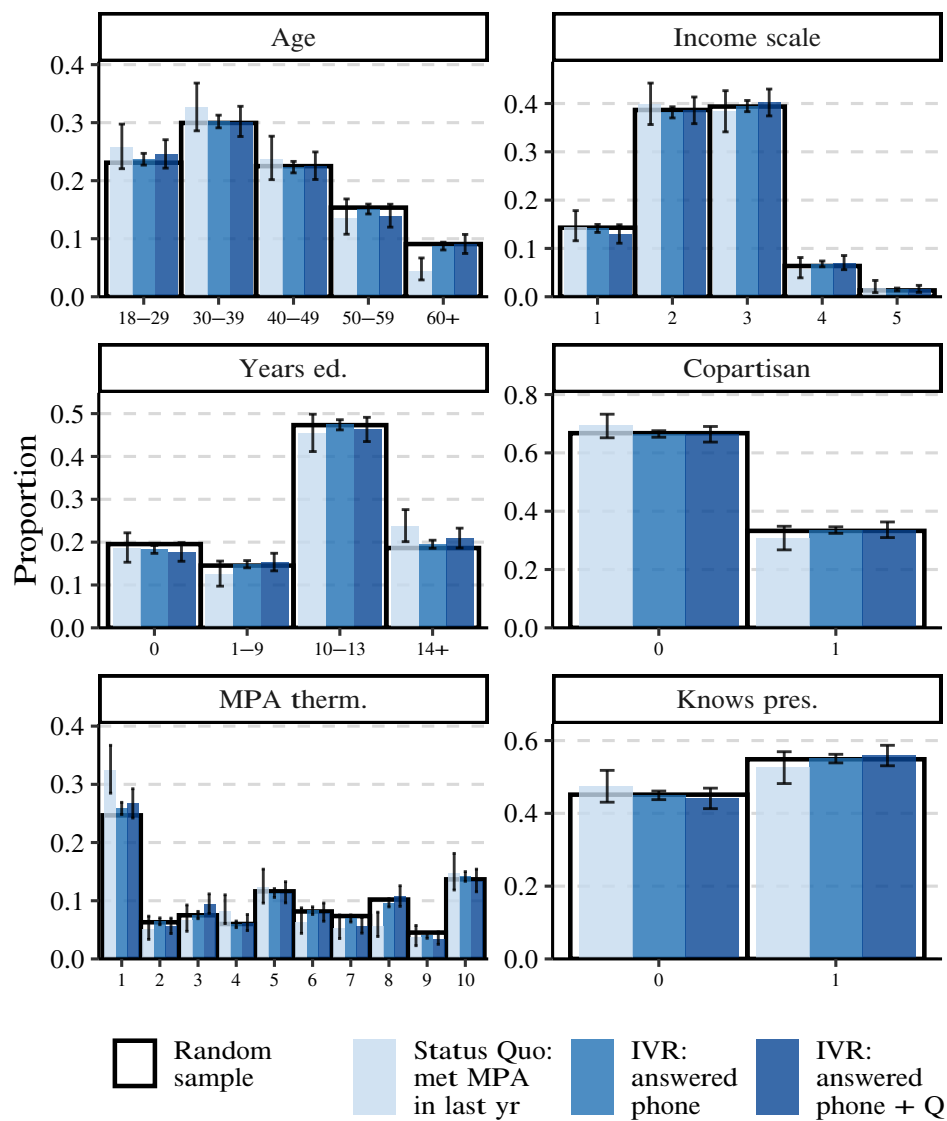
**Representativeness** We also consider which voters MPAs are able to reach using IVR across three groups, displayed in Figure 6: (a) people who report having met their MPA in person in the last year; (b) randomly sampled respondents enrolled in IVR who answer a question; and (c) randomly sampled respondents enrolled in IVR. These data show that both modes of communication reach constituents who are representative along various dimensions, including demographic (age), socio-economic status (income and education), partisanship (co-partisan and MPA thermometer scale), and political knowledge (whether the respondent accurately identifies Pakistan's president).[8] Face-to-face interactions are not elite-biased, and IVR replicates the pattern of reaching a representative sample. In particular, both modalities reach equal proportions of respondents as regards income and education, and politicians interact with more respondents who are not their supporters as they do with co-partisans.

# 5   Downstream Results of IVR

Results reported in the previous section show that IVR improves four specific dimensions of interaction between politicians and constituents: uptake, geographic scope, content, and scale. Next we examine whether IVR affects attitudes, self-reported political behavior, and politician service delivery. Except as noted below and detailed in Appendix A, outcome measures, equation specifications, and treatment effects reported were all pre-registered. Appendix H discusses power considerations in interpreting results.

---

[8]Both methods also are unsuccessful at reaching women as independent voters in the household (Cheema et al., 2023).

Figure 6: Constituent characteristics by mode of contact

*Notes:* Box plots include 95% confidence intervals.

## 5.1 Effects on individual attitudes and behavior

**Outcomes:** We have 13,988 individuals in our *endline* dataset which, combined with the *baseline* dataset, constitute the data analyzed in this section.[9] At the individual level, we focus on three main outcome indices from survey responses, measuring *evaluations of the incumbent*, *political participation*, and *prospects for electoral accountability*.[10]

**Estimation:** We estimate effects on the indices and constituent outcomes with:

$$Y_{hpm(t=1)} = \tau D_h + \alpha Y_{hpm(t=0)} + \beta' \mathbf{H}_{hpm(t=0)} + \lambda_p + \epsilon_h,$$

where $Y_{hpm(t=1)}$ is the outcome $Y$ for household $h$ in polling station $p$ at endline ($t = 1$), $D_h$ is a binary indicator for treatment status, $Y_{hpm(t=0)}$ is the pre-treatment outcome $Y$ collected in the baseline (if available), and $\lambda_p$ is a polling station fixed effect. The vector of household variables, $\mathbf{H}_{hpm(t=0)}$, is a set of pre-treatment variables selected from *age by decade, an income scale, education bins, an index of political knowledge, a set of indicators for the party the respondent supports, a binary measure of reported turnout in 2013, and a binary measure of support for the MPA's party* if they predict the outcome in the control group. We use heteroskedasticity-consistent standard errors, since treatments of interest are assigned at the household level.

**Individual level effects:** Table 3 shows results. The first column presents the standardized control mean, which is zero. The second set of columns presents the treatment effect of any IVR call and sample size. The absolute values of all treatment effects are smaller than 0.02 standard deviations, and all are statistically indistinguishable from zero. We find

---

[9]We were unable to recontact 411 individuals at endline. Results are similar with inverse probability of attrition weights in Table G.1.

[10]Details regarding indices construction appear in Appendix F

no effects of receiving any kind of IVR call on individual attitudes for any index, their sub-components (see Appendix I), or when we analyze effects among compliers (see Table G.3). In the third set of columns, we consider the effect of the fully responsive IVR intervention, whose respondents received an initial call with a question and a subsequent follow-up acknowledging first-stage responses. Although we did not pre-register this comparison, we note it because it is the deepest use of IVR and we have the greatest expectation of positive treatment effects. Nonetheless, we do not observe any treatment effect. Other results analyzing differences between various treatment arms are reported in Appendix I.

Table 3: **Effects of any IVR call and effects of full IVR treatment on household head outcomes**

|  | Control mean: no call | ITT: any call | | ITT: full responsive treatment | |
|  | {H0} | {H1, H2} vs. {H0} | | {H2R} vs. {H0} | |
| Outcome indices | $\mu$ | $\tau$ | N | $\tau$ | N |
| **Incumbent evaluations index** | 0.000 | -0.009 | 13757 | -0.016 | 6539 |
|  | (1.000) | (0.009) | | (0.013) | |
| **Political participation index** | 0.000 | -0.020 | 13780 | 0.004 | 6551 |
|  | (1.000) | (0.016) | | (0.025) | |
| **Prospects for accountability index** | 0.000 | 0.004 | 13759 | 0.025 | 6539 |
|  | (1.000) | (0.017) | | (0.026) | |

*Notes:* [†], p-value $< 0.1$; [*], p-value $< 0.05$; [**], p-value $< 0.01$; [***], p-value $< 0.001$. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses.

Pre-treatment control variables not displayed; see text for details. Because our preferred specification includes pre-treatment covariates and the baseline measure of the outcome may have some missingness and because there is some missingness on the outcomes themselves, the sample sizes in the tables do not represent the full $13,988$ individuals from whom we collect both baseline and endline data.

The letters in braces refer to the experimental groups described in Figure 1.

## 5.2 Effects on aggregate electoral outcomes

Our design allows identifying the effects of saturating one polling station area with IVR calls while leaving other areas untouched. This allows us to estimate whether respondents reward or punish their MPA for communicating via IVR with dozens of households in a polling station area, as well as whether increased communication changes the turnout rate at the polling station level. Again, we do not observe any treatment effects (see Appendix K).

Unsurprisingly, given the lack of individual-level effects, we find no evidence of intervention effects on vote shares for our partner MPA (or his party) or for turnout.[11]

## 5.3    Effects on politician behavior

We next evaluate whether IVR involvement affects the behavior of politicians. We examine the kinds of messages they craft as well as whether they direct more attention, spending, or public services to treated localities. We analyze information that comes from interviews with *key informants* in treated localities. For details, see Appendix J.

