Title

Flows, leaks and blockages in informational interventions: A field experimental study of Bangalore's water sector

Authors
Kumar, Tanu
Post, Alison E
Ray, Isha

Publication Date
2018-06-01

DOI
10.1016/j.worlddev.2018.01.022

Peer reviewed

World Development 106 (2018) 149–160

Contents lists available at ScienceDirect

World Development

journal homepage: www.elsevier.com/locate/worlddev

Flows, leaks and blockages in informational interventions: A field experimental study of Bangalore's water sector

Tanu Kumar, Alison E. Post *, Isha Ray

University of California, Berkeley, Berkeley, United States

ARTICLE INFO

Article history:
Accepted 23 January 2018

Keywords:
Information
Intermittent water supply
Transparency
Frontline worker
Stress
India

ABSTRACT

Many policies and programs based on informational interventions hinge upon the assumption that providing citizens with information can help improve the quality of public services, or help citizens cope with poor services. We present a causal framework that can be used to identify leaks and blockages in the information production and dissemination process in such programs. We conceptualize the "information pipeline" as a series of connected nodes, each of which constitutes a possible point of blockage. We apply the framework to a field-experimental evaluation of a program that provided households in Bangalore, India, with advance notification of intermittently provided piped water. Our study detected no impacts on household wait times for water or on how citizens viewed the state, but found that notifications reduced stress. Our framework reveals that, in our case, noncompliance among human intermediaries and asymmetric gender relations contributed in large part to these null-to-modest results. Diagnostic frameworks like this should be used more extensively in development research to better understand the mechanisms responsible for program success and failure, to identify subgroups that actually received the intended treatment, and to identify potential leaks and blockages when replicating existing programs in new settings.

1. Introduction

Following the 2004 *World Development Report's* influential call, economists and political scientists have analyzed many programs that provide citizens, especially the poor, with more information about public services (World Bank, 2004). Citizens armed with information about corruption or service deficiencies, it is argued, can vote underperforming politicians out of office and hold public sector bureaucracies more firmly to account. Providing information thereby sets in motion a virtuous cycle leading to improvements in service quality and access. A growing body of scholarship has evaluated the extent to which, and circumstances under which, these propositions hold (see Pande, 2011; Lieberman, Posner, & Tsai, 2014). Much less attention has been paid to the production and dissemination of the information itself, a process that is fundamental to the success or failure of such interventions.

In this paper, we present a causal framework that can be used to identify leaks and blockages in the "information pipeline" for programs that attempt to change household behaviors through informational interventions. This framework, we will show, can guide researchers and policymakers in understanding why interventions fail to produce their expected effects, as well as what difficulties to anticipate when scaling up programs and/or replicating them in new settings. In its focus on information and production, our framework complements Lieberman et al. (2014), which begins with households receiving information and examines why the information may not generate changes in political behavior. We expect the framework to help scholars to understand why many such informational interventions have proven to be ineffective in experimental evaluations.[1]

We apply our causal framework to an impact evaluation of a notification service intended to make intermittently-supplied urban water deliveries more predictable. Over 300 million people receive piped water intermittently, mostly in South Asia and Sub-Saharan Africa (Kumpel & Nelson, 2016, 543). In general, throughout Asia and Africa, intermittency is the hallmark of public service delivery: buses do not run on a standard schedule, water supplies stop and start again, and electricity blackouts occur regularly. Moreover, the poor state of the underlying infrastructure—

* Corresponding author at: University of California, Berkeley, 210 Barrows Hall, Berkeley, CA 94720-1950, United States.

E-mail address: aepost@berkeley.edu (A.E. Post).

[1] For example, Experiments in Governance and Politics launched an initiative composed of seven coordinated studies of informational interventions. The common informational interventions all had null estimated effects on voter behavior (Dunning, Grossman, Humphreys, Hyde, & McIntosh, Forthcoming).

https://doi.org/10.1016/j.worlddev.2018.01.022
0305-750X/

prone to pipe leaks and power outages—means that services are not only intermittent, but are often unpredictable. Yet few studies have considered how better information can reduce the coping costs of unpredictability.

We evaluated the household-level impact of a service developed by NextDrop, an Indian social enterprise, which sent households text-message notifications on water arrival times and supply cancellations. Utilities in India (and elsewhere) often do not possess sensors that allow them to monitor where water is flowing within their network. In larger systems, water flows are managed by "valvemen", or utility employees who operate the valves that channel water into hydraulically isolatable "valve areas" of 50–200 households at a time. The valves are supposed to be turned on and off according to a utility-prepared schedule, though there are both social and technical reasons for the valvemen to stray from the official schedule. NextDrop developed a novel system, successfully piloted in the twin cities of Hubli-Dharwad, in which the valvemen notified them whenever they were opening and closing valves. NextDrop then sent notifications to individual households, via cellphone text-messages, letting them know when their water would be turned on. These notifications went to households with their own taps as well as those depending on water from communal standpipes.[2] We examined the impact of its service through a cluster-randomized experiment as the organization rolled out its services in Bangalore, about 400 km from the pilot site. Our study is the first experimental evaluation of a program designed to make intermittent urban water supplies more predictable.

There were many reasons to expect that increasing service predictability would improve household welfare, particularly for women in low-income households. When water comes only once or twice a week, household members (typically women) must fill their storage containers the moment it arrives, often having waited long periods because timing is unpredictable. This waiting can prevent them from spending time at work or in the community, and induce stress because it is expensive and inconvenient to miss a day's supply. Service unpredictability may also weaken the bonds between individuals and the state, as citizens who cannot depend upon regular services may be less likely to view government service providers as competent. Moreover, if citizens see (or think) that those with stronger political connections have better services or better information, they may see the service providers as biased or discriminatory. They may direct their water-related complaints and inquiries to more approachable intermediaries, such as local leaders, rather than to state agencies themselves.

We found no impact of NextDrop's program on outcomes such as time spent waiting for water, expenditures on substitute sources, or citizen relations with the utility. We did find that the program triggered modest reductions in stress among low-income households. This result highlights the importance of examining the psychological aspects of household welfare in development interventions; these aspects are frequently neglected in impact evaluation research. To understand the null results for core outcomes such as wait times, we drew on our framework to diagnose weaknesses in NextDrop's "information pipeline." We found two key points of leakage in Bangalore: (i) valveman failure to submit accurate notifications, which meant that many households did not receive, or could not act on, them and (ii) the common practice of the adult male taking the family's cellphone to work, which meant that the (usually) female "waiter" at home often could not access the notifications.

In this paper, we first describe the informational intervention we evaluate, and outline our expectations regarding the potential effects of this effort. We draw mainly on literatures in behavioral economics and political science to develop our central hypotheses. We next describe our experimental research design and present our analysis of the program's impact. We draw on our causal framework to first diagnose where the information production and dissemination broke down when the NextDrop program was rolled out, and then to evaluate the effect of more predictable services on those who actually did receive accurate notifications. We then explain how our framework could be used to identify implementation difficulties *before* programs are introduced in new locations. We conclude with the broader implications of our findings for our specific experiment, the design of informational interventions more broadly, and the challenges of transferring even rigorously verified programs from one location to another.

2. The intervention: water arrival notifications

This study evaluates the impact of an SMS based program to provide prior notifications regarding water arrival times on household welfare and state-society relations. Intermittent water deliveries are sometimes predictable but are more often irregular and unreliable (Kumpel & Nelson, 2016). We examine the program's impact on household welfare and on political attitudes and behavior. The system was developed for urban India, where cell phone penetration rates are high, and the likelihood that water systems will soon be upgraded to continuous service is low. The service is also potentially useful in much of urban Asia and Africa, where intermittency is rife, local governments cannot adequately fund water systems, and cell phone penetration is increasing rapidly.

In NextDrop's system, the utility employees (valvemen) were asked to notify the company by calling a toll-free number whenever they opened and closed valves. NextDrop then sent free SMS notifications with expected water arrival times to individual households, which it had cataloged by valve area. Notifications could be sent before water flowed through household taps because it takes some time for water to flow into a valve area and fully pressurize that portion of the network. NextDrop reasoned that a notification 30–60 min prior to water arrival would be sufficient to allow customers to return from nearby locations, though not from distant workplaces; notifications arriving fewer than 30 min in advance would not help them unless they were already at home.

To correctly place households in valve areas, NextDrop collected GPS coordinates for households and created valve area maps, which Indian utilities typically do not possess. It drew on valvemen's tacit knowledge regarding the area boundaries, accompanying them on walks around the edges and taking GPS readings. Each polygon in Fig. 1 is an example valve area from Bangalore: the city has thousands of these, for which valvemen turn water off and on by manually adjusting a valve. After receiving a valve opening call, NextDrop could automatically let households know when their water would arrive based on their location.

