Massive Citizen Reporting is Too Inconsistent and Costly to Improve Public Services: A Framework and Field Experiment

Mark T. Buntaine[1], Patrick Hunnicutt[2], and Polycarp Komakech[3]

[1]University of California, Santa Barbara, buntaine@bren.ucsb.edu

[2]University of California, Santa Barbara, phunnicutt@bren.ucsb.edu

[3]University of California, Santa Barbara, k.polycarp@gmail.com

May 1, 2018

Abstract

Governments around the world are investing in technologies that allow massive, frequent, and localized contact with citizens, though there is little evidence about the impacts of the streams of data these technologies create on the delivery of public services. We report a large-scale randomized controlled trial that involved recruiting 50 citizens in each of 100 neighborhoods across Kampala, Uganda to provide weekly reports on the delivery of solid waste service via an SMS-messaging platform to a municipal government, resulting in 17,520 verified and usable reports during the study period. Citizen reporting did not reduce waste accumulation. More positively, reporting reduced the amount of burning and unmanaged waste piles for a time, but this positive result did not persist after an unexpected staff restructuring in the unit responsible for waste management. Waste collection did not improve in zones with more reports or more dissatisfied reporters. Using our observations as participants in development and deployment of the platform and interviews with key staff at the government agency receiving citizen reports, we show how the adoption of new technologies to collect data from citizens requires both new capacities and data consistent enough to reduce uncertainty about the allocation of effort. We provide a formal framework for analyzing the challenge of utilizing citizen-sourced data for the management of public services. Citizen-sourced data must be both low-cost relative to alternatives and consistent enough to reduce uncertainty about decisions related to public effort.

1 Introduction

Governments around the world are building or adopting platforms to collect and process feedback from citizens about public services. New information and communication technologies (ICTs) enable governments to collect dispersed observations from citizens, opinions about service quality, and ideas for improvements to services, in addition to the ability to communicate with the public about responses taken to address concerns, all at lower costs. These tools also increase the potential for public agencies to aggregate and track trends in relevant information and follow-up actions and to engage a broader set of citizens in collaborative management (Bertot, Jaeger and Grimes, 2010; Rotberg and Aker, 2013; Grossman, Humphreys and Sacramone-Lutz, 2014).

We provide both a field experimental test of the impact of citizen reporting on public services and first-hand qualitative analysis on the adoption and operation of such a platform, in addition to a general framework for understanding when these ICT tools have the greatest potential to improve public services. Despite optimism that ICT can contribute to public sector performance and the delivery of public services, limited empirical evidence exists about whether the adoption of new ICT tools actually fosters effective public engagement and improves the public services that citizens care about. Furthermore, a formal framework that outlines the conditions under which data from citizens will be useful for decision-making as compared to self-monitoring by government agencies is not available.

While data from citizens can be collected more broadly and in a more timely manner than traditional ways of monitoring public services, there are a number of reasons to be cautious about their potential. Processing new flows of data and turning them into information that can be used for decision-making requires new skills and capacities, potentially implying significant costs. Being responsive to new information may require a realignment of work effort, which can also be costly, politically contentious, or limited by existing procedure.

Besides practical considerations about acting on data, the quality and consistency of data from citizens about public services may be significantly lower than what can be collected using more institutionalized monitoring systems. Citizen-sourced data is often unstructured, noisy, and inconsistent, which creates additional challenges for using it to improve public

services. For instance, citizens who want governments to exert more effect improving public service have incentives to send information that will attract attention and improvements regardless of the actual state of affairs. If this is the case, trust in the veracity of citizen-sourced data might break down. Furthermore, citizen-sourced data may be inconsistent enough because of measurement error that it does not offer the basis for reducing uncertainty about decisions, eroding its usefulness for the management of public services.

In the present study, we offer two kinds of evidence about the impacts that the adoption of ICT tools to collect citizen-sourced data have had for the delivery of public services. In partnership with the Kampala Capital City Authority (KCCA), we study the adoption of a new SMS text-messaging platform to collect, process, and aggregate citizen feedback about waste collection services, which resulted in 17,520 verified and usable reports over the approximately nine-month study period. First, we present results from a large-scale randomized field experiment that involved recruiting reporters from randomly-assigned neighborhoods to send reports about waste collection services to the KCCA waste management unit. Because the accumulation of solid waste in informal piles is visible and results from a lack of collection services, we are able to independently audit whether citizen reporting decreases informal dumping and burning of waste by residents, as compared to neighborhoods where citizen reporters are not active.

Second, we use participant observations to interpret the opportunities and challenges of adopting new technologies and processing data to improve public services. Our research team was embedded at the KCCA, the agency that adopted the SMS platform, to assist with its development and use. We recorded our participant observations about the process of using citizen-sourced data systematically. We also conducted in-depth interviews with all of the KCCA staff who interacted with the platform to understand the opportunities and barriers of using citizen-sourced data to improve public services. We thus provide first-hand evidence about the challenges that governments will face when adopting these tools and the capacities that they should have in place to leverage citizen-sourced data for the management of public services.

To preview our results, we find that ICT-enabled citizen reporting did not significantly impacted waste accumulation in Kampala neighborhoods. In the nine-month study period

reported here, we do not find evidence that the amount of waste accumulation decreased in neighborhoods assigned to citizen reporting, as compared to neighborhoods without citizen reporting. We find some promising results in terms of the proportion of piles with burning and the amount of non-organic waste at the first post-treatment audit of informal waste piles five months after the baseline, but these results do not persist to the second post-treatment audit of informal waste piles four months later.

While the final results of the waste audits evince no impact, we gained additional insights about the process of using citizen-sourced data from an unexpected and disruptive reorganization at the KCCA, which included staff and management rotations that directly affected the waste management unit that developed and used the SMS text-messaging platform. This reorganization happened immediately after our first post-treatment data collection wave. The new director and team were much less committed to using the platform, considered the data that it produced to be unreliable and inconsistent, and preferred to develop their own systems for collecting data on the quality of collection services. We interpret observations about the transition between management teams to shed light on how changes in monitoring costs and willingness to update based on citizen-sourced data affect the potential of massive, high-frequency, and local citizen reporting to improve the management of public services.

Citizen-source data, even though it can be massive, timely, and localized, is no panacea for the problems facing public sector managers with limited budgets and costly options for monitoring the delivery of public services. While citizen reporting can save costs and allow for a greater proportion of available public resources to be spent on improving services, rather than providing monitoring and oversight, it is also likely to create a more inconsistent and unreliable data stream across many settings, requiring significant effort at processing and interpretation. Our formal framework helps illustrate how optimism about citizen-sourced data is likely misplaced across a range of realistic circumstances.

2 Literature and Theory

2.1 Related Research

The idea that information technologies can facilitate citizen monitoring of public services has sparked cautious enthusiasm (Oates, 2003; Grossman, Humphreys and Sacramone-Lutz, 2014; Charalabidis et al., 2012; Linders, 2012; Zurovac, Talisuna and Snow, 2012; Rotberg and Aker, 2013). Many researchers recognize great potential for new information technologies to make government agencies more responsive to citizen demands, but they also identify significant political, operational, and data-processing challenges to employing spatial information sourced from non-representative groups of citizens into the delivery of public services (Ntaliani, Costopoulou and Karetsos, 2008; Mossberger, Wu and Crawford, 2013; Evans and Campos, 2013).

While several prominent platforms have been established to generate citizen monitoring of public services in developed countries (e.g., SeeClickFix, FixMyStreet, NoiseTube), these platforms are not designed to facilitate research about foundational questions of citizen-sourced data provision, quality, and impact. To this point, the majority of research on citizen-sourced information deals with either disaster responses (e.g., Zook et al., 2010) or environmental monitoring (e.g., Connors, Lei and Kelly, 2012). Neither issue deals with eliciting long-term improvements to public services. Scholars from fields as varied as information science, technology studies, political science, and public administration recognize that research about mobile citizen feedback requires more focused empirical research approaches (Linders, 2012; Charalabidis et al., 2012; Saxton, Oh and Kishore, 2013; Seltzer and Mahmoudi, 2013). Recent research has begun this work related to participation in reporting on public services (Sjoberg, Mellon and Peixoto, 2017; Buntaine, Nielson and Skaggs, 2017), in addition to a recent experiment in Uganda that involved reporting deficiencies to politicians, rather than government agencies (Grossman, Platas and Rodden, 2017).

Results about the impacts of citizen monitoring of governments aside from ICT-platforms are mixed. Some studies indicate that monitoring alone is insufficient to generate substantial impact. For example, Olken (2007) finds that citizen monitoring of road construction projects in Indonesia through public meetings did not reduce the amount of funds lost. Banerjee,

Deaton and Duflo (2004) find that the assignment of individual monitors to health facilities in India did not increase the attendance of health workers. And Banerjee et al. (2010) report that increased public participation in monitoring education services in India did not increase educational achievement.

Research suggests that monitoring can have an impact on the delivery of public services when it is tied to credible enforcement mechanisms over providers. For example, Björkman and Svensson (2009) find that when a higher number of citizens formed a collective monitoring body for community health facilities in Uganda, they are able to increase the quality of services provided. In a follow-up paper (2010), they report that communities with many cleavages, which limits collective action, reduces the impacts of this kind of citizen monitoring. Reinikka and Svensson (2011) find that a large-scale newspaper campaign in Uganda reporting how local units spend educational grants from the central government reduced money lost to corruption. The campaign created shared knowledge among citizens, which in turn promoted more collective and political action demanding high-quality services.

Considering the mixed evidence to date about the impacts of non-governmental monitoring on the provision of public services, we offer formal framework for analyzing the potential of these citizen-sourced data based on costs, inconsistency, and uncertainty. We assume there is some amount budget and incentive to deliver public services in developing this framework.

2.2 Theoretical Preliminaries

Consider a public manager that is under a budget constraint C, such that her spending on improvements to public services can be divided between any number of tasks j such that $C = \sum c_j$. Her goal is to allocate this budget so that she maximizes improvements to public services. Her main problem is uncertainty about how to best allocate her resources to maximize payoffs, which we denote as $\sum \Theta_j$, where Θ_j is the payoff of task j that is stochastic in each period from an underlying probability.

For each task j, we assume the public manager has a belief about the underlying distribution of payoffs $p(\theta_j)$ that will be achieved when allocating budget toward that task. Without any additional monitoring, the manager chooses in order the tasks that have the highest payoff relative to the budget outlays that are required to accomplish them, until her

budget is used up.

The manager can also spend resources on monitoring m_j to acquire better information about the realized values Θ_j drawn from $p(\theta_j)$ in a period of effort. But the manager is also uncertain about the value of new information relative to using the underlying beliefs about the cost-effectiveness of actions. We assume the manager spends her entire budget constraint C in any scenario, such that her objective function is, with tasks $(1')$ through (k') being those chosen after monitoring:

$$U_m = \sum_{(1')}^{(k')} \Theta_j - \sum^{j} m_j \tag{1}$$

There are two primary problems that the manager must confront when attempting to maximize this objective function. She might pick the wrong tasks j for a given period of effort. And she might spend part of her budget on monitoring m_j that does not help her make better decisions, leaving less budget to actually deliver public services.

Monitoring technologies differ both in their costs and in their ability to reveal information about the realization of Θ_j such that they help make better decisions. It only makes sense to do any monitoring if the cost of acquiring information about the realization of Θ_j is less than the expected increase in payoffs attained from choosing tasks more effectively.

2.3 Allocating Effort and Monitoring

To make the problem more tractable, consider a manager with a set of tasks j that all have the same binary payoff structure $\Theta_j \in [0, 1]$, which in each instance is drawn from a Bernoulli process where there is an underlying true probability $p(\theta_j)$, which is known to the manager.

2.3.1 No monitoring

Consider the manager who does not have any access to monitoring technology, but must allocate effort toward tasks j when there exists a realized $\Theta_j \in [0, 1]$ drawn from the true distribution of $p(\theta_j)$. The costs of carrying out the tasks are fixed whether or not the payoff for each task is realized. For simplicity of exposition, we also assume that the cost of tasks j are constant.

The goal of the manager is to allocate tasks j such that $\sum_{(1)}^{(k)} \Theta_j$ is maximized in each period. When monitoring is not available that reveals information about the realizations of Θ_j, the manager will simply be guided by their beliefs and allocate effort to the tasks that have the highest payoffs in expectation. Specifically, she will order her beliefs about each task such that:

$$p(\theta_j)_{(1)} > p(\theta_j)_{(2)} > ... > p(\theta_j)_{(k)} > ... > p(\theta_j)_{(z)} \tag{2}$$

The manager will then choose tasks (1) through (k), where $U_m = \sum_{(1)}^{(k)} \Theta_j$. In this case, the payoff she expects is simply the sum of all probabilities through task (k):

$$E\left[\sum_{(1)}^{(k)} \Theta_j\right] = \sum_{(1)}^{(k)} p(\theta_j) \tag{3}$$

2.3.2 Perfect monitoring

If monitoring can be added that reveals information about the particular realization of Θ_j drawn from θ_j prior to the allocation of effort, then it becomes possible to make better choices. Consider first that the manager can pay some monitoring cost m_j that will reveal Θ_j, subject to the budget constraint such that $\sum m_j < C$. We assume that m_j is strictly less than c_j, otherwise it would never make sense to pay for monitoring.

The problem for the manager who has the option of perfect monitoring is when and where to spend resources on monitoring, leaving fewer resources for carrying out the public service tasks. If there are many tasks j, the costs of monitoring can be large and quickly consume her budget. The severity of the monitoring versus effort trade-off will depend on the relative costs of monitoring and action related to the public service; when monitoring is inexpensive the trade-off between searching for new information and acting is small, but when monitoring is expensive the trade-off is large.