We allowed partner MPAs to craft their own IVR messages. We anticipated that this would permit communications tailored to treated localities. However, politicians did not craft specific messages. The modal question (see Section 3) was, by our reading, general and imprecise. Indeed, 16 of 20 partner MPAs asked identical questions, recycling a prompt provided by our staff. That does not mean that politicians did not invest in the experiment, however. The general pattern is consistent with the interpretation that politicians are experimenting with a new technology to see for themselves how useful it might be for them. Despite limited time and resources, they were willing to engage with their constituents using a new mode of communication.

The politicians enrolled in our study interacted repeatedly with us in the nine months preceding an election. This might have encouraged them to improve service delivery in the polling station areas targeted for calls or perhaps to visit these locations more often to augment the political visibility offered by the calls. However, a quantitative evaluation of key informant data shows no effect on MPA effort (see Appendix J.)

---

[11]We note for the record that even in polling station areas with the highest saturation of calls, a sample size of 120 respondents would have required massive spillover to the more than 1,100 other voters to produce electoral effects visible at the polling station level, let alone across the 50 to 294 polling station areas that comprise a constituency. The intervention was not designed at a scale to potentially affect actual election outcomes.

## 5.4 Evidence of voter engagement

Despite null effects of treatment on voter attitudes and behavior, some components of the intervention successfully affected *voter engagement*. The data shows that when citizens receive a call from their MPA that includes a question — as opposed to an exclusively credit-claiming call — they are more likely to answer a subsequent call from the same politician (see Table I.2). This suggests that citizens appreciate being included in a policy discussions with a politician and would like to continue the conversation. Mere communication alone, however, is insufficient to change voting outcomes (see Table I.3).

## 5.5 What did we expect? A forecasting exercise

Were the results we observe obvious ex ante? We conduct forecasting exercises that poll potential policy consumers of this research in Pakistan and academics in the United States and Europe, asking them about their beliefs about the experimental impacts without seeing the results. To do so, we provide information on the pilot study, including the fact that the pilot did not result in statistically significant changes in respondents' evaluations of their MPAs. Nonetheless, the 400 forecasters significantly and substantively overestimate treatment effects across our three downstream respondent-level outcomes (Ahrenshop et al., 2023).[12]

The aim of the forecasting exercise is not to get a representative set of forecasters but rather to capture the ex ante beliefs of those who express an interest in the research by volunteering their time. The exercise follows DellaVigna and Pope (2018), which documents the utility of using expert forecasts to understand what we can learn from experiments. We assume interest and relative expertise among such volunteers.

---

[12]Forecasts were incentivized: respondents whose responses were closest to the truth received small Amazon giftcards.

We analyze results from two sets of forecasts: the first with 283 (out of 1,584 solicited) academic experts in or associated with the United States[13] and the second with 117 (out of 172 solicited) university students, faculty, and policy practitioners in Pakistan, for an overall total of 400 responses. The two waves allow examination of differences between local and global expertise (Casey et al., 2023). After briefly describing the context and the experiment, we ask respondents to forecast the take-up rates for the IVR calls as well as impacts on downstream results on the three key voter indices analyzed above. Details are available in (Ahrenshop et al., 2023).

**Results:** Panel A in Figure 7 shows that, on average, nearly 40 percent more people answer the phone than forecasters predict. However, conditional on whether the IVR call was answered, forecasters were more likely to think that respondents would answer an IVR question than the number that actually did. That is, forecasters underestimate whether respondents would answer the phone but overestimate whether respondents would answer an IVR question.
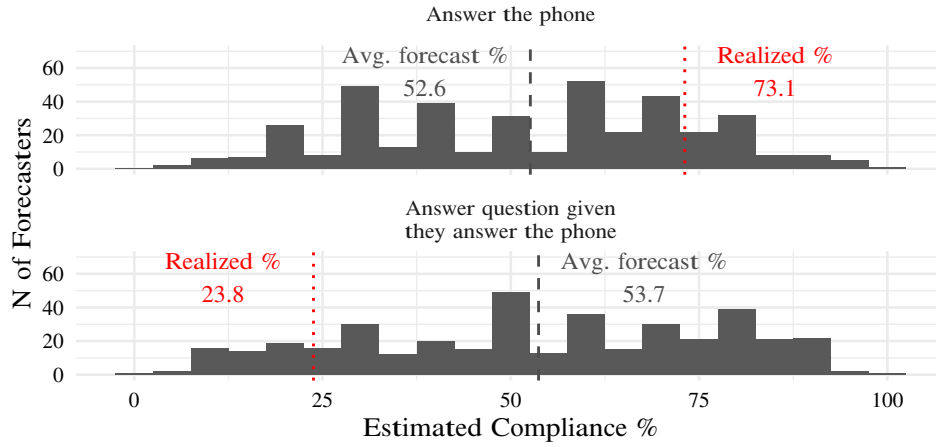
Panel B contrasts the forecasts against realized results on the three downstream outcome indices. Forecasters were asked to predict the intent-to-treat effect for each index. Forecasters overestimate the size of all effects: in every case, average forecasts were substantially (and statistically significantly) above realized average treatment effects. In other words, forecasters expect the intervention to produce relatively large results.

Of course, forecasts of field experiments might generally be overly optimistic. One reason lies with publication bias: almost all field experiments that get published report successful interventions. Those that produce null results are typically difficult or even impossible to publish (Christensen, Freese and Miguel, 2019). A second reason is that interested academics and policy experts might have a kind of intellectual vested interest in seeing in-
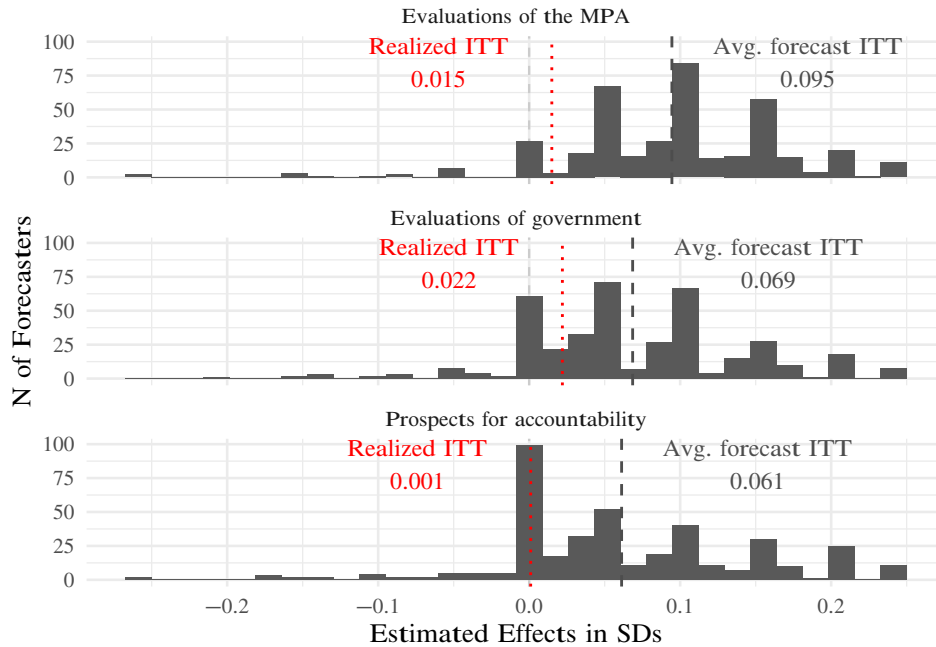
---

[13]Academics "associated" with the United States were members of the American Political Science Association's Organized Section in Comparative Politics.

Figure 7: Forecast and realized compliance and treatment effects

**Panel A: Forecasts of Compliance**



**Panel B: Forecasts of Treatment Effects**



*Notes:* All differences between the average forecast and realized estimates are statistically significant with p<0.001.

terventions produce results ███████████████████████.

Perhaps for both or other reasons, forecasting shows that the results of the intervention we report were not predicted by interested experts. Experts expect take-up to be less than was the case and they expect downstream results to be larger than those obtained.

These differences show the intervention produced unanticipated results. The main interpretation that we stress is that there were high expectations among research consumers of the intervention in advance of knowing the results. While after-the-fact reasoning may claim the null results were to be expected, the forecasting results show that they were not.

# 6  Interpretation and Conclusions

The experiment we report was embraced by senior politicians who were Members of the Khyber Pakhtunkhwa Provincial Assembly in Pakistan. Both academics and policy practitioners expected it to succeed. More MPAs volunteered to partner with us than we could accommodate not just in engagement, but also change in voter behavior. Yet MPAs failed to redirect efforts towards areas where voters received IVR communication, or to otherwise make observable efforts to leverage the intervention for electoral goals. On the other side, the evaluations that citizen respondents made of their representatives did not improve. What happened?