The company successfully piloted its system in Hubli-Dharwad, Karnataka, a city of approximately 1 million where most residents received water services every 4–5 days (Burt & Ray, 2014). By August 2013, the firm had 15,844 *paying* customers in the city, suggesting significant demand for its services. NextDrop surveys of their Hubli-Dharwad customers suggested that they valued the notifications;[3] this is believable given that the average time spent

[2] A "standpipe" is a (usually free) public tap, shared by several households.

[3] NextDrop conducted a survey of 200 households in June 2010 in Hubli-Dharwad; according to its records, respondents indicated a willingness to pay INR 8 per month on average for water arrival information (1USD = INR 45 in 2010). In June 2011, it surveyed 60 enrolled households in its pilot, asking whether they would prefer to remain enrolled or receive an INR 5 cell phone recharge. 94% preferred to continue with the service.

Fig. 1. Example valve areas in BWSSB Subdivision E3.

waiting for and collecting water in the city was estimated at ~30 h per month in 2011–2012 (Burt et al., Under review).

In 2013, NextDrop expanded to the megacity of Bangalore, with the support of the city's water utility, the Bangalore Water Supply and Sewerage Board (BWSSB). BWSSB had no way to monitor the valve opening regimen or the flow of water through its networks in real time before NextDrop approached them with their product. NextDrop made a number of changes to its program design to allow it to enter the new market and operate efficiently at a larger scale. First, notifications would be delivered free to the customers: NextDrop would obtain revenues from the utility and from sponsored advertising rather than charge those receiving messages. Second, in Hubli-Dharwad, a smaller city, NextDrop had developed personal relationships with individual valvemen; this, the company thought, would not be viable or cost-effective in larger cities. For its Bangalore rollout, therefore, NextDrop negotiated an agreement with the utility that made the notifications part of the valvemen's job description. The company tracked the rates at which valvemen submitted reports, as they had done in Hubli-Dharwad. NextDrop employees hand-delivered reports on valveman notification rates to their supervisors on a weekly basis, so managers had the information they needed to enforce compliance. From NextDrop's perspective, this was a scalable approach.

3. Expected program impacts

In this section, we outline our prior expectations regarding the potential effects of the NextDrop program in Bangalore. The causes and consequences of intermittent and unreliable water supplies have been more the domain of engineering and public health research than of the social sciences (Galaitsi et al., 2016). The water policy literature on intermittency has analyzed the effects and coping costs of intermittent water (e.g., Burt & Ray, 2014; Kumpel, Woelfle-Erskine, Ray, & Nelson, 2017; Kumpel & Nelson, 2016) but has not examined interventions that improve predictability within intermittent systems. We therefore build on behavioral economics, political science, and urban water policy insights to develop our hypotheses on the effects of providing accurate information to improve the predictability of water services on: a) household welfare; and b) citizens' relationship with the state.[4]

3.1. Household welfare effects

The water policy literature has shown that intermittent water supply imposes significant costs on households, especially with respect to water quality and human health (Kumpel & Nelson, 2013; Ercumen et al., 2015). *Unpredictable* supplies impose particularly large costs, especially upon household members tasked with managing the water (Pattanayak et al., 2005; Subbaraman et al., 2015; Zérah, 2000). In low-income households that cannot afford maids or automatically-filling storage tanks, household members—particularly women—may stay near the home and devote time to waiting that might otherwise be spent on work, or on community and family events[5]. We therefore hypothesized that accurate prior notifications regarding water delivery or service disruptions would reduce waiting times, allow for more participation in community and social activities, and result in fewer foregone earnings. Because municipal water typically costs less than substitutes such as vendor-supplied water (e.g. Estache, Gomez-Lobo, & Leipziger, 2001; Kjellén & McGranahan, 2006), we also hypothesized that notifications would reduce the reliance on substitutes, because they would decrease the probability of missing a supply.

All costs are not material in nature; supply unpredictability may impose psychological costs as well. Given that domestic water is a vital resource, the household member responsible for managing it may feel stress when services are unpredictable or water storage cannot be planned. This argument builds directly on empirical studies of psychosocial stress related to water insecurity (e.g. Wutich & Ragsdale, 2008; Stevenson et al., 2012), as well as the behavioral economics literature, which has shown that scarcity imposes cognitive stress (Mullainathan & Shafir, 2013). We therefore hypothesized that accurate notifications would lead to a reduction and stress and worry for the person waiting.

While these effects may be observable across the entire urban population in cities with water intermittency, we expected them to be particularly pronounced for *low-income households*, because the cost of substitutes as a fraction of household income would be greater, and because, as noted earlier, poverty itself may exacerbate stress. We also expected larger effects for households that do not have automatically-filling overhead or underground tanks

[4] Our specific hypotheses and research design were recorded in pre-analysis plan 20150514AA, registered with EGAP prior to the receipt of our baseline data. See also the Online Appendix, Section 2.

[5] A 2011–2012 survey in Hubli-Dharwad found that 25% of the respondents (n = 3922) reported sometimes missing such events because the water had not yet arrived (Burt et al., Under review).

("sumps"); overhead tanks in particular cannot be supported on structures of poor construction quality.

3.2. Political effects

There is reason to expect that – even in the absence of substantive service improvements – better information alone leads to a more favorable view of the local state and its agencies. Here we build on a broad literature investigating how citizens "see" the state (Corbridge, Williams, Srivastava, & Véron, 2005; Ferguson & Gupta, 2002; Evans, 2008) as well as the information and communication technologies literature, which has argued that better and cheaper information directly influences citizens' views of the state (though not necessarily in a positive direction) (e.g., Madon & Sahay, 2002; Tolbert & Mossberger, 2006).

We hypothesized that increasing the predictability of services would improve citizens' image of the state, given that NextDrop's messages were sponsored by, and branded as coming from, the state-run utility. Receiving accurate prior information could potentially make services easier to access, remove the need to acquire information through bribes or connections, convey greater state capacity, and – given the innovative dissemination of text messages to *all* citizens – could make the state seem more modern and universalistic (see Harriss, 2006; Kuriyan & Ray, 2009; Ghertner, 2011).

We also hypothesized that these notifications would shift perceptions regarding who is responsible for addressing citizens' concerns. The literature on citizen-state interactions in the developing world suggests that ordinary citizens often turn to political intermediaries or direct action when they have service problems.[6] With a universally administered notification system (like Nextdrop's) that connects citizens more directly to the service provider, citizens may view government agencies themselves, rather than local intermediaries, as responsible for addressing their problems.[7]

We expected that these effects, while relevant across the urban population, would be particularly strong among households that consider themselves marginal in religious, social (e.g. caste), or linguistic terms. These segments were less likely to have politically influential intermediaries prior to the intervention. We also expected effects to be stronger for households with less money to buy non-tap water, and those without automatically-filling storage tanks.

4. Research design and methods

We evaluated the effectiveness of the NextDrop system through a cluster-randomized experiment in Bangalore, a city of over 8 million that is often called India's Silicon Valley. Prior scholarship on domestic water, particularly on water intermittency, has not been field-experimental in nature. The handful of empirical studies on the coping costs and inefficiencies associated with unreliable water supply have either been observational or stated-preference based experiments (Akram & Olmstead, 2010; Baisa, Davis, Salant, & Wilcox, 2010; Dauda, Yacob, & Radam, 2014; Pattanayak et al., 2005; Subbaraman et al., 2015; Zérah, 2000; Kumpel et al., 2017).

To design our evaluation, we worked closely with NextDrop and BWSSB. We structured the study to evaluate the efficacy of NextDrop's system in a real world setting, i.e., their existing efforts to scale up in new cities. This meant deferring to some of NextDrop's implementation decisions, so long as these did not prevent randomized assignment and noninterference between treatment and control; we return to this point below. This section outlines the main features of our study, conducted in 2014–2015.

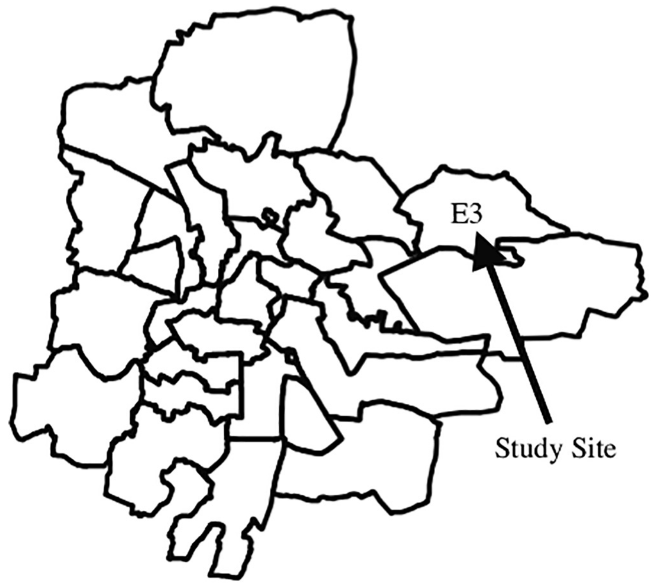

Fig. 2. BWSSB subdivision E3.