Monitoring leads to increased payoffs only when it changes the allocation of effort. In particular, the payoff to monitoring will be exactly equal to the number of $\Theta_{(k)} = 0$ tasks avoided and replaced with tasks for which $\Theta_{(k')} = 1$ from Eq. 1. From changes in the allocation of effort between the baseline where no monitoring information is available ((1)

through (k)), to the allocation of effort following monitoring ($(1')$ through (k')), the value of monitoring can be conceptually described. The manager will choose to pay for monitoring whenever there exists an additional unit m_j for which the value of monitoring $V(m_j)$ is positive:

$$V(m_j) = \sum_{(1')}^{(k')} \Theta_j - \sum m_j - \sum_{(1)}^{(k)} p(\Theta_j) \tag{4}$$

The problem for the manager is that this value cannot be solved because the payoffs to monitoring are unknown prior to the application of monitoring. This is an extremely complex problem that defies simple analytical solutions, because the payoffs to m_j are not independent of other decisions m_{-j}. A complete set of monitoring decisions chosen will determine the ordering of posterior beliefs that will drive the actual allocation decision.

This problem can be simplified based on the idea of regret, as applied to individual parts of the ordering of the prior beliefs given by Eq. 2. For each $p(\theta_j)$ where $(j) \leq (k)$, the expected regret of allocating effort is equal to the expected probability that effort will lead to zero payoff:

$$E[R_j] = c_j * (1 - p(\theta_j)) \tag{5}$$

It makes sense it engage in perfect monitoring in sequence for each task j whenever the cost of monitoring is less than the expected value of the regret, that is where $m_j < R_j$, until there exists an ordering of beliefs given by Eq. 2 such that it no longer makes sense to pay for monitoring.[1] The expected payoff of monitoring within this dynamic search process will depend on a number of factors, including the sequence of underlying probabilities $p(\theta)$, the cost of monitoring m_j, the benefits that can be attained through effort Θ_j, and the budget constraint C. Overall, however, evaluating monitoring costs in terms of expected regret helps make sense of when it is advantageous to invest in perfect monitoring. When monitoring costs are high, more regret will be tolerated and vice versa.

[1]To avoid discontinuities in effort based on the budget constraint, we assume that partial effort can be applied to task (k'), with the payoff equal to zero or the proportion of full effort exerted.

2.3.3 Imperfect monitoring

Technologies for citizen monitoring might drive down the costs of monitoring, broaden the number of tasks that are monitored, and reduce the trade-off between monitoring and effort. Indeed, the push towards more bottom-up and citizen-driven monitoring systems is largely premised on the idea of increasing the coverage of monitoring at reduced costs, preserving more budget for the actual delivery of public services. Additionally, there may be political or reputational benefits for being responsive to citizen concerns (McCubbins and Schwartz, 1984).

Monitoring done by citizens is imperfect, however, because it comes in the form of a noisy signal about a particular draw of Θ_j. In the context of citizen monitoring, people who report on the value of Θ_j might disagree, have faulty observations, or provide misinformation purposefully, all of which will harm the ability of a manager to draw clear inferences about the true realized value of Θ_j, which would help to make decisions about effort.

With imperfect monitoring, the manager has to extract signal from noisy, citizen-sourced data. Imperfect monitoring is time-bound and only provides information on a single instance of Θ_j. The value of citizen monitoring for decision-making is directly related to the consistency and amount of the information provided by citizens. Consider that the manager must calculate $p(\Theta_j = 1|y_j)$ for each task j, where y_j is a mix of reports containing binary information about whether $\Theta_j = 1$. By Bayes rule, this belief about the particular realization of Θ given y_j can be computed as:

$$p(\Theta_j = 1|y_j) = \frac{p(y_j|\Theta_j = 1)p(\Theta_j = 1)}{p(y_j|\Theta_j = 1)p(\Theta_j = 1) + p(y_j|\Theta_j = 0)p(\Theta_j = 0)} \tag{6}$$

One additional assumption is needed to compute a posterior probability $p(\Theta_j = 1|y_j)$, which is the proportion of reports that are incorrect. We assume that the manager can estimate the proportion of incorrect reports $w \in [0, 1]$, by examining globally the proportion of reports that deviate from the modal value. An important assumption at this point is that the manager will be able to determine that the reports tend toward being correct or incorrect on average, that if 30 percent of reports deviate from the modal response, this indicates that 30 percent of reports are incorrect, rather than 70 percent of reports being incorrect. With

this assumption, the proportion of reports that deviate from the modal value will transform $p(\Theta_j)$ into a probability model for realized reports.

$$p(\Theta_j = 1|y_j) = \frac{\binom{n}{k}(\Theta - w)^k w^{n-k} p(\theta)}{\binom{n}{k}(\Theta - w)^k w^{n-k} p(\theta) + \binom{n}{k} w^k (\Theta - w)^{n-k}(1 - p(\theta))} \tag{7}$$

From this posterior belief, the value of imperfect monitoring (I) alone can be expressed by the expected amount of regret that is avoided as compared to acting only on prior beliefs:

$$V(I) = \sum_{(1')}^{(k')} p(\Theta_j = 1|y_j) - c_I - \sum_{(1)}^{(k)} p(\theta_j) \tag{8}$$

Since regret is a directly a function of the probability of making the wrong decision when allocating effort (see Eq. 5), the value of imperfect monitoring will be realized when it increases the confidence in allocating effort among the chosen tasks, particularly such that further perfect monitoring can be avoided. This is equivalent to decreasing the amount of expected regret among the actions that are chosen.

2.4 Illustrations and Predictions

Using this framework, it is possible to illustrate the conditions under which imperfect monitoring is predicted to lead to improved public services. We conduct a small simulation loosely fitted to the conditions in our field study. In particular, we assume that 28% of reports deviate from the zone modal value and assume also that the number of reports per task is a random variable in each period with 10 percent of reports active from among 50 recruited reporters. We assume that the manager has the budget to implement 20 tasks out of a total of 100 possible tasks. We assume that the manager has beliefs for each period of effort $p(\theta_j)$ drawn from a random uniform distribution [0.2,0.8].

Under these parameters, Figure 1 shows that imperfect monitoring can avoid regret and improve the number of tasks successfully completed, as long as the cost of imperfect monitoring (e.g., data collection, processing, planning) is less than 25 percent of the total implementation budget each period. While imperfect monitoring does not help make all decisions more certain, it has the potential to make enough decisions more certain given that

effort is constrained by budget.

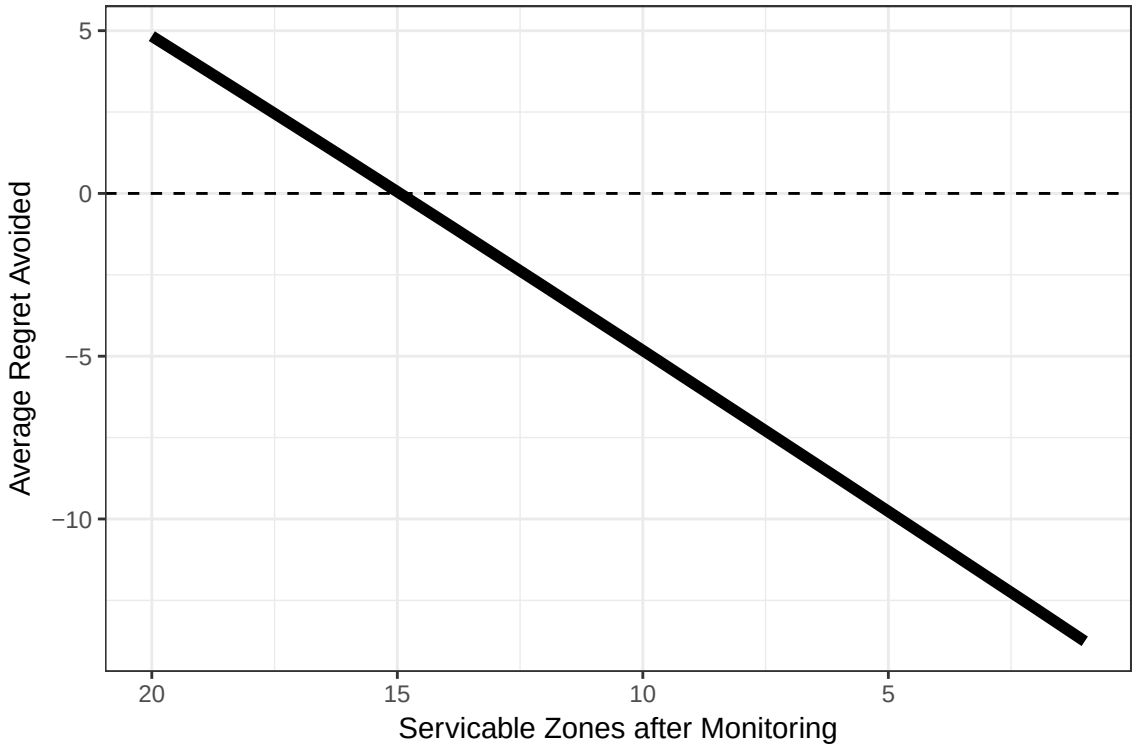


Figure 1: Regret avoided by the number of zones serviceable after monitoring.

This result highlights that even noisy and inconsistent data has value, if the authority in question has an ability to process and respond to the incoming data in ways that identify at least several problem areas that need services with a high degree of certainty. Of course, it may be particularly challenging to arrive at this kind of conclusion when the volume of information is very large (Hiltz and Plotnick, 2013). If the cost of processing bottom-up information larger than the reduction in regret, it does not make sense to use imperfect monitoring in the management of public services.

There are a number of assumptions underlying this framework that if relaxed reveal the potential for other functions of citizen monitoring. Bottom-up, citizen monitoring is a form of public pressure and reveals how much the public is tracking the performance of the government. This monitoring might cause the government unit receiving reports to work more efficiently (lowering the cost of completing tasks) or even to allocate more effort to tasks within the larger context of budgeting decisions.

2.5 Hypotheses

In light of this backdrop, we test whether the large-scale solicitation of feedback about the provision of public services from citizens will enable the more effective delivery of solid waste services in Kampala. We hypothesized that citizen monitoring could play an important role in improving waste services because it both provides information that is hard for our government partner to collect at a large scale and because citizen monitoring itself reveals the spatial location of need for the delivery or oversight of public services. Our field experiment thus tests:

H1. Zones assigned to citizen monitoring will experience a larger decrease in solid waste accumulation in the piles measured at than zones assigned to control.

Pre-registered measures (from photographs and field measurements)

- Area of total waste accumulation (primary outcome)
- Area of unmanaged waste accumulation
- Amount of burning
- Amount of non-organic waste

3 Research Design

3.1 Solid Waste Management in Kampala

Kampala city is currently undergoing a transformation as regards the provision of solid waste services. Previously, the collection and disposal of solid waste was the sole responsibility of the KCCA, which managed every aspect of solid waste collection, transportation, and disposal. The KCCA also bore the entire cost of providing solid waste services. The only role residents played was delivering their solid waste to collection locations.

Over the last few years, the KCCA adopted a Public-Private Partnership (PPP) approach to providing solid waste services. Under this approach, the KCCA contracts out the management of solid waste services to private concessionaires. These concessionaires are private companies that are given the responsibility to collect, transport, and dispose of

solid waste from particular areas of the city. In general, private companies contracted to remove solid waste provide services of lower quality to groups of people that are not able to share monitoring information about their performance with governments (Oteng-Ababio, 2010; Katusiimeh, Mol and Burger, 2012). This places city managers in a challenging position, especially given information asymmetries, pressures toward corruption, and wealth disparities across communities. Gaining information on where services are and are not being delivered can be used to allocate oversight and clean-up efforts.

In Kampala, the most noticeable change in the move to a PPP from the perspective of city residents is the new expectation that they will bear many of the costs of solid waste services. Unlike in the previous arrangement where the KCCA bore the full cost of solid waste collection, under the PPP, the private concessionaires are allowed to charge the residents a specified amount of money in return for collecting their solid waste on a door-to-door basis. At the same time, they are still contractually required to provide common collection points available to all residents regardless of ability to pay. The incentive to maximize revenue from citizens through door-to-door collection is at odds with requirements to make collection widely accessible, so contractors have mostly failed to establish common collection points.

The deterioration of solid waste services under this model has led to a spike in interest about waste management. Whereas previously some residents may have paid little attention to the quality of waste management services, the mere fact that they have to pay for the services is leading many of them to begin demanding better quality services. Yet, to this point, the KCCA has not had a way to collect data on such demands and observations about where contractors are not fulfilling their contractual obligations or even makings pickups among those households who are willing to pay for door-to-door collection. Under these conditions, the KCCA needs information about how to allocate oversight and supplementary clean-up efforts.

3.2 The Platform

The rapid proliferation of mobile phones in Kampala offer an opportunity to engage a much broader range of citizens in timely ways than has been possible previously. The latest statistics in Kampala indicate that more than 90% of adults own a mobile phone (of Statistics,

N.d.), creating the potential for significant interaction between the KCCA and citizens in ways that solve information problems related to the allocation of oversight and supplementary clean-ups.

Indeed, the KCCA faces similar problems of monitoring and accountability for solid waste management as many other parts of the world (Bhuiyan, 2010; Okot-Okumu and Nyenje, 2011). With Kampala growing rapidly like many developing cities (Vermeiren et al., 2012), the need to improve the quality and scale of services is pressing. A majority of solid waste in Kampala is disposed of in informal dumps or openly burned in streets and alleys. A large majority of residents are personally concerned and dissatisfied with solid waste services (see Buntaine, Nielson and Skaggs, 2017).

