NOVAFRICA
Working Paper Series

Coordination and the Poor Maintenance Trap: an Experiment on Public Infrastructure in India

Alex Armand
Nova School of Business and Economics and NOVAFRICA

Britta Augsburg
Institute for Fiscal Studies

Antonella Bancalari
School of Economics and Finance, University of St. Andrews, and Institute for Fiscal Studies

ISSN 2183-0843
Working Paper No 2110
July 2021

NOVAFRICA Working Paper

Coordination and the Poor Maintenance Trap: an Experiment on Public Infrastructure in India*

Alex Armand Britta Augsburg Antonella Bancalari

Abstract

Poorly maintained public infrastructure is common in poorer countries, but very little is known about the obstacles leading to such equilibrium. By experimentally identifying the impact of incentives for maintenance for both providers and potential users, this paper provides one of the first economic analyses of provider–user dynamics in the presence of local coordination failure. We randomly allocate shared sanitation facilities in two major Indian cities to either a control or two treatments: the first incentivizes maintenance among providers, while the second adds a sensitization campaign about the returns of a well-maintained facility among potential users. Using a wide range of survey, behavioral and objective measurements, we show that maintenance does not favor collective action. The treatments raise the quality of facilities and reduce free riding, but at the cost of user selection, with consequences for public health. While potential users' willingness to pay and cooperation are unaffected, their demand for public intervention increases. Sensitization raises awareness among potential users, but does not alter their behavior. (*JEL* D12, C93, I15, I18, O18, Q53)

Keywords: infrastructure, maintenance, free riding, willingness to pay, basic services, water and sanitation, information, health.

*[*] **Armand**: Nova School of Business and Economics – Universidade Nova de Lisboa, CEPR, NOVAFRICA, and Institute for Fiscal Studies (e-mail: alex.armand@novasbe.pt); **Augsburg**: Institute for Fiscal Studies (e-mail: britta_a@ifs.org.uk); **Bancalari**: School of Economics and Finance, University of St. Andrews, and Institute for Fiscal Studies (e-mail: antonella.bancalari@st-andrews.ac.uk). We would like to thank Nava Ashraf, Orazio Attanasio, Oriana Bandiera, Joan Costa-Font, Oliver Cumming, Greg Fisher, Sebastian Galiani, Michael Kremer, Eliana La Ferrara, Berkay Ozcan, Rohini Pande, Imran Rasul, Sandra Sequeira, Manisha Shah, Marcos Vera-Hernandez, and seminar/conference participants at the Econometric Society's DSE Winter School, London School of Economics and Political Science, Institute for Fiscal Studies, University College London, CSAE Oxford University, University of St. Andrews, Royal Economic Society, 3ie Delhi Evidence Week's WASH Challenges and Opportunities in the Urban Setting workshop, and the Asian Development Bank Institute / Bill and Melinda Gates Foundation workshop for helpful comments. This project would not have been possible without the generous and insightful support of Abhijit Banerji, Sneha Santwana, Bhartendu Trivedi, and the team of implementers, field supervisors and enumerators at FINISH Society and Morsel Research and Development. We are also indebted to study participants for their willingness to contribute to the study. We gratefully acknowledge financial support from 3ie – International Initiative for Impact Evaluation (grant no. DPW1-1105) and from the Economic and Social Research Councils Centre for the Microeconomic Analysis of Public Policy at the IFS (grant no. ES/M010147/1). Ethics approval was secured from University College London (application n. 2168/012). The pre-analysis plan was registered on the AEA RCT registry (Armand et al., 2018).

Poorly maintained public infrastructure is highly prevalent in low- and middle-income countries (LMICs). This condition characterizes the provision of a wide array of basic services, from unreliable electricity grids and over-strained transport networks to inadequate water and sanitation facilities (G20, 2017; United Nations, 2020). Limited investment by public and private providers, mixed with low private willingness to pay (WTP) for use and free riding, can induce a *poor maintenance trap*: the coordination failure between users and providers can drive effective prices for use below the marginal cost of operation, jeopardizing incentives to invest in maintenance, reducing private valuations further, and enduring public infrastructure in a lower quality equilibrium (Burgess et al., 2020). The World Bank estimates that 4.2 trillion US$ could be saved by investing in more resilient infrastructure to avoid costly upgrades or the abandonment of obsolete facilities in the future (Hallegatte et al., 2019).

This vicious cycle has serious consequences for economic development and public health in LMICs (Bartram et al., 2005). Low-quality basic services deriving from poorly maintained facilities often translate into low levels of human capital and persistent poverty (Currie and Vogl, 2013; Ghatak, 2015). This issue is particularly salient in densely populated informal settlements (or *slums*), home to more than 1 billion people worldwide, mostly in LMICs (Marx et al., 2013; United Nations, 2020). In these overcrowded urban spaces, coordination between users and providers that could guarantee improvements in maintenance is limited by a complexity of market failures. These are linked to severe health externalities and low valuations for environmental improvements among citizens (Greenstone and Jack, 2015; Ezeh et al., 2017; Berry et al., 2020), to informality of property rights (Field, 2005; Galiani et al., 2017) and to poor governance (Olken and Pande, 2010; Fox, 2014; Marx et al., 2019). In such settings, free riding is expected to be pervasive (Banerjee et al., 2008b; Chidambaram, 2020).

While a large literature highlights the substantial returns of upgrades of public infrastructure,[1] very little is known about the obstacles leading to poor maintenance and how to incentivize collective action to sustain a higher-quality equilibrium (Duflo et al., 2012). This paper fills this gap by studying how supply- and demand-side constraints contribute to the quality of infrastructure in LMICs. By experimentally identifying the impact of incentives for local maintenance for both providers and potential users, and mapping effects into behavioral responses from both sides of the market, we provide one of the few economic analyses of provider–user dynamics in the presence of severe coordination failure.

We focus on access to sanitation in urban areas. With more than 700 million urban residents worldwide lacking access, this basic service is of first order importance for the developing world (Ezeh et al., 2017; WHO-UNICEF, 2017).[2] A unique setting for studying the mechanics behind

[1]Evidence covers the communication (Jensen, 2007), electricity (Dinkelman, 2011; Rud, 2012; Lipscomb et al., 2013), water (Duflo and Pande, 2007; Kremer et al., 2011; Meeks, 2017) and transportation networks (Gonzalez-Navarro and Quintana-Domeque, 2016; Donaldson, 2018; Banerjee et al., 2020a; Asher and Novosad, 2020).

[2]India and sub-Saharan Africa lag behind with less than 40% of their populations with access to *improved sanitation*, which ensures hygienic separation of human excreta from human contact (WHO-UNICEF, 2017). In India, lack of access is responsible for 9% of the total infant mortality, and a loss of up to 6.4% of annual GDP (Geruso

the low-quality trap is represented by community toilets (CTs) in India, where maintenance is highly characterized by coordination at the local level between users and providers. These shared sanitation complexes typically operate on a pay-to-use basis, and revenues are used to fund operation and maintenance (O&M).[3] CTs are often found in slums, where large shares of residents frequently resort to the practice of open defecation (OD), with severe social and economic costs (Prüss-Ustün et al., 2014; Pickering et al., 2015; Augsburg and Rodríguez-Lesmes, 2018). In slums, CTs are considered the most appropriate medium-term solution for access to sanitation (WSUP, 2011).[4] A well-maintained facility can generate large social returns, but coordination fails. On the one hand, facilities are poorly maintained, dirty and with a widespread presence of bacteria harmful to human health. On the other hand, potential users have extremely low WTP for their use, and free riding is widespread.

The degrading status of CTs in India mirrors the low level of maintenance of a wide variety of alternative infrastructures, and the low quality of basic service delivery that results from it. Evidence about the mechanisms leading to such a low-quality equilibrium remains scarce. From a supply-side perspective, most of the evidence highlights the constraints associated with limited *last-mile* connections (Devoto et al., 2012; Duflo et al., 2015; Ben Yishay et al., 2017; Lee et al., 2020), unfinished projects (Bancalari, 2020), and nonpayment among users (Coville et al., 2020). More generally, the literature also highlights how corruption and institutional weakness can contribute (Ferraz and Finan, 2008; Burgess et al., 2015; Lewis-Faupel et al., 2016; Williams, 2017).

The role of maintenance at the micro level is largely ignored. In particular, we know very little about the incentives of actors who are directly involved in the maintenance of local facilities (or nodes) in the infrastructure network. First, if facilities are resource-constrained, a one-off external provision of funds to support maintenance can directly improve the quality of the facility, create value among potential users and reduce free riding, ultimately setting facilities on a trajectory towards a higher-quality equilibrium. Growing evidence shows that this instrument can help ultra-poor households or micro-entrepreneurs to escape poverty traps (Banerjee et al., 2015; Haushofer and Shapiro, 2016; Bandiera et al., 2017; Balboni et al., 2020), but it has not been considered in this setting. Second, extrinsic rewards can be used to increase work-

and Spears, 2018; World Bank, 2018).

[3]This model is present in a wide variety of LMICs, with varying degree of quality (UNICEF, 2019). In India, CT services are generally rendered on a public–private partnership with private contractors or NGOs, and differ from public toilets as they are provided for a defined group of residents lacking access to private toilets (Government of India, 2018).

[4]Safely managed individual household toilets are considered the first-best solution. Several approaches aimed at this solution have been shown successful, such as the Community-Led Total Sanitation campaign (Pickering et al., 2015; Cameron et al., 2019; Abramovsky et al., 2019), subsidies (Guiteras et al., 2015; Lipscomb and Schechter, 2018; Andres et al., 2020), and credit (Augsburg et al., 2021; Ben Yishay et al., 2017). The Indian Government combines awareness creation and subsidy provision to vulnerable households under the *Swachh Bharat Mission (SBM)*, with positive results (Spears and Lamba, 2016). However, under its urban component, toilets are pragmatically envisioned for only 80% of households engaging in OD, with the remaining 20% to be catered by CTs due to space constraints (Government of India, 2017).

ers' effort in maintaining the facility (Bénabou and Tirole, 2003, 2006). Financial rewards can positively affect performance on pro-socially motivated tasks, such as running CTs, especially when incentives are linked to own performance and in low-income settings (Muralidharan and Sundararaman, 2011; Ashraf et al., 2014a,b). However, rewards can also crowd out intrinsic motivation, generate multi-tasking problems or induce strategic behavior (Besley and Ghatak, 2018), as shown in the healthcare literature (Croxson et al., 2001; Basinga et al., 2011; Olken et al., 2014). To the best of our knowledge, there is no experimental evidence on how incentives for maintenance among providers determine infrastructure quality.

On the demand side, poor maintenance is recognized as the main constraint to CT use (Biran et al., 2011), highlighting the role of supply-side factors. Even small improvements in the facility are expected to increase WTP if slum residents assign value to them (Neal et al., 2016). However, the extent to which users value maintenance depends on whether they internalize the health risks deriving from using a poorly maintained infrastructure, and whether social norms that justify risky behaviors and free riding limit its demand (Dupas, 2011).[5] Where institutions are weak, while subsidies have been shown to reinforce a low-quality equilibrium and fines can lead to extortion (Mcrae, 2015; Ashraf et al., 2016), information and sensitization campaigns targeting risky behavior can be effective at raising private valuations.[6] In slums, campaigns can also generate information externalities (Dupas, 2014). Bottom-up incentives, such as beneficiary oversight, can be successful at improving provider performance by mobilizing citizens (Björkman and Svensson, 2009; Dupas and Miguel, 2017). Understanding why constraints to the demand for maintenance persist in the presence of high private and social returns of improved maintenance remains an open question.

