*American Economic Journal: Applied Economics 2021, 13(2): 52–82*
*https://doi.org/10.1257/app.20190783*

# Improving Last-Mile Service Delivery Using Phone-Based Monitoring†

*By* Karthik Muralidharan, Paul Niehaus, Sandip Sukhtankar, and Jeffrey Weaver*

*Improving "last-mile" public service delivery is a recurring challenge in developing countries. Could the widespread adoption of mobile phones provide a scalable, cost-effective means for improvement? We use a large-scale experiment to evaluate the impact of phone-based monitoring on a program that transferred nearly a billion dollars to 5.7 million Indian farmers. In randomly selected jurisdictions, officials were informed that program implementation would be measured via calls with beneficiaries. This led to a 7.8 percent reduction in the number of farmers who did not receive their transfers. The program was highly cost-effective, costing 3.6 cents for each additional dollar delivered.* (*JEL* O13, O33, Q12, Q18)

The low quality of public service delivery in developing countries adversely affects the lives of billions of people (World Bank 2003). As a result, improving "last-mile" service delivery has been a recurrent theme in recent research, from incentivizing employees to show up to work (Duflo, Hanna, and Ryan 2012) to ensuring that beneficiaries receive money that they are entitled to (Muralidharan, Niehaus, and Sukhtankar 2016). One common challenge is that, like any organization, a government can only manage its personnel to the extent that it can *measure* their performance (Bloom and Van Reenen 2007). Yet measuring service delivery is difficult, especially when frontline work takes place in many remote and dispersed communities.

Existing mechanisms for measuring and monitoring last-mile service delivery are limited in different ways. Internal reporting by the lower layers of bureaucracy is often distorted, as individuals exaggerate their own performance or overstate

* Muralidharan: UC San Diego, JPAL, NBER, and BREAD (email: kamurali@ucsd.edu); Niehaus: UC San Diego, JPAL, NBER, and BREAD (email: pniehaus@ucsd.edu); Sukhtankar: University of Virginia, JPAL, and BREAD (email: sandip.sukhtankar@virginia.edu); Weaver: University of Southern California (email: jbweaver@usc.edu). Benjamin Olken was coeditor for this article. We are grateful to officials in the Government of Telangana, especially Mr. K Ramakrishna Rao and Mr. C Parthasarathi. This paper would not have been possible without the efforts and inputs of the J-PAL South Asia/UCSD project team in the Payments and Governance Research Program, including Kartik Srivastava, Avantika Prabhakar, Frances Lu, Vishnu Padmanabhan, Surya Banda, Mayank Sharma, and Burak Eskici. We also thank Michael Callen, Gordon Dahl, Markus Goldstein, Abhijeet Singh, Adam Solimon, and several seminar participants for helpful comments. Finally, we thank the Strategic Impact Evaluation Fund (SIEF) at the World Bank (especially Alaka Holla) and the Bill and Melinda Gates Foundation (especially Dan Radcliffe) for the financial support that made this study possible. AEA Trial Registry RCT ID: 0002942. UCSD IRB: 180754S.

† Go to https://doi.org/10.1257/app.20190783 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

problems with initiatives they want to undercut.[1] Mechanisms through which citizens can report problems in service delivery, such as grievance redressal phone numbers, yield nonrepresentative data and are often heavily underutilized, perhaps because beneficiaries do not think that they will help.[2] Finally, periodic surveys such as the Living Standards Measurement Surveys or India's National Sample Survey yield independent and representative data, but they are typically too small, infrequent, and delayed in their release to be of use for management purposes.

In this paper, we test whether governments can improve last-mile service delivery using a simple approach to measuring whether people get what they are due: calling and asking. This approach leverages the rapid increase in mobile phone penetration in low-income countries, from 1 mobile subscription per 100 people in 2002 to 62 in 2017 (World Bank 2018). Phone calls bypass bureaucrats to obtain information directly from beneficiaries without the cost and delay associated with field surveys. Managers can use phone-based monitoring data to track and incentivize worker performance in close to real time. Governments have begun to use outbound phone calls to gather data on service delivery, with Pakistan's Citizen Feedback Monitoring Program and Andhra Pradesh's Real Time Governance initiative as prominent examples (Masud 2015; World Bank 2016; Gelb, Mittal, and Mukherjee 2019). However, there is limited evidence to date on the impact of phone-based monitoring by governments on the quality of service delivery.

We examine whether phone-based measurement can in fact improve service delivery in the context of a high-stakes government initiative in India. Beginning in May 2018, the government of Telangana state (GoTS) attempted to distribute \$0.9 billion, or around 3.5 percent of the state's annual budget, as lump-sum payments to farmers. Responsibility for implementing the scheme rested primarily with Mandal (subdistrict) Agricultural Officers (MAOs), who managed the distribution of physical checks to the 5.7 million farmers in the state. An important goal of the program was to reduce the debt taken on by farmers to finance the up-front costs of cultivation. The government therefore made it a priority to deliver transfers to farmers before the onset of the monsoon rains to facilitate timely agricultural investments, such as purchasing seeds and fertilizer and hiring workers to prepare their fields.

Working with the government, we implemented an experimental, at-scale test of phone-based performance measurement. We randomly assigned around a quarter of the 498 MAOs in the state to a treatment condition in which they were told that a call center would call at least 100 of the farmers for whom they were responsible, collect information on implementation outcomes (including whether and when farmers received their checks and corruption during the process), and produce reports visible to them and their supervisors. These details were communicated to treated MAOs by the state commissioner of agriculture personally via a video conference. This communication was then reinforced with a formal letter to treatment MAOs. The government contracted a call center that surveyed 22,127 farmers within two

[1] As an example of the former, Singh (2019) finds evidence of substantial overreporting of primary student learning levels in official data compared to an independent retest of the same students.

[2] For example, only 8,292 complaints were registered between 2012 and 2016 regarding the Public Distribution System in India (Department of Food and Public Distribution 2018) despite the fact that it serves hundreds of millions of people and has an all-India leakage rate of 32 percent (Drèze and Khera 2015).

weeks. The data from these calls were used to create reports on MAOs' absolute and relative performance.

To evaluate the impact of the announcement of phone-based monitoring to MAOs, we use farmer-level administrative bank records of whether and when these checks were encashed as a reliable measure of MAO performance. This high-quality administrative data on the entire universe of 5.7 million potential program beneficiaries allows us to precisely estimate the impact of the phone-based monitoring treatment on the entire population—including beneficiaries with and without cell phones. It also provides the opportunity to test the reliability of phone data as compared to administrative data. Thus, the data from the phone calls were used to *implement* the intervention (creating MAO reports), while administrative data were used to *evaluate* the intervention. Taken together, the design and data allow us to "experiment at scale."[3]

We find that phone-based monitoring significantly improved the likelihood of farmers ever receiving their transfer, as well as the likelihood of receiving it "on time," meaning prior to the arrival of the monsoon on June 8, 2018. On-time delivery of transfers was 2.4 percentage points higher in treated mandals (with a mean of 69 percent in control mandals), which is a 0.25 standard deviation increase relative to the distribution of mandal-level means in control mandals. The likelihood of checks ever being delivered was 1.3 percentage points higher in treated areas (the control mean was 83 percent), or a 0.16 standard deviation increase. Expressed differently, the intervention led to a 7.8 percent reduction in the number of beneficiaries who did not receive their benefits. These effects correspond to a $3.9 million increase in transfers that were delivered on time, a $1 million increase in amount ever delivered, and 17,771 additional farmers encashing their checks.[4] If phone-based monitoring had been extended to the entire state over both agricultural seasons, our estimates suggest that $33.1 million more would have been delivered on time, an additional $8.6 million would ultimately have been delivered, and around 155,000 additional farmers would have received their payments.

The incidence of the intervention was mildly progressive, although the transfer program itself was regressive (since check sizes were proportional to landholdings). For farmers in the bottom quartile of landholdings, there was a 2.2 percentage point (3.3 percent) increase in check encashment rate, around twice the overall effect. We do not find evidence that MAOs skewed their efforts toward farmers with phones. Although the phone-based monitoring system only assessed MAO performance by calling the approximately 60 percent of farmers with listed cell phone numbers, we cannot reject that the measured improvements in performance are the same across beneficiaries with and without phones.

[3] Specifically, the study meets all three aspects of "experimentation at scale" identified in Muralidharan and Niehaus (2017). The treatment is randomized across all 5.7 million landowning farm households in a state of 35 million people (and outcomes are observed for all of them); the intervention was implemented by government at scale, with 1.3 million farmers in treated areas; and the unit of randomization was large enough for treatment effects to be inclusive of spillovers.

[4] Among larger landowners, the treatment effect is larger for "on-time" delivery than "ever delivered." Since they received larger transfers, the impact on on-time delivery relative to ever delivered is bigger for *value* of funds than for *fraction* of checks.

Next, we estimate that phone-based measurement was highly cost-effective. Costing the intervention at the price paid to the call center vendor, we estimate that the incremental cost per additional dollar of benefits ever delivered to beneficiaries was 3.6 cents. Turning to on-time delivery, the cost per dollar of benefits delivered on time was less than one cent. These costs per additional dollar delivered are lower than the administrative costs of almost any antipoverty program for which such data is available (see Section IIIC). In addition, we also calculate the economic returns to phone-based monitoring. To do this, we define benefits as the difference between the estimated return on capital held by farmers rather than the government. Even under relatively conservative assumptions, we estimate a benefit of four times the cost.

Finally, we compare the phone and administrative data to assess the quality of the data obtained from the phone calls.[5] We find that the match rate between phone and administrative data is high (88.6 percent). Comparing measures of MAO performance using both phone and administrative data, our results suggest that the phone-based data is reliable enough to identify which MAOs should be pushed for more effort but perhaps not reliable enough for more serious personnel actions without data over multiple cycles.

Turning to mechanisms, monitoring systems of the kind we study could affect outcomes through two main channels. First, the anticipation of reports generated by the system could alter the *incentives* of the agents being monitored. In our context, there was no explicit link between the measurement and future benefits such as advantageous postings (Khan, Khwaja, and Olken 2019). However, the MAOs may have anticipated that the monitoring would increase the visibility of their performance to senior officials and potentially be used for future rewards. They may also care about how their performance looked relative to their peers. This would plausibly incentivize exertion of additional effort, although monitoring systems explicitly linked to formal incentives might provide an even stronger push.

Second, phone-based monitoring could affect service delivery by providing the principal monitoring the agent or the monitored agent themself with *information* that affects how they deliver services. For example, a principal could use the information generated by the system to determine which agents required nudging to work harder, or an agent could adjust his own effort based on what he learns about his performance as well as that of others.[6]

Our results are best explained by the incentive mechanism. Because the check distribution was implemented quickly, MAOs and their supervisors received performance reports only after nearly 95 percent of all checks that would be delivered had been delivered. At that point, it was too late for the information contained in the reports to affect MAO behavior. Consistent with this, we find that significant treatment effects emerge shortly after the *announcement* of the monitoring

[5] In this particular case, we were fortunate to have high-quality administrative data on outcomes. However, in other cases, the data from the phone calls may be the only source of data available, which makes it policy-relevant to compare phone and administrative data.

[6] Although this was not a component of the intervention we study, phone-based monitoring systems could in principle also provide new information to the monitored agent, such as geographically disaggregated data on service delivery. For example, the MAO may lack information on village-level delivery of services by agricultural extension workers whom they supervise. If the reports had contained village-level information, the MAO could use it to identify and allocate resources toward the most underserved communities.