One hint comes from the failure of MPAs to make repeated IVR recordings in period of the intervention, which covered the seven months preceding joint provincial and national elections. The research design was meant to leverage reelection aspirations of enrolled MPAs.[14] In the end, however, most MPAs recorded only a single question and a single follow-up response.

To understand why this occurred, we conducted open-ended follow-up interviews with 16 of the 20 treated MPAs, as well as another 20 MPAs and Members of the National Assembly. We asked open-ended questions about the IVR intervention and about political representation more broadly. MPAs reported that the information they gained when we delivered aggregate feedback from voters was politically helpful because it revealed the con-

---

[14]Appendix B discusses the ethics of working with incumbents alone.

stituency areas where they had more or less electoral support.[15] This verifies that MPAs considered the information useful for electoral goals. Many claimed that they lacked existing methods to collect this information.[16] Yet they also reported that voters tended to misinterpret the IVR questions as *commitments* by the MPA to provide new infrastructure.[17] When they went back to their village offices after the IVR calls, some MPAs confronted angry voters wanting to know when they would make good on their promises.[18] As a result, despite their initial enthusiasm, MPAs disengaged from the experiment and did not record additional messages. Asking citizens what they wanted only raised expectations without providing additional resources or capacity with which to satisfy them — and this in an environment where MPAs are unable to satisfy the existing requests that they receive in person.[19] The experiment thus generated misaligned expectations: respondents who answered a question deemed their representative unreliable due to his inability to follow up on their demands. It is perhaps unsurprising, given this, that respondents who engaged with their representative via IVR may have lowered their evaluations of him (see Table 3).

That the intervention generated a bottleneck indicates the types of dynamics that could potentially trigger a new, virtuous cycle of electoral accountability and improved performance. Since IVR communication raised expectations among some voters, this could potentially prod politicians to reconsider their use of existing resources. However, using political communication to break out of the current equilibrium characterized by low expectations and poor responsiveness might require more extensive programming.

---

[15]For instance, respondents 18, 19, 20, 21, and 22, group interview 25 April 2019.

[16]Reported for instance by respondent 27, interviewed 26 April 2019; respondent 26, interviewed 26 April 2019; respondent 11, interviewed 24 April 2019.

[17]Respondents 18, 19, 20, 21, and 22, group interview 25 April 2019.

[18]Respondents 18, 19, 20, 21, and 22, group interview 25 April 2019.

[19]A chronic inability to respond to citizen requests was noted by respondents 2, 3, 4, 5, 6, and 7, group interview 22 April 2019; respondent 10, 24 April 2019; respondents 12 and 13, group interview, 24 April 2019; and respondent 15, 25 April 2019.

Would the intervention have produced different results if MPAs had continued to call respondents and ask new, perhaps more finely tailored, questions? Of course, we have no way to know without undertaking additional research. It is possible that voters place intrinsic value on face-to-face interactions with their representatives and would not judge IVR calls genuinely responsive regardless of their depth and frequency. It is also possible that increasing political communication itself will have little impact on voter evaluations of their representatives without visible improvements in public goods. This is, we note again, part of what MPAs reported to us. The surprise evinced by Pakistani provincial authorities at the results of disseminating messages and questions to voters mimics the results of research conducted in the United States that finds that political elites are not very good at knowing what messages will be convincing to the public (Broockman et al., 2024). Even in high-information environments, political communication is a learning process.

The change in reactions by MPAs that occurred over time to the intervention reveals that they operate under hidden constraints that neither they nor we anticipated when we began the IVR treatment. We speculate that these constraints potentially come from two different sources. MPAs in Pakistan may lack autonomy because of the control that political party leaders exercise over their ability to deliver club goods to localities. For instance, there is some evidence that the governing party directs constituency spending to the areas that are most electorally helpful to it, thereby limiting the autonomy and discretion of lower levels of government (Malik, 2019). A second possibility is that MPAs simply lack adequate constituency development funds or a formal role in policy formulation outside of legislation in the first place. This can occur regardless of any higher-level political pressures, and it may hamper the capacity of MPAs to respond to bottom-up claims. Only additional research can disentangle these potential explanations. Both speak to the need for additional systematic on-the-ground research into the operation of distributive politics in Pakistan and other developing countries.

Recent literature encourages disseminating null results of research in order to reduce publication bias (Franco, Malhotra and Simonovits, 2014). We agree, and add that providing plausible — even if speculative — interpretations of null results adds to the case. Previous research on spring-boarding communication between politicians and voters reports low take-up by voters. Our research shows that a politician-led approach resolves the take-up challenge but introduces other, perhaps more fundamental and previously unobserved complications — complications that neither we nor our politician partners anticipated. Even if politicians wish to respond to voter preferences, in many developing countries they may lack the capacity to do so. A more robust communication infrastructure between politicians and voters may need a concurrent boost in the arrangements and resources that allows politicians to make promises they can keep. Politicians may back off from using new communication technologies if these expose them to voter disappointment. This response is reasonable, and suggests that many things must change simultaneously to improve political responsiveness where it is low.

# References

Ahrenshop, Mats, Miriam A. Golden, Saad Gulzar and Luke Sonnet. 2023. "Inaccurate forecasting of a randomized controlled trial." *Journal of Experimental Political Science* First view.

Allie, Feyaad. 2023. The Representation Trap: How and Why Muslims Struggle to Maintain Power in India. Technical report Working Paper.

Arias, Eric, Pablo Balán, Horacio Larreguy, John Marshall and Pablo Querubín. 2019. "Information Provision, Voter Coordination, and Electoral Accountability: Evidence from Mexican Social Networks." *American Political Science Review* 113:475–98.

Auerbach, Adam Michael. 2019. *Demanding development: The politics of public goods provision in India's urban slums.* Cambridge University Press.

Bafumi, Joseph and Michael C. Herron. 2010. "Leapfrog Representation and Extremism: A Study of American Voters and Their Members in Congress." *American Political Science Review* 104(3):519–42.

Barth, Fredrik. 1965. *Political Leadership among Swat Pathans.* London School of Economics Monographs on Social Anthropology (Book 19) London: Berg Publishers.

Besley, Timothy. 2006. *Principled Agents? The Political Economy of Good Government.* Oxford: Oxford University Press.

Blair, Graeme, Rebecca Littman and Elizabeth Levy Paluck. 2019. "Motivating The Adoption of New Community-Minded Behaviors: An Empirical Test in Nigeria." *Science Advances* 5(3).

Brookman, David E. and Christopher Skovron. 2018. "Bias in Perceptions of Public Opinion among Political Elites." *American Political Science Review* 112(3):542–63.

Brookman, David E., Joshua L. Kalla, Christian Caballero and Easton Matthew. 2024.

"Political practitioners poorly predict which messages persuade the public." *Proceedings of the National Academy of Sciences* 121(45):e2400076121.

Buntaine, Mark T., Daniel L. Nielson and Jacob T. Skaggs. 2019. "Escaping the Disengagement Dilemma: Two Field Experiments on Motivating Citizens to Report on Public Services." *British Journal of Political Science* 51(2):685–705.

Buntaine, Mark T., Patrick Hunnicutt and Polycarp Komakech. 2020. "The Challenges of Using Citizen Reporting to Improve Public Services: A Field Experiment on Solid Waste Services in Uganda." *Journal of Public Administration Research and Theory* 31(1):108–27.

Bussell, Jennifer. 2012. *Corruption and Reform in India: Public Services in the Digital Age.* New York: Cambridge University Press.

Bussell, Jennifer. 2019. *Clients and Constituents: Political Reponsiveness in Patronage Democracies.* New York: Oxford University Press.

Casey, Katherine, Rachel Glennester, Edwards Miguel and Maarten Voors. 2023. "Skill versus Voice in Local Development." *Review of Economics and Statistics* 105(2):311–26.

Cheema, Ali, Sarah Khan, Asad Liaqat and Shandana Khan Mohmand. 2023. "Canvassing the gatekeepers: A field experiment to increase women voters' turnout in Pakistan." *American Political Science Review* 117(1):1–21.

Chong, Alberto, Ana L. De La O, Dean Karlan and Leonard Wantchekon. 2015. "Does Corruption Information Inspire the Fight or Quash the Hope? A Field Experiment in Mexico on Voter Turnout, Choice and Participation." *Journal of Politics* 77(1):55–71.

Christensen, Garret, Jeremy Freese and Edward Miguel. 2019. *Transparent and Reproducible Social Science Research: How To Do Open Science.* Oakland: University of California Press.

Costa, Mia. 2017. "How Responsive are Political Elites? A Meta-Analysis of Experiments on Public Officials." *Journal of Experimental Political Science* 4:241–54.

Cruz, Cesi, Julien Labonne and Pablo Querubin. 2017. "Politician family networks and electoral outcomes: Evidence from the Philippines." *American Economic Review* 107(10):3006–37.

DellaVigna, Stefano and Devin Pope. 2018. "Predicting experimental results: Who knows what?" *Journal of Political Economy* 126(6):2410–56.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D. Hyde and Craig McIntosh, eds. 2019. *Information and Accountability: A New Method of Cumulative Learning.* New York: Cambridge University Press.