4.1. Study site characteristics

We conducted our impact evaluation in a socio-economically diverse section of Bangalore chosen to maximize similarities with other Indian urban centers and allow us to detect variation in impacts across income groups. NextDrop had recently received approval from BWSSB to introduce services across the city.[8] Because NextDrop was rolling out quickly, we agreed to restrict our evaluation to one of the utility's 32 subdivisions not scheduled for immediate expansion, areas mostly outside the city center. After review of the limited (and often inaccurate) government data on low-income settlements and population densities in Bangalore and extensive site visits throughout the city in 2014, we chose to conduct our evaluation in BWSSB subdivision E3 Fig. 2. Water services are typically more intermittent and unreliable further from the center, so this choice meant we were working in an area where the intervention was more likely to have an impact. It also meant that our study location was more typical of water service in urban India more broadly, as BWSSB performs well in relation to other South Asian utilities (Connors, 2005; McKenzie & Ray, 2009).

Subdivision E3 contained a sufficient number of low-rise residential neighborhoods along with several low- and middle-income neighborhoods. Data from our baseline survey suggests that approximately 33% of the area's residents – 14% of whom included recent migrants from states such as Tamil Nadu and Andhra Pradesh – could be classified as Bangalore's bottom third of the income distribution. Over 85% of residents received water services just once or twice a week, which is common in urban India, and also frequent enough that we could detect the effect of the notifications on our outcomes of interest, if they were indeed useful. Moreover, 28% of residents possessed neither an automatically filling overhead tank nor sump, requiring someone to be

[6] On Bangalore, see Ranganathan (2014); on similar dynamics elsewhere in India, see Berenschot (2010) and Krishna (2011). On intermediaries within clientelistic party systems in the developing world, see Stokes, Dunning, Nazareno, and Brusco (2013).

[7] Our emphasis here diverges from the political science and economics literatures on informational interventions, which examine whether information increases political participation or changes votes, thus fueling bottom-up pressure to make service providers more accountable (see Pande, 2011; Lieberman et al., 2014).

[8] May 2014 memorandum of understanding.

present when the water arrived to collect and store it for use between supplies.

4.2. Randomization and sampling strategies

Within BWSSB subdivision E3, we employed a cluster-randomized experimental design.[9] We opted for cluster- rather than household-level randomization to eliminate possible information sharing between treatment and control households. We separated clusters of households from one another by at least two streets so as to create physical buffers preventing information sharing between our treatment and control groups (Fig. A.1, Online Appendix). Moreover, inter-cluster spillovers were unlikely because information on water arrival times is relevant only to individuals within the same valve areas (typically 50–200 households).[10] Our study, powered using pilot data to detect a reduction of 30–45 min in the time spent waiting for water per week, included 120 clusters of 25 households each for a total sample of 3000 households.[11]

Because blocking on a variable associated with the outcome of interest can improve the precision of causal estimates in cluster-randomized experiments (Imbens, 2011), we employed a geographic approach to stratification. Blocking on socio-economic geography also enabled analyses of subsets corresponding to areas where we expected to observe stronger effects: those with poorer residents and with poorer quality water infrastructure. Based on extensive site surveys, we designated 30 geographic blocks with a particular socio-economic character, either low income (10 blocks) or mixed income (20 blocks). Each block included four clusters that we expected to be similar not only in socio-economic terms, but also with respect to the underlying water infrastructure. Within each block, we randomly assigned two clusters to receive treatment and two to the control condition.[12]

4.3. Data and measurement

We measured the impact of the intervention through two surveys administered to the treatment and control groups. A baseline survey was conducted prior to the intervention in April and May of 2015, and an endline survey was conducted in October and November 2015. We ran the trial for four months to give households enough time to adapt their daily routines to the notification service. Enumerators surveyed only those individuals who managed and stored water for the household, returning to the household if the "waiter" was unavailable at first. Because women typically manage water, 80% of our respondents were women.

Enumerators concluded the baseline survey by offering all households the opportunity to enroll in NextDrop services, when they became available in their area, by submitting their cell phone numbers and offering consent. Offering services to both treatment and control allowed us to employ a placebo design to help identify compliers—i.e., those who would accept treatment—in both groups. Respondents signed up for text or voicemail notifications in English, Kannada, Telugu, or Tamil. They were informed that the service was being provided by BWSSB, the state water utility, with NextDrop handling implementation.[13] NextDrop enrolled the households in our treatment group following the completion of the baseline survey, and waited until the end of our study to enroll the control group.

Enumerators also collected GPS coordinates (5 m precision) from each household. This allowed NextDrop to correctly place treatment group households in valve areas (so they received relevant information), and helped our team to verify that enumerators had not strayed outside cluster boundaries. Coordinates also assisted with returning to the same households in the second wave.

Comparing key characteristics between our treatment and control groups, we see that our cluster-randomized design achieved balance between treatment and control with respect to household characteristics, water supply conditions, and political factors (Online Appendix, Table A.2). As is to be expected with two-wave designs, our sample did experience attrition: we lost 16% of our initial sample between waves 1 and 2, often because households had moved. Attrition did not affect covariate balance (Table A.2).

4.4. Conditions at baseline and implications for aggregate impacts

A first condition for the intervention to generate an effect would be that treatment group members indeed faced costs due to unpredictable water services. We conducted pilot surveys in subdivision E3, prior to confirming our choice of this area for the impact evaluation, and these pilots suggested room for movement on key outcomes of interest.[14] Population means from our baseline survey confirmed that there was room for movement on waiting times for water, use of substitutes for piped water, and tendency to contact the utility directly regarding service problems.[15] Moreover, our baseline data suggest that many households faced difficult water supply conditions. A full 69% of our households reported that their water did not come at a specific time. Additionally, 43% reported that they simply learned that water had arrived when it began to come out of their taps, rather than knowing when to expect the water from the supply schedule, valvemen, or local leader. There was less room for movement on outcomes such as missing work due to waiting or attitudes towards the utility (these were already quite favorable; Table 1).[16]

5. Overall program impacts

Though baseline conditions provided ample room for NextDrop's notifications to have household-level impacts, intent-to-

[9] A cluster randomized controlled trial is a type of randomized controlled trial (or RCT) in which groups of subjects, as opposed to individuals, are randomized to treatment and control conditions.

[10] In an ideal world, we would have randomized assignment to valve areas rather than clusters we ourselves designated. After discussions with our survey team, we realized that, because valve area boundaries are not visible above ground and do not follow the street layout, survey enumerators would have had difficulty following even boundaries drawn on maps. Substituting cluster-level for valve area randomization led to only minimal spillovers (see below).

[11] For more details, see our pre-analysis plan (Online Appendix). These calculations presumed that we would lose approximately 20% of our sample through attrition and that 20% of households would refuse to sign up for services.

[12] Given the lack of accurate state data on the existence and location of the city's numerous and scattered small slums, identifying an area with a suitable demographic mix required significant on-the-ground legwork by our team. We included four clusters per block rather than two following Imbens (2011). Within each cluster, we followed a systematic sampling plan with a skip of three between households on every street. After piloting the survey in low-income areas, we decided that a skip of three would be sufficient to avoid group interview sessions in which neighbors "help" respondents answer survey questions.

[13] Forms describing the NextDrop service were also translated into these languages. Interviews were typically conducted in Kannada (the primary language spoken in Bangalore), but were conducted in Telugu or Tamil when relevant (usually for recent migrants).

[14] For outcomes such as wait time and expenditures on substitutes we also consulted survey data from Hubli-Dharwad, collected during one of the authors' previous research efforts, which suggested substantial room for movement (see previous section).

[15] Replication code and data can be found at dataverse.harvard.edu. Respondents contacted the utility, local leaders, and elected officials at extremely low rates at baseline.

[16] It may seem surprising that, despite BWSSB's unreliable service, its standing in these communities was good. This can be explained by recent improvements service quality. BWSSB had completed new infrastructure allowing it to supply this area with water from the Cauvery river, rather than brackish borewell water, roughly 6 years before our study. Satisfaction with the water supply and the utility was therefore high at baseline.

Table 1
Conditions at baseline.