We bring together both the demand and supply branches of the literature by studying how provider–user dynamics determine infrastructure quality in the context of urban CTs. We partnered with city governments and a wide variety of non-governmental organizations (NGOs) in Lucknow and Kanpur, the state capital and the second-largest city of Uttar Pradesh, the fourth-most-densely populated state of India (Government of India, 2011). Similar to several growing cities in South Asia and sub-Saharan Africa, both are experiencing a fast pace of urbanization, widespread presence of slums, and the need for a rapidly expanding urban infrastructure. Following extensive efforts to identify and map all the slums and CTs in the two cities and census all slum residents, we implemented a field experiment targeting residents with no access to private toilets and living in the catchment area of CTs.

A total of 110 catchment areas were randomly allocated to one of three groups. In a first group, labeled *maintenance* treatment, we focus on supply-side constraints by incentivizing maintenance of the CT. The intervention is targeted at the CT caretaker, the person in charge of the

[5] Demand can also be distorted by time-inconsistent preferences. See, e.g., Koszegi and Szeidl (2013).

[6] Information campaigns have proven effective in influencing the adoption of new products in other settings (Jalan and Somanathan, 2008; Luoto et al., 2011), including sanitation (Pickering et al., 2015; Briceño et al., 2017; Cameron et al., 2019; Abramovsky et al., 2019).

maintenance in the facility and of the collection of fees among users. A one-off maintenance grant was first offered to improve facilities. Following the grant, caretakers were incentivized through a large bimonthly financial reward (roughly 40% of their monthly salary) to keep the facility clean. In a second group, labeled *maintenance plus sensitization* treatment, the intervention provided in the *maintenance* treatment was supplemented by an intensive sensitization campaign to raise awareness among slum residents about the importance of a well-maintained facility and of avoiding free riding to support maintenance. The campaign included face-to-face sessions, the distribution of leaflets, monthly voice message reminders sent to mobile phones, and posters hung in the CTs. Finally, in a *control* group, no intervention was implemented.

The study incorporates a wide range of measurements for both slum residents and caretakers. From April 2018, over a period of 18 months, we collected a baseline survey, followed by five rounds of survey data at the CT level by interviewing caretakers, and three waves of a household panel survey. Multiple follow-up measurements at relatively short intervals allow us to estimate impacts with reduced noise and increased power (McKenzie, 2012), as well as gaining an insight into dynamics over time. Objective measurements of CT quality and free riding were collected using observers and laboratory tests to measure bacteria prevalence. Surveys were supplemented with behavioral measurements to capture real-world behaviors (Charness and Fehr, 2015; Gneezy and Imas, 2017). These included a structured community activity (SCA) to measure demand for public intervention (see, e.g., Casey et al., 2012) and lab-in-the-field experiments to measure incentive-compatible WTP for CT use, preference for maintenance and willingness to cooperate among slum residents, and pro-social motivation for the cause among caretakers. The design of the experiment and the analysis presented follow a pre-analysis plan registered in the AEA RCT Registry (Armand et al., 2018). Our estimates of treatment impacts are robust to a wide variety of estimation procedures, from standard OLS and ANCOVA specifications to machine learning techniques.

Relaxing supply-side constraints generates sustained improvements in the observed quality of facilities, accompanied by a significant reduction in free riding among users by 18% as compared with the control group. While slum residents perceive the improvements, relaxing supply-side constraints has no average effect on either the way they value use or their attitudes towards maintenance and cooperation. The WTP for CT use is only marginally reduced when the intervention is limited to the part that only offers the maintenance grant to caretakers, in line with a (temporary) crowding-out effect of the interventions.

The sensitization campaign was effective at raising awareness among slum residents, but did not affect the behavior of either slum residents or caretakers. Contrary to Ashraf et al. (2013) and Guiteras et al. (2015), we observe only a weak complementarity between the maintenance intervention and the sensitization campaign. In settings where targeted populations are already familiar with the service rendered or there are strong implicit norms in their behavior, stimulating both the demand and supply sides has limited effectiveness as compared with lifting supply-side constraints alone.

Relaxing supply-side constraints also increases the share of slum residents asking local politicians for public intervention in the O&M of CTs. Since such demand tends to be low, the increase is large at 50% over the control mean. This result sheds light on the political economy of public service delivery in urban slums, where political representation of citizens tends to be low (Fox, 2014). While citizen mobilization has been shown to be responsive to information (Ferraz and Finan, 2011; Banerjee et al., 2020b), we show that mobilization is increased when external funds are transferred to improve the local facility. Sensitization among slum residents has no additional effect on this outcome. Together with an unaffected WTP for use, a demand for public intervention suggests that slum residents treat access to sanitation as a right rather than a service to purchase. This has fundamental consequences for the design of policies targeting access to sanitation in urban areas (Burgess et al., 2020).

Altogether, externally incentivizing maintenance does not favor collective action between users and providers in a way that would allow for financial self-sustainability of the O&M of CTs. The intervention is mainly effective on the supply side, but, rather than through voluntary reductions in free riding, this is coming at the cost of user selection and with a small decrease in traffic. User selection is likely achieved by stronger enforcement of payment as caretakers respond strategically to incentives by allocating a larger share of their time to management activities (i.e., the collection of fees and the supervision of cleaning) and away from operating activities, leaving their total labor supply or pro-social motivation unchanged. In treated catchment areas, the average share of respondents who practice OD at endline is 18 percentage points larger than for the control group, and a larger share of slum residents report payment enforcement and refusals to enter. As the share of slum residents reporting positive health expenditures increases in response to the intervention, this result has important consequences for public health, since exclusion might generate larger externalities than the use of CTs, regardless of their quality. These novel findings provide insights not only into key market failures in public good provision in low-income settings, but also on how to improve it. Understanding such limitations is crucial for the economic development of cities in LMICs and for the welfare of their citizens (Bryan et al., 2020).

1 The interventions

Our focus is the slums of the two largest urban agglomerations in the Indian state of Uttar Pradesh, Lucknow and Kanpur. Appendix B discusses the external validity of the study area, and provides the timeline and operational details about the interventions. Similar to many expanding cities in South Asia and sub-Saharan Africa, Lucknow and Kanpur are characterized by a rapid population growth, a relatively large prevalence of informal settlements, and the lack of access to sanitation for their slum populations. In Uttar Pradesh, more than 40% of slum residents lack access to private toilets (Government of India, 2011). For these households, the only alternative to OD is the use of CTs. These are compounds of several defecation cubicles, urinals, hand-washing and bathing facilities arranged in gender-specific areas. They are available for a defined group of slum residents, and they are connected either to sewerage systems or to septic tanks, providing an upgrade in the sanitation ladder compared with rudimentary private sanitation facilities or OD (WHO-UNICEF, 2017). In Uttar Pradesh, CTs operate on a pay-to-use system, with a standard fee of 5 Indian rupees (INR, corresponding to US$ 0.07), which allows for the use of all services.[7] This infrastructure is constructed by municipal corporations, and services are generally rendered on a public–private partnership on a long-term basis.

In both cities, CTs serving slums are poorly maintained, reflected by poor quality of construction, lack of functioning hand-washing facilities, and dirtiness (Appendix E). Less than 40% of facilities have finished walls for the compound, and hand-washing for both genders is available in only half of the facilities. Female areas are more poorly maintained than male areas, with less functioning lighting, worse cleanliness, more unpleasant smell and higher prevalence of insects. Overall, more than 95% of facilities have a visible sewage leak, and the prevalence of bacteria harmful to humans is high (Appendix C.3). These conditions are not specific to our study area as the low quality of CTs all over India is widely covered by the media (National Geographic, 2017; The Times of India, 2020).

The daily O&M of CTs is performed by caretakers. They are in charge of collecting user fees and they implement routine maintenance, which includes cleaning the facility or supervising cleaners, maintaining the stock of cleaning agents, asking for and/or implementing repairs and the deep cleaning of sanitation systems. Caretakers are hired centrally by the organization managing the CT and are supervised by zone managers who are charged with multiple CTs. They receive a fixed salary, equal on average to INR 5,000 per month, and, as part of their benefits, they can live in the CT compound and run an independent business in the facility (e.g., a small retail shop). Caretakers can be fired or moved to another facility in the case of poor performance, which is often measured by the level of revenues collected.[8] In our sample, caretakers have on average 10 years of experience in the same job, and 4 years working in the same facility. The status of CTs is closely related to the caretaker's characteristics. Recent improvements, which positively correlate with the share of functioning toilets, are less prevalent where caretakers spend more time on managing the CT rather than in operations. Free riding is lower in facilities where the caretaker is male, is more pro-socially motivated, spends a higher share of time in managing the CT, does not work as the cleaner and chooses better inputs for the routine

[7]Nominal INR are converted to nominal US$ using the 2019 average exchange rate of 70.42 US$/INR (IMF, 2020).

[8]Revenues are used to hire cleaners and plumbers (hired by the central organization or by the caretaker), pay for water and electricity bills, implement repairs, and buy cleaning agents (generally provided by the organization managing the CT). Qualitative interviews with the city manager of the main NGO operating CTs in the study area highlight that caretakers are expected to collect in revenues an amount larger than their own salary, and any shortfall can be deducted from their salary at the end of each month (Armand et al., 2020a).

maintenance (Appendix E.3).

The low quality of facilities is perceived by slum residents. Only 15% of households report liking the facility and less than half report liking the services offered; 36% report that the CT is clean and 28% report that the CT is safe. WTP for using the CT is particularly low among potential users (Panel A in Figure 1). On average, slum residents are willing to pay INR 1.40 to use the CT, corresponding to just 28% of the official market rate of INR 5 per ticket. WTP is slightly higher for male respondents (INR 1.46 versus 1.36 for female respondents), and in households that always use the CT (INR 1.53 versus 1.33 in households that do not always use it). At current conditions, WTP is unrelated to the quality of the CT. However, for a hypothetical higher-quality CT, potential users are, on average, willing to pay above the market price of INR 5 (Appendix C.5).

In the absence of free riding, an average household of four members (excluding under-5s) would spend INR 600 per month to use the CT once per day, around 8% of their average household income and less than the amount they report spending on intoxicants (INR 817). However, free riding is often found to be rampant (Panel B in Figure 1). On average, only 66% of users pay the CT fee. Among female users, 50% do not pay the fee, with 30% of CTs receiving no payment at all, and everybody paying the fee in only 21% of CTs. Among male users, free riding is instead at 24%, with 34% of CTs receiving no payment at all; at the same time, though, only 3% of CTs have everybody paying the fee. Payments are only partly enforced by caretakers, with just 8% of slum residents reporting having been prevented from using the facility for not paying the fee. At lower levels of OD in the community, free riding and OD are negatively related, indicating that stricter payment enforcement may lead residents to practice OD. At high levels of OD in the community, however, the relationship turns positive, which characterizes areas with poor-quality CTs and rampant free riding (Appendix Figure E3).