(reflecting the incentive effect) but no change in the treatment-control gap after performance reports were issued. In this sense, we view the results as a lower bound on the long-term potential effects of the approach as compared to contexts where it would also be possible to use information from the reports for officials to do their jobs better.[7]

This paper complements recent work testing more specialized approaches for using technology to improve governance, such as monitoring worker attendance with time clocks (Banerjee, Duflo, and Glennerster 2008; Dhaliwal and Hanna 2017) or with custom smart-phone applications (Callen et al. 2018). Relative to these specialized approaches, measurement by phone has the advantages of (i) low fixed and variable costs and time to deploy, as call center services are typically cheap and available quickly; (ii) the flexibility to scale across an unusually wide number range of places, programs, and outcomes; and (iii) scope to adapt quickly as challenges and circumstances on the ground change. We also complement recent work by Aker and Ksoll (2018), which tests a phone-based monitoring intervention implemented by a nongovernmental organization (NGO) in an adult education program in 134 villages in Niger and finds learning gains.

Empirical evidence on service delivery in developing countries suggests that increased top-down administrative monitoring can be an effective tool for improving last-mile service delivery (Olken 2007, Muralidharan et al. 2017). However, a practical barrier to the broader use of top-down monitoring has been the challenge of obtaining credible high-frequency data on last-mile service delivery at a sufficiently spatially disaggregated level to hold appropriate officials accountable. Our results suggest that calling and surveying representative samples of beneficiaries, who increasingly have access to a phone, may provide a promising solution.[8]

Specifically, using outbound call centers for measuring the quality of last-mile service delivery can expand state capacity for measurement on several policy-relevant margins, including scale, cost, and speed. First, the intervention was successfully implemented by the government at a scale affecting nearly 6 million households.[9] Second, the cost per phone survey is substantially lower than the cost of a field survey, which allowed for a much larger sample size within a fixed budget. This made it possible to generate credible estimates of program performance at a lower-level unit of governance than would typically be possible and improve the accountability of officials who are closer to the last mile of service delivery. Third, the intervention was both simple to set up (taking one month from agreement to implementation) and quickly generated usable data on MAO performance within two weeks of starting to make the phone calls.

[7] In addition to these two mechanisms, there could also be a direct effect of the phone calls to farmers, such as encouraging them to encash their checks. We test for and find no evidence for such an effect.

[8] While prior studies of monitoring have looked at impacts on reducing corruption or provider absence, we focus on improving bureaucratic effort and service delivery quality. The value of doing so is seen, for instance, in the results from Bandiera, Prat, and Valletti (2009), which show that inefficiencies may be a much larger source of welfare loss in government than corruption per se.

[9] Testing at scale under government implementation is especially important for policy given recent evidence that interventions implemented by governments have systematically smaller effects than those implemented by academics or NGOs (Vivalt 2019).

Of course, this new measurement technology should only be expected to deliver improved governance when there is also political interest in doing so (as was true in this case). Thus, our results should be interpreted as demonstrating the value of expanding the feasible set of measurement for governments. Historically, innovations in measurement technologies have been a foundation for improved productivity in several settings by enabling better coordination, contracting, and management (Landes 1983; Baker and Hubbard 2004; Duflo, Hanna, and Ryan 2012). High-frequency and low-cost measurement of last-mile service delivery using phone-based monitoring could similarly enable productivity improvements in the delivery of public services.

The rest of the paper is organized as follows. Section I describes the setting and nature of the phone-based monitoring intervention. Section II describes the research design and data. Section III presents the results and a cost-effectiveness analysis. Section IV concludes.

## I. Setting and Intervention

Telangana is India's newest state, created in 2014 from Andhra Pradesh. It has a population of 35 million, with around 60 percent living in rural areas, and is relatively well-off, with per capita income 53 percent higher than the all-India average as of 2016–2017 (Government of Telangana 2016). It has maintained the same administrative infrastructure as when it was part of Andhra Pradesh, including many of the same bureaucrats and the same capital city. As a result, it is thought to be relatively well administered; Andhra Pradesh ranked third out of 19 major states in the most recent Government Performance Index (Mundle, Chowdhury, and Sikdar 2016).

Although the state is relatively well administered as compared to other Indian states, last-mile service delivery continues to be a challenge, as seen by Telangana's performance on India's two largest welfare programs—the Public Distribution System (PDS) for food security and the National Rural Employment Guarantee Scheme (NREGS) for employment security. Drèze and Khera (2015) estimates that 22 percent of subsidized food provided under the PDS does not reach intended beneficiaries, and Muralidharan, Niehaus, and Sukhtankar (2016) finds leakage of 18 to 30.7 percent of NREGS funds in the state. These facts reflect both corruption and weak state capacity to monitor and motivate effort by its workers. Thus, finding low-cost scalable ways of improving state capacity for measuring last-mile service delivery and using this to improve monitoring of workers is likely to be of broad use and interest.

### A. *The Rythu Bandhu Scheme*

The Government of Telangana (GoTS) introduced the "Rythu Bandhu Scheme" (meaning "Friend of the Farmer," hereafter abbreviated as RBS) in May of 2018 to provide capital for the purchase of agricultural inputs prior to the main agricultural season. The RBS was hailed by economists as a more efficient response to widespread farmer economic distress than common alternatives such as raising crop procurement prices or waiving farmers' loans (Subramanian 2018). It authorized payments of Rs 4,000 ($55) per acre to every farmer registered as a landholder in each of the two yearly agricultural seasons.[10] A farmer who owned half an acre of land, for example, would receive one payment of Rs 2,000 prior to the kharif cycle (monsoon/fall) and another payment of Rs 2,000 prior to the rabi season (winter/spring). The mean check amount was Rs 8,817, while the median was Rs 5,280 (see online Appendix Figure A.1 for the full distribution of check amounts).

In total, the expected outlay for RBS was approximately $0.9 billion per cropping season or $1.8 billion annually—equivalent to 7 percent of the annual state budget and 1.6 percent of the gross domestic product (GDP) of Telangana (PRS India 2018). As a fraction of GDP, this is more than three times as large as two of the most well-known cash transfer programs—Progresa in Mexico (0.4 percent of GDP) (Dávila Lárraga 2016) and Bolsa Familia in Brazil (0.5 percent of GDP) (Hellman 2015). The RBS has also contributed to a broader trend in India toward such large cash transfer programs for farmers, where there are now similar programs at the national level ($11 billion annually) and in two other Indian states (Odisha and Andhra Pradesh, worth $1.9 billion annually in each state) (Outlook India 2019, Press Trust of India 2019, Lasania 2019).

GoTS had never previously attempted this type of cash transfer program, so even aside from the unprecedented size of the transfer, figuring out implementation was nontrivial. Since GoTS did not have bank account information for landholders, they could not transfer money into the farmers' bank accounts. They instead elected to distribute physical checks to each farmer in the form of "order checks" associated with a particular bank. Farmers could go to any branch of the bank listed on the "order check" and exchange it for cash, regardless of whether or not they held an account with that bank (conditional on providing official identification matching the name on the check). The government allocated all the mandals in the state among eight banks, assigning all farmers in a given mandal to the same bank. The Department of Revenue managed the printing of checks (including confirming eligibility of beneficiaries), which occurred in April and May, after which the checks were transferred to the Department of Agriculture for distribution.

The government had originally intended to distribute checks in April and May so that farmers had enough time to apply them toward agricultural investments prior to the onset of the monsoon in June. Typical investments include purchasing seeds and fertilizer and hiring workers to prepare their fields. It is optimal for farmers to plant soon after the monsoon rains arrive; if they wait, then seeds fail to benefit from the rain that has already fallen and yields are lower.[11] However, due to delays in

[10] The government's land registry was updated and digitized in late 2017 and early 2018 to reflect the identity of current landholders.

[11] Giné, Townsend, and Vickery (2007) measure expenditure patterns of farmers in the Mahbubnagar district of Telangana and find that 37.4 percent of all expenditures for a season occur prior to the arrival of the monsoon. This includes 30–40 percent of all spending on bullock, manual, and tractor labor; 90.1 percent of all spending on manure; and 59.4 percent of spending on irrigation. Additional fertilizer is typically purchased and applied later in the season.

implementation, the check distribution process did not begin until May 8, creating additional pressure for speedy delivery.[12]

The Department of Agriculture managed the distribution of checks, with MAOs responsible for distribution in their respective mandals. In most cases, a MAO was responsible for one mandal, though due to vacant MAO positions, there were a few cases in which a single MAO oversaw implementation across multiple mandals (498 MAOs covered 548 mandals). An average mandal contains roughly 60,000 individuals living in 20 villages, of whom around 10,000 are landholders receiving checks. The program was widely discussed in the media, and beneficiary awareness about the program was high. MAOs were assisted by junior officials in informing village-level leaders regarding the dates of the village-level meetings for distributing these checks and in implementing the large-scale check distribution process.

The MAOs scheduled one meeting in each village. During those meetings, the MAOs supervised teams of government employees who verified the identities of landowners residing in that village and distributed their checks to them. The first village meetings were held on May 8. Each mandal then had one to two village meetings per day, with meetings every day of the week except Sunday, until all villages had been served. There was only one meeting per village, so if a landholder did not get their check then, they had to pick it up at the mandal-level Department of Agriculture offices. To do so, however, they needed signed approval from both the MAO and another bureaucrat in the revenue department, which could be a substantial barrier if for example one or both were absent; beneficiaries may also not have understood the process to follow for picking up a check outside of the village meetings.

Implementing RBS well was a major priority for the government given the sum disbursed, the number of recipients (5.7 million), the high media profile of the scheme, the fact that the government had never before done anything comparable, and upcoming elections. Anticipated risks included (i) nonissue of checks; (ii) nondelivery of checks; (iii) late delivery of checks, which would force farmers to reduce investment or borrow at high rates to finance time-sensitive agricultural inputs; and (iv) corruption during the distribution process (e.g., bribe demands).[13]

### B. *Phone-Based Monitoring Intervention*

The state government had previously collected phone numbers for farmers as part of land record digitization. Overall, 3.5 million (61 percent) of the 5.7 million entries in the registry listed a contact number. GoTS contracted a call center to collect data from beneficiaries between May 29 and June 15.[14] The call center attempted to reach a random sample of 46,263 farmers representative of those with

[12] The two main delays were in the printing of checks and ensuring that bank branches held sufficient cash to permit encashment of nearly a billion dollars in checks over a short period of time.

[13] While this is of course a particular type of service (delivery of cash), the problems are similar to the delivery of many other services in the last mile. Even when there is no corruption per se, limited state capacity for measuring and monitoring the effectiveness of service delivery staff can contribute to slack in their effort (as seen for instance in the absence of workers).

[14] The intervention was designed by the research team but implemented by GoTS.

listed phone numbers in the GoTS administrative records. A sample of 150 farmers per treatment mandal and 50 per control mandal was randomly selected, stratified by quintile of the mandal-specific land size distribution to reduce variability across mandals due to sampling.[15]

The call center placed calls to the mobile phone numbers of sampled farmers. If a call did not connect, the call center would attempt to reach that number up to five more times over the following two days before giving up. If connected, the call center operator verified the respondent's identity and identified themselves as conducting a survey on behalf of the Government of Telangana to understand the respondent's experience with the Rythu Bandhu Scheme. Calls collected information on whether, where, and when the farmer received their check; whether and when they encashed it; any problems receiving or encashing the check (including time costs and bribes); how they used the funds; suggestions for future rounds of RBS; and overall satisfaction with RBS. Sampled farmers were surveyed once, and the average completed survey lasted four minutes. The call center successfully completed calls with 22,127 (47.8 percent) of the sampled farmers, began but did not complete surveys with another 24 percent, had 10 percent decline to participate, and could not reach the remaining 18 percent for other reasons (e.g., phone number was no longer valid).[16] The rate of phone call completion did not differ across treatment and control mandals (47.6 percent in control, 48.0 percent in treatment; $p = 0.39$).