Outcome	Overall population	Low income blocks	Target group
Household conditions			
Time spent waiting for water (*hrs. per supply day*)	0.92	1.26	1.30
Missing community events (*fraction of respondents*)	0.20	0.25	0.26
Missing work (*hrs. missed last 6 months*)	2.36	3.95	2.20
Need for substitutes (*fraction of respondents unable to store enough on supply days*)	0.25	0.25	0.23
Psychological conditions			
Worrying about water *(ranges from 1 = often, to 4 = not at all)*	2.43	2.38	2.25
Thinking about water during the day*(ranges from 1 = often, to 4 = not at all)*	2.46	2.29	2.23
Political attitudes			
Perception that providers are competent *(ranges from 1 = agree to 3 = disagree)*	1.38	1.40	1.29
Perceptions that providers are innovative and modern (*same as above)*	1.38	1.41	1.32
Perception that providers care about "people like us" (*same as above)*	1.48	1.50	1.44
Contacting			
Contacting providers directly about problems with service *(fraction of respondents contacting utility rather than others)*	0.07	0.05	0.05
Holding state water providers directly responsible for service *(fraction of respondents naming utility rather than others)*	0.15	0.08	0.11

treat (ITT) estimates for average treatment effect show no statistically significant change for any outcome variables across the entire study population, except those related to worry and stress (Table 2).[17] They also show no effects when we restrict our analysis to our ten low-income blocks, which contain lower-income populations and less variability in water infrastructure. Moreover, they show no robust effects for our target group – low-income households without automatically filling tanks. Tests for many other heterogeneous effects outlined in our pre-analysis plan also do not yield statistically significant or substantively important effects.[18] In other words, the NextDrop program failed to generate discernible impacts on most indicators of household welfare and state-society relations in the treatment group. This could be because the estimated average treatment effect on (reported) time spent waiting for water per supply day is very small—a roughly 2.5 min reduction for the treated population. Further analysis suggests that we should have been able to detect reasonably-sized effects, had they been present: the minimum detectable effect for our wait time outcome is 9 min per supply day.[19] It is unlikely that reductions of this size would impact household wellbeing.[20]

We do detect a small but measurable decrease in worrying about water and thinking about water during the day among the overall population and within our low-income blocks.[21] It is possible that there was an effect for the target population, but we simply do not know, as the experiment was not powered for these outcomes.

6. Flows and leaks along the pipeline: a framework for analyzing information interventions

To explain these null-to-modest results, we turn to our causal framework that disaggregates informational interventions into their constituent production and dissemination processes. Fig. 3 displays six nodes at which informational interventions can break down during production and dissemination. First, the entity responsible for collecting the information may not obtain it, or obtain it only partially; for instance, frontline workers[22] could fail to supply the information requested of them (Hyun, Post, & Ray, 2018). Organizations often cannot fully control the agents charged with executing their assigned tasks (e.g., Gailmard & Patty, 2012), and frontline workers, or street level bureaucrats, can exercise significant autonomy in "making policy on the ground" (e.g., Lipsky, 1980). Second, the entity responsible for providing information may not be able to analyze or compile the information into a usable format. Third, the entity charged with disseminating the information may not, in fact, do so; there could be poorly-aligned incentives and/or technical difficulties. Fourth, information may be sent, but respondents may not receive it. Messages may go to the wrong person because phone numbers change or because the intended recipient does not keep the household cell phone. Fifth, the intended recipient may technically receive the messages, but not *register* receipt; information may be sent in the wrong language, or go unnoticed if a recipient is inundated with messages, or be so useless that the respondent stops paying attention. Finally, the information sent may actually be inaccurate. Inaccuracies could reflect deliberate efforts to conceal information, carelessness, or a lack of measurement ability. Researchers evaluating informational interventions should consider the strength or weakness of each node in the information production and dissemination process, though all nodes may not be relevant for all interventions.

6.1. Using the framework to understand null results

This "information pipeline" can act as a diagnostic tool to help researchers understand why a particular intervention may have, or (in our case) may not have, had the expected impacts. Using this framework, we identify several breakdowns in NextDrop's system: valvemen often failed to submit notifications to NextDrop (Node 1; Fig. 3), many household "waiters" either did not receive or did not

[17] ITT estimates are calculated based on observed differences between the entire population assigned to treatment and that assigned to control conditions, regardless of whether not treatment group members complied with the treatment. See the Online Appendix for results without covariate adjustment (Table A.3).

[18] Results available upon request. See the Online Appendix for our pre-analysis plan.

[19] We estimate power as Power = 1 − Φ (1.96 − Effect size/SE) + Φ (−1.96 − Effect size/SE), where Φ is the cumulative distribution function for a standard normal random variable, and SE is the standard error for the average effect size where standard errors are clustered at the cluster level. The reported power is the probability that the null hypothesis of a zero average treatment effect is rejected at the 5% level. The minimum detectable effect (MDE) is an estimate of the smallest effect size that would yield a test with 80% power (see e.g. Miguel, Kevin, Hicks, Eric, Kremer (2016)). We used the following formula: MDE = (1.96 + 0.84)σ, where σ is the standard error of the coefficient on the treatment indicator in a regression including the treatment indicator, relevant covariates, and clustered standard errors (the regression model we use in our analysis throughout the paper).

[20] These results do not appear to be driven by spillovers between the treatment and control groups. In our endline survey, we specifically asked respondents who had received notifications whether they had received them from the utility or from other sources. Only 11 respondents reported receiving notifications from anyone other than BWSSB.

[21] These decreases are slightly less significant statistically after adjustments for multiple hypothesis testing.

[22] A frontline worker, sometimes called the worker at the 'last mile', is the last point of connection between the service provider and the customer. Examples include electricity meter readers, telecommunications linemen, letter carriers, and, in this case, water valvemen.

Table 2
Intent-to-Treat Estimates (with covariate adjustment).[1]

Outcome[1]	Overall Population[2]			Low Income Group[3]			Target population[4]		
	Control mean[5]	ATE	P[6]	Control mean	ATE	P	Control mean	ATE	P
Household welfare effects									
Time spent waiting for water	0.51	−0.04	0.54 [0.94]	0.79	−0.05	0.71 [0.94]	0.71	0.01	0.91 [0.94]
Missing community events	0.13	0.00	0.89 [0.94]	0.16	0.03	0.46 [0.94]	0.14	0.00	0.91 [0.94]
Hours of work missed	2.28	−0.71	0.33 [0.94]	3.00	−0.05	0.94 [0.94]	3.30	0.15	0.90 [0.94]
Need for substitutes[6]	0.16	−0.02	0.18 [0.94]	0.23	−0.03	0.25 [0.94]	0.22	−0.07	0.03 [0.34]
Psychological effects[7]									
Worrying about water	2.63	0.11	0.04 [0.08]	2.51	0.12	0.07 [0.10]	2.53	0.05	0.31 [0.31]
Thinking about water during the day	2.77	0.10	0.04 [0.08]	2.52	0.22	0.00 [0.02]	2.60	0.07	0.23 [0.27]
Political effects									
Perception that providers are competent	1.35	−0.05	0.16 [0.99]	1.34	−0.08	0.25 [0.99]	1.33	−0.04	0.54 [0.99]
Perceptions that providers are innovative and modern	1.50	−0.02	0.61 [0.99]	1.47	0.00	0.97 [0.99]	1.48	−0.02	0.75 [0.99]
Perception that providers care about "people like us"	1.66	0.00	0.99 [0.99]	1.63	0.01	0.93 [0.99]	1.66	−0.05	0.41 [0.99]
Contacting									
Contacting providers directly about problems with service	0.09	0.00	0.84 [0.91]	0.05	−0.03	0.08 [0.23]	0.02	0.03	0.11 [0.23]
Holding state water utility directly responsible for service	0.22	0.00	0.91 [0.91]	0.16	−0.06	0.02 [0.12]	0.13	0.02	0.55 [0.83]
N	2440			848			642		
N Treated	1227			426			336		

Note: [1] Unless noted otherwise hypothesis testing has been conducted using two-tailed tests. [2] The covariates included are a block indicator, baseline outcome variable, an indicator for whether or not a household is low income, whether household receives Kaveri water supply, whether a household receives water supply every 2–4 days, whether a household receives supply without regularity (everyday supply is the omitted category), whether or not the household has an overhead tank/sump, and with the exception of row 1, the time reported waiting for water in wave 1. [3] Covariates included are the same as those included for the overall population. [4] Covariates included are the same as those included for the overall population, with the exception of whether the household has a tank and whether the household is low income. [5] Mean for control group in wave 2 of survey. [6] Calculated using Fisher's exact tests. P-values adjusted using Benjamini-Hochberg adjustments for multiple testing are in brackets. [7] Hypothesis testing based on one-tailed tests.

register receipt of NextDrop notifications (Nodes 4 and 5), and many notifications were inaccurate (Node 6). These observations suggest that "leaking" nodes in the production and dissemination of information played a large role in the program's failure.