Two interventions were implemented in conjunction with the local NGO FINISH Society with the objectives of improving CT maintenance and raising awareness among slum residents about the importance for the community of a well-maintained CT. The first intervention, labeled *maintenance*, was designed to promote the quality of the CT facility and the service provided using two separate components. The first component is a one-off grant to rehabilitate the infrastructure. The grant is offered to the CT caretaker who, according to the CT's needs, could choose for it to be spent on one of three packages: deep cleaning (i.e., septic tank sewage removal, unclogging latrines and sewerage pipes and cleaning walls, floors and inside toilets), repairs (i.e., sanitation/water connection repairs and/or infrastructure refurbishment) or a training session on cleaning coupled with tools and agents. FINISH Society would then implement the caretaker's choice. 41% of CT caretakers selected deep cleaning, 41% selected repairs and 18% selected the training. The average value of the grant is equal to INR 25,000 (US$ 355). Examples are provided in Appendix B.1.

The second component is a bimonthly financial reward for caretakers to improve the quality of the service provided by the CT. This component was announced to caretakers two months after the implementation of the maintenance grant. The announcement was made in person by a member of the implementing team, who also provided the selected caretakers with a summary page of the reward scheme, including a contact number for information. Payments were made every two months for a total of four times during the study. The financial reward was conditional on complying with various objectives measured during enumerators' visits. First, the availability of soap in the hand-washing facilities rewarded caretakers with INR 500 (US$ 7.10). Second, visible cleanliness of latrines rewarded caretakers with an additional INR 500 (US$ 7.10). Finally, bacteria counts in defecation cubicles kept to a minimum standard rewarded caretakers with a further INR 1,000 (US$ 14.20).[9] In total, caretakers could receive a maximum of INR 2,000 (US$ 28.40) in each round, equivalent to 40% of their average monthly salary, or INR 8,000 (US$ 113.60) in all rounds combined, equivalent to 13% of the annual salary.

In each payment round, we informed caretakers of their past performance to help them estimate the effort required to achieve the conditions. We linked payments to their own contemporary performance, rather than to relative performance or rankings, to minimize strategic behavior associated with past performance and irrational behavior, such as handicapping their own performance to preserve a high self-image (Bénabou and Tirole, 2002, 2003; Bandiera et al., 2015). Caretakers received on average INR 779 (US$ 11.06) in the first round of incentives (39% of the potential reward), INR 1,036 (US$ 14.71) in the second round (52% of the potential reward), INR 1,058 (US$ 15.02) in the third round (53% of the potential reward) and INR 972 (US$ 13.80) in the last round (49% of the potential reward).[10] At the end of the intervention, the caretakers were given a certificate signed by all implementing partners with the results achieved.

The second intervention complemented the maintenance with a *sensitization* campaign among slum residents. The campaign, titled *Awareness campaign to encourage CT use and maintenance in India*, aimed to increase awareness of the importance of paying the CT fee to fund O&M, the role of users in holding caretakers accountable for maintenance, and negative externalities resulting from unsafe sanitation behavior. The sensitization campaign was provided through four different means. First, door-to-door visits were implemented three times in April–June 2018, July–September 2018 and January–March 2019 to raise awareness about the campaign's main messages. This component targeted all household members, and was implemented using a flip chart with pictures to allow participants with low literacy to process key messages. Second, at the end of the visit, a leaflet summarizing the main messages was left with the households. Third, posters highlighting messages provided during the door-to-door campaign were

[9]We define the minimum standard as being below the median value of the bacteria count for E. coli computed among all CTs at baseline. Details about the collection of bacteria counts are provided in Appendix C.3.

[10]In CTs with more than one caretaker (20% of the sample), the reward was equally split between them. Caretakers work only in one CT, and therefore efforts can be targeted only at the specific facility where they work. To circumvent rotation of caretakers to different facilities, we provided city managers with the list of all of the CTs in the study without revealing the treatment group. During the study period, we do not observe rotation of caretakers to different study CTs. The implementing team paid regular visits to CTs in order to inform new caretakers about the intervention.

placed in the CTs. Fourth, voice message reminders were sent by mobile phone to households with a monthly frequency, which has been shown ti be adequate to induce behavioral change (Cortes et al., 2021). Details are provided in Appendix B.2.

2 Research design

The research design is a two-stage cluster randomized controlled trial. In the first stage, the unit of interest is the catchment areas of CTs serving slum residents. In the second stage, the unit of interest is residents of the catchment areas that are potential users of the facility. Appendix C provides detailed information about sampling.

To define catchment areas, in 2017, we first conducted a census of all CTs in the study area, followed by a mapping of the slums surrounding each facility. During this process, we collected data on the location, the physical characteristics, the management and maintenance practices, and the users of each facility. These data formed the basis for selecting as part of the study facilities that are pay-to-use and are used mostly by slum residents. In the slums surrounding the selected facilities, we then conducted a resident census during the second half of 2017. In total, we collected information on more than 30,000 households in both cities. The census covered household demographics, dwelling characteristics (including geo-location) and sanitation-related behavior of slum residents. Based on this information, we finalized the definition of catchment area. Since the use of the facility declined rapidly with distance from the facility (Appendix E.1), we define the catchment area as the space inside the slum borders and within a radius of 150 or 250 meters from the facility, depending on the spatial distribution of private toilet ownership in the slum. A total of 110 catchment areas were identified, 52 in Lucknow and 58 in Kanpur.

Within each catchment area, we sample for the study the potential users of the facility. These are defined as households residing in the catchment area where at least one member reported not using a private toilet during the resident census, and with no intention of migrating within the 18 months following the census. The study sample is, on average, highly comparable to the population of slum residents in Uttar Pradesh and in the other states of India (Appendix C).

To provide exogenous variation in the provision of the interventions discussed in Section 1, each catchment area was randomly allocated to one of three groups. A first group, the *maintenance* treatment, receives the maintenance intervention. A second group, the *maintenance plus sensitization* treatment, receives the maintenance intervention and the sensitization campaign. A third group, the control, does not receive any intervention.[11] To allocate catchment areas to

[11] Concerning voice message reminders, to disentangle the effects of receiving voice messages and the effect of receiving messages about the sensitization campaign, all study participants received voice messages with no content related to the sensitization campaign. These messages stated the opening times of the local CT. Participants in the maintenance treatment additionally received voice messages informing them that the CT had been granted aid to improve its service.

treatment arms, we first stratified catchment areas according to the main organization managing the facility and the city. Using the rich census information we collected, we built blocks of three CTs using Mahalanobis-distance relative proximity and we randomly allocated each catchment area within a block to a treatment group using a lottery with equal probability of assignment. To prevent treatment contamination, we allocated catchment areas to the same treatment arm when the distance to each other was below 300 meters.

As a result, 35 catchment areas were allocated to the maintenance treatment, 35 catchment areas were allocated to the maintenance plus sensitization treatment, and 40 catchment areas were allocated to the control group. Figure C3 shows the maps of Lucknow and Kanpur with the allocation of catchment areas to the control and the two treatment arms.

3 Data

To gather information on both the supply side, represented by CTs and caretakers, and the demand side, represented by slum residents, we gathered a substantial amount of original data. These include a series of household and caretaker surveys, objective measurements, incentivized behavioral measurements, and an SCA. Detailed descriptions of each measurement and the relative scripts are provided in Appendix C.

3.1 CT survey and objective measurements

For each facility selected for the study, we administered a panel survey with the caretakers. The baseline survey was administered in April–June 2018, followed by five waves of follow-up data collection to document bimonthly the behavior of the caretakers, starting one month after the baseline: in July–September 2018 (follow-up 1), October–November 2018 (follow-up 2), January–March 2019 (follow-up 3), April–May 2019 (follow-up 4) and July–September 2019 (follow-up 5). The questionnaire covered CT management, time allocated to different tasks and cleaning practices. Appendix Table D1 presents descriptive statistics of CTs and their caretakers at baseline. In 80% of facilities, O&M is performed by a single caretaker; caretakers are generally male (82%), have roughly 10 years of experience in their job, 44% are part of the local community, and 27% are also cleaners. Caretakers allocate 68% of their time to managing the facility (collecting fees and supervising cleaners), and 32% to conducting repairs, cleaning the facility, or spending time with the manager.

To address self-reporting bias, survey data were supplemented with objective measurements. First, for each wave of survey data, independent observers collected information for the duration of one hour at dawn when there is higher user traffic. They recorded the number of users, and the cleanliness and maintenance status of the infrastructure in the CT. In addition, to measure free riding, they noted whether users payed the fee or not. In addition, as the lack of cleanliness in the CT can drive exposure to pathogens through direct contact with contaminated surfaces (Montgomery and Elimelech, 2007; Flores et al., 2011; WHO-UNICEF, 2017), observers collected bacterial swabs from randomly selected spots on the floor of the CTs. These were then analyzed in a laboratory to identify the presence and counts of bacteria. On average, more than three types of dangerous bacteria were found in each facility in each round. Further details are provided in Appendix C.3.

Survey and objective measurements were supplemented with an adapted dictator game played with caretakers to measure pro-social motivation for the cause. In each survey round, caretakers were provided with an endowment of INR 100 (US$ 1.42) with the option to donate all or part of it to a project improving access to water, sanitation and hygiene in disadvantaged areas of India, implemented by our NGO partner. Having collected the contributions among all caretakers, the total amount was donated. Similar versions of this game have proven effective at identifying socially motivated workers (Ashraf et al., 2014a; Lagarde and Blaauw, 2014).

Attrition of CTs was kept to a minimum between baseline and follow-up surveys. We were able to take observations and collect bacterial swabs from almost all CTs and surveyed caretakers in 92% of CTs in all five follow-up rounds. We find no differential attrition across treatment groups (Appendix Table C3).

3.2 Slum resident surveys

In conjunction with the baseline CT survey, a baseline survey was administered for slum residents. This survey covered a sample of 1,575 households living in the 110 selected catchment areas. The main respondent was the household's main decision-maker, being in most cases the household head and always falling in the age range 18–64 years. The questionnaire covered the household's socio-demographic characteristics, such as dwelling characteristics, assets and expenditures, the health status of family members, hygiene- and sanitation-related behavior and beliefs. Appendix Table D2 presents descriptive statistics for households at baseline. On average, household heads are 45 years old, mostly men, with primary education or less, and Hindu.

To follow the behavior of slum residents over time, the baseline survey was followed by three waves of follow-up surveys corresponding to the CT survey's follow-up 1, follow-up 3 and follow-up 5. All follow-up surveys covered hygiene- and sanitation-related behaviors, beliefs, attitudes and expectations. In addition, follow-up 3 and follow-up 5 collected data about the health status of family members, sanitation-related expenditures and updated details on demographic characteristics. In follow-up 5, to avoid stigma in the reporting of sanitation behavior, the survey was supplemented with a list randomization question, a technique that enhances the reporting of sensitive behavior (see, e.g., Karlan and Zinman, 2012; Treibich and Lépine, 2019).[12]

[12]Respondents were asked to report the number of true statements from either a list of statements on general

Each baseline household was interviewed on average in 2.6 out of 3 follow-up measurements, with only 2% of baseline households that was never re-interviewed at follow-up, and with attrition rates for individual follow-up surveys ranging from 9 to 19%. Attrition from baseline to follow-up surveys was addressed with replacements randomly selected from the sampling frame used for the baseline survey. While there is no differential attrition across treatment groups (Appendix Table C4), we show the robustness of our findings to the use of ANCOVA specifications and to correcting for attrition using inverse probability weights (see, e.g., Wooldridge, 2002; McKenzie, 2012).