Prior to the distribution of checks and calls to beneficiaries, the Telangana Department of Agriculture informed treatment MAOs that their mandals had been selected by lottery to take part in a pilot of the phone-based monitoring system. During a special video conference with the treatment MAOs (on May 2, 2018), the state commissioner of agriculture explained the initiative and the data that would be collected. He informed them that reports from the phone call data would be provided to them and their supervisors, including a performance rating for their mandal. The MAOs were told which outcomes the report would cover but not the specific formula for calculating ratings. On May 10, 2018, the Department of Agriculture sent treatment MAOs a follow-up letter containing the same information. To reduce the risk of spillovers, treatment MAOs were explicitly told the identity of other treatment MAOs in their district and that no other MAOs in their districts were part of the pilot.

In order to give MAOs time to distribute checks and thus have their performance accurately reflected in the call data, phone calls to farmers commenced on May 29, three weeks after check distribution began, and continued through June 15. Reports based on the phone data were issued to treatment MAOs and their supervisors between July 9 and 13. The reports listed five metrics: the proportion of farmers

[15] Control MAOs were not informed about the phone calls to farmers in their mandals, since data from these calls were not used to prepare reports or even shared with the Department of Agriculture. The data from control mandals were collected solely for research purposes. As discussed in the preanalysis plan, we had originally anticipated using the phone data to compare between treatment and control areas on outcomes such as corruption and satisfaction that were not available in the administrative data. Although we report these outcomes for completeness in online Appendix Table A.2, this analysis turned out to be highly underpowered due to extremely low rates of corruption and high rates of beneficiary satisfaction.

[16] The vendor also piloted automated calls (IVR), but these had a high error rate in capturing responses and were therefore discontinued.

Table 1—Timeline of Project

| Date(s) | Activities |
|---|---|
| September to December 2017 | Land records in Telangana are updated and digitized |
| February 28, 2018 | Rythu Bandhu Scheme is announced |
| April 1, 2018 | First meeting between JPAL research team and Government of Telangana |
| April to May 2018 | Rythu Bandhu checks printed and distributed to MAOs |
| May 2, 2018 | Treatment MAOs informed of intervention (via video conference) |
| May 8, 2018 | Check distribution begins |
| May 23, 2018 | At this point, 50 percent of all checks have been distributed and encashed (60.2 percent of all checks that will ever be encashed) |
| May 29, 2018 to June 15, 2018 | Call center collects data |
| June 8, 2018 | Monsoon rains arrive in Telangana |
| June 9, 2018 to June 11, 2018 | Phone survey of MAOs is conducted |
| June 15, 2018 | At this point, 75 percent of all checks have been distributed and encashed (90.4 percent of all checks that will ever be encashed) |
| July 9, 2018 | Reports sent to treatment MAOs and their supervisors. At this point, 79.4 percent of all checks have been distributed and encashed (95 percent of all checks that will ever be encashed) |
| September 26, 2018 | Research team receives bank records of check distribution through this date |

who reported receiving their check, receiving it before May 20 (a measure of speed of delivery), successfully encashing it at the bank, being asked for a bribe, and being satisfied with the program overall. They rated performance on these metrics for the mandal in question relative to other mandals within the same district and relative to the state overall. They also showed a simple, color-coded categorical rating ("Poor," "Fair," "Good," or "Excellent") based on absolute performance, motivated in part by the finding of Callen et al. (2018) that "flagging" of high or low performers can make performance data more accessible. A redacted example report is in online Appendix B. Table 1 provides the full timeline of the Rythu Bandhu Scheme and phone-based monitoring intervention.

The treatment of phone-based monitoring in principle includes both a monitoring component (which potentially altered incentives) and an information component (which could help MAOs or the MAOs themselves do their jobs better). In practice, and because of the program's compressed time frame, the information provided by the reports came too late for District Agricultural Officers (DAOs) or MAOs to react to it. The program aimed to distribute all checks between early May and mid-June, whereas reports were issued in early July (by the time of which 95 percent of checks ever issued had already been issued). Our estimates thus reflect the impact of MAOs knowing they were being monitored but do not capture the potential additional benefits of DAOs and MAOs having timely information from the calls.

The Department of Agriculture did not inform control MAOs about the existence of the pilot. If asked, it said that the initiative might be extended to their areas in the future but not during the current season. While the call center collected phone data from control mandals, it did not generate reports using these data or inform control MAOs of their existence. Of course, the interpretation of reduced-form intent-to-treat effects depends on treatment and control MAOs' beliefs, which we discuss later.

As discussed in the introduction, MAOs may respond to this monitoring and the creation of reports, even in the absence of formal incentives, for several reasons. These include wanting to look good on rankings of performance relative to their peers and not wanting to get admonished for poor performance (which was made more likely by the availability of data on absolute and relative performance).

There were several ways in which MAOs and their staff could potentially improve their performance in response to this additional monitoring. They could improve processes to ensure that checks were distributed to all eligible beneficiaries. They could work harder to find recipients—either before the village meetings by publicizing them more thoroughly, during the meetings by extending the meeting length, or after them by following up with those who did not attend. They could also demand fewer bribes.[17]

Yet one might also reasonably expect phone-based monitoring to have limited effects in this setting, as government scrutiny of RBS implementation was already high: MAOs digitally recorded whether checks were distributed, and banks recorded check encashment in order to claim reimbursement. The availability of high-quality administrative data on outcomes makes the RBS an unusually low-cost setting in which to measure effects of phone-based monitoring, but it could also lead to the effects being lower than they would be in other settings where phone data are the *only* performance information available. Our estimates should thus be interpreted as the effects of adding an incremental, independent source of monitoring and making this salient to implementing officials.

## II. Research Methods

Our design and methods follow a registered preanalysis plan.[18]

### A. *Experimental Design*

The study population consists of nearly all households eligible to receive RBS, i.e., all landholding households in Telangana. We excluded 1 largely urban district (Hyderabad), as it had very few program beneficiaries, leaving 30 remaining districts. Since we randomized across nearly the universe of mandals in the state (outside Hyderabad), the study sample is representative of the rural population of the state.

Within these districts, we randomly assigned treatment at the level of the MAO, since some MAOs oversee multiple mandals. We randomly selected approximately 25 percent of MAOs for treatment, yielding a total of 122 treatment MAOs and 376 control MAOs. This corresponded to 131 treatment and 417 control mandals.

[17] MAOs were only responsible for check distribution. A different government department (Revenue) printed checks *before* our intervention after verifying farmer eligibility, and banks independently checked farmer identity before cashing the checks. Improvements in the delivery of benefits are thus unlikely to have come at the cost of lower scrutiny of eligibility requirements.

[18] See https://www.socialscienceregistry.org/trials/2942.

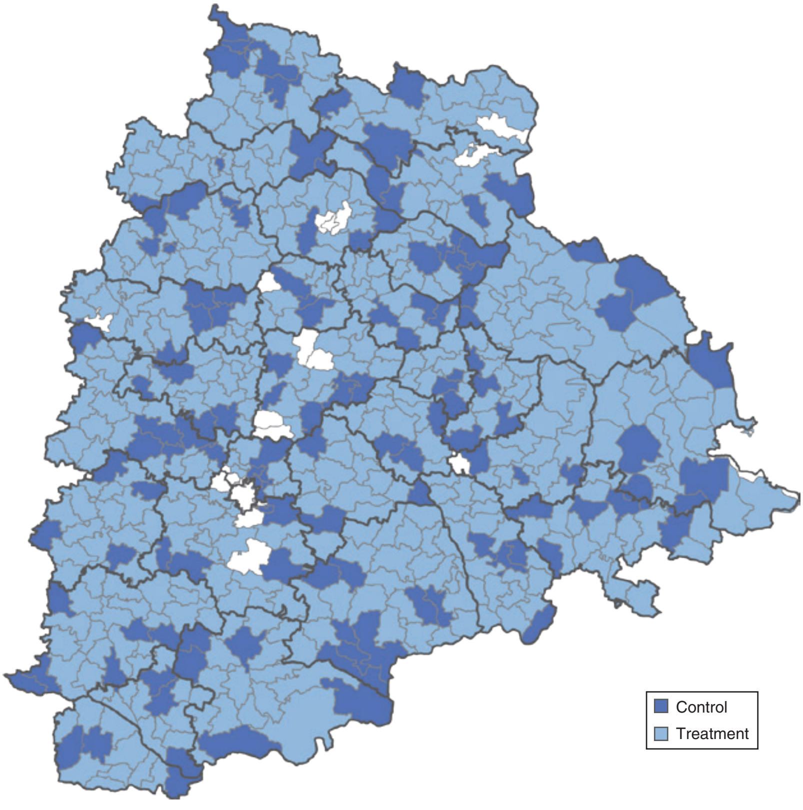


FIGURE 1. STUDY AREAS WITH TREATMENT AND CONTROL MANDALS

*Notes:* This map shows the geographical distribution of treatment and control mandals (subdistricts) across the entire state. Dark black lines indicate district boundaries, whereas gray lines are mandal boundaries. Randomization was stratified by district and occurred at the mandal agricultural officer level. Mandals in white were not included in the randomization and study. This typically occurred because the mandal is urban, such as those around Hyderabad, or did not have an MAO assigned to it, so it was not possible to implement the treatment. Note that since there are ten cases where a treatment MAO oversees multiple geographically contiguous mandals, there is slightly more geographical clustering of treatment mandals than would occur due to chance.

We stratified randomization within each district on an indicator for whether an MAO oversees multiple mandals, the only MAO-level covariate available to us at the time of randomization. (Further details of the randomization algorithm are in the preanalysis plan.) Figure 1 shows the geographical distribution of treatment and control mandals.

We test for balance on a broad range of characteristics and report them in Table 2, grouping covariates into four broad categories: (i) characteristics of Rythu Bandhu beneficiaries, (ii) general characteristics of the mandal (e.g., demographics, wealth), (iii) characteristics of the mandal that may be related to implementation of Rythu Bandhu, and (iv) MAO characteristics and past performance. For the characteristics of Rythu Bandhu beneficiaries, the unit of observation is at the individual farmer level, while for the other three categories, the unit of observation is either the mandal or MAO. We regress the variable of interest on treatment status and randomization strata fixed effects.