Erlich, Aaron, Danielle F. Jung, James D. Long and Craig McIntosh. 2018. "The double-edged sword of mobilizing citizens via mobile phone in developing countries." *Development Engineering* 3:34–46.

Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50(1–3):5–25.

Ferejohn, John A. 1974. *Pork Barrel Politics: Rivers and Harbors Legislation, 1947–1968.* Stanford: Stanford University Press.

Ferrali, Romain, Guy Grossman, Melina Platas Izama and Jonathan Rodden. 2016. "Explaining Uptake of the U-Bridge Program: Network Analysis Final Report." Unpublished paper.

Franco, Annie, Neil Malhotra and Gabor Simonovits. 2014. "Publication Bias in the Social Sciences: Unlocking the File Drawer Problem." *Science* 345(6203):1502–05.

Gallup Pakistan. 2013. "10th General Elections in Pakistan 2013.". Accessed: 2024-11-14.
**URL:** *https://gallup.com.pk/bb_old_site/election/10GIER2013.pdf*

Golden, Miriam A. and Brian Min. 2013. "Distributive Politics Around the World." *Annual Review of Political Science* 16:73–99.

Golden, Miriam A., Saad Gulzar and Luke Sonnet. 2017. "Improving Political Communication: Results of a Pilot Field Experiment in Pakistan." International Growth Center S-37412-PAK-2.

Grossman, Guy, Kristin Michelitch and Marta Santamaria. 2017. "Texting Complaints to Politicians: Name Personalization and Politicians' Encouragement in Citizen Mobilization." *Comparative Political Studies* 50(10):1325–57.

Grossman, Guy, Macartan Humphreys and Gabriella Sacramone-Lutz. 2014. ""I wld like u WMP to extend electricity 2 our village": On Information Technology and Interest Articulation." *American Political Science Review* 108(3):688–705.

Grossman, Guy, Macartan Humphreys and Gabriella Sacramone-Lutz. 2020. "Information Technology and Political Engagement: Mixed Evidence from Uganda." *Journal of Politics* 82(4):1321–36.

Gulzar, Saad. 2021. "Who Enters Politics and Why?" *Annual Review of Political Science* 24:253–74.

Kalla, Joshua L. and Ethan Porter. 2021. "Correcting Bias in Perceptions of Public Opinion Among American Elected Officials: Results from Two Field Experiments." *British Journal of Political Science* 51:1792–1800.

Kruks-Wisner, Garielle. 2018. *Claiming the State: Active Citizenship and Social Welfare in Rural India.* Cambridge: Cambridge University Press.

Larreguy, Horacio, John Marshall and Pablo Querubin. 2016. "Parties, brokers, and voter mobilization: How turnout buying depends upon the party's capacity to monitor brokers." *American Political Science Review* 110(1):160–79.

Leo, Ben, Robert Morello, Jonathan Mellon, Tiago Peixoto and Stephen Davenport. April 2015. Do Mobile Phone Surveys Work in Poor Countries? Working Paper N. 398 Center for Global Development.

Liaqat, Asad, Ali Cheema and Shandana Khan Mohmand. 2020. Who Do Politicians Talk To? Political Contact in Urban Punjab. In *Pakistan's Political Parties*, ed. Mariam Mufti, Sahar Shafqat and Niloufer Siddiqui. Washington, D.C.: Georgetown University Press.

Lyon, Stephen M. 2004. *An Anthropological Analysis of Local Politics and Patronage in a Pakistani Village.* Mellen Series in Anthropology (Volume 12) Lewiston: Edwin Mellen Press.

Malik, Rabia. 2019. "(A)Political Constituency Development Funds: Evidence from Pakistan." *British Journal of Political Science* 51(3):963–80.

Martin, Nicolas. 2014. The Dark Side of Patronage in the Pakistani Punjab. In *Patronage as Politics in South Asia*, ed. Anastasia Piliavsky. Delhi: Cambridge University Press.

Martin, Nicolas. 2016. *Politics, Landlords and Islam in Pakistan.* London: Routledge.

Mayhew, David. 1974. *Congress: The Electoral Connection.* New Haven: Yale University Press.

Mohmand, Shandana Khan. 2019. *Crafty Oligarchs, Savvy Voters: Democracy under Inequality in Rural Pakistan.* Cambridge: Cambridge University Press.

Pakistan Telecom Authority. 2021. "Telecom Indicators.".
**URL:** *https://www.pta.gov.pk/en/telecom-indicators*

Pereira, Miguel M. 2021. "Understanding and Reducing Biases in Elite Beliefs About the Electorate." *American Political Science Review* 115(4):1308–24.

Persson, Mikael and Anders Sundell. 2024. "The Rich Have a Slight Edge: Evidence from

Comparative Data on Income-Based Inequality in Policy Congruence." *British Journal of Political Science* 54(2):514–25.

Shaul-Cohen, Sara and Azi Lev-On. 2020. "Smartphones, Text Messages, and Political Participation." *Mobile Media and Communication* 8(1):62–82.

Siddiqui, Niloufer A. 2022. *Under the gun: Political parties and violence in Pakistan.* Cambridge University Press.

Soo, Nikki, James Weinberg and Katharine Dommett. 2020. "One moment, please: Can the speed and quality of political contact affect democratic health?" *The British Journal of Politics and International Relations* 22(3):460–484.

Stokes, Susan C. 2005. "Perverse accountability: A formal model of machine politics with evidence from Argentina." *American political science review* 99(3):315–325.

Walgrave, Stefaan, Arno Jansen, Julie Sevenans, Karolin Soontjens, Jean-Benoit Pilet, Nathalie Brack, Frédéric Varone, Luzia Helfer, Rens Vliegenthart, Toni van der Meer, Christian Breunig, Stefanie Bailer, Lior Sheffer and Peter John Loewen. 2023. "Inaccurate Politicians: Elected Representatives' Estimations of Public Opinion in Four Countries." *Journal of Politics* 85(1):209–22.

Wantchekon, Leonard. 2003. "Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin." *World Politics* 55(03):399–422.

Wilder, Andrew R. 1999. *The Pakistani Voter: Electoral Politics and Voting Behaviour in the Punjab.* Oxford: Oxford University Press.

Zaman, Adil and Muhammad Mushtaq. 2022. "Public Perspective About Party Identification in Peshawar Valley, Pakistan." *Pakistan Journal of Social Research* 4(2):500–513.

# Appendices

## A   Registry of changes from the pre-analysis plan (PAP)

Key informant interviews: We originally intended to conduct multiple waves of key informant interviews but were prevented from doing so when field activities were interrupted by government security services. As a result, we deviate from the PAP in two ways. We use heteroskedasticity-consistent standard errors (HC2) rather than clustered standard errors, since the treatments of interest are assigned at the polling station level, the same level as outcomes. We also remove wave fixed effects since we have none.

Measuring PS level effects: We deviate from our pre-specified analysis, where we had a post-treatment variable (whether the MPA ran again) on the right hand side. We removed that variable in the analysis.

## B   Ethics of the intervention

Prior to any field activities involving human subjects, we sought permission from Institutional Review Board (IRB) at the University of California at Los Angeles (UCLA). Stanford University's IRB had agreed to subordinate to UCLA Interviews with politicians conducted in 2019 were approved in an IRB amendment.

Our intervention raises at least three major ethical concerns: partisanship, possible interference in the democratic process, and the gender imbalance of our study.

- Partisanship: A first ethical concern is that we partner with incumbent MPAs and do not offer IVR communication to challengers. This might bias the political process in favor of incumbents or their parties.

  There were practical reasons for working exclusively with sitting MPAs. Our experiment was rolled out in the nine months prior to our best guess of when the next election would be held. Because of the lead time required for the experiment, it was not feasible to undertake it immediately prior to the election. At the time the experiment was conducted, the identity of assembly candidates in the upcoming 2018 elections was unknown. (Indeed, seven of the 40 incumbents we work with did not run again.) In addition, our funding agency (J-PAL) prohibited us from conducting research that could have been construed as interfering with or participating in the election campaign. For both of these reasons, it was not possible to work in the month just prior the election when campaigning was underway and when the candidates had been selected and their identities known.

  These logistical considerations explain the timing of our research, but do not necessarily speak directly to the ethics of working with sitting MPAs. Some might worry that our work would shore up their political power. We were relatively unconcerned about this, however. The reason is that the MPAs we work with did not constitute an entrenched

political elite. Indeed, 60 percent of those enrolled in the IVR treatment had never served previously in the provincial or national assembly, and 50 percent of them did not get reelected (either because they were not on the ballot or because they did not win the seat). Similarly, elections in KP are often closely fought; the average margin of victory in 2013 was 11 percent, considerably lower than in Pakistan's other three provinces. Indeed, we considered that providing a new way for MPAs to communicate with voters might encourage a largely inexperienced group of politicians to learn to do their jobs more effectively and thus might improve political representation in KP.

- Political interference: A second ethical concern is that our intervention might have affected political outcomes. However, our sample sizes are too small for this to have been possible. For evidence, see the discussion in Section 5.2.