Survey data were supplemented with incentive-compatible behavioral measurements. First, we elicited individual-level WTP for CT use during each wave of the slum resident survey.[13] Following extensive piloting and the low level of literacy in the sample, we opted for the incentivized version of the multiple price list (or take-it-or-leave-it) methodology (Andersen et al., 2006), which works better than the BDM method (Becker et al., 1964) in environments, like our setting, where market prices are well-known (Berry et al., 2020). The methodology prompts the participant to choose between different amounts of cash or a bundle of 10 tickets to use the CT in the catchment area where they live. One of the options is then randomly drawn and the decisions are realized. While the market value of 10 tickets is INR 50 (US$ 0.71), we offered different amounts of cash, ranging from INR 0 to 60 (US$ 0.85, above the current market price to deal with truncation) with increases of INR 5 (US$ 0.07). We define the WTP for CT use as the point at which the participant switches from preferring the bundle of tickets to preferring the cash. We divide this value by 10 to get WTP for a single use. WTP was elicited for both the most senior male and the most senior female decision-makers in the household, who are commonly spouses. The behavioral game was played with each member alone, without other senior members present, and households were revisited where necessary.

Second, similar to the game implemented with caretakers (Section 3.1), we played with all participants and in each survey round an adapted dictator game to measure preference for maintenance among slum residents. Slum residents were provided with an endowment of INR 50 (US$ 0.71) with the option to donate all or part of it to the purchase of cleaning products for the CT. Having collected all the contributions within each slum, the total amount was used to purchase cleaning products, which were then delivered to the caretaker.

Third, to measure willingness to cooperate among the slum residents, we implemented a standard public goods game (PGG) with the experiment participants. The game is based on the

behavior (short list) or the same list with an additional statement on OD, CT use or hand-washing with soap (long list). Allocation to different lists were randomized at individual level. The difference in the average number of true statements between the short and the long lists estimates the proportion engaging in the sensitive behavior. Appendix C.4.1 provides the lists of statements.

[13] Since an individual use is relatively cheap and highly recurrent, we do not focus on ability to pay. In the presence of new products and inability to pay, price subsidies have proven to be effective for adoption (Kremer and Miguel, 2007; Ashraf et al., 2010; Dupas, 2014). Ben Yishay et al. (2017) look specifically at credit constraints and WTP for the adoption of latrines.

voluntary contribution mechanism, in which participants receive an endowment of INR 100 (US$ 1.42), and they have to decide whether to keep the endowment or to invest part or all of it in a public pot. The contributions in the group are increased by a multiplier and shared equally among participants. The multiplier is randomly varied at catchment-area level to either double or triple the contributions. We played simultaneously with three groups of four or six participants in each community. As standard in the literature, we interpret contributions to the public pot as individual willingness to cooperate because the dominant strategy is not to contribute at all (Attanasio et al., 2009).

Finally, we measured incentivized demand for public intervention using an SCA. In this activity, labeled as the *voice-to-the-people* initiative, slum residents were provided with the opportunity to fill in an anonymous card to report the most pressing issue in their community from a set of pre-defined topics. In the list of potential issues, we introduced the cleanliness of the local CT to identify demand for public intervention with respect to the O&M of facilities. Respondents were informed that the content of these cards would be summarized and provided to their city's municipal corporation (or *Nagar Nigam*), i.e., the governmental institution responsible for community services in urban areas with more than 1 million inhabitants. We conducted this SCA after follow-up 3, and reported the summary of issues to municipal corporations at the end of the study. Similar activities were implemented to capture demand for political participation and accountability by Batista and Vicente (2011); Collier and Vicente (2014); Armand et al. (2020b). Details about this activity are provided in Appendix C.8.

4 Results

We report mean differences at baseline between the control group and the treatment groups for household- and CT-level characteristics in Appendix Tables D1 and D2. Randomization was successful in creating observationally equivalent groups in the experiment. We can therefore estimate treatment effects by restricting the sample to post-baseline observations and accounting for stratification. Because both treatment groups receive the maintenance intervention, we begin by pooling the two experimental arms and estimate treatment impacts of any intervention on the outcome Y_{ijt} of individual/household/CT i in catchment area j at time t using the following specification:

$$Y_{ijt} = \beta T_j + \alpha \mathbf{X}_{ij} + \delta_t + \epsilon_{ijt} \tag{1}$$

where T_j is an indicator variable equal to 1 if the catchment area j is in any treatment group (*maintenance* or *maintenance plus sensitization* groups), and 0 otherwise. $\mathbf{X}_{ij}$ are a set of indicator variables capturing randomization strata, while δ_t are survey round indicator variables. When the outcome of analysis is at individual or household level, the error term ϵ_{ijt} is assumed to be clustered by catchment area and data collection round. When the outcome of analysis is at CT level, the error term ϵ_{ijt} is assumed to be clustered by catchment area. In an alternative specification, we estimate equation (1) by distinguishing the effect of each intervention. Specifically, we estimate the following specification:

$$Y_{ijt} = \beta_1 T1_j + \beta_2 T2_j + \alpha \mathbf{X}_{ij} + \delta_t + \epsilon_{ijt} \tag{2}$$

where $T1_j$ is an indicator variable equal to 1 if the catchment area j is in the *maintenance* group and 0 otherwise, and $T2_j$ is an indicator variable equal to 1 if the catchment area j is in the *maintenance plus sensitization* group and 0 otherwise.

Results are robust to alternative specifications, such as using ANCOVA specifications or correcting for attrition using inverse probability weights (Appendix D.3), including control variables selected with the post-double selection LASSO procedure of Tibshirani (1996) and Belloni et al. (2013) and using the causal forest procedure of Wager and Athey (2018) and Athey et al. (2019) (Appendices D.4 and D.5). In addition, we show evidence against the presence of spillover effects across treatment arms (Appendix D.6). Finally, we present estimates of heterogeneous treatment effects based on machine learning techniques and on pre-specified dimensions in Appendices D.5 and D.7.

The parameters β in equation (1) and β_1 and β_2 in equation (2) identify the intent-to-treat (ITT) impact of each treatment. To estimate these parameters, as discussed in Section 3, we collected up to five follow-up measurements that can be used to estimate treatment effects (i.e., t ranges from 0 to 5, where 0 indicates the baseline measurement). We opted for multiple follow-up measurements because to understand dynamics, but also because we expected our outcomes of interest to be measured with a varying degree of noise and to present low serial correlation. Under these conditions, pooling multiple follow-up measurements allows the averaging out of noise in the outcome variables, and increases power (McKenzie, 2012). Following this strategy, we present main results in Tables 1–4 by pooling all available follow-up measurements.

In each table, Panel A presents estimates of treatment effects using equation (1), while Panel B shows estimates using equation (2). The hypotheses of interest are whether any intervention has an impact ($H_A\colon \beta = 0$), whether the *maintenance* treatment has an impact ($H_B\colon \beta_1 = 0$), whether the *maintenance plus sensitization* treatment has an impact ($H_C\colon \beta_2 = 0$) and whether the impact of the *maintenance* treatment is different from the impact of the *maintenance plus sensitization* treatment ($H_D\colon \beta_1 = \beta_2$). In Tables 1–4, we report coefficients and standard errors for hypotheses H_A–H_C. For statistical testing, we present both p-values for the significance of each individual coefficient and p-values adjusted for multiple hypothesis testing using the List et al. (2019) bootstrap-based procedure. In each table, separately for the hypotheses H_A, H_B, and H_C, we test that each hypothesis is jointly true for all outcomes considered.

Section 4.1 discusses the implementation of interventions, while the Sections 4.2–4.4 discuss the main results grouped by CT-level outcomes, caretakers' behavior, and slum residents' behavior. We supplement estimates of the treatment effects over the full study period with an analysis of how treatment effects vary over time. Aware of the timing of the different interventions (see Section 1), we pool follow-up measurements in two separate periods. First, the *grant period*, which includes only follow-up 1, is the period of the study in which only the grant is offered to caretakers. Second, the *incentive period*, which includes follow-up surveys 2–5, is the period that follows the grant period and in which financial incentives to caretakers are paid. Estimates for individual follow-up measurements, including the evolution of mean values of y_{it} in the control group, are presented in Appendix D.13.

4.1 Implementation of interventions across treatment groups

The successful implementation of interventions generated significant differences across treatment arms in terms of exposure to each component (Appendix B.3 shows detailed estimates for these differences). Transfers made to the CTs and to caretakers were significantly larger in the treatment groups as compared to the control group. The total transfer made to CTs includes the value of the initial grant to treated CTs, the subsidized use of tickets from the WTP game to both treated and control CTs, and the transfer of products donated by slum residents as part of the adapted dictator game to both treated and control CTs. During the whole study period, the program transferred, on average, a cumulative amount of INR 1,577 (US$ 22.39) to CTs in the control group, and INR 25,270 (US$ 358.84) to CTs in the treatment groups, 16 times larger than the amount transferred to control CTs. The total transfer made to caretakers includes the financial incentive provided in treated CTs and the amounts kept from each round of the adapted dictator game. During the whole study period, caretakers in the control group received on average INR 373 (US$ 5.30), while caretakers in the treatment groups received an additional INR 4,179 (US$ 59.34) or roughly 83% of their monthly salary. Transfers both to the CTs and to caretakers are not statistically significantly different across treatment arms.

Even though the exposure to water, sanitation and hygiene (WASH) campaigns is relatively high among slum residents, likely driven by the Governments SBM, the sensitization campaign was effective at reaching the target population. While 68% of participants in the control group report having been exposed to a WASH campaign before the interview, participants in treatment groups have a 7.3 percentage point higher probability of recalling exposure. We observe a significant effect on recall for all means used in the sensitization campaign, with the largest effect recorded for posters, for which we observe a 16 percentage point difference from the control group (Appendix B.3). Concerning voice messages, participants in the treatment groups have an exposure that is double as compared to the exposure in the control group (0.399 versus 0.188). These effects are driven by the maintenance plus sensitization treatment, while no significant effect is observed for the maintenance treatment.

4.2 CT-level outcomes

Table 1 presents estimates of treatment effects on CT-level outcomes. Column (1) focuses on physical maintenance, measured by an indicator variable equal to 1 if the CT received repairs and/or deep cleaning in the month previous to the visit, and 0 otherwise. Column (2) combines in an index indicators about the physical quality and cleanliness of the facility, as measured by observers, and about the lack of bacteria, as measured by laboratory tests. The outcome presented in Column (3) indicates whether the CT is of higher quality, defined as whether the quality index from Column (2) is above the 75$^{\text{th}}$ percentile of the sample distribution. Appendix C.10 presents details about the index and the individual components.

While the interventions did not lead to sustained increases in physical maintenance, they had consistent effects on the quality of the facility, shifting CTs towards the top of the sample distribution. On average, treatments resulted in an increase in quality of 5 percentage points, leading to a 12 percentage point increase on the share of facilities in the top quality group. The latter effect is robust to multiple hypothesis testing with a p-value of 0.04, while the effect on the quality index has a p-value of 0.13. Upgrades in quality are characterized by a push upwards of the whole distribution for treated CTs: a Kolmogorov-Smirnov test of the equality of the distributions in the control and in any treatment group is rejected at the 1% level of confidence (Appendix C.10.2). Upgrades are mainly driven by improvements in physical quality and in visible cleanliness, while no effect is observed for the lack of bacteria. Impacts on both physical maintenance and quality are not statistically different across treatment arms.