In the analysis that follows, we discuss how specific pipeline leaks and blockages contributed to reductions in the effective size of our treatment group (Fig. 4). First, valvemen had to submit information by calling NextDrop's automated voice mail system to log valve opening times—this was the key factor in NextDrop's ability to collect the information it aimed to disseminate. Comparing our geo-coded survey responses with logs of the valvemen's own reports to NextDrop, we find that valvemen sent reports to NextDrop approximately 70% of the time.[23] Reporting rates were equivalent for the treatment and control groups.[24] Valveman non-reporting reduced the effective size of our treatment group from 1193 to 854.

For information to have an impact, it must not only be sent, but must also be received. Only 38% of treatment group members (453 out of 1193) reported receiving notifications at least once every two weeks, a much lower percentage than the 70% rate at which valvemen were regularly submitting notifications (Fig. 4). This gap between treatment assignment and actual receipt of messages becomes larger for the populations for which we expected the intervention to have the greatest effect, i.e., lower income households without automatically-filling tanks; only 25% of this target group reported receiving notifications.

What explains the low rate at which treatment group households reported receiving messages? First, information could have been lost in the transmission process. This clearly happened in our study: many household "waiters" for water did not possess the cellphone registered with NextDrop. Women in poor and middle-class households did not possess the household cellphone in similar proportions. This was the case for 207 out of the 854 households that were regularly sent notifications, reducing our effective treatment group size to 647. Gender differentials in mobile access, then, diluted any impact that even accurate notifications might have had (Fig. 4). The drop-off associated with differential access was not as large as that associated with valveman non-reporting.

Given that cellular phone services are quite reliable in urban India, we infer that much of the remaining discrepancy between the number of households sent messages and those that registered receipt can be attributed to respondents simply not noticing

[23] We reached the same percentage through two different calculations. First, we analyzed the number of valvemen reports a week to NextDrop as a percentage of expected reports for each valve area, based on the official utility supply schedule, for 4 weeks prior to the endline survey. In addition, we compared household survey responses naming the last water supply day with valveman reports for each valve area for the week preceding the endline survey. The geo-coded nature of our data facilitated this analysis. Moreover, our parallel, ethnographic study of the valvemen in the NextDrop intervention (in a different subdivision) also found that they did not submit information regularly, or sometimes submitted inaccurate notifications—e.g., sending a round of notifications during tea breaks rather than when actually turning on water valves (Hyun et al., 2018).

[24] Reports were sent at least 70% of the time to 72% of the treatment group households, and to 75% of the control group households. Control group households did not receive NextDrop notifications. Household refusals totaled only 3% of our sample. Because of our placebo design for enrollment, we can identify the set of non-compliers for both the treatment and control groups.

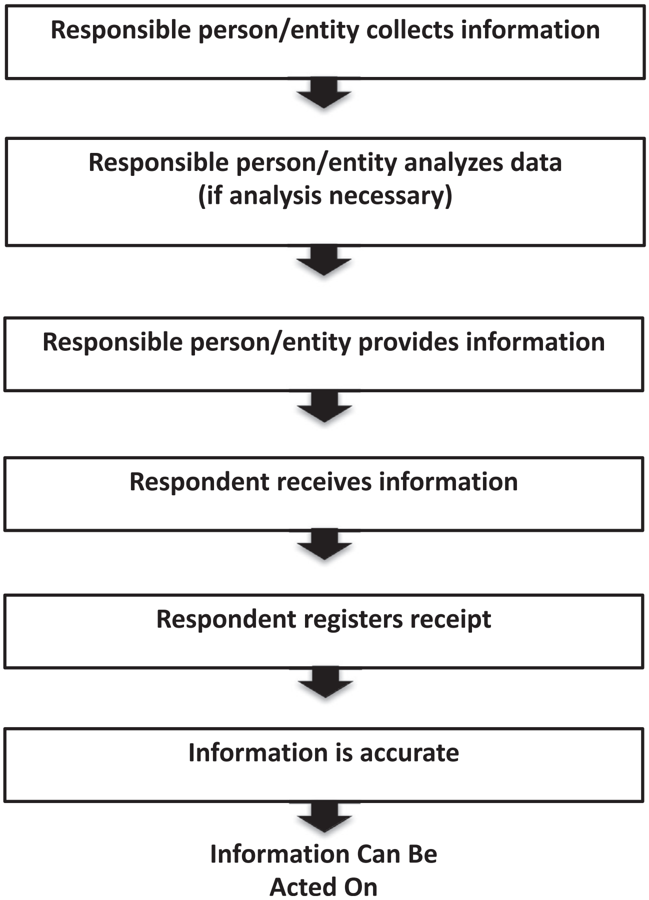

Fig. 3. Information pipeline: from collection to receipt.

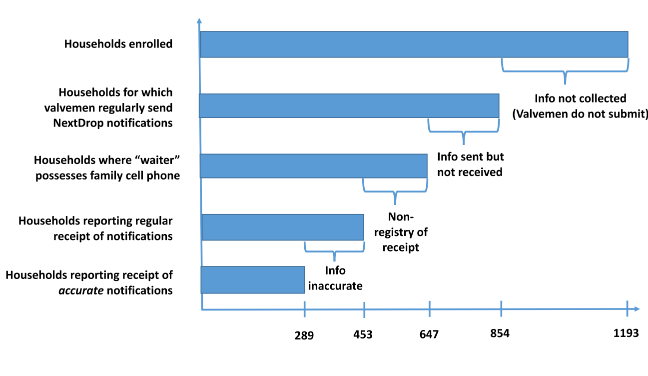

Fig. 4. NextDrop's leaky information pipeline and treatment group attrition.

NextDrop's texts or voicemails. They may have been inundated with text messages, or not bothered with subsequent notifications if the first one they received was not useful.[25] Non-registry of receipt appears to account for a reduction from 647 to 453 in our effective sample size (Fig. 4).

Finally, to be useful, information must be accurate. Only 289 of our 1193 treatment group respondents reported that the information received in NextDrop notifications was either always or usually accurate. To understand the extent of information inaccuracies, we compared household survey responses about the last day they had received water and the average time of water arrival to the time-stamped and geo-coded valveman reports for the relevant valve areas. Our analysis shows that, where valvemen had submitted reports in the week prior to our endline survey, 36% of households reported receiving water on a different day than that reported by the valveman. Furthermore, a comparison of household reports with valveman report data on water arrival times suggests that 62% of households received reports *after* the water had arrived (Fig. A.2, Online Appendix).[26] These inaccuracies further reduced our effective treatment group size from 453 to 289. They may also explain why so many households who were sent notifications did not register receiving them; those charged with waiting for water may have not understood their purpose or begun ignoring them. Thus serious problems of data accuracy compounded already significant problems with non-reporting.

6.2. Using the framework to identify subgroups receiving treatment

Given the extensive leakage in NextDrop's information pipeline, evaluating our original hypotheses, rather than simply assessing the impact of NextDrop services more broadly, requires calculating effects for those actually receiving *accurate* prior information.[27] Given the high rates of noncompliance, or "incomplete administration of treatment" (Angrist, 2006), a precise assessment of the impact of receiving notifications entails analyzing the causal effects for compliers. We define compliers (n = 289) as those who actually received the treatment of interest, namely accurate information.

Table 3 presents our Complier Average Causal Effect (CACE) estimates for the subset of households reporting that they had received accurate notifications (our intended treatment).[28] The tables include results for the entire study population, low-income blocks, and our target households. Similar to the ITT results, we observe no significant differences between the treatment and control groups for most outcomes.[29] Accurate notifications do, however, appear to have reduced levels of stress and worry among low-income households to a greater extent than suggested by the ITT results. A larger effect size among the (small) group receiving accurate information supports our original hypotheses about the effects of predictability on stress and worry.[30]

6.3. Using the framework to design preliminary research

Our "information pipeline" framework is not just a diagnostic tool for understanding the specific points at which an information production and dissemination process has broken down. It can also be used proactively before the launch of a major initiative, or

[25] We did not ask survey respondents how many text messages they received on an average day. Study participants could choose to receive notifications via text, or voicemail, in their chosen languages, so it seems unlikely that treatment group members did not understand the notifications they received.

[26] This percentage was calculated for the 33% percentage of households who were able to name a specific time when their water typically arrived residing in valve areas where valvemen had issued reports in the month prior to their survey interview.

[27] We registered our revised empirical strategy entailing the analysis of effects among those who received accurate messages as an amendment to our pre-analysis plan prior to the receipt of data from our endline survey (see the Online Appendix).

[28] For further discussions of noncompliance and obtaining CACE estimates, see Gerber and Green (2012).