Beginning in 2014, our research team approached the KCCA to investigate whether they would be interested in adopting and testing a platform that would enable them to collect information from citizens about the quality of waste collection services in real-time and at the scale of neighborhoods. The platform would be co-developed over time and be based on toll-free SMS-messaging from residents in randomly-assigned neighborhoods, who would be invited to sign up as reporters. Because we recruited these citizen-reporters in the field, all of the reports can be tagged to individual locations throughout Kampala, called "zones" or "LC1", which are the lowest-level administrative unit in both the city and throughout Uganda.

In phases from November 2015 to August 2017, we prompted citizens to send reports about various aspects of solid waste management to a single, toll-free SMS shortcode established for the project. To process citizen reports, we employ a customized application of *SMSOne* procured by the KCCA. This platform offers a tested and convenient way to manage messages from mobile phones and is currently being expanded by the KCCA to manage all types of incoming communication from citizens across their technical directorates. The prompts involved questions co-designed by our research team and the KCCA waste management unit about various aspects of waste management along with pre-defined response categories for most prompts. For example, we used the following prompt at various points throughout the study and implementation period:

When did the rubbish truck last collect your rubbish? A) never B) more than two weeks ago C) last week D) this week

Our research team processed the raw responses into zone-wise summaries before passing the resulting information to the KCCA waste management unit in a spreadsheet format. Our research team cleaned the responses that did not conform to the structured response categories when possible prior to create the zone-wise summaries (e.g., response of "never" instead of "A" for the prompt above). The KCCA waste management unit indicated that they preferred to receive the data in this way. In addition to the distribution of responses by zone, our research team appended any messages that contained additional useful information to the zone-wise spreadsheet reports (e.g., the truck was seen but was full and did not collect anything).

After receiving the weekly spreadsheet processed at the zone level, the KCCA was fully responsible for any further processing the zone-wise data and taking any action that they deemed necessary. For a time, this involved creating weekly action plans at the zone-level to address problems identified by reporters, though ultimately this practice was suspended when the number of incoming reports and zones increased. It is important to note that both the prompts and the way that the responses were aggregated and delivered to the waste management unit were discussed and agreed upon in advance. The information and format that our research team delivered was the information that the waste management unit requested. Over the entire study period, the reporters that we recruited in the field sent 24,720 verified and on-topic reports, 17,520 of which were sent during the study period considered here (the earlier phases of the project focused on motivating reporters to send reports, see (Buntaine, Nielson and Skaggs, 2017)).

3.3 Sample and Random Assignment

We randomly selected 200 zones (out of 755) in Kampala to form our experimental sample. We randomly selected an additional 50 zones to use as replacements for zones that were inaccessible to our enumerators, demolished at the time of enumeration, or for which at least two problematic waste piles could not be identified by residents of the zone at baseline. We

assigned half of the experimental sample of zones to the citizen monitoring treatment using complete randomization, as indicated in Figure 2.

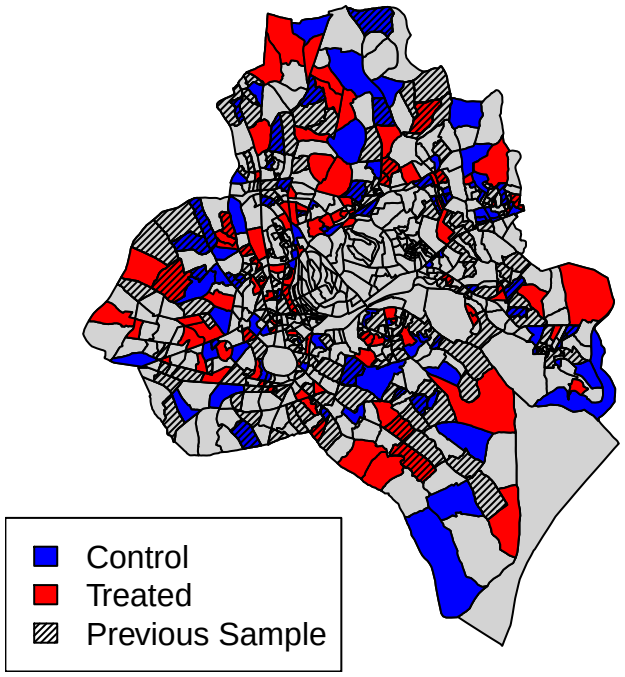


Figure 2: Experimental sample, including continuing reporting from previous phases.

We intended to select a sample for this field experiment that included entirely new zones without any previous reporting. Due to an indexing error, we selected a sample that overlapped with the Phase 1 and Phase 2 samples. This error was not caught until after baseline data had been collected. The resulting treatment still adds 50 new reporters to each of these zones, potentially boosting the number of reports considerably. Additionally, our baseline measure takes into account any treatment effects that emerged as a function of citizen monitoring in earlier phases of the project.

Our partners at the KCCA were blinded as to which zones with citizen reporting were being measured, since our design tests their ability to provide better oversight on the basis of citizen monitoring. The KCCA might re-direct attention to zones assigned to treatment apart from the information contained in reports if they were not blinded to treatment assignment. Thus, we continued to collect and pass along reports from hundreds of zones in previous phases.

3.4 Treatment

The treatment is the delivery of information about zone-level reports about waste management to the KCCA. Each week, a prompt was sent out from among a list of questions that the KCCA waste management unit identified as important for management. Our research team then compiled the responses by zone and delivered a spreadsheet containing that information to staff at the waste management unit, as requested. The processing tasks by our research team were done in expectation of an enhancement to the platform that would automate basic data summaries later. While we observed several of the plans that the waste management unit made with these data, our research team was not involved in planning or delivering any responses to the data.

3.5 Compliance with treatment

Overall, we observed expected rates of on-topic and usable reports from citizens, averaging around a 10 percent response rate during the reporting period. This rate matches what was observed in previous phases of this project that investigated how citizens could be motivated to provide monitoring.

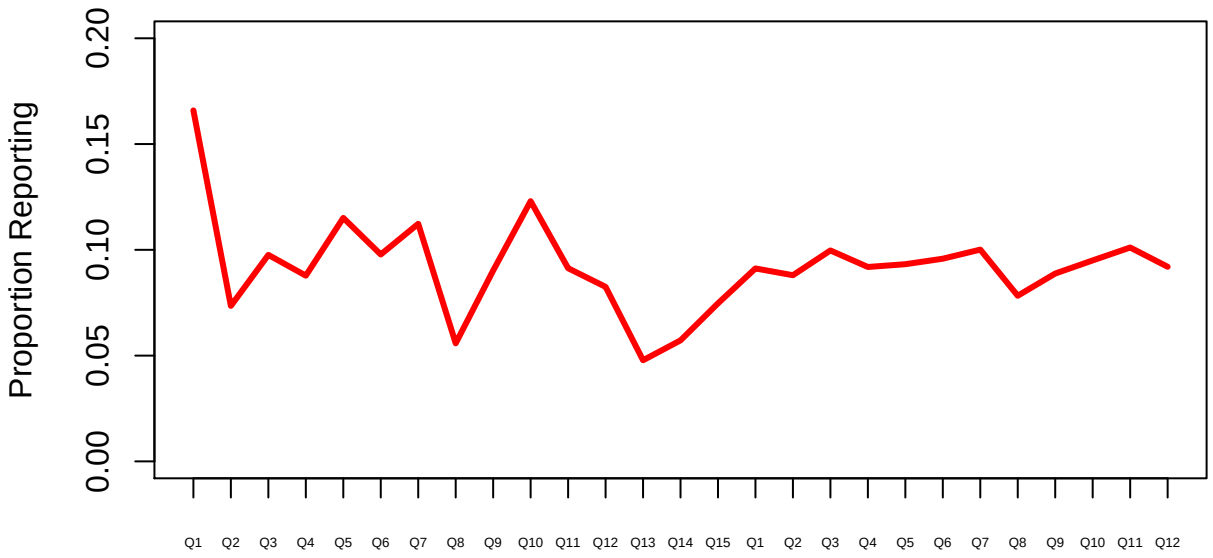


Figure 3: Experimental sample, including continuing reporting from previous phases.

We choose our sample size to ensure that almost all zones would be covered by reports

each week. Given the response rates came in about as expected, for each of the 100 treated zones, there would be an average of 3-7 reports per week, from among 50 recruited reporters. In total, the KCCA received 17,520 verified and usable reports. At one point our counterparts at the KCCA asked that we decrease the frequency of data deliveries, because they were overwhelmed by the volume and speed of data needing to be processed.

The internal consistency of the reports sent in by citizens within zones varied, but was generally low. Figure 4 displays the consistency of responses within zones on a standardized measure responses to prompts that would indicate poor service quality to the KCCA.[2] Zones reporting poor service quality on average are highlighted in red. Darker fills indicate a greater proportion of responses that deviate from the modal response within a zone. On average, 28 percent of citizen monitors recorded survey responses that deviated from the zone-level modal response – i.e., monitors indicating that service quality was poor when a majority of respondents in the zone indicated that service quality was acceptable, or vice versa.

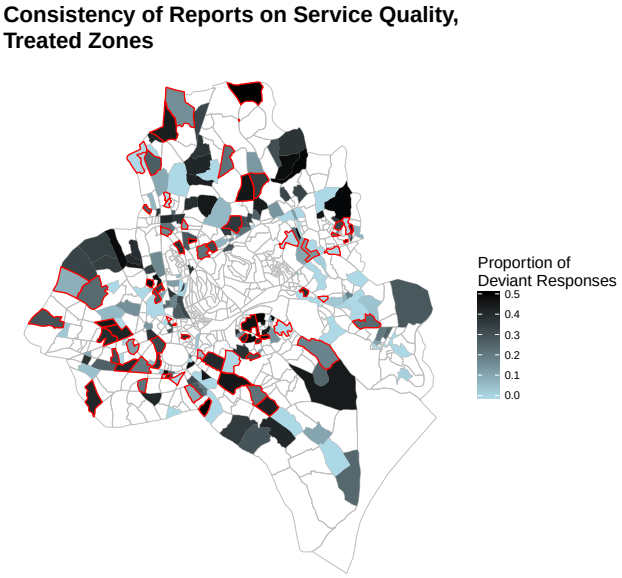


Figure 4: Consistency of reports from treated zones in the experimental sample. *Along a standardized measure of poor service provision, zones in red, indicated that KCCA service provision was poor on average. Zones with darker fills represent zones that inconsistently reported the quality of KCCA services, relative to the zone-level modal response.*

[2]The standardized measure of poor service provision combined citizen monitor responses on the following indicators: the frequency and accessibility of service provision, reported waste collector treatment of citizens outlined, and the amount of waste burning or litter.

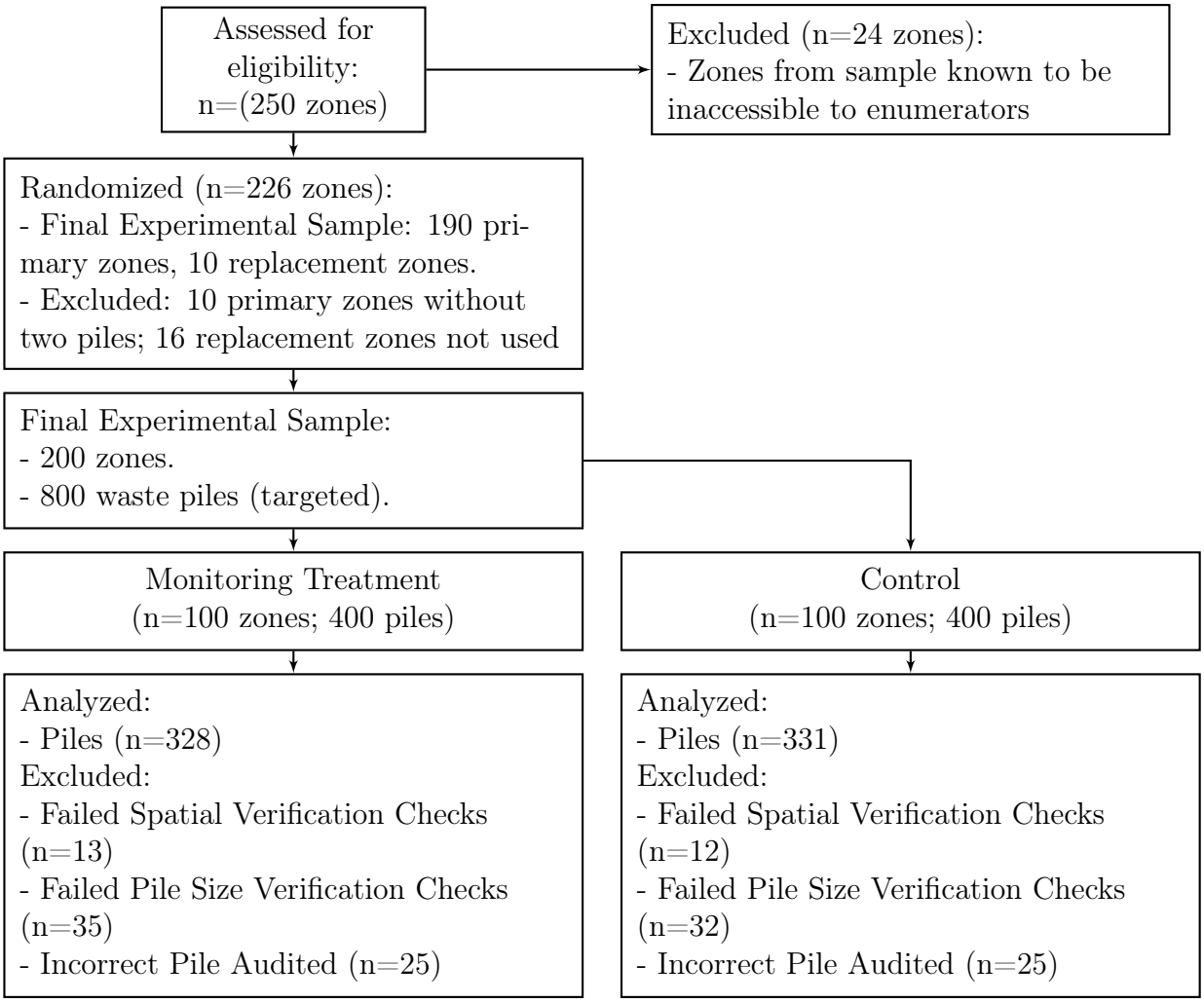


Figure 5: CONSORT diagram tracking study design.