Panels A and B in Figure 2 show how these impacts differ over time. The interventions were successful in increasing physical maintenance during the grant scheme, but these effects are limited to the short term and vanish during the incentive period. This suggests that improvements in physical maintenance are strongly tied to external funding. Physical maintenance does not translate immediately into an improvement in the quality of the facility, for which we do not observe any significant effect during the grant period. In contrast, quality significantly improves during the incentive period.

Outcomes related to *traffic during rush hour* use data from observers. Column (4) presents impacts on the number of users (in logarithms), while Column (5) reports impacts on free riding, measured as the percentage of users who do not pay while using the CT. We do not observe any significant effect on the number of users during rush hour, although a 10% reduction is observed for the maintenance treatment, significant at the 10% level. No particular pattern is observed over time for this variable (Appendix Figure D7). Free riding reduces by 8 percentage points on average, not statistically different between treatment arms. While this effect is not robust to multiple hypothesis testing in the full study period (corrected p-value is equal to 0.19), it is robust if we restrict the sample to the incentive period. In fact, the effect on free riding is driven by a significant reduction in free riding during this period only (Panel C in Figure 2).

In addition, estimates of treatment effects using machine learning techniques present lower p-values as compared to the OLS estimate presented in Table 1 (Appendices D.4 and D.5). Finally, while costs and revenues at the CT level are not observable to the researcher, we make use of observers' data on users and payments to estimate the effect of the interventions on revenues during rush hour. We observe only a small increase, not statistically significant (Appendix B.5).

Overall, the effects on both the quality of the facility and free riding are highly homogeneous. This is confirmed by an analysis of heterogeneous treatment effects based on machine learning and on pre-specified dimensions (Appendices D.5 and D.7).

4.3 Caretakers' behavior

CT-level results indicate an effect of the interventions especially in the presence of financial incentives for caretakers. In addition, no significant differential impact is recorded for the sensitization campaign, suggesting that observed improvements are mostly supply-side driven. To further understand these results, we study caretakers' behavior in Table 2. Column (1) focuses on the number of hours worked daily by the caretaker, while Column (2) analyses the share of time allocated to management activities, which include collecting fees and supervising cleaners. The remaining share is spent conducting repairs, cleaning the facility, or spending time with the manager. Focusing on routine maintenance, Column (3) captures the caretaker's awareness, measured through an indicator variable equal to 1 if the caretaker knows the recommended practices of the cleaning routine and the need for deep cleaning, and 0 otherwise, and Column (4) focuses instead on the inputs used in routine maintenance, measured with the number of tools, equipment and cleaners used during the last routine cleaning of the facility (details of the individual components are presented in Appendix D.10). Finally, Column (5) captures the caretaker's pro-social motivation, through the share of the endowment that is donated in the adapted dictator game (Appendix C.6).

Caretakers work on average 12 hours a day and allocate the vast majority of their time to managing the facility (69%). Interventions did not translate into a significant change in the number of hours worked. This is likely driven by labor supply being closely aligned with the opening times of the facility. However, caretakers significantly change their time allocation in response to the interventions. In the treatment groups, the share of time spent managing the facility increases by 5.2 percentage points, an effect that is observed only during the incentive period (Panel D in Figure 2). In the whole study period, this effect is significant at the 7% level, but only at the 21% level when correcting for multiple hypothesis testing. The increase in the time allocated to managing the facility, together with a joint reduction in free riding and in the number of users (Section 4.2), indicates that the caretaker likely implemented user selection at the entrance, enforcing payments among users and excluding potential users not willing to pay. For CTs that experienced high free riding at baseline, this finding is confirmed by a larger share of slum residents reporting that the caretaker refused them entry to the CT because they were not willing to pay the user fee (Appendix D.12).

Interventions also more than double the share of caretakers who are aware of recommended cleaning practices, with awareness increasing by 9 percentage points, as compared with an average share of 7% in the control group. This effect is robust to multiple hypothesis testing at the 10% level. The shift in time management towards fee collection and supervision of cleaners and an increased awareness of cleanliness are in line with the simultaneous increase in the quality of the facility and reduction in free riding in the incentive period (Section 4.2). These effects are not driven by changes in caretakers' pro-social motivation. On average, caretakers donate 35% of the transfer each time the game is played. Finally, none of the impacts on caretakers' behavior differs significantly by treatment arm, reinforcing the finding that the sensitization campaign among slum residents had no effect on supply-side outcomes.

4.4 Slum residents' behavior

We turn our attention to the demand side by exploring slum residents' behavior. Table 3 focuses on the valuation of CT use and on attitudes towards the public good among slum residents. Column (1) focuses on the incentivized WTP for a single CT use, reported in INR (see Section 3 for details about the measurement), while Column (2) reports impacts on whether respondents perceived any improvement in the CT infrastructure. Column (3) looks at the demand for public intervention in the O&M of CTs, measured using an indicator variable equal to 1 if the household asks for it during the voice-to-the-people SCA (Appendix C.8), and 0 otherwise. Column (4) shows impacts on the share contributed to the public pot in the public goods game, while Column (5) looks at preference for maintenance, measured as the share of the endowment that is retained by the respondent in the adapted dictator game.

Slum residents are willing to pay just INR 1.15 for a single CT use, compared with the market price of INR 5. Over the whole study period, interventions have on average no significant effect on the WTP, even though we observe a significant increase of 3 percentage points in the share of respondents who perceived an improvement in the CT over and above a share of 15% in the control group. However, effects on private valuation remains small, and effects are not robust to multiple hypothesis testing. When looking at the effect of different treatment arms, we observe a marginally significant reduction of INR 0.12 in the WTP in the maintenance plus sensitization treatment. This effect is statistically different from the effect in the maintenance treatment. Looking at how these effects evolve over time, WTP experiences a marginally significant reduction during the grant period, more pronounced for those that were subject to the sensitization campaign, while no significant effect is observed in the incentive period (Panel F in Figure 2). Figure 3 shows estimates of the inverse demand curve for CT use separately for the control and treatment groups. During the grant period (Panel A), the interventions shift the demand curve downwards as compared with the control group, with statistically significant differences concentrated at the lowest prices. This statistically significant effect is not observed at the upper end of the inverse demand curve, nor during the incentive period (Panel B). This pattern of results is driven by male residents and is not observed for female respondents (Appendix D.8). In terms of the price elasticity of demand for CT use, we do not observe any significant effect of treatments in either the grant or the incentive period (Appendix D.9).

Externally funding maintenance that is positively perceived by potential users has the capacity to crowd out users' private contributions. This conclusion is evident when looking at the demand for public intervention. The share of households reporting to local politicians that the O&M of CTs is the most pressing issue in their community during the voice-to-the-people initiative is 5.3 percentage points larger in the treatment groups than the 9.6% of respondents who report the same issue in the control group (impact estimates on other topics are shown in Appendix D.11). This represents an increase of 55% in the demand for public intervention in the O&M of CTs, and is comparable for both treatment arms.[14] The effect is significant at the 4% level, and at the 20% level when correcting for multiple hypothesis testing. In addition, this effect is mainly driven by households that have lower preferences for maintenance (Appendix D.7). The increase in the demand for public intervention is in line with the crowding out effect of public investments evidenced in both advanced economies and low- and middle-income countries (Peltzman, 1973; Cutler and Gruber, 1996; Bennett, 2012; Das et al., 2013; Armand et al., 2017). These effects are not accompanied by a change in preference for maintenance or in cooperative behavior. On average, slum residents in both treatment and control areas donate 21% of the transfer in the adapted dictator game and contribute 17%, of the endowment to the public pot in the public goods game.

To understand whether the interventions also impact hygiene- and sanitation-related behavior, Figure 4 shows the share of slum residents practicing OD, using CTs, and washing hands with soap at follow-up 5 estimated using the list randomization technique (see Appendix C.4.1 for details about measurement). In the control group, 21% of respondents practice OD, 59% use the CT and 82% wash their hands with soap. On average, the interventions significantly increase OD practice by 18%, while we do not observe a significant difference in CT use or in hand-washing with soap. These results are compatible with the selection of users who can enter the facility (Section 4.3). Similar to the effect on WTP, the effect on OD is driven by male respondents, who are also the category for which payment enforcement is higher (Appendix D.8).[15]

[14] The maintenance intervention was perceived by slum residents as externally funded. At follow-up 5, the majority of respondents believe that the improvement in the local facility had been funded by the government (65%).

[15] While not statistically significant, we observe a slight increase in CT use for the treatment arms. While the list randomization technique does not allow us to identify individual behavior, this result is possible as interventions might have led in aggregate to a shift away from CT use and into OD, balanced by a shift away from unimproved (private) sanitation to CT use. In addition, users could have practiced both (slum residents defecate 1-2 times a day on average). We do not observe any significant impacts of the intervention on self-reported sanitation behavior, which under-report OD as compared to the list randomization technique (Appendix Table D3).

We then turn our attention to the awareness of risky behavior and health (Table 4). Columns (1) and (2) refer to whether the respondent is aware of health and safety risks associated with unsafe sanitation, and of the health externalities associated with OD. Column (3) refers to the self-reported morbidity, measured with an indicator variable equal to 1 if any household member had fever, diarrhea or cough during the two weeks prior to the interview, and 0 otherwise. Columns (4) and (5) focus on the extensive and intensive margins of household health expenditures.

Slum residents are highly aware of risks associated with unsafe sanitation. Eighty-four percent of households in the control group are aware of health and safety risks associated with OD and with using low-quality CTs, while 66% are aware of health externalities that OD generate for their family. While in previous sections we observe limited effects of the sensitization campaign for both slum residents' and caretakers' behaviors, the sensitization campaign was effective at raising awareness of risks among slum residents. As compared with the control group, participants in the maintenance plus sensitization treatment have a 1.7 percentage point higher probability of being aware of health and safety risks, and a 4.9 percentage point higher probability of being aware of health externalities associated with OD. Both effects are robust to multiple hypothesis testing at the 3% and 6% level respectively. The effect of the maintenance treatment is insignificant, while the difference between the two treatment arms is significant at the 11% level of confidence for the awareness of health and safety risks, and at the 6% level of confidence for awareness of externalities.

Focusing on health, the high-disease environment that characterizes slums is confirmed by a large share of residents reporting positive expenditure on health (64%). While we do not observe any significant effect on morbidity, the interventions cause an increase in health expenditures on the extensive margin by 5 percentage points. This effect is mainly driven by the maintenance plus sensitization treatment, whose treatment impact is robust to multiple hypothesis testing. These effects are not found on the intensive margin. For all health-related variables, we do not find any significant difference across treatment arms. In terms of timing, we observe that the effect on expenditure is driven by a significant increase in the first follow-up of the incentive period, confirmed by corresponding significant increase in morbidity (Appendix Figure D9).

5 Should sanitation in slums be fully subsidized?

A long-run objective of any government is to provide universal access to private toilets (WHO-UNICEF, 2017). Evidence showing that property rights and housing improvements raise investments and household well-being supports this objective (Galiani et al., 2017; Field, 2005). India's large governmental policy, the *SBM*, aims at reaching this target by fully subsidizing the construction of private latrines. In urban areas, the subsidy per household varies from INR 15,000 to 18,000 (US$ 213–256) (Government of India, 2017). However, space constraints, lack of technology and skills to implement existing solutions, lack of legal status among slum residents, and constrained eligibility for the subsidy remain important limitations of this strategy in urban areas. These constraints are officially acknowledged by the Indian Government (Government of India, 2017). Furthermore, since a large part of the subsidy is provided only post-construction, liquidity constraints remain a hurdle (Augsburg et al., 2019). Larger-scale investments might therefore be needed in order to guarantee an adequate level of basic services in slums, potentially requiring the relocation of slum residents, which has been shown to create limited benefits for residents (Barnhardt et al., 2017).