Table 2—Balance Tests

| Variable | Control mean | Treatment mean | Difference | Standard error | $p$-value |
|---|---|---|---|---|---|
| *Panel A. Characteristics of Rythu Bandhu beneficiaries* | | | | | |
| Land ownership (acres) | 2.22 | 2.18 | −0.02 | 0.05 | 0.72 |
| Median land size | 1.57 | 1.54 | −0.00 | 0.05 | 0.92 |
| Land size—twenty-fifth percentile | 0.65 | 0.65 | 0.02 | 0.04 | 0.63 |
| Land size—seventy-fifth percentile | 2.97 | 2.87 | −0.06 | 0.06 | 0.33 |
| Registered mobile numbers | 0.61 | 0.61 | −0.01 | 0.02 | 0.67 |
| Farmers per MAO | 10,311.52 | 10,016.79 | −287.57 | 362.70 | 0.43 |
| *Panel B. General mandal characteristics* | | | | | |
| Mandal population | 46,714.75 | 49,894.53 | −3,264.32 | 2,603.48 | 0.21 |
| Share of female population | 0.50 | 0.50 | −0.00 | 0.00 | 0.57 |
| Share of SC population | 0.18 | 0.18 | −0.00 | 0.01 | 0.74 |
| Share of ST population | 0.13 | 0.14 | 0.02 | 0.01 | 0.06 |
| Literacy rate in mandal | 0.59 | 0.59 | −0.00 | 0.01 | 0.56 |
| Share of working population | 0.52 | 0.52 | 0.01 | 0.00 | 0.24 |
| Share of villages with paddy as main crop | 0.68 | 0.65 | −0.01 | 0.04 | 0.86 |
| Share of irrigated land | 0.52 | 0.51 | −0.01 | 0.04 | 0.76 |
| Share of electrified villages | 0.95 | 0.94 | −0.02 | 0.02 | 0.36 |
| Average village distance from Hyderabad | 135.91 | 134.76 | 0.32 | 2.09 | 0.88 |
| *Panel C. Mandal characteristics related to Rythu Bandhu* | | | | | |
| Number of banks in mandal | 3.52 | 4.12 | −0.26 | 0.35 | 0.47 |
| Average distance to nearest ATM | 12.72 | 12.43 | −0.18 | 0.47 | 0.70 |
| Share of HHs using banking services | 0.45 | 0.43 | −0.01 | 0.02 | 0.45 |
| Average distance to nearest bank | 7.51 | 7.22 | −0.15 | 0.31 | 0.62 |
| Share of villages with all-weather road | 0.91 | 0.91 | 0.00 | 0.01 | 0.79 |
| Share of HHs owning mobile phones | 0.52 | 0.50 | −0.01 | 0.01 | 0.70 |
| Average rainfall in mandal 2013–2017 (mm) | 707.35 | 714.26 | 8.76 | 10.19 | 0.39 |
| *Panel D. MAO characteristics* | | | | | |
| Age of MAO | 35.57 | 36.36 | 0.89 | 0.76 | 0.24 |
| Gender of MAO (female = 1) | 0.30 | 0.33 | 0.02 | 0.05 | 0.65 |
| Number of SHC samples (2016) | 993.66 | 911.41 | −68.33 | 69.74 | 0.33 |
| Number of farmers covered by SHCs (2016) | 4,321.85 | 4,268.36 | −64.72 | 310.09 | 0.83 |
| SHC tests conducted (2016) | 211.28 | 180.75 | 1.05 | 39.21 | 0.98 |
| SHCs available on portal (2016) | 806.44 | 770.22 | 50.26 | 154.57 | 0.75 |
| Number of SHC samples (2017) | 1,030.01 | 961.11 | −64.65 | 54.10 | 0.23 |
| Number of farmers covered by SHCs (2017) | 4,470.67 | 4,572.42 | 34.66 | 251.47 | 0.89 |
| SHC tests conducted (2017) | 976.04 | 924.91 | −44.77 | 52.32 | 0.39 |
| SHCs available on portal (2017) | 4,176.72 | 4,332.26 | 85.98 | 240.81 | 0.72 |
| Joint $p$-value | | | | | 0.15 |
| Observations | 417 | 131 | 548 | | |

*Notes:* Differences in panel C are estimated through regressions on a treatment indicator, with fixed effects at the randomization strata level. Standard errors are clustered at the MAO level and reported in parentheses. For distance to nearest bank or ATM, the census records the distance to the nearest bank and ATM as one of three categories (less than 2.5 kilometers (km), 5–10 km, or more than 10 km). We substitute 2.5 km, 7.5 km, and 20 km, respectively, for each of these categories and calculate the average distance, weighting each village according to population. This produces the average distance across all farmers in the mandal.

The first category takes farmer-level characteristics from the Government of Telangana landholder registry (Government of Telangana 2018). The registry was updated in late 2017, shortly before the implementation of Rythu Bandhu, and includes information on all the land-owning farmers in the state (e.g., landholdings, percent of farmers with mobile phone numbers in the land registry and thus who could be called for the intervention).

In the second category, we use mandal-level data from the 2011 census of India, the most recent round of India's decennial census of all its residents (Government of India 2011).[19] Most of these characteristics (e.g., demographics, literacy, levels of irrigation) are slow to change, so 2011 levels should provide a reasonably good approximation of levels at the time of Rythu Bandhu.

In the third category, we use both the 2011 census of India and some more recent data available from the Government of Telangana that may be directly relevant to Rythu Bandhu implementation. The 2011 census of India contains information on the total number of banks in a mandal, the average distance from each village in the mandal to a bank or ATM, the share of households using banking services, the share of households owning cellphones, and the share of villages with an all-weather road, all of which may affect the ease of receiving and encashing a check.[20] Since income levels may have shifted since 2011, we also check for and reject differences across treatment and control in total rainfall between 2013 and 2017 (Government of Telangana 2019 ), as rainfall is an important determinant of income in these areas.

Finally, we test for balance on MAO gender, age (as a proxy for experience), and competence. Since there is no prior period in which we can observe their performance in distributing Rythu Bandhu checks, we measure MAO competence based on their performance in implementing another flagship program of the Government of India: the provision of soil health cards to farmers.[21] We use mandal-level data available from years prior to the start of Rythu Bandhu (2016 and 2017) on how many soil samples were collected, how many farmers were covered by those samples, the number of samples that were tested, and the number of soil health cards made available to farmers through the government's online portal (Department of Agriculture, Cooperation, and Farmers Welfare 2018).

Overall, across the 33 tests, only 1 demographic measure (Scheduled Tribe population share) is significant at the 10 percent level, and the joint $p$-value is equal to 0.15. As a robustness check, we have run all the main regressions with scheduled tribe population share included as a control, and it does not affect the results. Thus, we can be confident that any observed differences between treatment and control mandals in Rythu Bandhu performance outcomes can be attributed to the treatment of announcing phone-based monitoring to randomly selected MAOs.

[19] Some mandal boundaries shifted after the creation of the state of Telangana. To match present-day mandals to 2011 census data, we take the GPS location of 2011 census villages and towns as well as present-day mandal boundaries. The 2011 census characteristics of mandals are created from 2011 census data for those villages and towns that fall within their boundaries.

[20] While there has been an expansion of banking services across India since 2011, the 2011 levels are still informative, as the relative level of banking services across space, which is what matters for balance, has remained fairly static: there is a correlation of 0.98 between the number of banks per district in Telangana in 2011 and 2018 (Reserve Bank of India 2018).

[21] Under this program, the Department of Agriculture collects a large number of soil samples from across each mandal, tests the samples to assess soil health (e.g., macro and micronutrients, pH levels), and provides farmers with Soil Health Cards that contain recommendations on usage of fertilizer and other agricultural inputs. To date, over 1 million soil samples have been collected from Telangana farmers, with an average of slightly less than 1,000 per mandal annually.

### B. *Data*

We primarily measure outcomes using administrative data, including (i) the register of all agricultural landholders in the state, including names, village, acres held, and a contact phone number; (ii) a farmer-level record of check distribution maintained by the MAOs;[22] and (iii) farmer-level bank records of check encashment. Our analysis focuses on encashment, as getting the money is the ultimate outcome of interest to policymakers. Bank reports of encashment were recorded in real time and were the basis for reimbursement from the government; manipulating them would constitute serious fraud and could jeopardize a bank's operating license. We find that they closely match encashment as reported by the surveyed farmers, with agreement in 88.6 percent of cases.

We focus on check encashment status at two dates. The first (June 8) captures on-time delivery. This was exactly a month after the start of distribution and is the date on which the monsoon arrived in Telangana in 2018 according to the Indian Meteorological Department (Express News Service 2018). Most planting activities begin with the arrival of the monsoon, and the government's goal was to ensure that farmers had funds in place at the start of planting to buy seeds and hire labor. This was a high priority for GoTS since a key goal of the program was to break the cycle of farmer debt, which was widely believed to be a driver of farmer suicides. The second (September 26) captures whether the checks were *ever* encashed. This is after the last date (August 15) on which the checks were valid for encashment and thus should well approximate the final distribution of checks.[23]

We also use data from phone calls conducted by the call center as a secondary data source. These data were collected over the phone from 22,127 program beneficiaries, as described above. This source of data provides a picture of encashment up until the phone survey was completed (June 15th) but not after that. The administrative data, which covers all encashments through September 26th, is more reliable in determining whether and when checks were encashed, but the phone data contains information on additional outcomes of interest, such as beneficiary satisfaction and corruption.[24]

Finally, we use data from a short phone survey of MAOs. We surveyed 88 of 122 treatment MAOs and a sample of 54 control MAOs.[25] Surveys covered their awareness of the pilot and beliefs about their treatment status. We were concerned that surveying control MAOs might cause them to believe incorrectly that they

[22] We treat the records of check distribution maintained by the MAOs with caution, as there are a number of problems with this data. See online Appendix C for further description of the issues with the MAO data.

[23] Checks were printed in four tranches, on April 19, May 1, May 10, and May 15, and were valid for three months from the date of printing. It was possible to get checks reprinted, but by September 26, encashment activity had largely ceased.

[24] In principle, the bank data could have also been used for accountability purposes, but, in practice, such high-quality administrative data are unlikely to be available for other programs. Thus, while the bank data is clearly better for research (since it has data on the universe of checks issued), the phone-based measurement of beneficiary experience is a more generalizable intervention for governments to deploy in other sectors. Thus, our *intervention* focuses on the more scalable phone-based measurement and monitoring, while our *evaluation* uses the higher-quality bank data.

[25] We attempted surveys with all of the treatment MAOs (among whom the response rate was 72 percent) and a random sample of 2 control MAOs per district (60 percent response rate).

were being monitored by the phone-based monitoring system and thus affect their behavior (Hawthorne effects). As a result, we only conducted these surveys with a small sample of randomly selected control MAOs after the distribution was mostly complete.[26]

### C. *Estimation*

We report intent-to-treat estimates, comparing mean outcomes in treatment and control areas. We discuss MAO beliefs and their implications for interpretation in our cost-benefit analysis below. We thus estimate

$$y_{ivmsd} = \alpha + \beta T_{msd} + \delta_{sd} + \gamma \mathbf{X}_{ivmsd} + \epsilon_{ivmsd}, \tag{1}$$

where $y$ is an outcome, $T$ an indicator for assignment to treatment, and $\mathbf{X}$ a vector of prespecified covariates. In practice, there is only one covariable: the size of landholdings, binned into 40 evenly sized bins (i.e., 0 to 2.5th percentile of landholdings, 2.5th to 5th percentile, etc.).[27] Indices denote individual $i$ in village $v$ in mandal $m$ in stratum $s$ in district $d$. Treatment is strictly exogenous conditional on the randomization stratum fixed effects $\delta_{sd}$. We cluster standard errors at the level of treatment assignment (the MAO) and conduct randomization inference as a robustness check. When using call center data, we reweight estimates by the inverse probability of being sampled.

## III. Results

### A. *Effects on Overall Program Performance*

Overall, RBS implementation was imperfect but still fairly successful compared to many other similar programs. 4.03 million farmers (69 percent of target) encashed their checks before monsoon onset on June 8 (Table 3, column 2). After 5 months, 4.8 million farmers, 83 percent of the total, had encashed their checks (column 4). It appears that corruption was not a major issue, with only 2 percent of farmers reached by phone reporting that they had to pay a bribe to obtain their checks.[28] The

[26] On the survey date, 84 percent of checks that would ever be encashed had already been encashed. Online Appendix Table A.1 tests for and rejects the presence of Hawthorne effects, as control MAOs randomly selected for the survey (or actually surveyed among the set who were selected) do not have any higher or lower rates of encashment than other control MAOs. Given that being surveyed has no effect on outcomes, and that the surveys were conducted after the treatment effects are seen (see Section IIIA), the small difference in survey response rates between treatment and control MAOs for the survey cannot have affected overall encashment outcomes.

[27] Our preanalysis plan specifies that we will control for the size of landholdings. We implement this with indicators for percentile of landholding, as opposed to exact landholding, to preserve anonymity of individual farmers in the data. The results are nearly unchanged if we include the exact landholding as a control instead.

[28] We also find slightly higher reported encashment rates (75.5 percent) in our phone call data than in the corresponding administrative records (73.6 percent), suggesting that officials did not collude with banks to encash beneficiary's checks without their knowledge. The call center provides further evidence against such collusion, with only 0.02 percent of beneficiaries reporting that someone else was given their check and only 1.5 percent stating that the check wasn't available when they tried to pick it up (which could be consistent with either fraud or poor organization by the MAOs). There is no effect of the treatment on the prevalence of either of these problems.