3.6 Outcomes

The core of our measurement strategy involved a field-based audit of waste piles, since the presence of informal dumping is a direct outcome of the unavailable or inaccessible waste collection services. We went to each zone in the experimental sample and asked residents to show us up to four informal waste piles that were of greatest concern to them. We measured the spatial extent of these waste piles, recorded their locations by GPS, and mapped the easiest way to return to them for re-measurement. The core outcome of our field experiment is whether waste piles in treatment zones change more positively than those in control zones.

Figure 6: Small pile

Figure 7: Medium pile

Figure 8: Large pile

Baseline measurement. We measure the size, take a photograph, and record the exact GPS coordinates of four waste piles that residents identify as a concern for each zone in all 200 experimental zones, according to the measurement protocol in Annex 1. We measure the length and width of each waste pile, evidence of burning, and the proximity of the waste heap to residences, active businesses, and public roads. Figure 6, figure 7, and figure 8 are representative examples of small, medium, and large waste piles, respectively.

Post-treatment measurement. After the monitoring platform operated for five months, we re-measured and re-photograph the waste sites identified at baseline. After the monitoring platform operated for nine months, we again re-measured and re-photograph the waste sites identified at baseline.

3.7 Estimation

We test our hypotheses about pile size by regressing the size of piles on treatment status in the current phase, treatment status in previous phases, the pre-treatment measures of

the pile size collected at baseline, and a number of zone-level covariates. We compute standard errors from the sampling distribution of the relevant parameter estimate, derived from randomization inference assuming the sharp null hypothesis and estimating the parameters under 10,000 permutations of the complete randomization procedure that we used to assign treatment. The core estimating equation for measures with both baseline and endline values is Equation 9. This estimation deviates from our pre-registered strategy in that it takes the pile, rather than the zone as the unit of analysis

$$y_{ij,t=b+n} = \alpha + \tau M_j^+ + \gamma y_{ij,t=b} + \boldsymbol{\beta}\mathbf{X_j} + \nu_h + \epsilon_h \tag{9}$$

Where y is the relevant size measure for pile i in zone j at time b baseline plus some follow-up period n, τ is the treatment effect of interest, M_j^+ is a binary indicator of treatment assigned at the zone-level j, γ is the parameter estimating the relationship of baseline size measure $y_{ij,t=b}$ to the follow-up outcome measure, $\boldsymbol{\beta}\mathbf{X_j}$ is the estimated adjustment for pre-treatment, zone-level covariates including the treatment status of zones during previous phases, ν_h is a fixed effect for division, and ϵ_h is an error term clustered at district. We drop piles that were not measured at baseline. We use the following zone-level covariates in each of the estimating equations, as pre-registered: zone area, density of improved roads, and luminosity. Further notes about data cleaning and organization are in the SI.

4 Results

4.1 Pile Sizes

We find no evidence that citizen monitoring reduces the number of existing waste piles in sampled zones (Figure 9). Even when changing the definition of a cleaned pile to include sites for which all waste was collected into a single, transportable container, we find no difference in the proportion of waste piles cleaned up among treatment and control groups that are inconsistent with the null hypothesis (Figure 10). Speaking directly to our primary hypothesis (H1), we find no evidence indicating that treated zones experienced greater reductions in waste accumulation than did control zones (Table 1). Table 1 shows that under

Table 1: Estimated Effects of Treatment on the Clean-up and Size of Waste Piles

	DV: Pile Cleaned (0/1) or Change in Waste Pile Size (m^2)			
	M1 Cleaned	M1 Change	M2 Cleaned	M2 Change
	(1)	(2)	(3)	(4)
Treatment	−0.014	3.565	−0.005	5.877
	(0.028)	(5.397)	(0.036)	(11.049)
Baseline Pile Area	0.015	6.419	−0.012	−0.117
	(0.030)	(5.691)	(0.037)	(11.575)
P1/P2 Monitoring	0.0003	−0.768***	0.0004	−0.802***
	(0.0002)	(0.047)	(0.0003)	(0.095)
Covariates	Yes	Yes	Yes	Yes
Observations	623	623	623	623
R^2	0.035	0.319	0.021	0.126
Adjusted R^2	0.017	0.307	0.003	0.110
Residual Std. Error	0.328	62.123	0.408	126.358
F Statistic	1.994**	26.019***	1.178	8.004***

Note: two-tailed tests *p<0.1; **p<0.05; ***p<0.01

all specifications, treatment assignment is signed counter to expectations of more cleaned piles and a reduction in pile size.

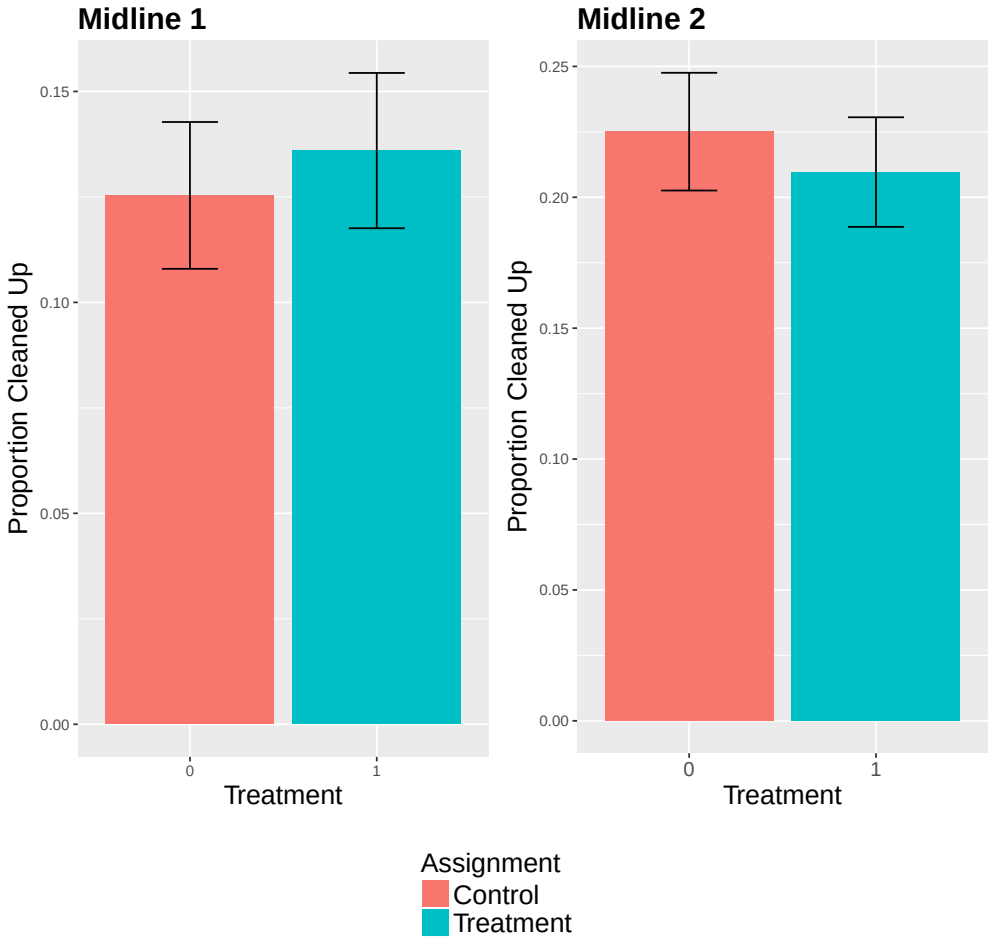


Figure 9: Proportion of waste piles cleaned up. *A waste pile is considered cleaned up when an enumerator is able to locate the exact spot of the previous pile using the project maps and no waste is present at the site.*

We observe similar results across various measures of pile size and characteristics. Citizen monitoring does not reduce the total area of unmanaged waste – estimated both as the level of waste storage and the level of organization per pile – in treated zones (Table 2). This effect is robust to various specifications of unmanaged waste.[3]

[3]Responses from both audits were used to estimate sizes organized, burned, or contained. Multiple scales were employed to ensure that our findings were robust against different specification of the outcome measure. For example, when a pile exhibited evidence of burning, enumerators reported whether more or less than 50 percent of the pile appeared to be burnt. Responses indicating that more than 50 percent of the pile had been burned were assigned values of 0.55 for the first specification and 0.67 for the second specification. Responses indicating that less than 50 percent of the pile had been burned were assigned values of 0.45 for

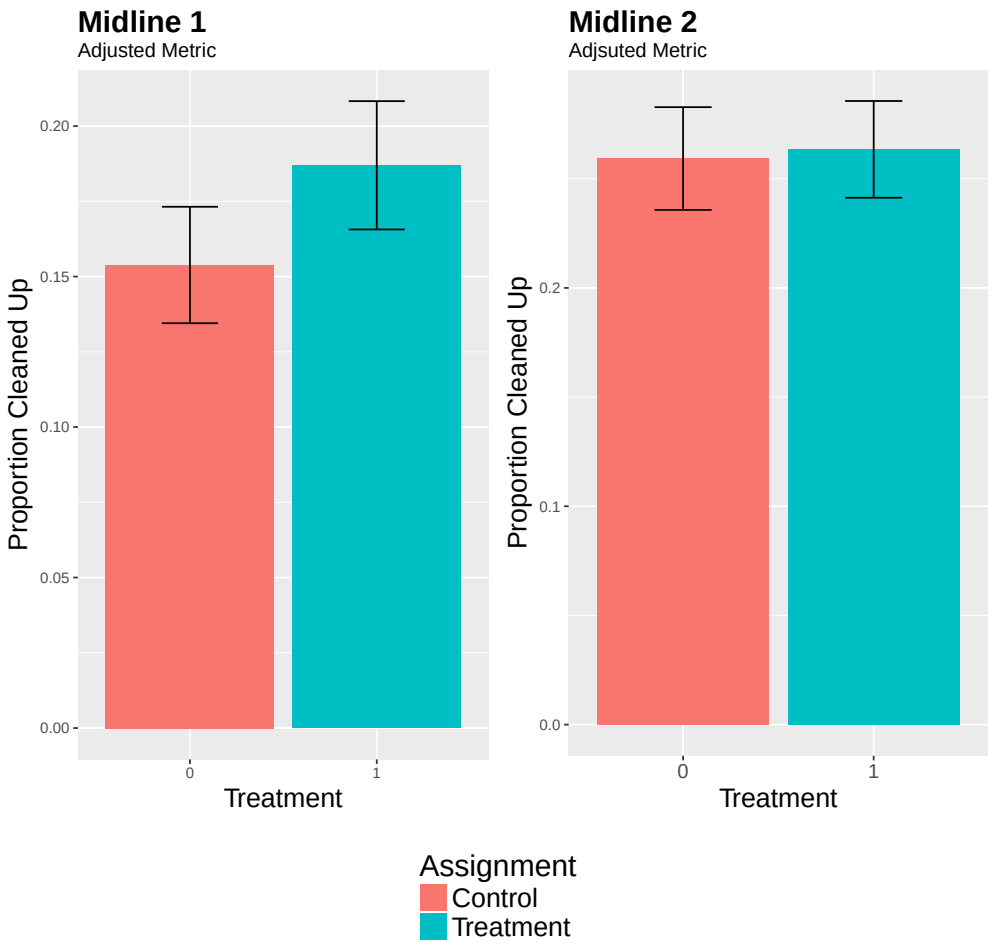


Figure 10: Proportion of waste piles cleaned up, including well-organized piles. *Cleaned piles under the adjusted specification include all piles for which waste was collected into a single pile and stored in transportable containers.*

Table 2: Estimated Effects of Treatment on Waste Pile Characteristics

	DV: Unmanagaed Waste Pile Area, m^2	
	Storage	Organization
	(1)	(2)
Treatment	1.854	3.399
	(11.553)	(11.679)
Baseline Pile Area	−5.715	−5.486
	(12.366)	(12.429)
P1/P2 Monitoring	−0.040	−0.076
	(0.102)	(0.102)
Covariates	Yes	Yes
Observations	623	623
R^2	0.020	0.021
Adjusted R^2	0.003	0.003
Residual Std. Error	135.000	135.680
F Statistic	1.142	1.193
Note: two-tailed tests		*p<0.1; **p<0.05; ***p<0.01

4.2 Pile Characteristics

While we see no pronounced effect of treatment on the reduction of waste pile size, we find marginal evidence that citizen monitoring temporarily accelerates specific components of the waste clean-up process. This effect is particularly pronounced in the data from the first post-treatment round of pile audits.