[29] For results without covariate adjustment, see the Online Appendix (Table A.6). CACE estimates for the somewhat larger subset of households simply reporting that they received notifications—accurate or inaccurate—are similar (Tables A.4 and A.5). As a robustness check, we calculated the average treatment effect for those receiving accurate messages based on a different measure of message accuracy: whether or not geo-coded survey responses regarding the last day water had arrived, and the time water usually arrived, corresponded with valveman reports from the preceding week. These analyses suggested modest impacts at best (Online appendix, Tables A.6 and A.7.).

[30] Further details, as well as power, of our CACE calculations are in the Online Appendix (Tables A9 & A10).

Table 3
CACE for Households Receiving Accurate Notifications (with covariate adjustment).

Outcome[1]	Overall Population[2]			Low Income Group[3]			Target population[4]		
	Control mean[5]	CACE	SE[6][P][7]	Control mean	CACE	SE[P]	Control mean	CACE	SE[P]
Household welfare effects									
Time spent waiting for water	0.49	−0.03	0.21 [0.90]	0.75	−0.09	0.70 [0.90]	0.67	0.61	0.95 [0.90]
Missing community events	0.13	−0.01	0.06 [0.90]	0.16	0.14	0.19 [0.90]	0.14	−0.03	0.19 [0.90]
Hours of work missed	2.01	−1.68	2.28 [0.90]	2.75	1.73	5.41 [0.90]	2.68	2.54	9.41 [0.90]
Need for substitutes[8]	0.15	−0.06	0.07 [0.80]	0.23	−0.18	0.22 [0.80]	0.22	−0.60	0.24 [0.07]
Psychological effects[8]									
Worrying about water	2.64	0.40	0.21 [0.07]	2.51	0.65	0.40 [0.08]	2.54	−0.41	0.64 [0.31]
Thinking about water during the day	2.78	0.37	0.20 [0.07]	2.54	1.16	0.43 [0.02]	2.61	−0.23	0.58 [0.34]
Political effects									
Perception that providers are competent	1.35	−0.18	0.12 [0.41]	1.34	−0.45	0.30 [0.41]	1.32	−0.25	0.33 [0.90]
Perceptions that providers are innovative and modern	1.50	−0.06	0.12 [0.90]	1.47	−0.02	0.34 [0.95]	1.48	−0.22	0.36 [0.90]
Perception that providers care about "people like us"	1.66	0.02	0.15 [0.95]	1.63	0.07	0.39 [0.95]	1.66	−0.58	0.35 [0.41]
Contacting									
Contacting providers directly about problems with service	0.09	0.01	0.04 [0.97]	0.05	−0.16	0.08 [0.10]	0.02	0.28	0.14 [0.10]
Holding state water utility directly responsible for service	0.22	0.00	0.08 [0.97]	0.16	−0.27	0.12 [0.10]	0.13	0.20	0.16 [0.29]
N[9]	2364			811			612		
N Treated	1193			403			319		
N compliers	289			71			45		

Note: [1] Unless noted otherwise hypothesis testing has been conducted using two-tailed tests. [2] The covariates included are a block indicator, baseline outcome variable, an indicator for whether or not a household is low income, whether household receives Kaveri water supply, whether a household receives water supply every 2–4 days, whether a household receives supply without regularity (everyday supply is the omitted category), whether or not the household has an overhead tank/sump, and with the exception of row 1, the time reported waiting for water in wave 1. [3] Covariates included are the same as those included for the overall population. [4] Covariates included are the same as those included for the overall population, with the exception of whether the household has a tank and whether the household is low income. [5] Mean for control group in wave 2 of survey. [6] Standard errors clustered at the cluster level. [7] P-values adjusted using Benjamini-Hochberg multiple testing corrections shown in brackets. [8] Hypothesis testing based on one-tailed tests. [9] Only those units in both treatment and control groups that agreed to sign up for NextDrop's services have been included.

before replicating a program in a new location, to identify potentially weak nodes in an intervention. Nodes 1, 2, 3 and 6 highlight the roles of those charged with collecting, analyzing and disseminating the relevant information. For our study, it would have been ideal to have conducted a pilot study in Bangalore, examining the extent to which valvemen submitted regular and accurate water arrival notifications, given that NextDrop had convinced BWSSB to add these notifications to valveman job descriptions.[31] If problems were detected, on account of the valvemen, NextDrop or the utility, we could have tried to understand the causes (and extent) of their noncompliance. In our specific case, we might have explored the incentives and sanctions in the system or whether valvemen faced conflicting mandates from different "principals" (see Gailmard & Patty, 2012; Shapiro, 2005; Maynard-Moody & Musheno, 2012).[32]

The fourth and fifth nodes in our information pipeline center on technical and human barriers to information receipt. Using our framework to proactively identify potential problems for NextDrop's rollout in Bangalore would have required investigating key differences between its smaller, less wealthy pilot location and its new site. NextDrop (and we) could have investigated if women possessed cell phones at lower rates in Bangalore, or whether rates of text spamming varied significantly between the two cities. Certainly, a framework like ours cannot help in anticipating every potential roadblock through pilot exercises; rolling out in a new locations inevitably leads to unexpected barriers as well as opportunities. It is possible, however, that the larger leakages along the information pipeline could have been anticipated.

7. Discussion

In this study, we examined the impact of an informational intervention designed to reduce the coping costs associated with water intermittency. Our experimental evaluation of the impact of NextDrop's water arrival notification system during its rollout in Bangalore suggests that the main impact of the program was a modest reduction in stress levels associated with managing household water among low-income households.[33] These results surprised

[31] While NextDrop did not anticipate problems with message accuracy based on its experiences or customer feedback in Hubli-Dharwad, a framework such as ours would have encouraged them to track not only submission rates, but message accuracy, during a Bangalore-based pilot.

[32] We did, in fact, explore how these factors affected valveman incentives and behavior in a mixed-methods study conducted in parallel with – not prior to – our impact evaluation, and in a different part of the city. We found that valvemen viewed NextDrop notifications as an additional burden imposed on top of a large existing workload, and that the utility did not provide sufficient incentives for them to prioritize notifications above preexisting responsibilities. Meanwhile, the utility needed valvemen enough that their noncompliance with NextDrop requests would have few to no consequences (Hyun et al., 2018).

[33] This somewhat significant effect is visible in our CACE analysis as well, with larger effects for households actually receiving accurate notifications Table 3.

us because the intervention did not affect wait times for water, the variable through which we expected psychological effects to be mediated. It may be that individuals worried less because they *felt* more informed with NextDrop notifications. Indeed, about 85% of our treatment group reported in our endline survey that they found the NextDrop notifications "useful"; almost 74% of those who claimed that notifications were rarely or never accurate nevertheless found them useful. This suggests that (some) informational interventions provide psychological gains for households even when the information is not used or even usable. These results also indicate that stress as an indicator of household welfare should be examined more frequently in development interventions; this is an important, and often gendered, outcome in its own right.

Our ability to evaluate the effect of receiving accurate notifications on household coping costs was attenuated because of NextDrop's leaky information pipeline. The leakiest nodes in our case were related to the valvemen, or the frontline workers, who, for many reasons, did not send timely notifications to NextDrop. Frontline workers are a crucial node not only in informational interventions, but in development programs more broadly. Many influential experimental evaluations in the development literature mention the roles of the human last mile in their methods sections, but do not return to their potentially critical roles when explaining their findings. This is especially the case for studies that report a "successful" result (see, for example, Cohen and Dupas (2010) in Kenya; Banerjee, Duflo, Glennerster, and Kothari (2010) in India; Blattman et al. (2014) in Liberia). None of these papers discusses the skills or motivations of the NGO workers or local intermediaries in explaining their positive results; this trend leaves the evaluation literature with an implicit message that failed experiments need to be explained (Karlan & Appel, 2016), while successful interventions somehow do not.

That we did not conduct enough research to identify these leaks in advance is an obvious limitation of our study. Yet, many researchers conducting impact evaluations of existing programs will, and do, face constraints similar to those we faced: namely, the necessity of negotiating a research design with partners, and the inability to control program implementation. Donors and foundations have increasingly called for, and often required, rigorous evidence that programs actually achieve their intended effects before funding organizations to scale up their operations; evidence from a randomized controlled trial (RCT) is viewed as most compelling. These same agencies now fund researchers to study existing programs in the field and provide the necessary evidence; many find less convincing the results of artificial interventions that are tightly controlled by researchers.[34] This approach makes sense for funders concerned with understanding if real-world programs deserve investment. For academic evaluators, this means that study designs must be negotiated, and researchers cannot (and arguably should not) control many aspects of program implementation.[35]

The need to work with partners who are not simply implementing a program conceived of by academics, but who have their own organizational ambitions and constraints, means that the program actually evaluated under an RCT can differ in important respects from pilots. For instance, to avoid interference between treatment and control, RCTs may have to be conducted with new populations, in new locations, where the organization is still planning to expand. These new sites will never completely resemble the original locations, and it is hard to fully anticipate the relevant differences, even with a framework-derived "checklist." Similarly, a pilot location in which an organization is anxious to succeed may generate a highly responsive and vigilant implementation effort. However, NGOs, social enterprises, and governments eventually face strong incentives to use standardized, lower-touch administrative structures that they can manage, and afford, across multiple locations. NextDrop's decision to ask BWSSB to make the valvemen's notifications part of their job descriptions (the Bangalore scale-up), rather than to incentivize them individually as they had originally done (the Hubli-Dharwad pilot), represents a concrete instance of this general phenomenon.[36] As external evaluators, we did not incentivize valvemen to cooperate with NextDrop, though we were aware that they were a possible weak node. We wished to study the actual intervention that NextDrop and BWSSB planned to implement (see Hyun et al. (2018) for a companion paper on valveman motivations in NextDrop's program). We also did not anticipate the rates at which men would take the family cell phone to work – we saw that once the study was already under way – but we would not have considered "incentivizing" them to leave their phones at home. These are common experiences, we imagine, for those conducting development research.