Table 3—Effect on Encashment Outcomes

| | Encashed before June 8th | | Ever encashed | | |
|---|---|---|---|---|---|
| | Treatment (1) | Control mean (2) | Treatment (3) | Control mean (4) | Observations (5) |
| Overall | 0.0240 (0.00774) | 0.69 | 0.0134 (0.00640) | 0.83 | 5,645,937 |
| *Land quartiles* | | | | | |
| Quartile 1 | 0.0277 (0.00913) | 0.52 | 0.0223 (0.00885) | 0.68 | 1,449,482 |
| Quartile 2 | 0.0248 (0.00791) | 0.71 | 0.0144 (0.00632) | 0.85 | 1,460,294 |
| Quartile 3 | 0.0241 (0.00755) | 0.76 | 0.0112 (0.00601) | 0.88 | 1,443,788 |
| Quartile 4 | 0.0209 (0.00803) | 0.77 | 0.00710 (0.00620) | 0.89 | 1,443,836 |
| Test of $H_0$: $\beta_{Q1} = \beta_{Q2} = \beta_{Q3} = \beta_{Q4}$ | 0.72 | (0.54) | 2.03 | (0.11) | |
| *Phone coverage* | | | | | |
| Non-phone-owners | 0.0252 (0.00989) | 0.57 | 0.00913 (0.0103) | 0.72 | 2,254,142 |
| Phone owners | 0.0207 (0.00810) | 0.76 | 0.0131 (0.00553) | 0.90 | 3,543,258 |
| Test of $H_0$:$\beta_{No\text{-}Phone} = \beta_{Phone}$ | 0.12 | (0.70) | 0.29 | (0.59) | |

*Notes:* All specifications include fixed effects at the randomization strata fixed level. Standard errors are clustered at the MAO level and reported in parentheses. The bottom row of each panel reports the $F$-statistic and $p$-value from a test of the null that coefficients are statistically similar across categories. Models are estimated using administrative data at the individual check level, as a handful (0.8 percent) of individuals in the database were issued multiple checks. According to the Revenue Department, amounts above Rs 50,000 (12.5 acres of land) were split into multiple checks. Outcomes are essentially perfectly correlated within individuals, as farmers either picked up and encashed all or none of their checks, which is accounted for by clustering at the mandal level. Farmers with less than 0.025 acres of land (less than 1 percent of the sample) were still issued checks but in the amount of Rs 100.

median farmer had a lag of only six days between receiving a check and successfully encashing it at the bank.

Despite good program implementation to begin with, phone-based monitoring led to further improvements. Treatment increased the probability of "on-time" check encashment (defined as prior to the arrival of the monsoon on June 8) by 2.4 percentage points, ($p = 0.002$, Table 3),[29] which is equal to a 0.25 standard deviation improvement (standard deviation of the mandal-level means of on-time check encashment in control mandals was 9.7 percentage points). The probability that farmers ever encashed their checks also increased by 1.3 percentage points ($p = 0.037$), a 0.16 standard deviation improvement (standard deviation of mandal-level means in control mandals was 8.1 percentage points).[30] Given the

[29] The onset of the monsoon changes from year to year—if we instead used the dates of monsoon onset from 2015 to 2017 (June 13, 2015, June 19, 2016, and June 12, 2017), then the $p$-values would be nearly the same ($p = 0.002, 0.005, 0.002$, respectively) and our conclusion would not change.

[30] For completeness, we also report effects on check distribution in online Appendix C, based on MAO records of whether each check was distributed and the date of distribution. We treat these data with caution, as they were uploaded by MAOs with substantial lags, causing date of distribution to be mismeasured, and were not subject to

relatively high-quality implementation in the control group, an alternative way of quantifying the impact of the program is that it reduced the fraction of farmers not receiving their checks on time by 7.9 percent (2.4 percentage points on a base of 31 percent) and ever receiving their checks by 7.8 percent (1.3 percentage points on a base of 17 percent). These are nontrivial rates of improvement—especially given the "light touch" nature of the intervention and its low cost.

Multiplied by the large number of affected farmers (1.3 million), this translates into 31,828 additional farmers encashing their checks prior to the monsoon in treatment areas, or more than 1 additional farmer per phone call completed by the call center. Furthermore, an additional 17,771 farmers ever encashed their transfers as a result of the treatment. If the program had been extended to the entire state, our estimates suggest that an additional 77,657 farmers would have received the transfer to which they were entitled during this agricultural cycle (and double that amount annually). These gains highlight the policy impact of programs like phone-based monitoring that can be feasibly implemented at large scale, where even modest rate improvements translate into meaningful absolute gains.

As seen in Table 4, conditional on ever encashing, the treatment lowered the mean number of days that passed before recipients encashed their checks by three-fourths of a day ($p = 0.039$). Figure 2 summarizes the main effects visually. The top panel plots the proportion of checks encashed by date in the treatment and control groups separately, while the bottom panel plots regression estimates of the treatment effect by date. The gap between treatment and control peaks at 2.8 percentage points (5.2 percent) on May 25 ($p = 0.004$) and then narrows, asymptoting to 1.3 percentage points by the end of September.

The larger gap between treatment and control at the end of May is likely a better estimate of the effect of phone-based monitoring on bureaucratic effort, because MAOs had the greatest ability to affect check delivery outcomes during the meetings in each village, nearly all of which occurred in May. After that, households who did not get their checks had to visit government offices to pick them up. Thus, the shrinking gap over time reflects a natural catch-up in which control households who had not received their checks in village meetings had to exert costly effort to get their checks.

Data from the call center only capture the experiences of a small and nonrepresentative sample of the farmers (0.4 percent) but sheds light on the process changes underlying these gains. Consistent with the largest gap between treatment and control being at the end of May, respondents in treated areas were more likely to receive their checks during village meetings ($p = 0.09$) (online Appendix Table A.2). Since this was far less costly than the alternative (traveling to mandal headquarters and going through an additional layer of bureaucracy to get their check), this is another channel through which the treatment improved beneficiary outcomes. We find no treatment effects on either the likelihood that phone call respondents were asked to pay a bribe or their overall satisfaction with the scheme (columns 2 and 3). This is not surprising given that baseline rates of corruption were

penalties for misreporting like those banks faced. See online Appendix C for further description of the issues with the MAO data. That said, the results are broadly consistent with those for encashment.

Table 4—Effect on Time to Encashment

| | Days to encashment | | |
|---|---|---|---|
| | Treatment (1) | Control mean (2) | Observations (3) |
| Overall | −0.785 | 20.16 | 4,663,678 |
| | (0.381) | | |
| Land quartiles | | | |
| Quartile 1 | −0.689 | 24.00 | 984,273 |
| | (0.508) | | |
| Quartile 2 | −0.675 | 20.08 | 1,239,638 |
| | (0.383) | | |
| Quartile 3 | −0.841 | 18.71 | 1,278,121 |
| | (0.359) | | |
| Quartile 4 | −0.989 | 18.80 | 1,284,764 |
| | (0.368) | | |
| Test of $H_0$: $\beta_{Q1} = \beta_{Q2} = \beta_{Q3} = \beta_{Q4}$ | 1.33 | (0.26) | |
| Phone coverage | | | |
| Non-phone-owners | −1.340 | 22.14 | 1,614,191 |
| | (0.453) | | |
| Phone owners | −0.493 | 19.13 | 3,172,605 |
| | (0.392) | | |
| Test of $H_0$:$\beta_{No\text{-}Phone} = \beta_{Phone}$ | 5.02 | (0.03) | |

*Notes:* All specifications include fixed effects at the randomization strata fixed level. Standard errors are clustered at the MAO level and reported in parentheses. The bottom row of each panel reports the $F$-statistic and $p$-value from a test of the null that coefficients are statistically similar across categories. Models are estimated using administrative data at the individual check level, as a handful (0.8 percent) of individuals in the database were issued multiple checks. According to the Revenue Department, amounts above Rs 50,000 (12.5 acres of land) were split into multiple checks. Outcomes are essentially perfectly correlated within individual, as farmers typically encashed all of their checks at the same time, which is accounted for by clustering at the mandal level.

extremely low (2 percent of respondents) and satisfaction with the Rythu Bandhu Scheme was high (93 percent of respondents).[31]

Recall that there are two channels through which phone-based monitoring may affect beneficiary outcomes: an incentive channel, where MAOs exert more effort when they know they are being monitored, and an information channel, where they use the information provided by the report to reallocate effort more optimally. Given that reports were issued after nearly all encashment had occurred (July 5–9), it seemed a priori unlikely that it would affect outcomes through information

[31] We can also examine the treatment effect on check encashment in the call center data. While this is a nonrepresentative and substantially smaller sample than the administrative records, it does let us check for consistency between estimates of impact in the two data sources. Taking the sample of farmers who were reached by phone, the estimated effect on encashment as measured by phone calls is not significantly different from the estimated treatment effect on encashment as measured using administrative records (online Appendix Table A.8, $p = 0.46$). Note that for this test, we adjust the administrative records to the date of the call (e.g., whether the call center collected information from the respondent on June 7th, whether the check was recorded as encashed in the administrative data on or before June 7th). The estimated treatment effect is not statistically significantly different from zero in either dataset, reflecting both the selected nature of the sample reached by phone and its much smaller size (22,537 as opposed to 5.7 million). It is reassuring to note that the control means are nearly identical, suggesting that the administrative data and phone call data are of comparable quality.

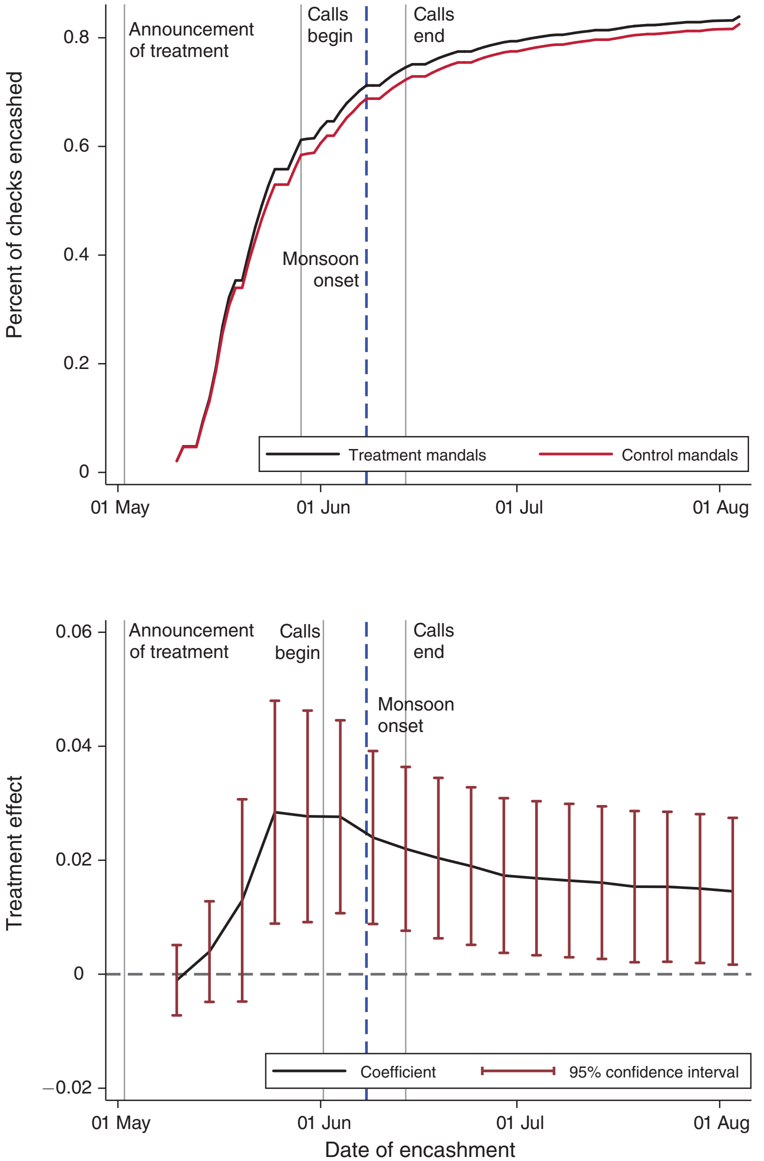


Figure 2. Treatment Effect, By Date

*Notes:* The two graphs in this figure report (a) the cumulative rate of encashment in treatment and control mandals by day and (b) the coefficient of treatment effect on the cumulative rate of encashment over the period of check distribution in our data. The coefficients in the bottom graph are estimated through regressions with fixed effects at the randomization strata level and standard errors clustered at the MAO level. Less than 1 percent of checks were encashed after August 4 or before May 10, so the axis is restricted to those time periods. The monsoon arrived on June 8th, so we consider this the date of "on-time" encashment.

effects. As shown in Figure 2, the treatment effects are concentrated during May and June, well before the phone-based monitoring reports were shared. There is also no evidence of a differential change in encashment rates following the distribution of the reports themselves.[32] Overall, the timing of the effects implies that they were driven by treated MAOs' knowledge that they were being monitored and their anticipation of future reports being made available to their superiors rather than by the information in the reports themselves.