Given these limitations, CTs will remain an important solution in the foreseeable future and we need to consider whether day-to-day running and upkeep of this infrastructure service should be provided as fully subsidized by the government through general taxation with free access for residents, or under the current model based on recovery tariffs and funded by user fees.[16] To understand the sustainability of each model, we analyze the monthly cost of O&M for a CT with median characteristics in the status quo scenario, characterized by poor maintenance. We then consider two alternative scenarios that complement the status quo with increased maintenance: the first, a hypothetical scenario, increases the amounts available under the different budget lines of the status quo scenario (e.g., increasing the number of cleaners); the second complements the status quo scenario with the maintenance intervention implemented as part of the study. Assumptions and details are provided in Appendix B.4. The monthly cost in the status quo scenario is estimated at INR 10,200 (US$ 144.85) per CT. Providing additional inputs to O&M more than double these costs to INR 28,800 (US$ 408.97), while adding the maintenance intervention generates a total monthly cost of INR 13,544 (US$ 192.33) per CT. The cost of these scenarios are 2.8 and 1.3 times the cost of O&M in the status quo.

In a model financed by user fees, an eligible household has the potential to provide monthly revenues of at least INR 600 if all members over 5 years of age use the CT once per day and pay the market fee of INR 5. Without considering the use of the CT by non-eligible slum residents, these revenues would cover the current O&M costs plus adding the maintenance intervention, and would cover 71% the costs of the hypothetical scenario.[17] As free riding is not zero, these estimates represent an upper bound, and they would mean some slum residents practicing OD where payments are enforced.[18]

While reducing free riding favors financial sustainability, coordination does not seem feasible in the slum setting. Reductions in free riding achieved by incentivizing maintenance come at the cost of user selection, without large increases in revenues, but with an increase in the share of

[16] We do not consider the funding for future system upgrades in this analysis.

[17] For urban areas of Punjab (India), Sridhar (2007) estimates that service revenues for the provision of water and sewerage cover approximately 50% of the total expenditure.

[18] A different model involves the use of monthly family passes, which provide unlimited entry to the CT for all family members at a fixed price. Eight percent of CTs in our sample were providing monthly passes, with a median price of INR 80 (US$ 1.14). The Government of India is considering the introduction of a monthly pass at the price of INR 200 (US$ 2.84). While this could favor reductions in free riding, monthly passes would cover a lower share of O&M costs and could introduce nonpayments for liquidity constrained households. At follow-up 5, we elicit WTP for family passes in an incentivized setting and we observe very low valuations, at INR 25 (US$ 0.35) per household.

residents practicing OD. We confirm the constraints to coordination using mediation analysis, following Gelbach (2016). Figure 5 shows the decomposition of the ITT effect of any of our treatments on free riding during the incentive period (equivalent to -9.3 percentage points) into supply- and demand-side mediators. While supply-side mediators mainly decrease free riding (upper bar), demand-side factors predominately increase it (middle bar). Combining demand and supply mediators (lower bar), the caretakers labor supply contributes most to the reduction in free riding (11.7% of the main effect), followed by the improvement in maintenance and quality (8.9%), and routine maintenance (4.9%). Slum residents awareness of unsafe sanitation is the strongest mediator increasing free riding (11.1%), and also attitudes towards the public good pull in the same direction (2.8%). Only morbidity and health expenditures are demand side factors that reduce free riding (4.6%). A large share of the impact on free riding remains unexplained, possibly due to unmeasured mediators, or due to the fact that free riding was measured only during 1 hours, and counted all users, not only those also captured in our sample. The effects of the interventions on free riding are further highly homogeneous, suggesting that the difficulties in coordination are at play across CTs (Appendices D.5 and D.7).

Assuming no slum resident would voluntarily practice OD over the use of a (potentially better-maintained) CT, the government could consider providing CTs for free to slum residents, especially in light of the large health externalities of OD and the poor health conditions of slum residents. This model of free provision to residents is not only closely associated with the general ability of the local government to raise tax revenues (see, e.g., Besley and Persson, 2013), but also with its ability to redistribute them towards categories that have very little political representation (Fox, 2014). At the same time, providing CTs for free to all residents can disincentivize private sanitation, and can have negative effects in terms of overcrowding. In our study area, free-to-use facilities and facilities with higher free riding (which mimic a free CT) were also found in degraded status (Armand et al., 2020a). As the prevalence of private access to sanitation is low and coordination fails in the presence of overcrowding (Banerjee et al., 2008b; Chidambaram, 2020), a model with fully subsidized CTs should consider imposing restrictions on the number of households that can use a CT and/or enacting additional monitoring mechanisms to ensure that facilities are preserved by users.

6 Conclusion

Understanding the mechanisms leading to the poor maintenance trap in LMICs is fundamental to unleashing the economic development of cities and increasing the well-being of citizens. We explore how to break a vicious cycle found in the context of public infrastructure in India, where coordination failure leads to an extremely poor level of maintenance of the facilities. We provide novel insights by studying both supply- and demand-side incentives. Externally incentivizing maintenance does not favor collective action. It does allow to achieve sustained improvements in the observed quality of facilities and significant reductions in free riding among users, but these come at the cost of user selection, with subsequent increases in OD among slum residents and in demand to local politicians for public intervention. Supporting external funding for the maintenance of the local facility with an intense sensitization campaign among slum residents was ineffective at improving coordination between slum residents and caretakers.

These results provide three major implications for the provision of public services in low-income areas, and open new questions for future research. First, in the presence of strong constraints to behavior change among residents, policies integrating user-provider dynamics have limited effectiveness, with fundamental consequences for the design of policies aiming to achieve a higher-quality infrastructure. If citizens treat access to basic services as a right, then standard market mechanisms of public service delivery fail. As the international community aims to achieve universal access to safe and affordable basic services as part of the Sustainable Development Goals (United Nations, 2020), governments need to consider these constraints when stimulating infrastructure investments. More broadly, it is important to understand which cost-effective policies can achieve sustainable improvements in environmental quality in LMICs by achieving collective action. In overcrowded and lower income areas, where stimulating private investments is not feasible and where health externalities have severe consequences, enforcing financial sustainability through threats of disconnection might not be socially desirable (see, e.g., Coville et al., 2020).

Second, while the supply side is more responsive to incentives, social norms or other constraints limiting behavior change among slum residents make it hard to achieve collective action. There is a need to investigate how to change social norms that make free riding acceptable, while not deterring safe behavior. In this urban setting, an intense information campaign proved to have limited effectiveness as compared with supply-side interventions. Alternative mechanisms, such as Community-Led Total Sanitation (CLTS) Campaigns, which use psycho-social levers of shame and disgust and appoint local monitoring committees, have proven to be effective in improving sanitation behavior but focus on rural areas (Cameron et al., 2019; Abramovsky et al., 2019). There is scope to understand whether an approach adapted to urban areas could be effective at stimulating coordination between slum residents and public service providers.

Finally, further research is needed to understand the effectiveness of monitoring technologies in low coordination environments. Incentive schemes at the community level could be effective at creating and reinforcing a new local norm of payment and respect for the public good (Neal et al., 2016). Newer technologies should aim to guarantee monitoring without fully excluding users and limiting bureaucratic collusion and extortion (Banerjee et al., 2008a; Jack et al., 2015; Ashraf et al., 2016).

References

Abramovsky, L., Augsburg, B., Lührmann, M., Oteiza, F., and Rud, J. P. (2019). Community Matters: Heterogeneous Impacts of a Sanitation Intervention. IFS working paper, Institute for Fiscal Studies, London, United Kingdom.

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, 9(4):383–405.

Andres, L., Deb, S., Joseph, G., Larenas, M. I., and Grabinsky Zabludovsky, J. (2020). A Multiple-Arm, Cluster-Randomized Impact Evaluation of the Clean India (Swachh Bharat) Mission Program in Rural Punjab, India. World Bank Policy Research working paper 9249, World Bank.

Armand, A., Augsburg, B., and Bancalari, A. (2018). Community Toilet Use in Slums: Willingness to Pay and the Role of Informational and Supply Side Constraints. AEA RCT Registry. June 20. https://doi.org/10.1257/rct.3087-1.0.

Armand, A., Augsburg, B., Bancalari, A., and Trivedi, B. (2020a). Community toilet use in slums: Willingness-to-pay and the role of informational and supply side constraints. *3ie Impact Evaluation Report 113*.

Armand, A., Carneiro, P., Locatelli, A., Mihreteab, S., and Keating, J. (2017). Do public health interventions crowd out private health investments? Malaria control policies in Eritrea. *Labour Economics*, 45:107–115.

Armand, A., Coutts, A., Vicente, P. C., and Vilela, I. (2020b). Does information break the political resource curse? Experimental evidence from Mozambique. *American Economic Review*, 110(11):3431–3453.

Asher, S. and Novosad, P. (2020). Rural roads and local economic development. *American Economic Review*, 110(3):797–823.

Ashraf, N., Bandiera, O., and Jack, B. K. (2014a). No margin, no mission? A field experiment on incentives for public service delivery. *Journal of Public Economics*, 120:1–17.

Ashraf, N., Bandiera, O., and Lee, S. S. (2014b). Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior and Organization*, 100:44–63.

Ashraf, N., Berry, J., and Shapiro, J. M. (2010). Can higher prices stimulate product use? Evidence from a field experiment in Zambia. *American Economic Review*, 100(5):2383–2413.

Ashraf, N., Glaeser, E. L., and Ponzetto, G. A. M. (2016). Infrastructure, Incentives, and Institutions. *American Economic Review*, 106(5):77–82.

Ashraf, N., Jack, B. K., and Kamenica, E. (2013). Information and subsidies: Complements or substitutes? *Journal of Economic Behavior and Organization*, 88:133–139.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *Annals of Statistics*, 47(2):1179–1203.

Attanasio, O., Pellerano, L., and Reyes, S. P. (2009). Building trust? Conditional cash transfer programmes and social capital. *Fiscal Studies*, 30(2):139–177.

Augsburg, B., Caeyers, B., Giunti, S., Malde, B., and Smets, S. (2021). Labelled loans and human capital investments. IFS working paper W21/09, Institute for Fiscal Studies, London, United Kingdom.

Augsburg, B., Caeyers, B., and Malde, B. (2019). *Can Micro-Credit Support Public Health Subsidy Programs?* World Bank.

Augsburg, B. and Rodríguez-Lesmes, P. A. (2018). Sanitation and child health in India. *World Development*, 107:22–39.

Balboni, C., Bandiera, O., Burgess, R., Ghatak, M., and Heil, A. (2020). Why Do People Stay Poor? CEPR working paper, Centre for Economic Policy Research.

Bancalari, A. (2020). Can white elephants kill? Unintended consequences of infrastructure development in Peru.

Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I., and Sulaiman, M. (2017). Labor markets and poverty in village economies. *Quarterly Journal of Economics*, 132(2):811–870.

Bandiera, O., Larcinese, V., and Rasul, I. (2015). Blissful ignorance? A natural experiment on the effect of feedback on students' performance. *Labour Economics*, 34:13–25.