8. Conclusion

The development community has been enthusiastic about informational interventions, but, as Pande (2011) notes, we still know little about how broad shifts in the information environments of most countries can be achieved. Our information pipeline highlights six potential leaks or blockages between information generation and actionable knowledge: failure to collect the intended information, failure to perform required analyses, failure to disseminate the information; citizen non-receipt of information due to technical or other factors; citizen non-registration of information received because of logistical or linguistic factors; and the provision of inaccurate information. Future studies of informational interventions should be structured to collect data on each of these nodes. This will not only allow researchers to explain their results, but also to proactively identify potential problems in programs before they roll out in new locations. A framework such as ours is necessarily specific to a particular type of intervention as causal pathways related to project implementation will vary with the type of program under study. Diagnostic frameworks like this, however, should be used more extensively in development research to understand the causal mechanisms in successful and unsuccessful programs, as well as inform preliminary research.

We illustrate the usefulness of our framework for evaluating an intervention designed to reduce the coping costs associated with water intermittency. Piped water for domestic use is perhaps *the* most important local service for human health and development outcomes. We measure the costs imposed by intermittent and erratic water services; we also emphasize the need for scholars of development, and political economy more broadly, to include measures of service quality such as frequency and predictability in their analyses of service access. These are familiar analytical variables in engineering, public health, and energy and water policy, but much less so for the social sciences.

---

[34] Many influential development-oriented experiments from the 1990s to the present day have been initiated by researchers. When researchers conceive of, design, and direct the implementation of an intervention (i.e., the researcher moves "from the role of the evaluator to the role of a coexperimenter" (Banerjee and Duflo, 2009, 154), they are more free to shape the parameters of an experiment. However, the experimental context then becomes a cross between the lab and the real world.

[35] To complicate matters, the timeframes offered by funders for program evaluations is quite compressed, with typical windows of two-three years to design, execute, and report on the results of a study. The time realistically needed to obtain permissions to conduct studies, negotiate the study site and design with the implementer, enter contracts with survey companies, etc., leaves little additional time for extensive pilots.

[36] After struggling with valveman noncompliance in Bangalore and Mysore during 2015, and further experimentation with business models that reduced their dependence on the valvemen, NextDrop discontinued their service in May 2016.

Aside from reductions in stress levels, an important outcome, our experimental results suggest that NextDrop's program failed to trigger changes in household welfare or state-society relations in Eastern Bangalore. However, our power to detect the impact of accurate notifications, and indeed the long run viability of NextDrop's programmatic model, was eroded by failures in the production and dissemination of the information around which this program was conceived. To the extent that our findings result partly from the one cell phone in the household being kept mainly by the men, our study reminds us that, too often, development interventions still treat households as a unitary construct, undifferentiated by internal gender dynamics (see Alderman, Chiappori, Haddad, Hoddinott, & Kanbur, 1995). To the extent that our findings result partly from non-cooperation by valvemen, our study serves as a reminder that frontline workers in public sector bureaucracies will play key roles whenever small-scale interventions are brought to scale. It makes more sense to study interventions in real-world settings, with government bureaucracies and in light of existing household structures, than in non-replicable settings with implementers or respondents answering to the research team.

Our findings highlight the importance of systematically investigating the extent to which both null *and* positive field experimental results depend upon the process of information production and dissemination. If an experiment suggests that a program is effective, funders may decide to replicate it elsewhere (to test for external validity) or to roll it out more generally (in a context deemed similar to that of the experiment). Yet as we demonstrate, replication in a different context is fraught with uncertainties (see also Ananthpur, Malik, and Rao, 2017; Bold, Mwangi Kimenyi, & Alice Ng'ang'a, 2013). While preliminary research can, and should wherever possible, be used to identify potential barriers ex ante, not all problems can be anticipated, or corrected even if anticipated, before implementation. This is simply the reality of field-based development research.

Acknowledgements

This project was funded through a "DIL Innovate" grant through the Development Impact Laboratory, Blum Center for Developing Economies, U.C. Berkeley (USAID Cooperative Agreement AID-OAA-A-13-00002). We thank C. Sharada Prasad, Zachary Burt, Francesc Pardo, Mary Rose Fissinger, and Megan Otsuka for their considerable contributions to this project; Maria Chang, Eugenia Giraudy, S. Krishna Kumar, Rajani Sampath, and Christine Thao Tyler for research assistance; and Ben Arnold, Adam Auerbach, Jennifer Bussell, Pradeep Chhibber, Thad Dunning, Christopher Hyun, Francesca Jensenius, Katerina Linos, Aila Matanock, Edward Miguel, Gareth Nellis, Malini Ranganathan, Jasjeet Sehkon, Laura Stoker, Rebecca Weitz-Shapiro, Erik Wibbels, Matthew Winters, seminar participants at U.C. Berkeley, Duke and Claremont McKenna, and especially Joel Middleton and Guadalupe Tunon for advice. We thank the anonymous reviewers of this journal whose comments we found very helpful. We also thank NextDrop, the Public Affairs Foundation (esp. Lalita Pulavarti and Narendra LV), and the Bangalore Water Supply and Sewerage Board (BWSSB) for their collaboration with the project. The project has been approved through IRB protocol 2014-04-6259 and approved by BWSSB. Pre-Analysis Plan 20150514AA registered with egap.org.

Appendix A. Online Appendix

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.worlddev.2018.01.022.

References

Akram, Agha Ali., & Olmstead, Sheila M. (2010). The value of household water service quality in Lahore, Pakistan. *Environmental and Resource Economics, 49*(2), 173–198. https://doi.org/10.1007/s10640-010-9429-7.

Alderman, H., Chiappori, P. A., Haddad, L., Hoddinott, J., & Kanbur, R. (1995). Unitary versus collective models of the household: Is it time to shift the burden of proof? *World Bank Research Observer, 10*, 1–19.

Ananthpur, Kripa, Malik, Kabir, & Rao, Vijayendra (2017). The anatomy of failure: An ethnography of a randomized controlled trial to deepen democracy in rural India. *World Development, 99*, 481–497.

Angrist, Joshua D. (2006). Instrumental variables methods in experimental criminological research: What, why and how. *Journal of Experimental Criminology, 2*(1), 23–44.

Baisa, Brian, Davis, Lucas W., Salant, Stephen W., & Wilcox, William. (2010). The welfare costs of unreliable water service. *Journal of Development Economics, 92*(1), 1–12.

Banerjee, Abhijit V., & Duflo, Esther. (2009). The experimental approach to development economics. *Annual Review of Economics, 1*, 151–178.

Banerjee, Abhijit Vinayak, Duflo, Esther, Glennerster, Rachel, & Kothari, Dhruva. (2010). Improving immunisation coverage in rural India: Clustered randomised controlled evaluation of immunisation campaigns with and without incentives. *BMJ, 340*, c2220.

Berenschot, Ward (2010). Everyday mediation: The politics of public service delivery in Gujarat, India. *Development and Change, 41*(5), 883–905.

Blattman, Christopher, Hartman, Alexandra C., & Blair, Robert A. (2014). How to promote order and property rights under weak rule of law? An experiment in changing dispute resolution behavior through community education. *American Political Science Review, 108*(01), 100–120.

Bold, Tessa, Kimenyi, Mwangi, Mwabu, Germano, Ng'ang'a, Alice, & Sandefur, Justin (2013). *Scaling up what works: experimental evidence on external validity in kenyan education* Working Paper 321. Washington D.C.: Center for Global Development.

Burt, Zachary, Ercumen, Ayse, Billave, Narayan, & Ray, Isha. (Under review). "From intermittent to continuous service: Costs, benefits, equity and sustainability of water system reforms in Hubli-Dharwad, India". Berkeley, CA.