One concern is that these estimates may be biased due to spillovers across mandals. Knowledge of the intervention may reallocate the attention of those who supervise MAOs toward treatment mandals and away from control mandals, if supervisors do not wish to have their subordinates who are being monitored look bad. We directly test for this possibility by taking advantage of random variation in treatment intensity within bureaucratic units. Districts in Telangana are divided into "revenue divisions" (up to 5 per district) and then into mandals (4 to 15 per revenue division). Although roughly the same fraction of mandals were treated in each district, we did not stratify the randomization at the revenue division level. As a result, there is random variation in the fraction of MAOs within each revenue division that are treated. If there were diversion of revenue division supervisor-level attention, we should expect worse performance among control MAOs with more treated MAOs in their revenue division, as these control MAOs would get less attention paid to them. We find no evidence of this (online Appendix Table A.4).

A second concern is that we are picking up a mechanical effect of the phone calls themselves rather than the impact on officials. Receiving a call may have acted as a "reminder" for respondents to pick up their checks, so the treatment effect could be a result of more individuals in treatment mandals receiving phone calls. Any such effect would necessarily be too small to fully explain our results, as calls only reached 0.82 percent of treatment farmers and 0.27 percent of control farmers, a difference that is smaller than the observed treatment effect.[33] Since farmers were randomly selected for phone calls, we directly test the effect of being called in online Appendix Table A.5. Farmers who were randomly selected for calls are not more likely to encash their checks or pick up their checks more quickly; if anything, they are slightly less likely to encash their checks on time, though the difference is economically small. This is most likely because most calls occurred post-encashment, so they would not have been possible to act as a reminder. We also confirm that the results are unchanged when we drop the individuals who were called in both treatment and control groups from the administrative data sample (online Appendix Table A.6).

A third concern is that the improved performance of treatment MAOs may have caused worse performance on other tasks (multitasking). In practice, this was not likely to be a concern, because MAOs were primarily focused on RBS check distribution during the month of May 2018 and other tasks were deprioritized by the government. We also check that an increase in effort intensity during this period did not lead to increased slacking at other periods by testing for differences in the issue of soil health cards (as described earlier, this was an effort-intensive task that MAOs were responsible for). In Table 5, we find that treatment MAOs perform no worse

[32] In hazard models, an indicator for post–July 9 is not a significant predictor of encashment (online Appendix Table A.3).

[33] Calls were attempted with approximately 150 farmers per treatment and 50 per control mandal.

Table 5—Effect on Other MAO Activities

| | Number of SHC samples entered (1) | Number of farmers covered by SHCs (2) | SHC tests conducted (3) | SHCs available on portal (4) |
|---|---|---|---|---|
| Treatment | −43.54 | −138.8 | −26.55 | −124.9 |
| | (47.03) | (205.9) | (45.31) | (203.5) |
| Control mean | 906.02 | 4,259.53 | 873.73 | 3,891.78 |
| Observations | 514 | 514 | 514 | 514 |

*Notes:* All specifications include fixed effects at the randomization strata fixed level. Standard errors are clustered at the MAO level and reported in parentheses. Models are estimated using administrative data on Soil Health Cards at the mandal level from the Department of Agriculture, Cooperation, and Farmers Welfare.

on those tasks. Although we cannot observe all of the tasks that MAOs perform, this is at least suggestive that there was limited effort diversion away from other tasks.

### B. *Distributional Consequences*

The baseline allocation of benefits under RBS was regressive, as check size was proportional to registered landholdings. A wealthier farmer who owned 10 acres of land, for example, would have received a check for Rs 40,000 ($570), while a poorer farmer who owned only half an acre would have received a check for only Rs 2,000 ($30). This pattern was exacerbated by differences in distribution and encashment rates. As of September 26, 89 percent of farmers in the top quartile of the landholding distribution (holding more than 3.1 acres of land) had encashed their checks, declining monotonically to 68 percent of farmers in the bottom quartile (holding fewer than 0.4 acres). This could reflect differences in the effort made by government officials or differences in farmers' motivation and ability to collect and encash their checks. For example, a farmer at the fifth percentile of the land size distribution (0.09 acres of land) would receive a check worth just Rs 370 ($6), possibly less than the time and money costs of encashment.

We find that the effect of announcing phone-based monitoring to MAOs on on-time delivery was significant and nearly identical across farmers of different landholding sizes. Figure 3 plots the distribution of treatment effects by quartile of landholding, and the middle panel of Table 3 reports the corresponding estimates. However, by the end of September, some suggestive distributional effects begin to emerge. The treatment effect on encashment continued to be statistically significant for farmers in the lower three quartiles of landholdings but not for farmers in the top quartile.[34]

Our interpretation is that the treatment initially improved delivery of the transfer among all types of farmers. Among the set of wealthier farmers, the control group eventually caught up with their treatment counterparts as they used their resources and better networks to ensure that they got their transfers. It may also be more likely to be worth it for them to put in time and effort to track down the checks (since the amounts were larger). Among the set of poorer farmers, the control group did not

[34] We reject equality of treatment effects between the top and bottom quartiles ($p = 0.045$) but do not reject a joint test of equality across all four quartiles at conventional significance levels ($p = 0.11$).

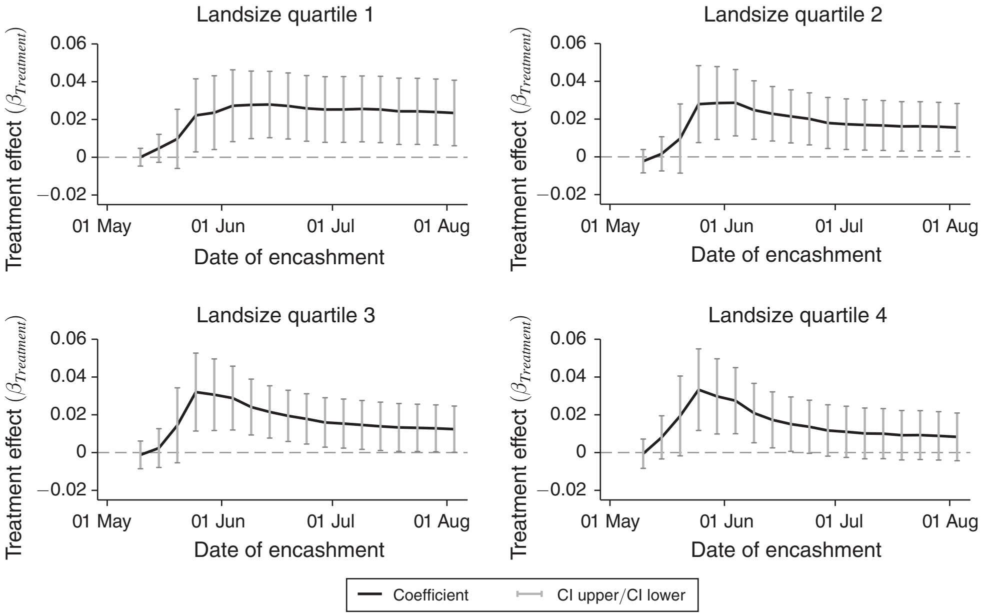


Figure 3. Treatment Effect over Time, by Landsize Quartile

*Notes:* The graphs in this figure report the coefficient of treatment effect on the cumulative rate of encashment over the period of check distribution across the four land size quartiles as well as the 95 percent confidence interval. The first graph is of treatment effects among the quartile of farmers with the smallest farms, while the fourth is for the quartile of farmers with the largest farms. The coefficients in the bottom graph are estimated through regressions with fixed effects at the randomization strata level and standard errors clustered at the MAO level. Less than 1 percent of checks were encashed after August 4 or before May 10, so the axis is restricted to those time periods.

catch up, and treatment effects remain significant over time. Wealthier farmers still benefited from the treatment, which sped up the process and lessened their cost of accessing the transfer (Table 4), but the gains along the margin of ever getting the transfer (which may matter more for the poor) are concentrated among the poor.

A further distributional concern about measurement by phone is that it could skew MAO effort toward those who own phones or have phone numbers. MAOs had access to the land registry, so they could see which farmers had numbers listed and thus would be able to provide information to the call center. However, we find significant positive impacts on on-time encashment for both those with and without phones, and we cannot reject that these effects are the same (Table 3, bottom panel). The difference in "ever encashed" between those with and without phones is also not statistically significant ($p = 0.59$), but time to encashment seems to have improved more for the group without a phone ($p = 0.03$). Thus, despite MAO performance being measured only in the population with mobile phones, the resulting increase in MAO effort appears to have led to improvement in program performance for all beneficiaries.[35]

[35] As seen in the control dependent variable means in the table, overall encashment rates were higher for those with phones (90 percent) than those without (72 percent). This reflects in part the higher level of wealth among

### C. *Tallying Costs and Benefits*

We next examine the cost-effectiveness of the intervention at delivering money to farmers and its overall welfare consequences. We cost the intervention at Rs 2.5 million ($36,000), the price GoTS paid the call center vendor prorated by the proportion of calls made to treatment areas. This is a conservative estimate, as the government paid a premium to complete the procurement process quickly, and conversations with the vendor indicate that the call center could be operated for roughly half this cost. On the other hand, this figure does not include the relatively small sunk costs of time spent by government employees or members of the research team designing the intervention (e.g., sampling protocols).[36]

At this cost, we estimate that phone-based monitoring was a highly cost-effective means of delivering cash transfers. The estimated impact on money ultimately delivered to farmers was Rs 69 million, or roughly $1 million, and its impact on money delivered on time was $3.9 million.[37] Since the funds for the program were already earmarked, the relevant benchmark for cost-effectiveness of an intervention seeking to improve program implementation (like the one we study) is the administrative cost of delivering antipoverty programs. The cost per incremental dollar delivered was 3.6 cents, which compares favorably to the administrative costs of well-implemented government cash transfer programs such as Progresa in Mexico (estimated administrative costs of 8.9 cents for every dollar delivered (Coady 2000)).[38] Focusing on the government's objective of getting transfers to farmers on time, the cost per dollar delivered on time was less than one cent.

To estimate an economic cost-benefit ratio, we must price the value of putting capital in the hands of farmers during the planting season as opposed to leaving it on the government's books. In this framework, higher rates of distribution and faster distribution create social benefits by increasing the amount of capital earning the higher rate of return among farmers. We assume that prior to receiving the transfer, farmers finance the purchase of inputs at a rate $r_f$, either by borrowing or substituting capital from alternative uses. Capital held by the government earns a lower return $r_g$.

individuals with listed phone numbers, as controlling for land size shrinks this gap. The remainder of the gap is likely explained by unobserved higher levels of ability among this group, as well as the fact that even in the absence of the treatment, if a farmer had a phone number listed, then they could be contacted and encouraged to pick up their check.