- Gender: A final ethical concern is that we work only with men and exclude women. As we note however (see fn. 7), even when we sent female staff to approach women, they were unwilling to provide their phone numbers and participate in the study. Although our study thus excluded women, we do not think that our activities harmed them. The reason is that we do not believe that providing a representative sample of male heads-of-households new ways to express political and policy opinions is necessarily detrimental to female voters. We would have considered the intervention to carry more problematic ethical implications if it had given voice to an unrepresentative sample of wealthier, older, or more conservative male respondents, for instance. That said, we remain troubled by the failure to enroll female respondents in the study and hope to examine how these barriers can be overcome in future work.

## C   Timeline of the intervention

Owing to weather and security constraints, we implemented the intervention in five phases. Our implementing partner put four of the 20 sampled constituencies into treatment in each stage, and then rolled out the baseline and recordings of messages with MPAs on a staggered basis. The timing of activities — baseline survey, two stages of phone calls, and endline survey — is reported in Table C.1. In each constituency, households were recruited, provided informed consent, and baselined over a two-week period. Simultaneously, meetings with MPAs were ongoing to enroll them in the project, introduce the technology to them, and work with them to script and record their initial calls to constituents. Recordings were made either with field staff on site or later on the MPA's own time. Because of the nature of the meetings and because some MPAs recorded their calls later when they found more time, many recordings were made on mobile phones, although we edited the files for clarity afterwards. Follow-up calls were made in similar fashion. The endline survey was rolled out after the general elections were held on July 25, 2018, also in phases due to the weather as well as security concerns.

Table C.1: Phased implementation of intervention and surveys

| Phase | Baseline survey | Initial calls | Follow-up calls | Endline survey |
|---|---|---|---|---|
| 1 | December 2017 - February | February | May - July | August - September |
| 2 | January - February | March | May - June | September - October |
| 3 | February | April - May | June - July | August - October |
| 4 | February - April | May | July | September - October |
| 5 | March - April | May | July | August - October |

*Notes:* All activities took place in 2018 unless noted otherwise. Constituency numbers (e.g. "PK-50") were assigned after redelimitation for the 2002 elections. These numbers were changed following a 2018 redelimitation.

Table D.1: Responses to first-stage IVR questions: descriptive data

| Response | Count | Percent |
|---|---|---|
| Education and Health | 432 | 35 |
| Development Works | 552 | 44 |
| Legislation | 250 | 20 |
| Total | 1,247 | 99 |

*Notes:* Excludes 13 respondents who pressed 4 in response to "If you are satisfied with my services, Press 4," recorded by one MPA. Total reported includes these subjects and thus does not add up to 100%. Reports numbers only in relation to number of persons who answered the IVR call and question.

# D   Stage One Script

**Introduction and Credit Claiming:** Assalam U Alaikum. I am [MPA NAME], your elected MPA. In the past nine years, girls' degree college, boys' degree college, and hundreds of primary, middle, high and higher secondary schools have been built. I have constructed a link road. In addition to this, gas lines to each house have been or will be completed. Furthermore, I have worked hard to speak for the people and their rights on the floor of the assembly.

**Question:** What do you think I should focus on going forward?
- Education and health, press 1
- Development works, press 2
- Legislation, press 3
- If you want to listen again, press 9

Thank you for taking your valuable time. Good bye.

# E   Additional Descriptive Results

## E.1   How contact varies by distance

Table E.1 reports results from two sets of polling station level regressions, distinguished according to whether the respondent reports having met his MPA in person in the last year

Table E.1: Distance to MPA's constituency office and average interaction rate, by PS area

| | Outcome | | | |
| | Status quo<br>% who met MPA in last yr | | IVR<br>% who answer IVR question | |
|---|---|---|---|---|
| Intercept | 16.699*** | | 17.750*** | |
| | (2.854) | | (2.461) | |
| Middle 2 PS | −1.441 | | −0.250 | |
| | (1.252) | | (1.646) | |
| Furthest 2 PS | −3.026† | | 0.750 | |
| | (1.699) | | (1.709) | |
| Distance (km) | | −0.225* | | 0.004 |
| | | (0.104) | | (0.102) |
| Constituency FEs | | Yes | | Yes |
| Num. Obs. | 120 | 120 | 120 | 120 |
| R2 | 0.009 | 0.704 | 0.002 | 0.710 |

*Notes:* †, p-value $< 0.1$; *, p-value $< 0.05$; **, p-value $< 0.01$; ***, p-value $< 0.001$. Standard errors clustered by constituency in parentheses. The first and third models have indicators for polling station distance by tercile within constituency. The omitted category is the first tercile, which includes the two closest polling stations.

or whether he answered an IVR question. The regressions study the relationship between distance to an MPA's constituency office and each type of contact with the MPA. We measure distance as the geodesic distance between the respondent's polling station and the MPA's constituency office.[20] We specify distance in two different ways, controlling for constituency level effects in two different ways to deal with the variation in constituency size in our sample: in the first column for each outcome we regress the percent of respondents reporting contact with the MPA on distance with constituency fixed effects; in the second column we instead aggregate respondents into terciles of the distance of the respondent's polling station within each constituency. In both cases, there is evidence that distance to the MPA more strongly predicts status quo contact rates than IVR, and that IVR thus reaches a more geographically diverse set of constituents. In the first column, the coefficient on distance shows that for each kilometer further from the MPA's constituency office, 0.2 percentage point fewer respondents meet in person with the MPA; moving from the 25th percentile (4km) to the 75th (15km) percentile on distance corresponds to a 2.5 percentage point decrease in the percent of respondents who report meeting face-to-face with the MPA. Thus, voters who live further from the MPA's constituency office are disadvantaged by the need to travel for face-to-face meetings.

---

[20]For IVR communication, the MPA is located at his 2013 constituency office whereas for status quo communication, he is located at his 2018 constituency office. Using the 2018 MPA's constituency office for the distance calculations for IVR produces substantively similar results.

# F  Household level index construction

At the household level we build each index following Kling, Liebman and Katz (2007). We first standardize all of the component outcomes by the mean and standard deviation of the outcome in the group that received none of the IVR intervention calls {H0}. We then impute all missing component outcome means to the average of that component outcome in the stage one household treatment group. Note that if there is missingness on all component outcomes for an index for a particular individual, no outcome is imputed and the individual is dropped. We restandardize the indices with respect to the control group (always defined as {H0}), so that effects are interpretable in standard deviation units of the index.[21] Furthermore, because we restandardize only once and with respect to control households, when we make comparisons between treatment arms, the standard deviation in the comparison group is not always equal to one.

We first study effects of the intervention on *evaluations of the incumbent*. The index comprises four outcomes: (1) a feeling thermometer for the MPA himself (1–10); (2) a feeling thermometer for the MPA's party (1–10); (3) a binary indicator for whether the respondent reports having voted for the MPA (or the incumbent MPA's party if the incumbent did not run again); and (4) the inverse of the ranking of the MPA among his top four challengers in the 2018 elections. Second, we study effects of the intervention on *political participation*. This index consists of three outcomes: (1) a binary indicator for whether or not the respondent voted in the 2018 election; (2) a binary indicator for whether or not the respondent attended a rally in the period leading up to the 2018 elections; and (3) a binary indicator for whether or not the respondent attended a political meeting before the 2018 elections. Third, we study effects of the intervention on *prospects for electoral accountability*. This index is made up of three outcomes: (1) a measure of self-stated political efficacy (1–5); (2) how important incumbent performance is in an individual's vote choice (1–6); (3) the number of conversations the respondent had about politics in the two weeks before the endline survey. The first item is included because we think voters are more likely to attempt to evaluate the performance of the incumbent when they have higher levels of political efficacy. We include the second item to measure whether the voter thinks that performance criteria should be used when deciding for whom to vote. The final item is included because we contend that voters cannot enforce political accountability without some political engagement that includes discussing issues.[22]

---

[21]The original indices were interpretable as an average of standard deviation unit treatment effects on the component measures, rather than as a standard deviation treatment effect on the index itself (Kling, Liebman and Katz, 2007).

[22]We also show that adding this outcome to the political participation index does not change any of the conclusions reported throughout the paper (see Table G.2).

# G  Alternative specifications for downstream household results

In this section we present the robustness of the null experimental results presented in Table 3 to (i) attrition, (ii) alternative specifications of two indices, and (iii) considering compliance rates by estimating local average treatment effects among compliers.

First, we present robustness of the main results to attrition. The results in Table G.1 replicate the main, downstream household level results in Table 3 but account for attrition using inverse probability of attrition weights. Using the full experimental sample we first estimate the probability a respondent attrited and then use these estimated probabilities to weight the non-attritors to overrepresent the respondents who have similar characteristics to those respondents who attrited. If the model estimating the probability of attrition is well-specified, then these weights will unbiasedly estimate the treatment effect among the full sample, including attriters. While we do not expect our model to be perfect, this is a common approach to dealing with attrition, especially when treatment status does not predict attrition and when it is not severe.

The weights for non-attriters used in the analysis below are

$$w_i = \frac{1}{\hat{p}_i}$$

where $\hat{p}_i$ is predicted probability of non-attrition from a regression of non-attrition on co-partisanship, age bins, income scale groups, education bins, political knowledge, and MPA feeling thermometer bins.