Banerjee, A., Duflo, E., Glennerster, R., and Duflo, E. (2008a). Putting a band-aid on a corpse: Incentives for nurses in the Indian public health care system. *Journal of the European Economic Association*, 6(2-3):487–500.

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Pariente, W., Shapiro, J., Thuysbaert, B., and Udry, C. (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, 348(6236):1260799.

Banerjee, A., Duflo, E., and Qian, N. (2020a). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics*, 145(102442).

Banerjee, A., Enevoldsen, N. T., Pande, R., and Walton, M. (2020b). Public information is an incentive for politicians: Experimental evidence from Delhi elections. NBER working paper 26925, National Bureau of Economic Research.

Banerjee, A., Iyer, L., and Somanathan, R. (2008b). Public action for public goods. *Handbook of Development Economics*, 4:3117–3154.

Barnhardt, S., Field, E., and Pande, R. (2017). Moving to opportunity or isolation? Network effects of a randomized housing lottery in urban India. *American Economic Journal: Applied Economics*, 9(1):1–32.

Bartram, J., Lewis, K., Lenton, R., and Wright, A. (2005). Focusing on improved water and sanitation for health. *Lancet*, 365(9461):810–812.

Basinga, P., Gertler, P. J., Binagwaho, A., Soucat, A. L., Sturdy, J., and Vermeersch, C. M. (2011). Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: An impact evaluation. *The Lancet*, 377(9775):1421–1428.

Batista, C. and Vicente, P. C. (2011). Do migrants improve governance at home? Evidence from a voting experiment. *World Bank Economic Review*, 25(1):77–104.

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

Belloni, A., Chernozhukov, V., and Hansen, C. (2013). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2):608–650.

Ben Yishay, A., Fraker, A., Guiteras, R., Palloni, G., Shah, N. B., Shirrell, S., and Wang, P. (2017). Microcredit and willingness to pay for environmental quality: Evidence from a randomized-controlled trial of finance for sanitation in rural Cambodia. *Journal of Environmental Economics and Management*, 86:121–140.

Bénabou, R. and Tirole, J. (2002). Self-confidence and personal motivation. *Quarterly Journal of Economics*, (1):871–914.

Bénabou, R. and Tirole, J. (2003). Intrinsic and extrinsic motivation. *Review of Economic Studies*, 70:489–520.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678.

Bennett, D. (2012). Does clean water make you dirty? Water supply and sanitation in the Philippines. *Journal of Human Resources*, 47(1):146–173.

Berry, J., Fischer, G., and Guiteras, R. (2020). Eliciting and utilizing willingness to pay: Evidence from field trials in Northern Ghana. *Journal of Political Economy*, 128(4):1436–1473.

Besley, T. and Ghatak, M. (2018). Prosocial motivation and incentives. *Annual Review of Economics*, 10:411–438.

Besley, T. and Persson, T. (2013). Taxation and development. *Handbook of Public Economics*, 5:51–110.

Biran, A., Jenkins, M. W., Dabrase, P., and Bhagwat, I. (2011). Patterns and determinants of communal latrine usage in urban poverty pockets in Bhopal, India. *Tropical Medicine and International Health*, 16(7):854–862.

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

Briceño, B., Coville, A., Gertler, P., and Martinez, S. (2017). Are there synergies from combining hygiene and sanitation promotion campaigns: Evidence from a large-scale cluster-randomized trial in rural Tanzania. *PLoS ONE*, 12(11):e0186228.

Bryan, G., Glaeser, E., and Tsivanidis, N. (2020). Cities in the developing world. *Annual Review of Economics*, 12:273–297.

Burgess, R., Greenstone, M., Ryan, N., and Sudarshan, A. (2020). The consequences of treating electricity as a right. *Journal of Economic Perspectives*, 34(1):145–169.

Burgess, R., Jedwab, R., Miguel, E., Morjaria, A., and Padró I Miquel, G. (2015). The value of democracy: Evidence from road building in Kenya. *American Economic Review*, 105(6):1817–1851.

Cameron, L., Olivia, S., and Shah, M. (2019). Scaling up sanitation: Evidence from an RCT in Indonesia. *Journal of Development Economics*, 138:1–16.

Casey, K., Glennerster, R., and Miguel, E. (2012). Reshaping institutions: Evidence on aid impacts using a preanalysis plan. *Quarterly Journal of Economics*, 127(4):1755–1812.

Charness, G. and Fehr, E. (2015). From the lab to the real world. *Science*, 350(6260):512–513.

Chidambaram, S. (2020). How do institutions and infrastructure affect mobilization around public toilets vs. piped water? Examining intra-slum patterns of collective action in Delhi, India. *World Development*, 132:104984.

Collier, P. and Vicente, P. C. (2014). Votes and Violence: Evidence for a Field Experiment in Nigeria. *The Economic Journal*, 124(574):327–355.

Cortes, K. E., Fricke, H., Loeb, S., Song, D. S., and York, B. N. (2021). Too little or too much? Actionable advice in an early-childhood text messaging experiment. *Education Finance and Policy*, 16(2):209–232.

Coville, A., Galiani, S., Gertler, P. J., and Yoshida, S. (2020). Pipe Dreams: Enforcing Payment for Water and Sanitation Services in Nairobi's Slums. NBER working paper 27569, National Bureau of Economic Research.

Croxson, B., Propper, C., and Perkins, A. (2001). Do doctors respond to financial incentives? UK family doctors and the GP fundholder scheme. *Journal of Public Economics*, 79(2):375–398.

Currie, J. and Vogl, T. (2013). Early-life health and adult circumstance in developing countries. *Annual Review of Economics*, 5(1):1–36.

Cutler, D. M. and Gruber, J. (1996). Does public insurance crowd out private insurance? *Quarterly Journal of Economics*, 111(2):391–430.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., and Sundararaman, V. (2013). School inputs, household substitution, and test score. *American Economic Journal: Applied Economics*, 5(2):29–57.

Devoto, F., Duflo, E., Dupas, P., Parienté, W., and Pons, V. (2012). Happiness on tap: Piped water adoption in urban Morocco. *American Economic Journal: Economic Policy*, 4(4):68–99.

Dinkelman, T. (2011). The effects of rural electrification on employment: New evidence from South Africa. *American Economic Review*, 101(7):3078–3108.

Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4–5):899–934.

Duflo, E., Galiani, S., and Mobarak, A. M. (2012). Improving access to urban services for the poor: open issues and a framework for a future research agenda. *J-PAL Urban Services Review Paper. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab*.

Duflo, E., Greenstone, M., Guiteras, R. P., and Clasen, T. F. (2015). Toilets can work: short and medium run health impacts of addressing complementarities and externalities in water and sanitation. NBER working paper 21521, National Bureau of Economic Research.

Duflo, E. and Pande, R. (2007). Dams. *Quarterly Journal of Economics*, 122(2):601–646.

Dupas, P. (2011). Health behavior in developing countries. *Annual Review of Economics*, 3:425–449.

Dupas, P. (2014). Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment. *Econometrica*, 82(1):197–228.

Dupas, P. and Miguel, E. (2017). Impacts and determinants of health levels in low-income countries. *Handbook of Economic Field Experiments*, 2:3–93.

Ezeh, A., Oyebode, O., Satterthwaite, D., Chen, Y. F., Ndugwa, R., Sartori, J., Mberu, B., Melendez-Torres, G. J., Haregu, T., Watson, S. I., Caiaffa, W., Capon, A., and Lilford, R. J. (2017). The history, geography, and sociology of slums and the health problems of people who live in slums. *The Lancet*, 389(10068):547–558.

Ferraz, C. and Finan, F. (2008). Exposing corrupt politicians: The effects of Brazil's publicly released audits on electoral outcomes. *Quarterly Journal of Economics*, 123(2):703–745.

Ferraz, C. and Finan, F. (2011). Electoral accountability and corruption: Evidence from the audits of local governments. *American Economic Review*, 101(4):1274–1311.

Field, E. (2005). Property rights and investment in urban slums. *Journal of the European Economic Association*, 3(2–3):279–290.

Flores, G. E., Bates, S. T., Knights, D., Lauber, C. L., Stombaugh, J., Knight, R., and Fierer, N. (2011). Microbial biogeography of public restroom surfaces. *PLoS ONE*, 6(11):1–7.

Fox, S. (2014). The political economy of slums: Theory and evidence from sub-Saharan Africa. *World Development*, 54:191–203.

G20 (2017). Global Infrastructure Outlook – Infrastructure investment needs in 50 countries, 7 sectors to 2040. *A G20 Initiative. Sydney: Global Infrastructure Hub*.

Galiani, S., Gertler, P. J., Undurraga, R., Cooper, R., Martínez, S., and Ross, A. (2017). Shelter from the storm: Upgrading housing infrastructure in Latin American slums. *Journal of Urban Economics*, 98:187–213.

Gelbach, J. B. (2016). When do covariates matter? And which ones, and how much? *Journal of Labor Economics*, 34(2):509–543.

Geruso, M. and Spears, D. (2018). Neighborhood sanitation and infant mortality. *American Economic Journal: Applied Economics*, 10(2):125–162.

Ghatak, M. (2015). Theories of poverty traps and anti-poverty policies. *World Bank Economic Review*, 29(suppl_1):S77–S105.

Gneezy, U. and Imas, A. (2017). Lab in the field: Measuring preferences in the wild. In *Handbook of Economic Field Experiments*, volume 1, pages 439–464.

Gonzalez-Navarro, M. and Quintana-Domeque, C. (2016). Paving streets for the poor: Experimental analysis of infrastructure effects. *Review of Economics and Statistics*, 98(2):254–267.

Government of India (2011). Census of India 2011. Office of the Registrar General and Census Commissioner, Ministry of Home Affairs, Government of India, New Delhi, India.

Government of India (2017). Guidelines for Swachh Bharat Mission – Urban. Swachh Bharat Mission, Ministry of Housing and Urban Affairs, Government of India.

Government of India (2018). Advisory on Public and Community Toilets. Ministry of Housing and Urban Affairs, Central Public Health and Environmental Engineering Organisation (CPHEEO).

Greenstone, M. and Jack, B. K. (2015). Envirodevonomics: A research agenda for an emerging field. *Journal of Economic Literature*, 53(1):5–42.

Guiteras, R. P., Levinsohn, J., and Mobarak, A. M. (2015). Encouraging sanitation investment in the developing world: A cluster-randomized trial. *Science*, 348(6237):903–906.

Hallegatte, S., Rentschler, J., and Rozenberg, J. (2019). *Lifelines: The Resilient Infrastructure Opportunity*. World Bank, Washington, DC.

Haushofer, J. and Shapiro, J. (2016). The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya. *Quarterly Journal of Economics*, 131(4):1973–2042.

IMF (2020). Official Exchange Rate (LCU per US$, period average). Washington, D.C.: International Monetary Fund, International Financial Statistics.

Jack, B. K., Smith, G., Kelsey Jack, B., and Smith, G. (2015). Pay as you go: Prepaid metering and electricity expenditures in South Africa. *American Economic Review: Papers & Proceedings*, 105(5):237–241.

Jalan, J. and Somanathan, E. (2008). The importance of being informed: Experimental evidence on demand for environmental quality. *Journal of Development Economics*, 87(1):14–28.

Jensen, R. (2007). The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector. *Quarterly Journal of Economics*, 122(3):879–924.