[36] We also do not cost incremental MAO effort, which is likely to be small or at least below the wage premium enjoyed by public employees (Finan, Olken, and Pande 2017). As a conservative and rough estimate, if we assumed the increased cost of effort to MAOs amounted to 15 percent of MAO monthly salary, then the cost of effort would amount to a total of approximately $9,000 among the 132 treatment MAOs, which does not noticeably change the cost-benefit calculations. Similarly, the total cost of analyzing the data and sending the reports was minor at less than $2,000. Finally, the marginal cost of collecting beneficiary phone numbers was virtually zero, since it had been done as an incidental part of the process of land record digitization. In other settings where phone numbers are not available, this cost would need to be included; however, it is now fairly common for governments to have databases of citizen phone numbers, as in this context.

[37] The treatment effects on amount ever delivered and amount delivered on time were Rs 54 ($0.77) and Rs 203 ($2.9) per farmer, respectively, so multiplied by the 1.3 million farmers in the treated mandals, this gives the estimates $1 million and $3.9 million.

[38] Many other programs are much less cost-effective. For example, Caldés, Coady, and Malussio (2006) estimates administrative costs of between 6.8 to 16.1 percent and 21.2 to 24.5 percent for 2 poverty alleviation programs similar to Progresa. In locations with high rates of corruption, the implied costs of social protection schemes are even higher (Olken 2006, Niehaus and Sukhtankar 2013).

Time runs from the start of the program ($t = 0$) to the date $T$ on which farmers' investments pay off and debt is repaid. The total value of a unit of capital held by the government until time $t$ and then by the farmer from time $t$ until $T$ is thus

$$v(t) = e^{r_g t} e^{r_f (T-t)}. \tag{2}$$

Given a distribution $F$ of check encashment dates, total social value is

$$W(F) = \int v(t)\, dF(t). \tag{3}$$

Faster and broader distribution of transfers shifts $F$ (as seen in Figure 2), increasing the amount of capital earning the higher rate $r_f$. We calculate $W(F)$ for both treatment and control groups using administrative records and conduct hypothesis testing using randomization inference.

We value capital on the government's books at the rate it earns on deposits ($r_g = 5\%$ annually),[39] and capital held by farmers at the going rate for short-term farm loans ($r_f = 25\%$).[40] We conservatively assume that investments are realized and debt is repaid immediately at harvest, so $T$ equals four months; in practice, farmers may continue to earn higher returns on capital that has been transferred to them, so this is a conservative assumption.

Using these estimates, phone-based monitoring generated Rs 10.6 million (\$140,000) in benefits, or roughly 4 times its cost. We reject the null of no benefit ($p = 0.04$) using randomization inference. This result is reasonably robust to variation in $T$ and $\delta$. At $\delta = 20\%$, benefits exceed costs for any $T$ longer than 26 days, while at $T = 4$ months, benefits exceed costs for any $\delta \in [5\%, 25\%]$ (Figure 4). Even under conservative parameter assumptions, the intervention was cost-effective.

These calculations may also be conservative in the sense that they reflect intent-to-treat estimates, while awareness in the treatment group was incomplete. Among treatment MAOs we surveyed, 90 percent had heard of the intervention, but only 28 percent were sure that the initiative had rolled out in their area; 28 percent were unsure, and 35 percent thought it had not. This may partly reflect strategic misrepresentation, such as if MAOs believed they could excuse poor results by feigning ignorance. In the control group, 52 percent of MAOs had heard about the intervention, but only 4 percent believed themselves treated, with another 8 percent unsure. While the control group was relatively "uncontaminated" by misperceptions of being treated, treatment effects may have been even larger if awareness of phone-based monitoring were universal in the treatment group.[41] Overall, these benefit-cost estimates suggest that phone-based monitoring can be cheaply applied to large-ticket programs at scale and create substantial economic value.

[39] In principle, the government could use funds for other productive investments. In practice, funds appropriated for the program would not be reallocated until the next fiscal year and would only earn interest.

[40] This is the rate charged by registered microfinance organizations that are subject to a regulated interest rate cap; informal moneylenders typically charge much higher rates.

[41] Despite potential confusion about treatment status, MAOs understood the nature of the intervention well. Among MAOs who had heard of the intervention, 89 percent correctly identified that the intervention would collect information from farmers over the phone, and 93 percent said that this data would be used for issuing reports on performance to MAOs/DAOs.

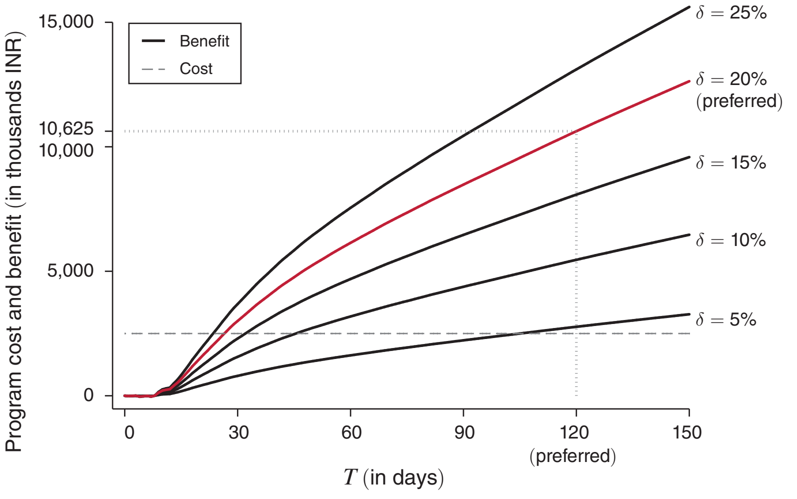


Figure 4. Sensitivity of Cost-Effectiveness Estimates

*Notes:* Sensitivity of cost-effectiveness estimates tested with respect to the total time period of consideration ($T$) and the differential rate of return ($\delta$, i.e., $r_f - r_g$). The interest earned by the government ($r_g$) is 5 percent annually, and the short-term annual interest rate for farmers ($r_f$) varies from 10 percent to 30 percent annually. The preferred specification for these parameters is $T = 120$ days and $\delta = 20\%$ since we think that $T = 120$ corresponds most closely to the planting season and that value of $\delta$ is relatively conservative, as described in the text.

### D. *Comparing Call Center with Administrative Records*

One concern with phone-based monitoring is whether this type of data provides an accurate picture of bureaucratic performance. Although the fact that MAOs responded to phone-based monitoring implies that they believed that it would at least partially reflect their true performance, this incentive could become stronger or weaker in the longer run as they learn more about the accuracy of phone-based monitoring. Further, the ability of senior officials to take meaningful follow-up action based on phone-based data will depend crucially on its reliability and accuracy. In most cases, it would be difficult to assess how accurately phone-based data measure performance without an independent data source to compare it to. In this case, we can take advantage of the existence of administrative data and examine the accuracy of phone data by comparing measured MAO performance in phone call data to the administrative data.

As a first pass, we check whether phone call and administrative data agree on whether a given check was encashed, which they do in 88.6 percent of cases.[42] However, to see if the data can be reasonably used for personnel management, the key metric of interest is whether aggregated phone and administrative data agree on performance at the MAO level. We calculate how often phone and administrative data rank the relative performance of a pair $(m, m')$ of MAOs within a district the

[42] To calculate the rate of agreement, we take the encashment status reported over the phone and check whether it matches the administrative record of whether the check was encashed by that date.

same way. For example, suppose the call center rates MAO A as third and MAO B as fourth best. If the administrative data rates them as second and third best, respectively, then since both sources ordered MAO A as performing better, we consider the sources to be in agreement. On the other hand, if the administrative data rated them as third and second best, respectively, then we would not consider them to be in agreement: MAO A is ranked better than MAO B by the call center but not by the administrative data.

These rankings disagree in 31 percent of cases. This disagreement is largely due to the relatively small sample of calls per mandal rather than inaccuracy in the phone call data per se—22 percent can be explained by sampling variation, with a relatively low true rate of disagreement between the two data sources of 9 percent (see notes to online Appendix Table A.7 for details on calculation). The key implication is that much of the disagreement could be eliminated if the government chose to increase the number of phone calls placed. We illustrate this in online Appendix Figure A.2, where we calculate the rate of disagreement due to sampling variation for different sample sizes—increasing the sample size to around 300 calls per mandal reduces the disagreement from sampling variation to only 10 percent.[43]

Overall, these results suggest that managers can reasonably use phone data to help decide which officials to push for more effort or acknowledge for good performance. However, the reliability of phone-based data may not be high enough to justify using them to determine more serious administrative actions (e.g., suspensions or promotions) with samples of the size that we observe here. Future work should consider the use of data over multiple years as well as decision-theoretic modeling of the costs and benefits of various follow-up actions under different levels of precision of measurement.[44]

## IV. Conclusion

We find evidence that the cheap, simple, and flexible approach of monitoring beneficiary experiences using outbound phone calls can be a cost-effective tool for improving last-mile service delivery. Further, unlike smaller-scale studies of interventions to improve monitoring of bureaucrats, our results suggest that phone-based monitoring can be rapidly deployed at the scale of an entire state of 35 million people. Beyond the case studied here, a unique feature of phone-based monitoring is its potential to be scaled across an unusually wide range of locations, programs, and outcomes. In general, phone-based monitoring can provide a rich picture of service delivery, as it can collect information on any outcome that can be observed by beneficiaries and described verbally.

[43] Another way of assessing the accuracy of the phone-based data is to compare their level of agreement on the worst MAOs. For the 20 percent of MAOs who were ranked as the worst performers in the phone data, 47 percent are also among the worst 20 percent of MAOs in the administrative data, while 80 percent are in the bottom 50 percent.

[44] Creating protocols for the optimal use of sample-based performance measures for high-stakes personnel action is a complex topic. The relevant issues are well illustrated by the literature on using estimates of teacher value added for teacher retention and tenure decisions (Chetty, Friedman, and Rockoff 2014; Koedel, Mihaly, and Rockoff 2015; Rothstein 2017).

In principle, the specific intervention we studied (of measuring MAO performance, creating report cards, and announcing to MAOs that their performance would be measured and reported) may not have required the phone calls, because the key metric in the report cards could have been created with the administrative data on check encashment from the banks.[45] However, in most cases, there is no similarly high-quality administrative data on last-mile service delivery. Thus, the attraction of phone-based monitoring is its wide applicability as a treatment that can be easily deployed and scaled across settings and programs. Consistent with wanting to study a scalable and widely applicable intervention, we use only the data from the phone calls to *implement* the intervention, and we only use the administrative data to *evaluate* it. Our setup has the dual advantage of studying an implementation protocol that is inclusive of practical problems (including nonrepresentative phone ownership and responses) while evaluating it using a different administrative dataset that does not have these problems and also has adequate power to detect treatment effects.

While the approach we studied here is itself adaptable to other settings and programs, this does not mean that its effects will be the same. It would therefore be useful to test phone-based monitoring in other settings. For instance, it may perform better for outcomes that beneficiaries experience more directly (e.g., check distribution) than indirectly (e.g., public good maintenance). It would also be useful to test this approach in a setting where the scope for improvement is greater than in the RBS, which was relatively well implemented. In general, Telangana is a relatively well-administered state, which may mean that the effects are larger than other areas, as the bureaucracy is more responsive. But it may also mean that the effects would be larger if phone-based monitoring were implemented in other locations where there is more margin for improvement.