Enumerators from the first audit recorded an overall reduction in the amount of non-organic waste found per pile in treated zones. We see that the proportion of treated piles with greater than ten pieces of organic waste is significantly lower than the proportion of similar control piles (left pane, Figure 11; $te = -0.11$, p=0.001). However, the amount of non-organic waste in treated piles increases by the second post-treatment audit. Treatment and control piles have statistically indistinguishable proportions of piles containing more or less-than ten pieces of non-organic waste at that stage (right pane, Figure 11; $te = -0.006$, p=0.46)

Data from the first post-treatment audit on waste burning indicates a similar temporary improvement. In the first midline audit, treated zones contained a larger proportion of piles with no evidence of burning (left pane, Figure 12; te=0.07, p=0.04. Enumerators also found a smaller proportion of piles with evidence of widespread burning in treated zones than in control zones ($te = -0.07$, p=0.07). However, both treatment effects attenuate for the second post-treatment audit (p=0.17 and 0.11, respectively).

Waste containment and pile organization saw a gradual but enduring improvement throughout both midline audits. While the difference in proportion of fully-contained piles—where all rubbish is stored in transportable sacks or containers—among treated and control zones was not significantly different from zero following midline one, treated zones in midline two contained a greater proportion of fully-contained piles than did control zones(Figure 13; te=0.03, p=0.058). Reports on general pile descriptions from the second midline audit corroborate this effect. As in midline one, treated zones in midline two contain a larger proportion of piles with waste stored for transport than do non-treated zones (right pane, Figure 11; $te = -0.03$, p=0.001).

the first specification and 0.33 second specification. These values were then multiplied by the estimated pile area, generating two sets of two measurements of the estimated area of burning per pile.

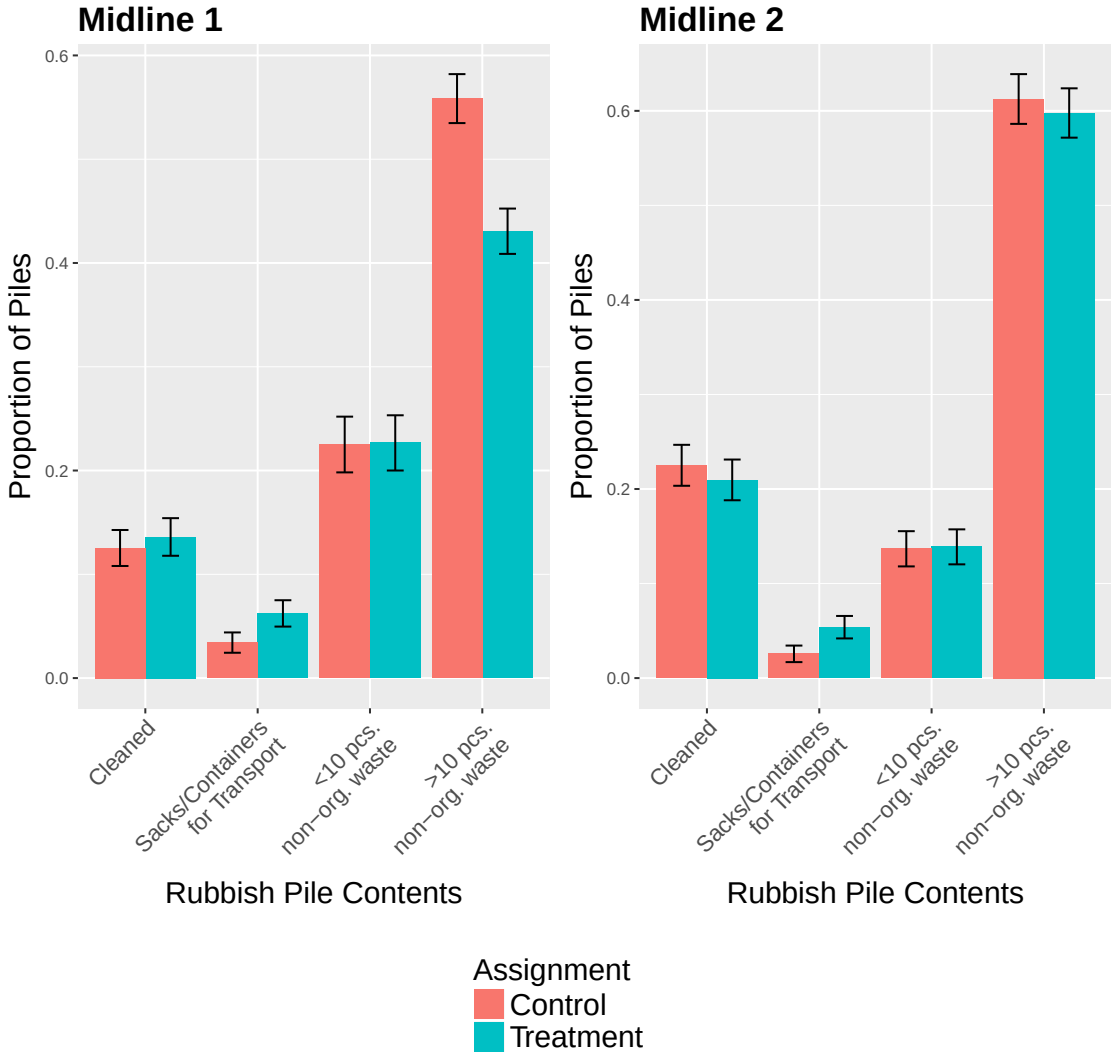


Figure 11: Descriptions of waste piles. *Seventy-one enumerators from the first post-treatment audit reported that a waste pile existed at a given location but did not report on the pile's general description. These missing responses are excluded from the above figure.*

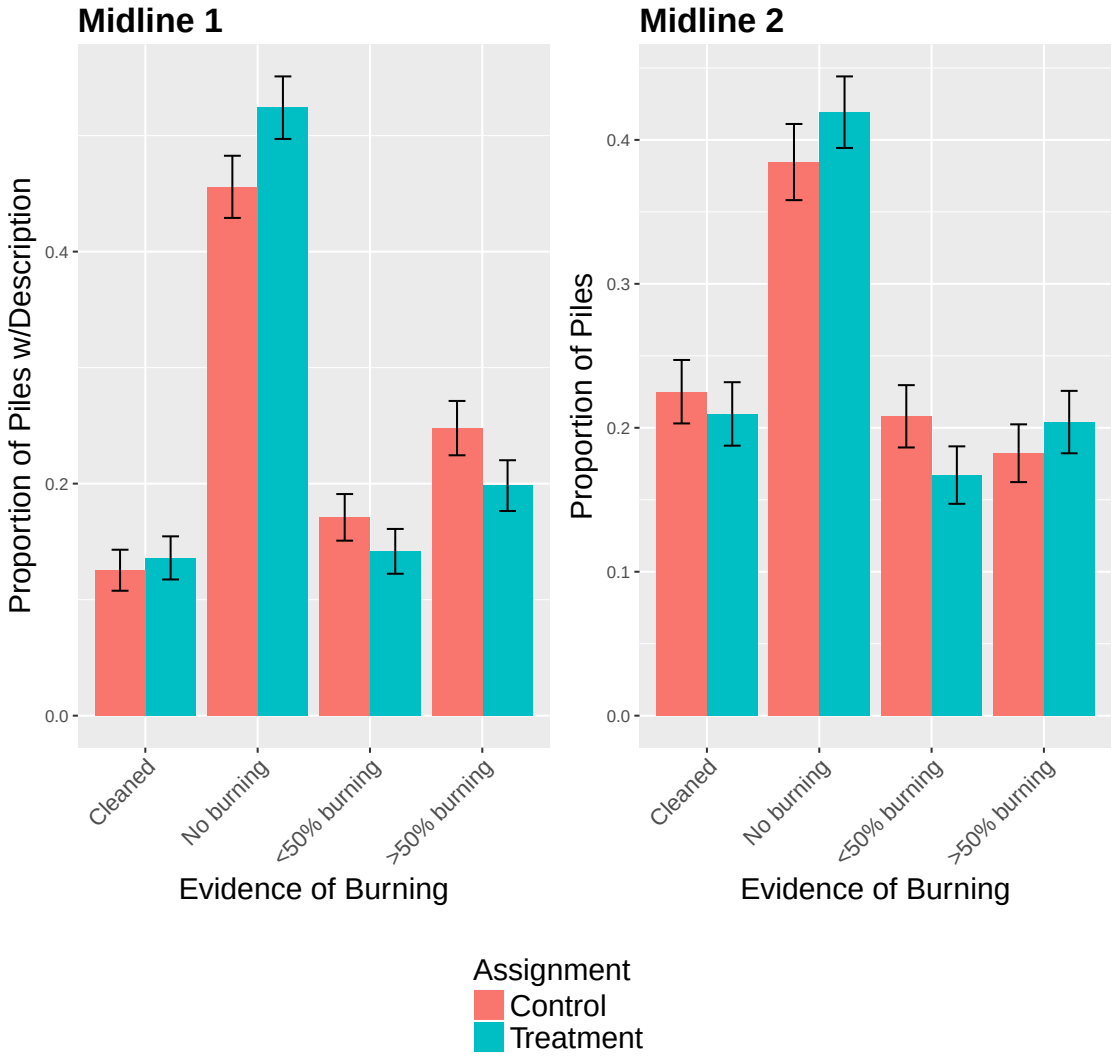


Figure 12: Proportion of piles with evidence of burning.

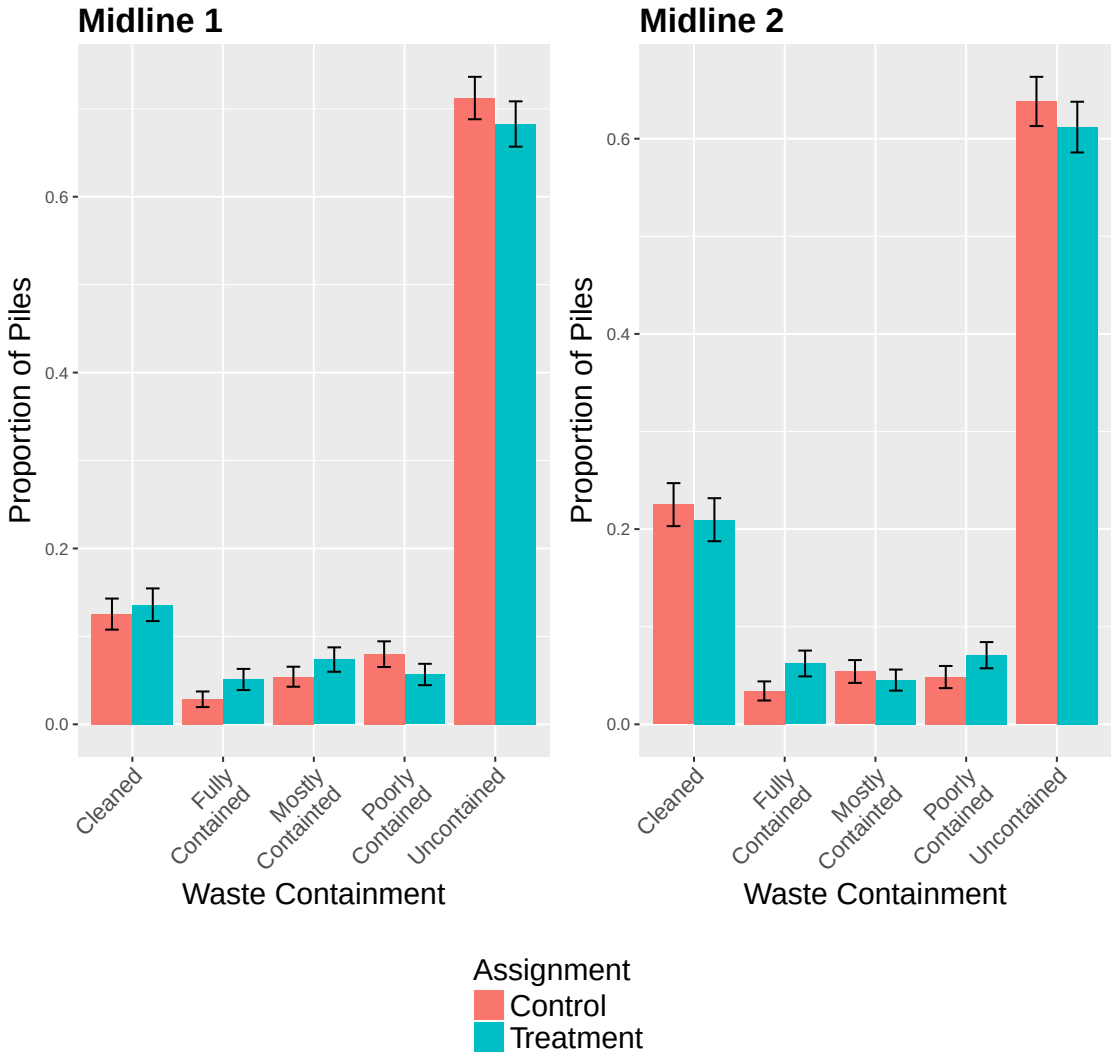


Figure 13: Proportion of waste piles with different levels of containment.