Burt, Zachary, & Ray, Isha (2014). Storage and non-payment: Persistent informalities within the formal water supply of Hubli-Dharwad, India. *Water Alternatives, 7*(1), 106–120.

Cohen, Jessica, & Dupas, Pascaline (2010). Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment. *The Quarterly Journal of Economics, 125*(1), 1–45.

Connors, Genevieve (2005). When utilities muddle through: Pro-Poor governance in Bangalore's public water sector. *Environment and Urbanization, 17*(1), 201–218.

Corbridge, Stuart, Williams, Glyn, Srivastava, Manoj, & Véron, René (2005). *Seeing the state: Governance and governmentality in India*. Cambridge: Cambridge University Press. http://ebooks.cambridge.org/ref/id/CBO9780511492211.

Dauda, Suleiman Alhaji, Yacob, Mohd Rusli, & Radam, Alias (2014). Household's willingness to pay for heterogeneous attributes of drinking water quality and services improvement: An application of choice experiment. *Applied Water Science, 5*(3), 253–259. https://doi.org/10.1007/s13201-014-0186-6.

Dunning, Thad, Grossman, Guy, Humphreys, Macartan, Hyde, Susan, & McIntosh, Craig. (Forthcoming). Metaketa I: the limits of electoral accountability. Cambridge University Press.

Ercumen, Ayse., Arnold, Benjamin F., Kumpel, Emily, Burt, Zachary, Ray, Isha, Nelson, Kara, & Colford, John M. (2015). Upgrading a piped water supply from intermittent to continuous delivery and association with waterborne illness: A matched cohort study in Urban India. *PLoS Medicine, 12*(10). https://doi.org/10.1371/journal.pmed.1001892.

Estache, Antonio, Gomez-Lobo, Andres, & Leipziger, Danny (2001). Utilities privatization and the poor: Lessons and evidence from latin america. *World Development, 29*(7), 1179–1198.

Evans, Peter. (2008). In search of the 21st century developmental state. The Centre for Global Policity Economy, University of Sussex Working Paper 4.

Ferguson, James, & Gupta, Akhil (2002). Spatializing states: Toward an ethnography of neoliberal governmentality. *American Ethnologist, 29*(4), 981–1002.

Gailmard, Sean, & Patty, John W. (2012). Formal models of bureaucracy. *Annual Review of Political Science, 15*(1), 353–377.

Galaitsi, S. E., Russell, Robert, Bishara, Amahl, Durant, John L., Bogle, Jennifer, & Huber-Lee, Annette (2016). Intermittent domestic water supply: A critical review and analysis of causal-consequential pathways. *Water, 8*(274), 1–25.

Gerber, Alan S., & Green, Donald P. (2012). *Field experiments: Design, analysis, and interpretation*. New York, NY: W.W. Norton.

Ghertner, D. Asher (2011). Gentrifying the state, gentrifying participation: elite governance programs in Delhi: gentrifying the local state in India. *International Journal of Urban and Regional Research, 35*(3), 504–532.

Harriss, John (2006). Middle-class activism and the politics of the informal working class: A perspective on class relations and civil society in Indian cities. *Critical Asian Studies, 38*(4), 445–465.

Hyun, Christopher, Post, Alison E., & Ray, Isha (2018). Frontline worker compliance with transparency reforms: Barriers posed by family and financial responsibilities. *Governance, 31*, 65–83.

Imbens, Guido W. (2011). *Experimental design for unit and cluster randomized trials* Working Paper. Harvard University.

Karlan, Dean, & Appel, Jacob (2016). *Failing in the field: What we can learn when research goes wrong*. Princeton, N.J.: Princeton University Press.

Kjellén, Marianne, & McGranahan, Gordon (2006). *Informal water vendors and the urban poor. Human settlements discussion paper series*. London: International Institute for Environment and Development London. http://www.mumbaidp24seven.in/reference/informal_water_vendors.pdf.

Krishna, Anirudh. (2011). Gaining access to public services and the democratic state in India: Institutions in the middle. *Studies in Comparative International Development, 46*(1), 98–117.

Kumpel, Emily., & Nelson, Kara L. (2013). Comparing microbial water quality in an intermittent and continuous piped water supply. *Water Research, 47*(14), 5176–5188.

Kumpel, Emily, & Nelson, Kara L. (2016). Intermittent water supply: Prevalence, practice, and microbial water quality. *Environmental Science & Technology, 50*(2), 542–553.

Kumpel, Emily, Woelfle-Erskine, Cleo, Ray, Isha, & Nelson, Kara L. (2017). Measuring household consumption and waste in unmetered, intermittent piped water systems. *Water Resources Research, 53*(1), 302–315. https://doi.org/10.1002/2016WR019702.

Kuriyan, Renee, & Ray, Isha (2009). Outsourcing the state? Public-private partnerships and information technologies in India. *World Development, 37*(10), 1663–1673.

Lieberman, Evan S., Posner, Daniel N., & Tsai, Lily L. (2014). Does information lead to more active citizenship? Evidence from an education intervention in Rural Kenya. *World Development, 60*(August), 69–83.

Lipsky, Michael (1980). *Street-level bureaucracy: Dilemmas of the individual in public service*. New York, NY: Russell Sage Foundation.

Madon, S., & Sahay, S. (2002). An information-based model of NGO mediation for the empowerment of slum dwellers in Bangalore. *The Information Society, 18*(1), 13–19.

Maynard-Moody, Steven, & Musheno, Michael (2012). Social equities and inequities in practice: Street-level workers as agents and pragmatists. *Public Administration Review, 72*(s1), S16–S23.

McKenzie, David, & Ray, Isha (2009). Urban water supply in India: Status, reform options and possible lessons. *Water Policy, 11*(4), 442–460.

Miguel, Edward A., Kevin, Croke, Hicks, Joan Hamory, Eric, Hsu, & Kremer, Michael Robert (2016). *Does mass deworming affect child nutrition ? Meta-analysis, cost-effectiveness, and statistical power* WPS7921. The World Bank. http://documents.worldbank.org/curated/en/926411482169164466/Does-mass-deworming-affect-child-nutrition-meta-analysis-cost-effectiveness-and-statistical-power?cid=DEC_PolicyResearchEN_D_INT.

Mullainathan, Sendhil, & Shafir, Eldar (2013). *Scarcity: Why having too little means so much*. New York, NY: Times Books, Henry Holt and Company.

Pande, Rohini (2011). Can informed voters enforce better governance? Experiments in low-income democracies. *Annual Review of Economics, 3*(1), 215–237.

Pattanayak, Subhrendu K., Yang, Jui-Chen, Whittington, Dale, & Bal Kumar, K. C. (2005). Coping with unreliable public water supplies: Averting expenditures by households in Kathmandu, Nepal. *Water Resources Research, 41*(2). n/a-n/a.

Ranganathan, Malini (2014). Mafias' in the waterscape: Urban informality and everyday public authority in Bangalore. *Water Alternatives, 7*(1). http://search.proquest.com/openview/8bd2275e590fe373749d8d00c2d0ee3e/1?pq-origsite=gscholar.

Shapiro, Susan P. (2005). Agency theory. *Annual Review of Sociology, 31*(1), 263–284.

Stevenson, Edward G. J., Greene, Leslie E., Maes, Kenneth C., Ambelu, Argaw, Tesfaye, Yihenew Alemu, Rheingans, Richard, & Hadley, Craig (2012). Water insecurity in 3 dimensions: An anthropological perspective on water and women's psychosocial distress in Ethiopia. *Social Science & Medicine, 75*(2), 392–400.

Stokes, Susan C., Dunning, Thad, Nazareno, Marcelo, & Brusco, Valeria (2013). *Brokers, voters, and clientelism: The puzzle of distributive politics*. New York, NY: Cambridge University Press.

Subbaraman, Ramnath, Nolan, Laura, Sawant, Kiran, Shitole, Shrutika, Shitole, Tejal, Nanarkar, Mahesh, et al. (2015). Multidimensional measurement of household water poverty in a mumbai slum: Looking beyond water quality. *PLOS ONE, 10*(7), e0133241.

Tolbert, Caroline J., & Mossberger, Karen (2006). The effects of e-government on trust and confidence in government. *Public Administration Review, 66*(3), 354–369.

World Bank (2004). *World development report 2004: Making services work for poor people*. Washington D.C: World Bank.

Wutich, Amber, & Ragsdale, Kathleen (2008). Water insecurity and emotional distress: Coping with supply, access, and seasonal variability of water in a Bolivian squatter settlement. *Social Science & Medicine, 67*(12), 2116–2125.

Zérah, Mar.-Hélène (2000). Household strategies for coping with unreliable water supplies: The case of Delhi. *Habitat International, 24*(3), 295–307.