Similarly, it would be valuable to examine how effects evolve over time in settings where bureaucrats perform similar functions repeatedly. As with all monitoring technologies, the officials being monitored would learn about the consequences of performing at different levels and might develop new strategies—both productive and counterproductive—to influence their ratings. But over time, phone-based monitoring could also inform officials in real time about what locations are most in need of their targeted intervention, as well as motivate them to increase effort. It could provide inputs for improving personnel management, which has been identified as the most important component of organizational management quality and is systematically worse for public organizations (Bloom and Van Reenen 2010). It could be tuned in many ways to improve performance, evolving statistical protocols for different types of follow-up action reflecting the cost of different kinds of Type I and Type II errors. Optimal monitoring protocols would take into account the need for whistleblower protection in small samples (Chassang and Padró I Miquel 2019) and the motivations of the respondents answering the phone (Fiorin 2019). One could

[45] The report cards also included information on beneficiary satisfaction and bribe requests, which would not have been possible with only administrative data. It may be that MAOs increased effort in order to satisfy these multiple performance targets, leading to better outcomes than would have been possible with just administrative data. However, we have no evidence on the potential differences in impact based on reports using phone versus administrative data.

even consider making the results publicly available, trading off the costs and benefits of transparency. Such public availability of data may also increase the effectiveness of bottom-up community and citizen-monitoring interventions (such as the ones studied by Reinikka and Svensson 2011).

Overall, phone-based measurement of beneficiary experiences opens up a broad set of possibilities for improving the quality of service delivery in developing countries. There is much exciting future research to be done in improving the design of such measurement systems, developing protocols for the optimal follow-up actions based on such measurement, and studying the impact of doing so. Social scientists can play an important role in improving the quality of governance by doing such work (Duflo 2017).

## REFERENCES

**Aker, Jenny C., and Christopher Ksoll.** 2018. "Call Me Educated: Evidence from a Mobile Monitoring Experiment in Niger." Center for Global Development Working Paper 406.

**Baker, George P., and Thomas N. Hubbard.** 2004. "Contractibility and Asset Ownership: On-Board Computers and Governance in U.S. Trucking." *Quarterly Journal of Economics* 119 (4): 1443–79.

**Bandiera, Oriana, Andrea Prat, and Tommaso Valletti.** 2009. "Active and Passive Waste in Government Spending: Evidence from a Policy Experiment." *American Economic Review* 99 (4): 1278–1308.

**Banerjee, Abhijit V., Esther Duflo, and Rachel Glennerster.** 2008. "Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System." *Journal of the European Economic Association* 6 (2–3): 487–500.

**Bloom, Nicholas, and John Van Reenen.** 2007. "Measuring and Explaining Management Practices across Firms and Countries." *Quarterly Journal of Economics* 122 (4): 1351–1408.

**Bloom, Nicholas, and John Van Reenen.** 2010. "Why Do Management Practices Differ across Firms and Countries?" *Journal of Economic Perspectives* 24 (1): 203–24.

**Caldés, Natàlia, David Coady, and John A. Maluccio.** 2006. "The Cost of Poverty Alleviation Transfer Programs: A Comparative Analysis of Three Programs in Latin America." *World Development* 34 (5): 818–37.

**Callen, Michael, Saad Gulzar, Ali Hasanain, Muhammad Yasir Khan, and Arman Rezaee.** 2018. "Data and Policy Decisions: Experimental Evidence from Pakistan." Stanford Institute of Economic Policy Research (SIEPR) Working Paper 1022.

**Chassang, Sylvain, and Gerard Padró I Miquel.** 2019. "Crime, Intimidation, and Whistle-Blowing: A Theory of Inference from Unverifiable Reports." *Review of Economic Studies* 86 (6): 2530–53.

**Chetty, Raj, John N. Friedman, and Jonah E. Rockoff.** 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104 (9): 2593–2632.

**Coady, David P.** 2000. *The Application of Social Cost-Benefit Analysis to the Evaluation of PROGRESA*. Washington, DC: International Food Policy Research Institute (IFPRI).

**Dávila Lárraga, Laura G.** 2016. *How Does Prospera Work?: Best Practices in the Implementation of Conditional Cash Transfer Programs in Latin America and the Caribbean*. Washington, DC: Inter-American Development Bank (IADB).

**Department of Agriculture, Cooperation and Farmers Welfare.** 2018. "Soil Health Cards Database." Ministry of Agriculture and Farmers Welfare, Government of India. https://soilhealth.dac.gov.in/ (accessed October 24, 2019).

**Department of Food and Public Distribution.** 2018. *Grievance Analysis and Systemic Reforms Recommendations 2017*. New Delhi: Department of Administrative Reforms and Public Grievances, Government of India.

**Dhaliwal, Iqbal, and Rema Hanna.** 2017. "The Devil Is in the Details: The Successes and Limitations of Bureaucratic Reform in India." *Journal of Development Economics* 124: 1–21.

**Drèze, Jean, and Reetika Khera.** 2015. "Understanding Leakages in the Public Distribution System." *Economic and Political Weekly* 50 (7): 39–42.

**Duflo, Esther.** 2017. "The Economist as Plumber." *American Economic Review* 107 (5): 1–26.

**Duflo, Esther, Rema Hanna, and Stephen P. Ryan.** 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102 (4): 1241–78.

**Express News Service.** 2018. "Monsoon Finally Arrives in Telangana, but June to Get Subdued Rain." *New Indian Express*, June 9. http://www.newindianexpress.com/states/telangana/2018/jun/09/monsoon- finally-arrives-in-telangana-but-june-to-get-subdued-rain-1825711.html.

**Finan, F., B.A. Olken, and R. Pande.** 2017. "Chapter 6—The Personnel Economics of the Developing State." In *Handbook of Economic Field Experiments*, Vol. 2, edited by Abhijit Vinayak Banerjee, and Esther Duflo, 467–514. Amsterdam: North Holland.

**Fiorin, Stefano.** 2019. "Reporting Peers' Wrongdoing: Experimental Evidence on the Effect of Financial Incentives on Morally Controversial Behavior." https://cega.berkeley.edu/wp-content/uploads/2020/03/Fiorin_PacDev2020.pdf.

**Gelb, Alan, Neeraj Mittal, and Anit Mukherjee.** 2019. "Towards Real-Time Governance: Using Digital Feedback to Improve Service, Voice, and Accountability." https://www.cgdev.org/sites/default/files/towards-real-time-governance-using-digital-feedback-improve-service-voice.pdf.

**Giné, Xavier, Robert M. Townsend, and James Vickery.** 2007. "Rational Expectations? Evidence from Planting Decisions in Semi-arid India." Bureau of Research and Economic Analysis of Development (BREAD) Working Paper 166.

**Government of India.** 2011. "Census Database." https://censusindia.gov.in/2011census/dchb/DCHB.html (accessed August 15, 2018).

**Government of Telangana.** 2016. "Telangana Socio Economic Outlook 2017." Planning Department, Government of Telangana. https://www.telangana.gov.in/PDFDocuments/Socio-Economic-Outlook-2017.pdf.

**Government of Telangana.** 2018. "Farmer Registry Database." (accessed September 26, 2018).

**Government of Telangana.** 2019. "Telangana Open Data Portal." https://data.telangana.gov.in/search/type/dataset (accessed October 24, 2019).

**Hellman, Aline Gazola.** 2015. *How Does Bolsa Familia Work?—Best Practices in the Implementation of Conditional Cash Transfer Programs in Latin America and the Caribbean*. Washington, DC: Inter-American Development Bank (IADB).

**Khan, Adnan Q., Asim I. Khwaja, and Benjamin A. Olken.** 2019. "Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings." *American Economic Review* 109 (1): 237–70.

**Koedel, Cory, Kata Mihaly, and Jonah E. Rockoff.** 2015. "Value-Added Modeling: A Review." *Economics of Education Review* 47: 180–95.

**Landes, David S.** 1983. *Revolution in Time: Clocks and the Making of the Modern World*. Cambridge, MA: Harvard University Press.

**Lasania, Yunus Y.** 2019. "Jagan Reddy Announces 'Rythu Bharosa' Input Subsidy Scheme for Andhra Farmers." *Mint*, June 6. https://www.livemint.com/politics/policy/jagan-reddy-announces-rythu-bharosa-input-subsidy-scheme-for-andhra-farmers-1559835306312.html.

**Masud, Mohammad Omar.** 2015. *Calling the Public to Empower the State: Pakistan's Citizen Feedback Monitoring Program, 2008–2014*. Princeton, NJ: Innovations for Successful Societies, Princeton University.

**Mundle, Sudipto, Samik Chowdhury, and Satadru Sikdar.** 2016. "Governance Performance of Indian States 2001–02 and 2011–12." National Institute of Public Finance and Policy Working Paper 16/164.

**Muralidharan, Karthik, Jishnu Das, Alaka Holla, and Aakash Mohpal.** 2017. "The Fiscal Cost of Weak Governance: Evidence from Teacher Absence in India." *Journal of Public Economics* 145: 116–35.

**Muralidharan, Karthik, and Paul Niehaus.** 2017. "Experimentation at Scale." *Journal of Economic Perspectives* 31 (4): 103–24.

**Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar.** 2016. "Building State Capacity: Evidence from Biometric Smartcards in India." *American Economic Review* 106 (10): 2895–2929.

**Muralidharan, Karthik, Paul Niehaus, Sandip Sukhtankar, and Jeff Weaver.** 2018. "High-Frequency Program Monitoring and Bureaucratic Performance: Experimental Evidence from India." AEA RCT Registry. https://www.socialscienceregistry.org/trials/2942.

**Muralidharan, Karthik, Paul Niehaus, Sandip Sukhtankar, and Jeffrey Weaver.** 2021. "Replication data for: Improving Last-Mile Service Delivery Using Phone-Based Monitoring." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.38886/E117431V1.

**Niehaus, Paul, and Sandip Sukhtankar.** 2013. "Corruption Dynamics: The Golden Goose Effect." *American Economic Journal: Economic Policy* 5 (4): 230–69.

**Olken, Benjamin A.** 2006. "Corruption and the Costs of Redistribution: Micro Evidence from Indonesia." *Journal of Public Economics* 90 (4–5): 853–70.

**Olken, Benjamin A.** 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115 (2): 200–249.

**Outlook India.** 2019. "Budget 2019." February. https://www.outlookindia.com/budget-2019.

**Press Trust of India.** 2019. "Odisha Allocates Additional Rs 3,234 Crore for Kalia Scheme." *Business Standard*, June 7. https://www.business-standard.com/article/pti-stories/odisha-allocates-additional-rs-3-234-crore-for-kalia-scheme-119060701025_1.html.

**PRS India.** 2018. "Telangana Budget Analysis 2017–18." New Delhi: PRS Legislative Research.

**Reinikka, Ritva, and Jakob Svensson.** 2011. "The Power of Information in Public Services: Evidence from Education in Uganda." *Journal of Public Economics* 95 (7–8): 956–66.

**Rothstein, Jesse.** 2017. "Measuring the Impacts of Teachers: Comment." *American Economic Review* 107 (6): 1656–84.

**Singh, Abhijeet.** 2020. "The Myths of Official Measurement: Auditing and Improving Education Data in Developing Countries." https://www.dropbox.com/s/knrc9mvx5gdplw3/Data_Integrity_Aug2020.pdf.

**Subramanian, Arvind.** 2020. "QUBI Can Wipe Off Farmers' Tears." *Hindu Business Line*, July 10. https://www.thehindubusinessline.com/opinion/qubi-can-wipe-off-farmers-tears/article24381521.ece.

**Vivalt, Eva.** 2019. "How Much Can We Generalize from Impact Evaluations?" *Journal of the European Economics Association*, forthcoming.

**World Bank.** 2003. *World Development Report 2004: Making Services Work for Poor People*. Washington, DC: World Bank.

**World Bank.** 2016. *Case Study from the Global Report—Engaging Citizens to Improve Service Delivery: The Citizen Feedback Monitoring Program in Pakistan*. Washington, DC: World Bank.

**World Bank.** 2018. "Mobile Cellular Subscriptions (per 100 People)." World Bank. https://data.worldbank.org/indicator/IT.CEL.SETS.P2?end=2017&locations=XM&name_desc=true&start=1998 (accessed September 26, 2018).