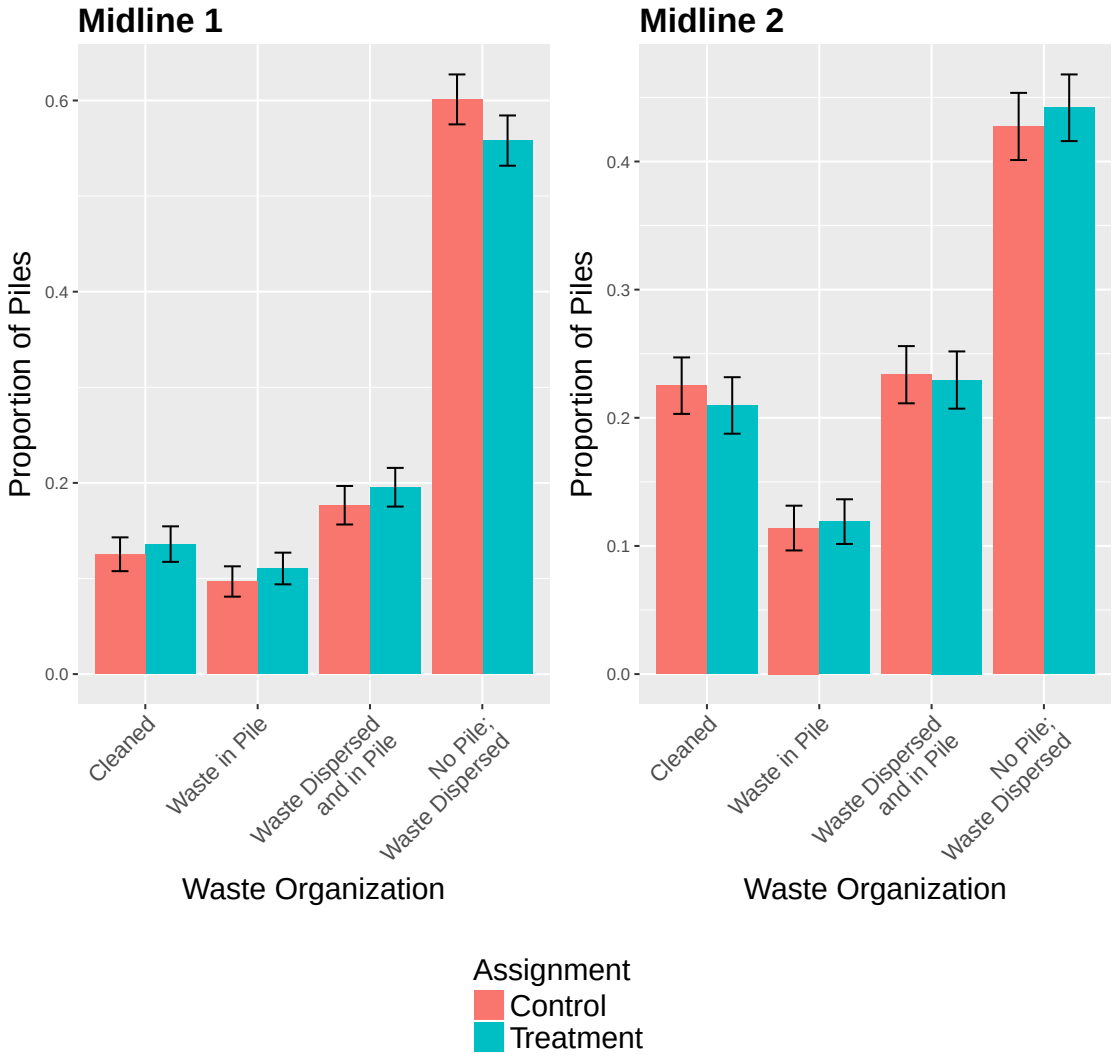


Figure 14: Proportion of waste piles with different levels of dispersion.

In summary, our results indicate that treatment assignment neither reduces the area of waste piles nor increases the frequency of waste pile clean-up. Some evidence from from the first post-treatment audit suggests that citizen monitoring improves specific aspects of waste service provision, though these advances generally do not persist over time. Piles in treated zones contain fewer pieces of non-organic waste and exhibit less evidence of burning following the first post-treatment audit; however, these effects are absent at the second post-treatment measurement. More durable – albeit gradual – improvements occurred with regard to pile organization. In both audits, treated zones contained more piles where waste was fully contained and readied for transport, though this is still a small proportion of the piles and the estimated effect is substantively small.

5 Mechanisms and Heterogeneous Treatment Effects

5.1 Political Targeting

Several recent papers show that politicians often use public goods and services as a way to reward supporters in elections (Jablonski, 2014; Drazen and Eslava, 2010; Baldwin, 2013; Briggs, 2012). Using public goods in this way is often an effective strategy to build political support. In the setting of our study, the National Resistance Movement (NRM) is the ruling party nationally, but faces generally low levels of political support within the city of Kampala. In 2011, aiming to reverse the trend of entrenched opposition within the capital city, the municipal government was nationalized and responsibility for services transfered from the elected city council to the KCCA. On might expect under these circumstances for the KCCA to use there discretion is targeted to reward areas of the city that vote for NRM candidates, as compared to opposition or independent candidates.

We test for this possibility by examining both the baseline amount of waste accumulation and whether the reporting treatment was more effective where the winning candidate in 2016 division elections for the parish constituency was a member of the NRM ruling party. As displayed in Table 3, we fail to find evidence that either the status or change in waste pile sizes is conditional on the party of the Division councillor, or that the treatment effect is

conditional on this party affiliation. These findings largely rules out the possibility that political targeting is driving the allocation of effort.

5.2 Reporting Rates and Message Content

Conventional wisdom holds that the most vocal stakeholders often receive the most attention from public service providers. Under public pressure, KCCA officials might disproportionately respond to zones that report more frequently. Alternatively, the KCCA might respond to zones that report attention is needed consistently.[4] Indeed, the theory that we present indicates that managers are averse to allocating effort towards zones that provide inconsistent information on the quality of service provision. KCCA officials might similarly scale up service provision to zones expressing extreme dissatisfaction with KCCA services or reporting severe waste problems.

Leveraging the content of reports collected prior to the first post-treatment audit, we test these hypotheses. We take a zone-level count of responses to examine the "squeaky wheel" hypothesis; and use the content of reports to create measures of zone-level dissatisfaction and problem severity. On all measures, we fail to detect a consistent and substantive effect on waste pile size at the conventional level of statistical significance (Tables 4, 5, 6). Zone-level response rates, dissatisfaction with the KCCA, and information indicating the initial quality of KCCA service provision have no effect on waste pile size.[5] Thus, these results dismiss the possibility that the weak effect of treatment on service provision is a consequence of a heterogeneous treatment effect based on the response rates or response content. Finally, we find no evidence that the consistency of zone-level responses on service quality reduces the

[4]For maps visualizing the consistent reporting of poor service provision over the span of the two midline audits, see figure 15. Figure 15 plots the proportion of zone-level responses that deviate from the zone-level modal response on a standardized indicator of poor service quality. Higher proportions of deviant responses indicate that citizen-monitors provided inconsistent feedback to KCCA officials on the quality of KCCA services.

[5]We operationalize "severity of waste management problems" in Table 5 using reports from an item asking citizen-monitors to report if a rubbish-collection truck visited their neighborhood. Possible responses include: yes, no, don't know. The latter two responses were coded as indicative of severe waste management problems. Following the logic outlined above, one would expect the KCCA to deploy trucks disproportionately to zones reporting that they had not received pick-up services recently. We additionally operationalize waste management problem severity using citizen-monitor reports commenting on rubbish burning, litter and illegal piles, rubbish spilling from KCCA trucks, and mistreatment by KCCA waste collectors. [Please refer to SI for a complete set of tables reporting these effects.]

size of waste piles in treated zones (Table 7).

6 Participant Observations and Staff Interviews

Our team was embedded in the KCCA waste management unit for close to nine months. During this time, we had numerous interactions with staff members at different levels from managers to staff at the frontline of executing day-to-day waste management activities. We also reviewed documents (including policy documents, internal action plans, and even court documents), participated in meetings and observed interactions among some stakeholders in the provision of waste management. We later conducted in-depth interviews with all individuals who interacted with the ICT platform built to facilitate communication between the KCCA team and citizen reporters. This section summarizes what we learned during these interactions and interviews.

6.1 Allocating effort to waste services

The KCCA allocates waste management services on a zone-by-zone basis. The allocation of services is based on KCCA's assessment of the waste management needs in that particular zone. For instance, one zone with large number of illegal dumps will require garbage clearing services, while another zone with inconsistent truck visits will require implementing a consistent schedule. The major challenge for the KCCA is assessing the specific needs of specific locations. There are three sources of information that the KCCA has used to assess the kinds of actions needed in each zone.

First, the KCCA uses administrative records that are specific to the Kampala for waste management. For instance, the KCCA can use census reports produced by the Uganda Bureau of Statistics, which contain information on the populations of different areas in Kampala and also population projections. The KCCA uses these records to predict the amount of waste which is likely to be produced from different areas. Relatedly, the KCCA has administrative records of solid waste deposited at the KCCA dumpsite at Kiteezi. Each waste delivery is recorded in terms of date delivered, the weight, and area(s) from which it was collected. The KCCA has such records dating back over a decade. The trends in the

waste deliveries from any particular area over time are used to assess the current and future needs of any area, though this is imperfect data since waste collection effort is not uniform across the city.

Second, the KCCA leverages the knowledge of staff who operate within different areas of the city. There are a number of KCCA staff who occasionally visit different parts of Kampala to report on various aspects of sanitation and waste collection. For instance, according to the former Supervisor of the Solid Waste Unit, KCCA employs up to 200 casual workers to report on waste and sanitation issues (e.g., illegal dump sites, open sewers, drainage pipe leakages). Among the core tasks of these workers is collecting information on different waste management issues in their areas of operation and reporting them up the chain of management.

These workers enable KCCA to keep up-to-date with the changing waste management needs of different areas in Kampala, though data from them has never been fully systematized. They also provide actionable information to KCCA staff. In the words of the Supervisor of the waste management unit: “I have a problem of illegal dumping. And I have my scouts. When they find a suspect, they use WhatsApp to send a message, I send a car to pick the suspect and take them to court."

The third source of information comes from Kampala citizens. The KCCA solid waste unit has a number of ways through which citizens can contact them. These include Twitter, WhatsApp, a toll-free line, office walk-ins, and community outreach events. All of these channels provide information which can help to identify the specific waste management needs of different areas.

6.2 The potential of citizen reporting

In light of this context, citizen reporting was initially deemed essential to identifying and keeping up-to-date with the changing waste management needs of different areas throughout Kampala. Although citizens provide KCCA with information which can help identify areas requiring attention, a systematic means of collecting, processing and utilizing this information had been lacking. It was the systematic approach to integrating information which citizen reporting through an ICT platform offered.

Citizen reporting was especially helpful for collecting information from areas which were beyond the reach of KCCA officials. Given their numbers, it was impossible for KCCA officials to be present in all areas of Kampala at once. As one Waste Management Officer reported: “My area of supervision contains 77 parishes and about 200 zones. It is impossible for me to be in all those places at the same time. The citizen monitors enable me to keep tabs in those areas by keeping me up-to-date with what is going on.” Thus, the great potential of citizen monitoring was perceived to be its expanded coverage relative to alternatives.

6.3 Using citizen reporting to improve services

Citizen reporting activated as part of the experimental treatment was used to gather information which formed the basis for evidence-based action. The citizen monitors were sent weekly prompts on various aspects of service provision in their areas. Their responses were compiled, processed and sent over to the KCCA waste management unit. The responsible officers then studied the reports and prepared Action Plans which were proposed interventions aimed at solving the problems identified through the information from citizens. The interventions were then implemented. In the words of a KCCA Solid Waste Officer, "We get the info, go back to those areas where people have complaints and rectify the problems." Serious effort was initially expended to respond to problems identified.

Each prompt to reports had response options which were pointers to particular problems. To illustrate this, consider question “Does a rubbish truck come into your neighborhood? A) No B) Yes C) I don’t know”. Here, the response options “No” and “I don’t know” indicate a problem if they emanate from an area which has an assigned truck which is supposed to visit periodically. They indicate that either the truck is not reaching the zone or that the residents are unaware of its visits. All the responses from a particular zone were aggregated and decisions made on the most prevalent problems reported by citizens.

Citizen monitoring initially approached at more than just an instrumental level. It was used as a means to improve citizen satisfaction with waste management services. This is because the monitoring operation involved a continuous exchange of information between KCCA and reporters. This engagement was used by KCCA to improve the experience of citizens with service delivery by ensuring that residents understood that their concerns

mattered and were being addressed. As clearly put by the Supervisor of the KCCA waste management unit: "As KCCA, one of the core values of the institution is client care. And client care cannot be actualized if the client is not satisfied. Public service delivery is directed towards the clients. If the clients send any feedback, it is incumbent upon KCCA to respond to this feedback (and if need be re-align its priorities/operations/services in line with the client feedback)."

Communicative responsiveness to citizen feedback through the platform was prioritized by the KCCA staff because the KCCA often lacked the resources to immediately act on citizen concerns. Even where resources were available to respond to concerns, KCCA's operational schedules sometimes restricted their capacity to act immediately. Responding to citizens was important to let the citizens know that KCCA is listening to them and making effort to respond to their concerns, creating the potential to build trust and foster data sharing.

Citizen monitoring helped KCCA to provide oversight to private contractors. Under the PPP arrangement, the mandate of KCCA was to monitor and regulate how private contractors operate within specific areas in Kampala. Part of this monitoring function is ensuring that the private contractors provide services in terms of their contractual agreements. The feedback from citizens was crucial on a number of occasions. For instance, according to the PPP contract, private companies are supposed to charge citizens Shs 3,000 (USD 0.83) per month for basic services. On February 2nd 2017, a citizen monitor sent a message "Why have they started charging each household 5,000 for garbage collection and yet they spend up to two months without coming to collect garbage?" This message alerted KCCA to the fact that private contractors were overcharging citizens, and prompted them to intervene to stop the practice. Similar complaints about other aspects of waste management (e.g., inconsistent visits by trucks and unprofessional conduct by employees of the private contractors) were used by KCCA to hold private contractors accountable and ensure that their service delivery is improved.