Second, political conversations could instead be considered political participation rather than a precursor to accountable electoral politics. As such, we rebuild the indices with political conversations moved to the political participation index and present the results in Table G.2.

Third, compliance with the full IVR treatments was around 17 percentage points. As such, local average treatment effects among compliers will be larger than intent-to-treat effects. We present local average treatment effects among compliers in Table G.3 where we define compliance with receiving any call {H1, H2} as answering the first stage phone call and we define compliance with the full, responsive treatment {H2R} as answering the first stage IVR question and answering the second stage phone call. For all analyses, we use the same specifications as the main results and instrument for the binary indicator of compliance with the treatment assignment. Even with these fairly restrictive definitions of compliance, the treatment effects remain substantively small and, unsurprisingly given the nature of the estimating local treatment effects among compliers, they remain statistically insignificant.

Table G.1: Effect of any IVR call and effect of full IVR treatment on household head outcomes - including index component measures and weighted for attrition

| Outcome | Control mean: no call {H0} | ITT: any call {H1, H2} vs. {H0} | | ITT: full responsive treatment {H2R} vs. {H0} | |
|---|---|---|---|---|---|
| | $\mu$ | $\tau$ | N | $\tau$ | N |
| **Incumbent evaluations index** | 0.000 (1.000) | −0.009 (0.009) | 13757 | −0.016 (0.013) | 6539 |
| MPA feeling thermometer (1-10) | 4.864 (3.340) | −0.056 (0.038) | 13753 | −0.087 (0.058) | 6536 |
| MPA party feeling thermometer (1-10) | 4.536 (3.501) | −0.018 (0.035) | 13758 | −0.019 (0.056) | 6538 |
| Voted for MPA (0/1) | 0.337 (0.473) | −0.004 (0.004) | 13753 | −0.010 (0.007) | 6538 |
| Inverse rank of MPA (1-5) | 2.661 (1.478) | −0.001 (0.018) | 13309 | −0.012 (0.027) | 6307 |
| **Political participation index** | 0.000 (1.000) | −0.021 (0.016) | 13780 | 0.004 (0.025) | 6551 |
| Voted (0/1) | 0.985 (0.122) | −0.001 (0.002) | 13260 | 0.002 (0.003) | 6282 |
| Attended rally (0/1) | 0.239 (0.427) | −0.008 (0.007) | 13760 | −0.001 (0.011) | 6539 |
| Attended political meeting (0/1) | 0.180 (0.385) | −0.007 (0.006) | 13780 | −0.002 (0.010) | 6551 |
| **Prospects for accountability index** | 0.000 (1.000) | 0.004 (0.017) | 13759 | 0.025 (0.026) | 6539 |
| Political efficacy (1-5) | 3.781 (1.163) | 0.003 (0.020) | 13930 | 0.058† (0.030) | 6618 |
| Vote choice based on performance (1-6) | 4.267 (1.684) | −0.007 (0.025) | 13703 | −0.020 (0.039) | 6514 |
| N political conversations | 3.739 (2.466) | 0.030 (0.040) | 13978 | 0.029 (0.062) | 6642 |
| **Global index** | 0.000 (1.000) | −0.015 (0.013) | 13950 | 0.005 (0.020) | 6629 |

Notes: †, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. Notes: Heteroskedasticity-consistent (HC2) standard errors in parentheses. Pre-treatment control variables not displayed; see Section 5.1 for details. Control means, treatment effects, and all estiamtes of uncertainty are weighted using inverse probability of attrition weights. These weights were generated using a linear model incorporating income, education, age, political knowledge, MPA feeling thermometers, and copartisanship with the MPA.

Table G.2: Effect of any IVR call and effect of full IVR treatment on household head outcomes - moving conversations to participation index

| Outcome | Control mean: no call {H0} | ITT: any call {H1, H2} vs. {H0} | | ITT: full responsive treatment {H2R} vs. {H0} | |
|---|---|---|---|---|---|
| | $\mu$ | $\tau$ | N | $\tau$ | N |
| **Incumbent evaluations index** | 0.000 | −0.009 | 13757 | −0.016 | 6539 |
| | (1.000) | (0.009) | | (0.013) | |
| MPA feeling thermometer (1-10) | 4.864 | −0.056 | 13753 | −0.087 | 6536 |
| | (3.340) | (0.038) | | (0.058) | |
| MPA party feeling thermometer (1-10) | 4.536 | −0.018 | 13758 | −0.019 | 6538 |
| | (3.501) | (0.035) | | (0.056) | |
| Voted for MPA (0/1) | 0.337 | −0.004 | 13753 | −0.010 | 6538 |
| | (0.473) | (0.004) | | (0.007) | |
| Inverse rank of MPA (1-5) | 2.661 | −0.001 | 13309 | −0.012 | 6307 |
| | (1.478) | (0.018) | | (0.027) | |
| **Political participation index (w/ convs.)** | 0.000 | −0.012 | 13780 | 0.009 | 6551 |
| | (1.000) | (0.016) | | (0.024) | |
| Voted (0/1) | 0.985 | −0.001 | 13260 | 0.002 | 6282 |
| | (0.122) | (0.002) | | (0.003) | |
| Attended rally (0/1) | 0.239 | −0.008 | 13760 | −0.001 | 6539 |
| | (0.427) | (0.007) | | (0.011) | |
| Attended political meeting (0/1) | 0.180 | −0.007 | 13780 | −0.002 | 6551 |
| | (0.385) | (0.006) | | (0.010) | |
| N political conversations | 3.739 | 0.030 | 13978 | 0.028 | 6642 |
| | (2.466) | (0.040) | | (0.062) | |
| **Prospects for accountability index (no convs.)** | 0.000 | −0.004 | 13759 | 0.023 | 6539 |
| | (1.000) | (0.016) | | (0.025) | |
| Political efficacy (1-5) | 3.781 | 0.003 | 13930 | 0.059$^{\dagger}$ | 6618 |
| | (1.163) | (0.020) | | (0.030) | |
| Vote choice based on performance (1-6) | 4.267 | −0.007 | 13703 | −0.020 | 6514 |
| | (1.684) | (0.025) | | (0.039) | |

*Notes:* $^{\dagger}$, p-value $< 0.1$; *, p-value $< 0.05$; **, p-value $< 0.01$; ***, p-value $< 0.001$. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses.
Pre-treatment control variables not displayed; see Section 5.1 for details.

Table G.3: Effect of any IVR call and effect of full IVR treatment on household head outcomes — local average treatment effects among compliers

| Outcome | Control mean: no call {H0} $\mu$ | LATE: any call (answered phone) {H1, H2} vs. {H0} $\tau$ | N | LATE: full responsive treatment (answered first q and second call) {H2R} vs. {H0} $\tau$ | N |
|---|---|---|---|---|---|
| **Incumbent evaluations index** | 0.000 (1.000) | −0.012 (0.012) | 13757 | −0.108 (0.091) | 6539 |
| MPA feeling thermometer (1-10) | 4.864 (3.340) | −0.077 (0.052) | 13753 | −0.596 (0.399) | 6536 |
| MPA party feeling thermometer (1-10) | 4.536 (3.501) | −0.024 (0.048) | 13758 | −0.128 (0.385) | 6538 |
| Voted for MPA (0/1) | 0.337 (0.473) | −0.006 (0.006) | 13753 | −0.066 (0.047) | 6538 |
| Inverse rank of MPA (1-5) | 2.661 (1.478) | −0.001 (0.024) | 13309 | −0.084 (0.187) | 6307 |
| **Political participation index** | 0.000 (1.000) | −0.028 (0.022) | 13780 | 0.027 (0.171) | 6551 |
| Voted (0/1) | 0.985 (0.122) | −0.001 (0.003) | 13260 | 0.011 (0.023) | 6282 |
| Attended rally (0/1) | 0.239 (0.427) | −0.011 (0.009) | 13760 | −0.004 (0.073) | 6539 |
| Attended political meeting (0/1) | 0.180 (0.385) | −0.009 (0.009) | 13780 | −0.014 (0.067) | 6551 |
| **Prospects for accountability index** | 0.000 (1.000) | 0.006 (0.023) | 13759 | 0.174 (0.177) | 6539 |
| Political efficacy (1-5) | 3.781 (1.163) | 0.004 (0.027) | 13930 | $0.401^{\dagger}$ (0.209) | 6618 |
| Vote choice based on performance (1-6) | 4.267 (1.684) | −0.010 (0.034) | 13703 | −0.137 (0.267) | 6514 |
| N political conversations | 3.739 (2.466) | 0.041 (0.055) | 13978 | 0.196 (0.428) | 6642 |
| **Global index** | 0.000 (1.000) | −0.020 (0.018) | 13950 | 0.037 (0.135) | 6629 |

*Notes:* $^{\dagger}$, p-value $< 0.1$; $^{*}$, p-value $< 0.05$; $^{**}$, p-value $< 0.01$; $^{***}$, p-value $< 0.001$. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses.