6.4 Abandoning citizen monitoring as a source of information

There are a number of reasons why citizen monitoring was abandoned over time as a solution to service problems. These include the high cost of monitoring, the amount of inconsistency in the citizen reports, and the availability of a more efficient and reliable alternative to citizen monitoring. We explore these reasons in turn.

6.4.1 Cost of monitoring

The major reason for the abandonment of the citizen monitoring was the relatively high costs involved in running it. At the initial implementation of the citizen monitoring program, the costs recruiting citizens, communicating with them and monitoring them were not borne by KCCA. Our research team covered the cost of messaging in previous phases of the project (see Buntaine, Nielson and Skaggs (2017)).

Once the program was handed over to KCCA in the current study, the costs had to be shouldered by the authority. This is when questions started to rise about the cost-effectiveness of citizen monitoring. The main point of consternation was the low response rate from the citizen monitors. On most occasions, the response rate to the SMS text messages sent ranged from 7 percent to 12 percent, with the average being about 10 percent. Given that KCCA was billed per SMS sent, it seemed to staff that 90 percent of the budget allocated for sending prompts was being wasted.

Questions about the low response rate of citizen monitors had already been raised by KCCA on several occasions during meetings with the research team. A case in point was during a presentation of the Phase I and Phase II results to the Deputy Executive Director of KCCA on July 28th 2016. The DED pointed out that the response rate seemed extremely low. This sentiment was echoed on numerous other occasions during presentation of progress results to the KCCA.

However, it was the cost implications of a low response rate which caused most concern. This partly arose from the setup of the SMS-based citizen monitoring process. Each engagement cycle involved sending a three of four SMS-text messages i.e. an introductory message, a prompt, a responsiveness message and a lottery winner message (the introductory and

lottery winner messages could be fused into one).

Given that there were 7,500 citizen monitors, and each SMS cost UGX 33 (USD 0.009), a minimum of UGX 742,500 (USD 206) was spent on each engagement cycle just to send messages. The cost of receiving each response was UGX 110 for SMS and 230 for USSD. A 10 percent response rate meant that on average 750 responses were received. On average, 80 percent responded via USSD and 20 percent via SMS leading to an average cost of UGX 172,500 (USD 48) for the responses. The combined cost of sending and receiving messages in one engagement cycle was a minimum of UGX 915,000 (USD 254).

Comparatively, the cost of facilitating KCCA scouts (the self-monitoring agents) was much cheaper on two fronts. First of all, shilling for shilling, it cost less to facilitate the KCCA scouts to complete monitoring. This is because the scouts were already KCCA employees and were carrying out sanitation-related duties within different communities in Kampala. All they required in terms of facilitation was Internet data bundles which they would use to report their observations via the messaging application WhatsApp.

The average cost of facilitating each KCCA scout with internet bundles was UGX 50,000 (USD 14) per month. The total cost for all 72 scouts per month was UGX 3,600,000 (USD 1000). Comparatively, the citizen monitors were engaged a minimum of once a week, therefore, the total cost of engaging each month was about equal.

Secondly, it was easier for managers to justify the expenses made on facilitating KCCA scouts than those spent on communicating with citizen reporters. For KCCA managers, this justification was of concern since accountability had to be made to their superiors. It was quite difficult to make a cost for a program of which 90 percent of the expenses were not producing any tangible results. This fact also came into play when decisions were made to shift from citizen monitoring to self-monitoring. This is precisely why the Supervisor who made the decision to drop citizen monitoring said: "For me, these messages are very expensive for nothing. That is why I was saying, 'Why don't we buy the scouts airtime and communicate on WhatsApp?'"

6.4.2 Data quality

One major consideration of managers when considering different monitoring mechanisms was the quality of information produced. Eventually, managers of the waste management unit came to question whether reports by citizen monitors accurately represented the realities of what was going on. Managers and staff were concerned about whether they could trust the information enough to use it as a basis for decision-making about the allocation of scarce resources.

The question of accuracy is one which perturbed KCCA managers who were tasked with acting on information sent by citizen monitors. This is because the information often had lots of inconsistency, even within zones. Indeed, the global average for report inconsistency within zones on binary items as compared to the mode was around 30 percent. Other times, the same individual would send two or three different responses to the same question. The responses were often contradictory. In some instances, reporters used the SMS platform to send abusive messages, which turned the team against the platform as a collaborative space.

Similarly, there instances where different people from the same location sent information which in totality was quite confusing. For example, some citizen monitors would report that a garbage truck visited the area within the past week. Other citizen monitors from the same area would report that the garbage truck had not appeared in more than a month. Such information would leave the KCCA staff responsible for planning interventions confused. This caused the Supervisor who made the decision to abandon citizen monitoring to state the following about the data collected through citizen monitoring: "The data is not useful because its authenticity or accuracy cannot be verified."

The managers initially employed a number of strategies in order to cope with the inconsistency coming in with the citizen information. These included (1) attempting to contact particular citizens who sent in multiple contradictory responses in order to understand which option to take; (2) following up with citizens from areas with contradictions from various citizens to get a clearer understanding of what is going on; (3) utilizing staff knowledge of those areas to interpret the information from the citizens; and (4) following up with other stakeholders in the service provision process e.g. where the information was about operations of

private contractors, the companies operating within the locations would be contacted. In cases where it was impossible to clarify or straighten out the inconsistencies in information, the information would simply be discarded. The process of managing and following up on inconsistency was costly and time-consuming from a management perspective, given the small staff size at the waste management unit and the extremely demanding mandate to deliver services to millions of residents.

The process of sifting through the inconsistencies in the information added a layer of complexity – as well as extra costs in processing the information. The time required to sift through the noise also slowed down the speed at which KCCA responded to citizen concerns – thereby having a negative impact on overall service delivery. According to KCCA staff, it took between one and two working days to sift through the noise in citizen monitoring reports.

When compared with information collected through self-monitoring, the information from citizens was far less usable. The KCCA scouts sent to specific areas were given strict reporting parameters, which minimized the noise from their reports. In case of any noise, the process of clarification was much easier, since the WhatsApp platform allowed the sending of photographs and there were fewer scouts needing follow-up. This made self-monitoring a preferable alternative to citizen monitors because the information they provided more accurate, authentic and trustworthy. As such managers could make critical decisions basing on the information because they trusted its accuracy, which was not the case with citizen-sourced data.

6.5 Interpretation

The rationale for monitoring is the acquisition of information necessary to facilitate decision making and ultimately improve service provision. The decision to invest in monitoring, opt for a given type of monitoring, or switch one form of monitoring for another relies on a number of factors, like cost, baseline beliefs about what tasks will lead to the highest payoffs, and consistency of the information.

In terms of the cost of monitoring, for any form of monitoring to make sense, the cost of acquiring information must be lower than the benefits that arise from making more informed

decisions. KCCA's decision to drop citizen monitoring clearly illustrates that whenever any form of monitoring becomes too costly not only monetarily, but also related to processing effort, public service providers will abandon it.

In terms of baseline knowledge and beliefs, and as noted in the section on no monitoring, when there is no information available through monitoring, managers are guided by their beliefs to allocate resources where they expect the highest payoffs. Information through monitoring is more useful when managers are unable to order their beliefs about what actions are needed clearly. While we do not have direct observations about the baseline beliefs and uncertainty about them for the KCCA in this study, we know that staff are regularly traveling in Kampala and have other sources of data that can reveal where waste services are most in need (e.g., projections of waste management needs based on population growth patterns or records of previous solid waste collections).

In terms of the consistency of information, citizen monitoring is often noisy. Before being able to utilize the incoming information, there has to be an extraction of the usable facts from the noise. The usefulness of the information is dependent on the level of consistency of the information. The KCCA decision to drop citizen monitoring because it was producing lots of noisy signals that limited assessments of whether it was useful or worth the scarce investment of resources.

7 Discussion

Communication technologies create new spaces for governments and citizens to come together to improve the delivery of public services and they offer the potential for governance and public management in the decades ahead. Yet, we lack solid evidence that the tools offered by emerging communication technologies can translated into the improved management of public resources. On the one hand, these tools have the direct potential to solve information problems for public agencies that deliver frontline services. After all, citizens directly experience these services or the lack of these services as part of their daily lives and have information on their experience. On the other hand, integrating high-frequency, high-volume, and hyper-local data streams into the active management of public services requires

considerable commitment and capacity on the part of public managers. It also requires the information to be relatively inexpensive compared to alternatives and to be of a quality and consistency that is useful for decision-making.

We fail to find improvements in the amount of waste accumulation in zones assigned to citizen reporting. We find some indications of promising results that later disappear, but at most the evidence is marginal that any improvements to solid waste services were made based on citizen monitoring.

The results of this study point out the many challenges of moving from citizen reporting to improved public services, like waste management. Citizen monitoring of public services is noisy, inconsistent, and costly to process. It can be frustrating for managers to follow-up on information when clarifications are needed prior to acting, since volunteer reporters are not at the disposal of managers. Additionally, the volume of data can be overwhelming, with managers scarcely having enough time to process one period of data before more data comes in requiring processing and action. Indeed, the waste management team even stopped producing weekly action plans in response to the data, because they felt they did not have enough time to act on each one and were spending more effort processing data, as compared to actually responding to the information that they received.

Overall, citizen-sourced data is promising because of its potential to expand the scope of monitoring, while at the same time offering localized and timely data. We find that this promise is likely to be overstated because of the complexities involved in processing citizen-source data and the inconsistencies that are inherent to citizen reporting. We frame the conditions under which citizen reporting will be helpful, that is when the data is produces is easy to process, consistent, low-cost relative to alternatives, and is brought to bear on decisions with high degrees of uncertainty. These conditions are unlikely to exist across a range of realistic circumstances.

References

Baldwin, Kate. 2013. "Why vote with the chief? Political connections and public goods provision in Zambia." *American Journal of Political Science* 57(4):794–809.

Banerjee, Abhijit, Angus Deaton and Esther Duflo. 2004. "Health, health care, and economic development: Wealth, health, and health services in rural Rajasthan." *The American economic review* 94(2):326–330.

Banerjee, Abhijit V, Rukmini Banerji, Esther Duflo, Rachel Glennerster and Stuti Khemani. 2010. "Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India." *American Economic Journal: Economic Policy* 2(1):1–30.

Bertot, John C, Paul T Jaeger and Justin M Grimes. 2010. "Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies." *Government information quarterly* 27(3):264–271.

Bhuiyan, Shahjahan H. 2010. "A crisis in governance: Urban solid waste management in Bangladesh." *Habitat International* 34(1):125–133.

Björkman, Martina and Jakob Svensson. 2009. "Power to the people: evidence from a randomized field experiment on community-based monitoring in Uganda." *The Quarterly Journal of Economics* 124(2):735–769.

Björkman, Martina and Jakob Svensson. 2010. "When is community-based monitoring effective? Evidence from a randomized experiment in primary health in Uganda." *Journal of the European Economic Association* 8(2-3):571–581.

Briggs, Ryan C. 2012. "Electrifying the base? Aid and incumbent advantage in Ghana." *The Journal of Modern African Studies* 50(4):603–624.

Buntaine, Mark T, Daniel L Nielson and Jacob T Skaggs. 2017. Escaping the Valley of Disengagement: Two Field Experiments on Motivating Citizens to Monitor Public Goods. Technical Report 41 AidData.

Charalabidis, Yannis, Anna Triantafillou, Vangelis Karkaletsis and Euripidis Loukis. 2012. "Public policy formulation through non moderated crowdsourcing in social media." *Electronic participation* pp. 156–169.

Connors, John Patrick, Shufei Lei and Maggi Kelly. 2012. "Citizen science in the age of neogeography: Utilizing volunteered geographic information for environmental monitoring." *Annals of the Association of American Geographers* 102(6):1267–1289.

Drazen, Allan and Marcela Eslava. 2010. "Electoral manipulation via voter-friendly spending: Theory and evidence." *Journal of development economics* 92(1):39–52.

Evans, Angela M and Adriana Campos. 2013. "Open government initiatives: Challenges of citizen participation." *Journal of Policy Analysis and Management* 32(1):172–185.

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

Grossman, Guy, Melina Platas and Jonathan Rodden. 2017. Crowdsourcing Accountability: ICT for Service Delivery. Technical report.
URL: *https://ssrn.com/abstract=3027923*

Hiltz, Starr Roxanne and Linda Plotnick. 2013. Dealing with information overload when using social media for emergency management: Emerging solutions. In *ISCRAM*.

Jablonski, Ryan S. 2014. "How aid targets votes: the impact of electoral incentives on foreign aid distribution." *World Politics* 66(2):293–330.

Katusiimeh, Mesharch W, Arthur PJ Mol and Kees Burger. 2012. "The operations and effectiveness of public and private provision of solid waste collection services in Kampala." *Habitat International* 36(2):247–252.

Linders, Dennis. 2012. "From e-government to we-government: Defining a typology for citizen coproduction in the age of social media." *Government Information Quarterly* 29(4):446–454.

McCubbins, Mathew D and Thomas Schwartz. 1984. "Congressional oversight overlooked: Police patrols versus fire alarms." *American Journal of Political Science* pp. 165–179.

Mossberger, Karen, Yonghong Wu and Jared Crawford. 2013. "Connecting citizens and local governments? Social media and interactivity in major US cities." *Government Information Quarterly* 30(4):351–358.

Ntaliani, Maria, Constantina Costopoulou and Sotiris Karetsos. 2008. "Mobile government: A challenge for agriculture." *Government Information Quarterly* 25(4):699–716.