# H  Robustness of null findings in downstream results

One concern with the null results is that the experiment might have been statistically underpowered and therefore unable to detect effects even if they exist. Although this may be the case at the polling station level, it is unlikely to be true at the household level, where we have a large sample. We can formalize this by computing equivalence confidence intervals, as proposed in Hartman and Hidalgo (2018). We estimate equivalence confidence intervals that contain treatment effects that are small enough that we cannot reject the null that they are too large. In other words, the values within the equivalence confidence intervals that we report in Figure H.1 are small enough to be consistent with the data; larger treatment effects can be rejected as too large given the data at our given significance level (here, 0.05). Therefore, instead of relying on failing to reject the null to establish a null effect, these confidence intervals allow us to find the largest treatment effects — those on the ends of the equivalence confidence intervals — that we cannot reject as too large. For the effect of receiving any call in the left panel, all treatment effects on individual attitudes and self-reported behavior larger than |0.05| sds can be rejected as too large, while effects larger than |0.08| sds are inconsistent with the observed effect of the full treatment (as seen in the right panel). These "largest possible effects" are quite small and reflect the substantial power of our design to detect meaningful individual level effects.

# I  Additional downstream results

We preregistered other analyses between various treatment arms: the marginal effect of receiving an initial call with questions (H2) versus receiving the initial call with no questions (H1); the marginal effect of receiving a responsive follow-up call (H2R) versus a generic follow-up call (H2G); and the marginal effect of receiving *any* follow-up call (H1G + H2G + H2R) versus no follow-up call (H1C + H2C). The first two analyses report similar treatment effects, where we find no large substantive effects with no statistically significant treatment effects. The only treatment effect that is statistically significant at even the 0.1 level comes when considering the marginal effect of receiving any follow-up call.

Below we expand tables reported in the main paper to include index components, and we include additional treatment group comparisons.

Table I.1: **Effect of any IVR call and full IVR treatment on household head outcomes including index components**

| Outcome | Control mean: no call {H0} | ITT: any call {H1, H2} vs. {H0} | | ITT: full responsive treatment {H2R} vs. {H0} | |
|---|---|---|---|---|---|
| | $\mu$ | $\tau$ | N | $\tau$ | N |
| **Incumbent evaluations index** | 0.000 (1.000) | −0.009 (0.009) | 13757 | −0.016 (0.013) | 6539 |
| MPA feeling thermometer (1-10) | 4.864 (3.340) | −0.056 (0.038) | 13753 | −0.087 (0.058) | 6536 |
| MPA party feeling thermometer (1-10) | 4.536 (3.501) | −0.018 (0.035) | 13758 | −0.019 (0.056) | 6538 |
| Voted for MPA (0/1) | 0.337 (0.473) | −0.004 (0.004) | 13753 | −0.010 (0.007) | 6538 |
| Inverse rank of MPA (1-5) | 2.661 (1.478) | −0.001 (0.018) | 13309 | −0.012 (0.027) | 6307 |
| **Political participation index** | 0.000 (1.000) | −0.020 (0.016) | 13780 | 0.004 (0.025) | 6551 |
| Voted (0/1) | 0.985 (0.122) | −0.001 (0.002) | 13260 | 0.002 (0.003) | 6282 |
| Attended rally (0/1) | 0.239 (0.427) | −0.008 (0.007) | 13760 | −0.001 (0.011) | 6539 |
| Attended political meeting (0/1) | 0.180 (0.385) | −0.007 (0.006) | 13780 | −0.002 (0.010) | 6551 |
| **Prospects for accountability index** | 0.000 (1.000) | 0.004 (0.017) | 13759 | 0.025 (0.026) | 6539 |
| Political efficacy (1-5) | 3.781 (1.163) | 0.003 (0.020) | 13930 | 0.059$^{\dagger}$ (0.030) | 6618 |
| Vote choice based on performance (1-6) | 4.267 (1.684) | −0.007 (0.025) | 13703 | −0.020 (0.039) | 6514 |
| N political conversations | 3.739 (2.466) | 0.030 (0.040) | 13978 | 0.028 (0.062) | 6642 |
| **Global index** | 0.000 (1.000) | −0.014 (0.013) | 13950 | 0.005 (0.020) | 6629 |

*Notes:* $^{\dagger}$, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses.
Pre-treatment control variables not displayed; see Section 5.1 for details.

Table I.2: **ITT effects of initial call type on follow-up pickup rates**

| Outcome | Control mean: call only {H1} | ATE: effect of getting asked IVR question vs. call only {H2} vs. {H1} | |
|---|---|---|---|
| | $\mu$ | $\tau$ | N |
| Answered follow-up phone call (0/1) | 0.787 (0.410) | 0.036* (0.015) | 3718 |

*Notes:* $^{\dagger}$, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses. Because nothing about a respondent's treatment condition is revealed before picking up the phone, initial call treatment status (e.g. {H1} or {H2}) only affects the respondent once he answers the initial call. Therefore, we subset the analysis to respondents who answer the first call.

Table I.3: **Marginal effect of IVR question in initial call on household head outcomes including index component measures**

| Outcome | Control mean: credit claiming call only {H1} | ITT: marg effect of IVR q {H2} vs. {H1} | |
| --- | --- | --- | --- |
| | $\mu$ | $\tau$ | N |
| **Incumbent evaluations index** | −0.018 (1.007) | 0.004 (0.012) | 9164 |
| MPA feeling thermometer (1-10) | 4.778 (3.377) | 0.014 (0.051) | 9162 |
| MPA party feeling thermometer (1-10) | 4.480 (3.486) | 0.031 (0.046) | 9166 |
| Voted for MPA (0/1) | 0.331 (0.471) | −0.002 (0.005) | 9160 |
| Inverse rank of MPA (1-5) | 2.646 (1.473) | 0.009 (0.023) | 8866 |
| **Political participation index** | −0.014 (0.987) | −0.009 (0.021) | 9179 |
| Voted (0/1) | 0.985 (0.123) | −0.001 (0.003) | 8834 |
| Attended rally (0/1) | 0.234 (0.423) | −0.001 (0.009) | 9167 |
| Attended political meeting (0/1) | 0.176 (0.381) | −0.003 (0.008) | 9179 |
| **Prospects for accountability index** | −0.003 (0.995) | 0.016 (0.022) | 9166 |
| Political efficacy (1-5) | 3.789 (1.146) | −0.003 (0.026) | 9283 |
| Vote choice based on performance (1-6) | 4.257 (1.681) | 0.010 (0.033) | 9129 |
| N political conversations | 3.723 (2.357) | 0.054 (0.052) | 9312 |
| **Global index** | −0.022 (1.009) | 0.006 (0.017) | 9293 |

*Notes:* [†], p-value $< 0.1$; [*], p-value $< 0.05$; [**], p-value $< 0.01$; [***], p-value $< 0.001$. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses.
Pre-treatment control variables not displayed; see Section 5.1 for details.
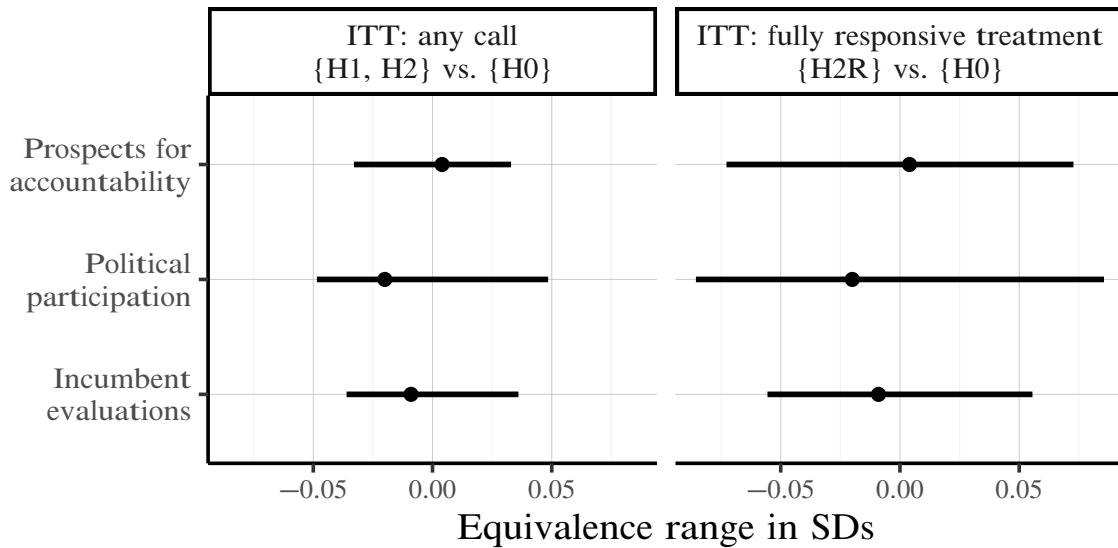
Table I.4: **Marginal effect of responsive vs. generic follow-up call on household head outcomes including index components**