Oates, Briony J. 2003. "The potential contribution of ICTs to the political process." *Electronic Journal of e-Government* .

of Statistics, Uganda Bureau. N.d. National Population and Housing Census 2014 Area Specific Profiles: Kampala Capital City Authority. Technical report.
URL: *http://www.ubos.org/onlinefiles/uploads/ubos/2014CensusProfiles/KAMPALA-KCCA.pdf*

Okot-Okumu, James and Richard Nyenje. 2011. "Municipal solid waste management under decentralization in Uganda." *Habitat International* 35(4):537–543.

Olken, Benjamin A. 2007. "Monitoring corruption: evidence from a field experiment in Indonesia." *Journal of political Economy* 115(2):200–249.

Oteng-Ababio, Martin. 2010. "Private sector involvement in solid waste management in the Greater Accra Metropolitan Area in Ghana." *Waste Management & Research* 28(4):322–329.

Reinikka, Ritva and Jakob Svensson. 2011. "The power of information in public services: Evidence from education in Uganda." *Journal of Public Economics* 95(7):956–966.

Rotberg, Robert I and Jenny C Aker. 2013. "Mobile phones: uplifting weak and failed states." *The Washington Quarterly* 36(1):111–125.

Saxton, Gregory D, Onook Oh and Rajiv Kishore. 2013. "Rules of crowdsourcing: Models, issues, and systems of control." *Information Systems Management* 30(1):2–20.

Seltzer, Ethan and Dillon Mahmoudi. 2013. "Citizen participation, open innovation, and crowdsourcing: Challenges and opportunities for planning." *CPL bibliography* 28(1):3–18.

Sjoberg, Fredrik M, Jonathan Mellon and Tiago Peixoto. 2017. "The effect of bureaucratic responsiveness on citizen participation." *Public Administration Review* 77(3):340–351.

Vermeiren, Karolien, Anton Van Rompaey, Maarten Loopmans, Eria Serwajja and Paul Mukwaya. 2012. "Urban growth of Kampala, Uganda: Pattern analysis and scenario development." *Landscape and Urban Planning* 106(2):199–206.

Zook, Matthew, Mark Graham, Taylor Shelton and Sean Gorman. 2010. "Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake." *World Medical & Health Policy* 2(2):7–33.

Zurovac, Dejan, Ambrose O Talisuna and Robert W Snow. 2012. "Mobile phone text messaging: tool for malaria control in Africa." *PLoS medicine* 9(2):e1001176.

Supporting Information

Data Cleaning

The results reported in the following section are those arrived at after an extensive effort to clean the data. We had two kinds of cleaning that are particularly important to mention as part of the present analysis. First, the data collection reported in this paper depended on the ability to visit the same area repeatedly to assess the area of waste accumulation. Because the amount of waste that people would add to unmanaged piles is directly related to the availability and use of formal pick ups, understanding changes in pile sizes that community members identified as most important should be a strong measure of waste services. We cross-checked the GPS locations of all piles in baseline and both midline waves and excluded from the data any pile location that was more than 100m from the baseline location, based on the field-tested accuracy of the tablets that we used for enumeration. Second, there appears to be unit errors in some of the baseline audit files, with pile sizes recorded that are implausible given the associated photographs. For the analysis reported below, we have completed a double-review of all piles that indicate large size changes between phases and excluded piles where unanimous or a major of reviewers believe the recorded pile sizes are implausible.

Additional Figures and Tables

Table 3: Treatment Effect of Citizen Reporting Conditional on the Party Affiliation of the Division Councillor Elected in 2016.

	DV: Pile Cleaned (0/1) or Change in Waste Pile Size (m^2)			
	M1 Cleaned	M1 Change	M2 Cleaned	M2 Change
	(1)	(2)	(3)	(4)
Treatment	2.964	5.823	9.421	12.691
	(10.945)	(12.260)	(5.736)	(8.515)
Independent	0.876	−4.053	2.041	−3.544
	(13.170)	(14.748)	(6.907)	(10.251)
Opposition	1.455	9.184	−3.181	5.629
	(13.398)	(14.987)	(7.023)	(10.413)
Baseline Pile Area	0.258***		0.156***	
	(0.075)		(0.039)	
Treatment X Independent	−3.953	−4.841	−5.725	−6.752
	(19.767)	(22.150)	(10.374)	(15.407)
Treatment X Opposition	28.514	10.640	−9.198	−29.536**
	(18.968)	(21.158)	(9.957)	(14.720)
Covariates	Yes	Yes	Yes	Yes
Observations	391	391	392	392
R^2	0.063	0.011	0.073	0.031
Adjusted R^2	0.028	−0.023	0.039	−0.002
Residual Std. Error	74.698	83.705	39.223	58.252
F Statistic	1.805**	0.335	2.124**	0.944

*Note:*two-tailed tests *p<0.1; **p<0.05; ***p<0.01

Table 4: Estimated Effects of Treatment Conditional on Zone-Level Response Rate

	DV: Pile Cleaned (0/1) or Change in Waste Pile Size (m^2)			
	M1 Cleaned	M1 Change	M2 Cleaned	M2 Change
	(1)	(2)	(3)	(4)
Zone-Level Response Rate	0.126	−35.858	−0.035	114.571
	(0.244)	(58.304)	(0.296)	(126.081)
Baseline Pile Area	0.0002	−0.563***	−0.0003	−0.740***
	(0.001)	(0.121)	(0.001)	(0.261)
P1/P2 Monitoring	0.068	23.004*	−0.033	−5.903
	(0.050)	(11.997)	(0.061)	(25.944)
Covariates	Yes	Yes	Yes	Yes
Observations	313	313	313	313
R^2	0.052	0.112	0.026	0.064
Adjusted R^2	0.017	0.079	−0.010	0.030
Residual Std. Error	0.338	80.844	0.411	174.823
F Statistic	1.491	3.446***	0.723	1.874**

Note: two-tailed tests *p<0.1; **p<0.05; ***p<0.01

Table 5: Estimated Effects of Treatment, Conditional on Baseline Quality of Service Provision

	DV: Pile Cleaned (0/1) or Change in Waste Pile Size (m^2)			
	M1 Cleaned	M1 Change	M2 Cleaned	M2 Change
	(1)	(2)	(3)	(4)
Service Quality	−0.041	2.227	−0.059	−0.059
	(0.042)	(10.166)	(0.051)	(0.051)
Baseline Pile Area	0.0002	−0.563***	−0.0003	−0.0003
	(0.001)	(0.121)	(0.001)	(0.001)
P1/P2 Monitoring	0.073	22.659*	−0.027	−0.027
	(0.050)	(12.050)	(0.061)	(0.061)
Covariates	Yes	Yes	Yes	Yes
Observations	313	313	313	313
R^2	0.054	0.111	0.030	0.030
Adjusted R^2	0.019	0.078	−0.006	−0.006
Residual Std. Error	0.338	80.888	0.410	0.410
F Statistic	1.556	3.412***	0.843	0.843

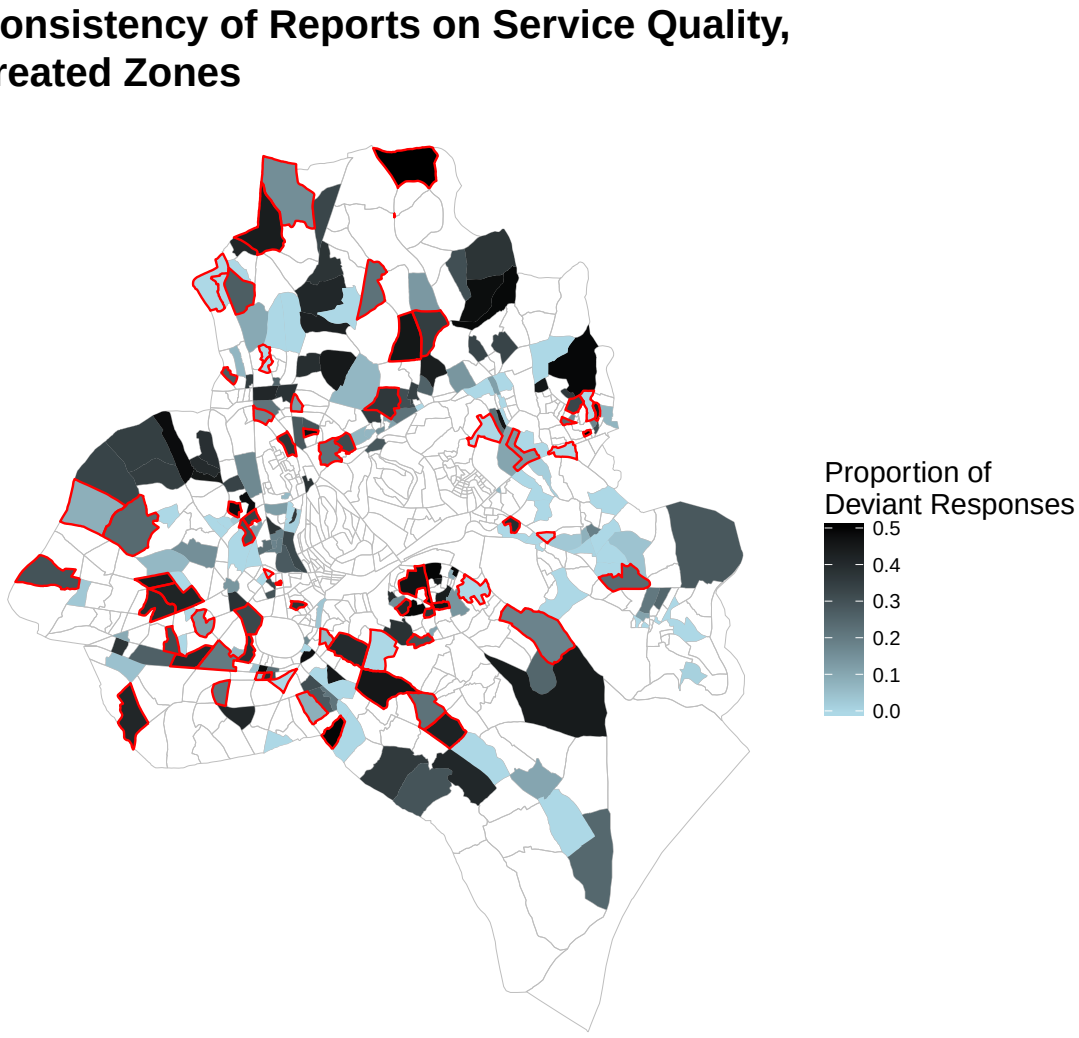
Note: two-tailed tests *p<0.1; **p<0.05; ***p<0.01

Table 6: Estimated Effects of Treatment Conditional on Zone-Level Dissatisfaction

	DV: Pile Cleaned (0/1) or Change in Waste Pile Size (m^2)			
	M1 Cleaned	M1 Change	M2 Cleaned	M2 Change
	(1)	(2)	(3)	(4)
Dissatisfaction	0.068	19.286*	0.036	−0.092
	(0.048)	(11.585)	(0.058)	(25.081)
Baseline Pile Area	0.0003	−0.584***	−0.0003	−0.818***
	(0.001)	(0.124)	(0.001)	(0.268)
P1/P2 Monitoring	0.056	17.863	−0.044	−7.023
	(0.052)	(12.539)	(0.063)	(27.146)
Covariates	Yes	Yes	Yes	Yes
Observations	293	293	293	293
R^2	0.065	0.130	0.029	0.070
Adjusted R^2	0.028	0.096	−0.009	0.033
Residual Std. Error	0.339	81.995	0.411	177.510
F Statistic	1.767*	3.824***	0.752	1.911**

Note: two-tailed tests *p<0.1; **p<0.05; ***p<0.01

Figure 15: Consistency of zone-level reporting on KCCA service quality.



Along a standardized measure of poor service provision, zones in red, on average, indicated that KCCA service provision was poor. The standardized measure of poor service provision combined citizen monitor responses on the following indicators: the frequency and accessibility of service provision, reported waste collector treatment of citizens outlined, and the amount of waste burning or litter.

Table 7: Treatment Effect of Citizen Reporting Conditional on Consistency of Zone-Level Reports on Service Quality

	DV: Pile Cleaned (0/1) or Change in Waste Pile Size (m^2)			
	M1 Cleaned	M1 Change	M2 Cleaned	M2 Change
	(1)	(2)	(3)	(4)
Consistency	−0.053	18.930	0.042	41.544
	(0.125)	(30.883)	(0.152)	(65.384)
Baseline Pile Area	0.0003		−0.0003	
	(0.001)		(0.001)	
P1/P2 Monitoring	0.070	23.115*	−0.034	−5.607
	(0.050)	(12.419)	(0.061)	(26.293)
Covariates	Yes	Yes	Yes	Yes
Observations	313	313	313	313
R^2	0.051	0.047	0.026	0.038
Adjusted R^2	0.017	0.016	−0.010	0.006
Residual Std. Error	0.339	83.585	0.411	176.962
F Statistic	1.483	1.504	0.729	1.188
Note: two-tailed tests				*p<0.1; **p<0.05; ***p<0.01

We use data from baseline surveys deployed in treated zones to measure the consistency of citizen-monitor reports within zones. Here, consistency is operationalized as the proportion of zone-level responses deviating from the zone-level modal response on an overall indicator of KCCA service quality. Higher proportions of deviant responses indicate that citizen-monitors from a given zone were providing inconsistent information to the KCCA regarding the quality of baseline waste services. We construct our overall measure of service quality using information from the baseline surveys on the following: the frequency and accessibility of service provision, reported waste collector treatment of citizens outlined, and the amount of waste burning or litter.