| Outcome | Mean: Generic Follow-up {H3G} | ITT: Marg. Effect of Responsive Follow-up {H3R} vs. {H3G} | |
| --- | --- | --- | --- |
| | $\mu$ | $\tau$ | N |
| **Incumbent Evaluations Index** | 0.000 | −0.016 | 6539 |
| | (1.000) | (0.013) | |
| MPA Feeling Thermometer (1-10) | 4.864 | −0.087 | 6536 |
| | (3.340) | (0.058) | |
| MPA Party Feeling Thermometer (1-10) | 4.536 | −0.019 | 6538 |
| | (3.501) | (0.056) | |
| Voted for MPA (0/1) | 0.337 | −0.010 | 6538 |
| | (0.473) | (0.007) | |
| Inverse Rank of MPA (1-5) | 2.661 | −0.012 | 6307 |
| | (1.478) | (0.027) | |
| **Political Participation Index** | 0.000 | 0.004 | 6551 |
| | (1.000) | (0.025) | |
| Voted (0/1) | 0.985 | 0.002 | 6282 |
| | (0.122) | (0.003) | |
| Attended rally (0/1) | 0.239 | −0.001 | 6539 |
| | (0.427) | (0.011) | |
| Attended political meeting (0/1) | 0.180 | −0.002 | 6551 |
| | (0.385) | (0.010) | |
| **Prospects for Accountability Index** | 0.000 | 0.025 | 6539 |
| | (1.000) | (0.026) | |
| Political efficacy (1-5) | 3.781 | 0.059$^{\dagger}$ | 6618 |
| | (1.163) | (0.030) | |
| Vote choice based on performance (1-6) | 4.267 | −0.020 | 6514 |
| | (1.684) | (0.039) | |
| N political conversations | 3.739 | 0.028 | 6642 |
| | (2.466) | (0.062) | |
| **Global Index** | 0.000 | 0.005 | 6629 |
| | (1.000) | (0.020) | |

*Notes:* $^{\dagger}$, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses.
Pre-treatment control variables not displayed; see Section 5.1 for details.

Figure H.1: Equivalence confidence intervals for main household level treatment effects



This figure contains the realized treatment effects and equivalence confidence intervals (Hartman and Hidalgo, 2018) for the two main analyses we report in Table 3. The points are the realized treatment effects and the equivalence confidence intervals are built at the 0.05 level. Our three main outcome indices are on the y-axis and the treatment effects in the original units of the outcomes, standard deviations, are on the x-axis. The equivalence confidence intervals represent the range of hypothetical treatment effects that are consistent with our data and estimated treatment effects. Any hypothetical treatment effect outside these intervals can be rejected by an equivalence test as too large at the 0.05 level.

## J    Effects on politician behavior

To measure politician behavior, we administer structured *key informant* interviews in the six treated polling station areas in each constituency where voters receive IVR and in six of the 14 control polling station areas where they do not. We conduct key informant interviews before and after the intervention. Key informants were generally salaried individuals (school teachers, for instance) who had no personal or professional relationship with the MPA and who were not involved in campaigning for any political party. Enumerators selected two key informants in each polling station area, both of whom were deemed likely to be available for repeated interviews.

We ask key informants whether politicians visit the polling station area: the goal is to evaluate whether they visit areas more where we direct their IVR phone calls. We also ask key informants whether politicians exert effort in delivering public services, including schools, roads, health facilities, employment conditions, electricity provision, gas provision, water provision, rubbish collection, and general security.

We estimate effects on these outcomes with OLS using the following specification:

$$Y_{pm(t=1)} = \tau D_p + \alpha Y_{pm(t=0)} + \lambda_m + \epsilon_{pm},$$

where $Y_{pm(t=1)}$ is outcome $Y$ at polling station $p$ at endline ($t=1$), $D_p$ is a binary indicator for treatment status, $Y_{pm(t=0)}$ is the pre-treatment outcome $Y$ collected in the baseline (if available), and $\lambda_m$ is a constituency fixed effect. We use heteroskedasticity-consistent standard errors (HC2) since the the treatments of interest here are assigned at the polling station level, the same level as the outcomes.

Table J.1 shows that there is little evidence of effects on politician effort in places they administer IVR, although estimates are not very precisely estimated due to the relatively small number of observations. Key informants report that, in control areas, MPAs made some effort in only 0.48 of the nine public goods domains. In addition, on average, politicians are reported as making some kind of effort in just over a quarter of areas (the control mean is 0.27), and there is almost no effect of treatment on this. Very few areas (only 0.05) saw their MPA visit in June, again virtually unaffected by treatment. Overall, we observe little effect on politician behavior from being enrolled in treatment: their low levels of effort remain as before.

## K    Polling station level treatment effects

Here we analyze downstream results at an aggregate level: polling station level electoral returns. As these results are realized after both voters and politicians have acted in response to treatment, results represent short-term equilibrium experimental outcomes. We compare outcomes for the six treated polling stations to those for the 14 polling stations that we did not treat within the set of 20 most competitive polling stations for each MPA. Thus, all polling stations are within a partner MPA's constituency although the MPA only used IVR

Table J.1: **ITT effects of IVR calls on MPA effort in polling station areas**

| Outcome | Control mean: control PS {P0} $\mu$ | ITT: treated PS {P1} vs. {P0} $\tau$ | N |
|---|---|---|---|
| N of domains where MPA made effort | 0.483 (1.004) | 0.017 (0.090) | 240 |
| Any MPA effort (0/1) | 0.267 (0.444) | 0.050 (0.047) | 240 |
| Any MPA visit in June (0/1) | 0.053 (0.153) | 0.035 (0.022) | 240 |

*Notes:* $^{\dagger}$, p-value < 0.1; *, p-value < 0.05; **, p-value < 0.01; ***, p-value < 0.001. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses. Pre-treatment control variables not displayed; see text for details.

to households in the six treated polling stations.

**Data:** Since we randomize across 20 polling stations in 20 MPA areas, we should have outcome data from 400 polling stations. We successfully collected official election data from only 341 constituencies. The remaining data are missing because: (i) some .psf format returns released by the Election Commission are illegible; (ii) in some instances we were unable to match our polling stations with polling stations resulting from a subsequent redelimitation; and (iii) initial results for 20 polling stations in one constituency were annulled because of low female turnout and the later results have not been made publicity available by the Election Commission.

**Estimation:** We estimate effects using OLS and the following specification:

$$Y_{pm(t=2018)} = \tau D_p + \alpha Y_{pm(t=2013)} + \lambda_m + \epsilon_{pm},$$

where $Y_{pm(t=2018)}$ is the outcome $Y$ for polling station $p$ in MPA constituency $m$ in the 2018 election, $D_p$ is a binary indicator for treatment status, $Y_{pm(t=2013)}$ is the pre-treatment outcome $Y$ in the 2013 election, and $\lambda_m$ is an MPA constituency fixed effect. As before, we use HC2 standard errors, since the treatment assignment is at the polling-station level.

We estimate effects on two outcomes at the polling station level: the vote share for the incumbent (partner) MPA and the turnout rate. Because of re-delimitation and because some of our partner MPAs did not seek office again, our partner MPAs were not candidates in 2018 in every polling station in our sample. In cases where the partner MPA was not a candidate for any party, we code for the candidate from the party with which our partner MPA was last associated.

Table K.1: ITT effects of IVR calls on polling station voting outcomes

| Outcome | Control mean: control PS {P0} | ITT: treated PS {P1} vs. {P0} | |
|---|---|---|---|
| | $\mu$ | $\tau$ | N |
| Incumbent MPA vote share | 0.332 (0.165) | 0.002 (0.016) | 341 |
| Turnout share | 0.477 (0.109) | 0.006 (0.012) | 288 |

*Notes:* [†], p-value $< 0.1$; [*], p-value $< 0.05$; [**], p-value $< 0.01$; [***], p-value $< 0.001$. *Notes:* Heteroskedasticity-consistent (HC2) standard errors in parentheses.

*Notes:* Results presented here are ITT effects estimated using OLS.

**Effects within treated constituencies:** Table K.1 presents intervention effects from treated polling stations compared to control polling stations within treated MPA constituencies.

On average, incumbents' vote share and turnout in elections in control areas remain low; about 33 percent and 47 percent respectively. The low vote shares received by incumbent MPAs are consistent with the generally low reelection rates of incumbents across the developing world (Golden and Nazrullaeva, 2023). Elections in KP often have more than two competitive candidates, meaning the local political environment is unstable and highly competitive, which feeds into low reelection rates.

In general, we do not find evidence to suggest that treatment affected election results, either in terms of stated incumbent vote share — whose point estimate is very close to zero — or voter turnout. Due to high attrition in the sample of polling stations, we are unable to state that we have estimated a precise null effect of the intervention on voting behavior, however.

# Appendix References

Golden, Miriam A. and Eugenia Nazrullaeva. 2023. *The Puzzle of Clientelism: Elections and Political Discretion Around the World.* Elements in Political Economy New York: Cambridge University Press.

Hartman, Erin and F. Daniel Hidalgo. 2018. "An Equivalence Approach to Balance and Placebo Tests." *American Journal of Political Science* 62:1000–13.

Kling, Jeffrey R, Jeffrey B Liebman and Lawrence F Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica* 75(1):83–119.