Karlan, D. S. and Zinman, J. (2012). List randomization for sensitive behavior: An application for measuring use of loan proceeds. *Journal of Development Economics*, 98(1):71–75.

Koszegi, B. and Szeidl, A. (2013). A model of focusing in economic choice. *Quarterly Journal of Economics*, 128(1):53–104.

Kremer, M., Leino, J., Miguel, E., and Peterson Zwane, A. (2011). Spring cleaning: Rural water impacts, valuation, and property rights institutions. *Quarterly Journal of Economics*, 126(1):145–205.

Kremer, M. and Miguel, E. (2007). The illusion of sustainability. *Quarterly Journal of Economics*, 122(3):1007–1065.

Lagarde, M. and Blaauw, D. (2014). Pro-social preferences and self-selection into jobs: Evidence from South African nurses. *Journal of Economic Behavior and Organization*, 107(PA):136–152.

Lee, K., Miguel, E., and Wolfram, C. (2020). Experimental evidence on the economics of rural electrification. *Journal of Political Economy*, 128(4):1523–1565.

Lewis-Faupel, S., Neggers, Y., Olken, B. A., and Pande, R. (2016). Can electronic procurement improve infrastructure provision? Evidence from public works in India and Indonesia. *American Economic Journal: Economic Policy*, 8(3):258–283.

Lipscomb, M., Mobarak, A. M., and Barham, T. (2013). Development effects of electrification: Evidence from the topographic placement of hydropower plants in Brazil. *American Economic Journal: Applied Economics*, 5(2):200–231.

Lipscomb, M. and Schechter, L. (2018). Subsidies versus mental accounting nudges: Harnessing mobile payment systems to improve sanitation. *Journal of Development Economics*, 135:235–254.

List, J. A., Shaikh, A. M., and Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, 22(4):773–793.

Luoto, J. E., Levine, D. I., and Albert, J. (2011). Information and Persuasion: Achieving Safe Water Behaviors in Kenya. RAND working paper WR-885, RAND Corporation.

Marx, B., Stoker, T. M., and Suri, T. (2013). The economics of slums in the developing world. *Journal of Economic Perspectives*, 27(4):187–210.

Marx, B., Stoker, T. M., and Suri, T. (2019). There is no free house: Ethnic patronage in a Kenyan slum. *American Economic Journal: Applied Economics*, 11(4):36–70.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, 99(2):210–221.

Mcrae, S. (2015). Infrastructure quality and the subsidy trap. *American Economic Review*, 105(1):35–66.

Meeks, R. C. (2017). Water works: The economic impact of water infrastructure. *Journal of Human Resources*, 52(4):1119–1153.

Montgomery, M. A. and Elimelech, M. (2007). Water and sanitation in developing countries: Including health in the equation - Millions suffer from preventable illnesses and die every year. *Environmental Science and Technology*, 41(1):17–24.

Muralidharan, K. and Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy*, 119(1):39–77.

National Geographic (2017). Nearly a billion people still defecate outdoors. Here's why. National Geographic Magazine. London, UK: August 2017.

Neal, D., Vujcic, J., Burns, R., Wood, W., and Devine, J. (2016). Nudging and Habit Change for Open Defecation: New Tactics from Behavioral Science. Technical report, World Bank's Water and Sanitation Program, Washington D.C.

Olken, B. and Pande, R. (2010). Corruption in developing countries. *Annual Review of Economics*, 4:479–509.

Olken, B. A., Onishi, J., and Wong, S. (2014). Should aid reward performance? Evidence from a field experiment on health and education in Indonesia. *American Economic Journal: Applied Economics*, 6(4):1–34.

Peltzman, S. (1973). The effect of government subsidies-in-kind on private expenditures: The case of higher education. *Journal of Political Economy*, 81(1):1–27.

Pickering, A. J., Djebbari, H., Lopez, C., Coulibaly, M., and Alzua, M. L. (2015). Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali: A cluster-randomised controlled trial. *The Lancet Global Health*, 3(11):e701–e711.

Prüss-Ustün, A., Bartram, J., Clasen, T., Colford, J. M., Cumming, O., Curtis, V., Bonjour, S., Dangour, A. D., De France, J., Fewtrell, L., Freeman, M. C., Gordon, B., Hunter, P. R., Johnston, R. B., Mathers, C., Mäusezahl, D., Medlicott, K., Neira, M., Stocks, M., Wolf, J., and Cairncross, S. (2014). Burden of disease from inadequate water, sanitation and hygiene in low- and middle-income settings: A retrospective analysis of data from 145 countries. *Tropical Medicine and International Health*, 19(8):894–905.

Rud, J. P. (2012). Electricity provision and industrial development: Evidence from India. *Journal of Development Economics*, 97(2):352–367.

Spears, D. and Lamba, S. (2016). Effects of early-life exposure to sanitation on childhood cognitive skills: Evidence from India's Total Sanitation Campaign. *Journal of Human Resources*, 51(2):298–327.

Sridhar, K. S. (2007). Reforming delivery of urban services in developing countries: Evidence from a case study in India. *Economic and Political Weekly*, 42(33):3404–3413.

The Times of India (2020). Why millions of Indians don't use the toilets they have. Published on February 20, 2020. Delhi, India: The Times Group.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Treibich, C. and Lépine, A. (2019). Estimating misreporting in condom use and its determinants among sex workers: Evidence from the list randomisation method. *Health Economics*, 28(1):144–160.

UNICEF (2019). Global Framework for Urban Water, Sanitation and Hygiene. United nations children's fund, New York: US.

United Nations (2020). Progress towards the Sustainable Development Goals. United Nations Economic and Social Council, New York.

Wager, S. and Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

WHO-UNICEF (2017). Progress on DrinkingWater, Sanitation and Hygiene 2017 Update and SDG Baselines. Geneva, Switzerland.

Williams, M. J. (2017). The political economy of unfinished development projects: Corruption, clientelism, or collective choice? *American Political Science Review*, 111(4):705–723.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

World Bank (2018). Understanding Poverty. Sanitation overview, World Bank, Washington D.C., US.

WSUP (2011). When are communal or public toilets an appropriate option? Technical report, Topic Brief 001, Water & Sanitation for the Urban Poor, London, UK.

Table 1: CT-level outcomes

	Maintenance and quality of the facility			Traffic during rush hour	
	Physical mainte-nance	Quality	Higher quality	Users	Free riding
	(1)	(2)	(3)	(4)	(5)
Panel A					
Any treatment (T)	0.045	0.050	0.116	-0.059	-0.075
	(0.049)	(0.023)	(0.043)	(0.050)	(0.042)
	[0.36 ; 0.37]	[0.03 ; 0.13]	[0.01 ; 0.04]	[0.24 ; 0.43]	[0.07 ; 0.19]
Panel B					
Maintenance (T1)	0.022	0.057	0.139	-0.100	-0.060
	(0.053)	(0.028)	(0.055)	(0.060)	(0.046)
	[0.68 ; 0.69]	[0.05 ; 0.18]	[0.01 ; 0.06]	[0.10 ; 0.29]	[0.20 ; 0.36]
Maintenance + sensitization (T2)	0.068	0.042	0.093	-0.018	-0.091
	(0.056)	(0.028)	(0.049)	(0.058)	(0.049)
	[0.22 ; 0.40]	[0.13 ; 0.38]	[0.06 ; 0.23]	[0.76 ; 0.77]	[0.06 ; 0.27]
T1 = T2 (p-value)	0.319	0.647	0.439	0.193	0.497
Mean (control group)	0.623	0.639	0.183	3.477	0.422
Std. dev. (control group)	0.486	0.219	0.388	0.429	0.290
Observations	542	542	542	542	542
Catchment areas	110	110	110	110	110
Follow-ups	1–5	1–5	1–5	1–5	1–5

Note. Estimates based on CT-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B. Standard errors clustered by catchment area are reported in parentheses. P-values are presented in brackets, the first from individual testing, the second adjusting for testing that each treatment is jointly different from zero for all outcomes presented in the table (see Section 4 for details). Dependent variables by column: (1) *Physical maintenance*: indicator variable equal to 1 if the CT received physical maintenance (repairs and/or deep cleaning intervention) in the month previous to the visit, and 0 otherwise; (2) *Quality*: index aggregating indicator variables about the physical quality of the facility, its cleanliness and the lack of bacteria (details about individual components are presented in Appendix C.10.2); (3) *Higher quality*: indicator variable equal to 1 if the quality index is above the 75th percentile, and 0 otherwise; (4) *Users*: total number of users observed during 1 hour at rush hour (reported in logarithms); (5) *Free riding*: share of users who do not pay the entry fee observed during 1 hour at rush hour. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. Additional details about the variables are presented in Appendix A.

Table 2: Caretaker's behavior

	Labor supply		**Routine maintenance**		**Motivation**
	Hours worked	Time allocated to managing	Awareness	Inputs	Pro-social motivation for the cause
	(1)	(2)	(3)	(4)	(5)
Panel A					
Any treatment (T)	0.175	0.052	0.089	0.034	-0.024
	(0.356)	(0.028)	(0.035)	(0.017)	(0.025)
	[0.62 ; 0.62]	[0.07 ; .21]	[0.01 ; 0.07]	[0.05 ; 0.16]	[0.33 ; 0.55]
Panel B					
Maintenance (T1)	-0.025	0.041	0.076	0.037	-0.021
	(0.419)	(0.031)	(0.043)	(0.019)	(0.030)
	[0.95 ; 0.95]	[0.20 ; 0.48]	[0.08 ; 0.32]	[0.05 ; 0.24]	[0.50 ; 0.76]
Maintenance + sensitization (T2)	0.378	0.062	0.102	0.032	-0.028
	(0.381)	(0.030)	(0.049)	(0.021)	(0.029)
	[0.32 ; 0.32]	[0.04 ; 0.19]	[0.04 ; 0.20]	[0.13 ; 0.33]	[0.33 ; 0.54]
T1 = T2 (p-value)	0.281	0.403	0.663	0.805	0.815
Mean (control group)	11.754	0.686	0.068	0.608	0.345
Std. dev. (control group)	3.690	0.226	0.253	0.162	0.219
Observations	542	542	542	542	542
Catchment areas	110	110	110	110	110
Follow-ups	1–5	1–5	1–5	1–5	1–5
Incentivized measurement	-	-	-	-	Yes

Note. Estimates based on CT-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B. Standard errors clustered by catchment area are reported in parentheses. P-values are presented in brackets, the first from individual testing, the second adjusting for testing that each treatment is jointly different from zero for all outcomes presented in the table (see Section 4 for details). Dependent variables by column: (1) *Hours worked*: number of hours worked by the caretaker; (2) *Time allocated to managing*: share of worked hours allocated by the caretaker to collecting fees and supervising cleaners, rather than conducting repairs, cleaning the facility, or spending time with the manager; (3) *Awareness*: indicator variable equal to 1 if the caretaker knows the recommended practices of the cleaning routine and the need for deep cleaning, and 0 otherwise; (4) *Inputs*: number of tools, equipment and cleaners used during the last routine maintenance for the CT, normalized to be between 0 and 1 (details about individual components are presented in Appendix D.10); (5) *Pro-social motivation for the cause*: share of the endowment that is donated by the caretaker in the adapted dictator game (Appendix C.6). All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. Additional details about the variables are presented in Appendix A.

Table 3: Valuation and attitudes towards the public good

	Valuation		Attitudes towards the public good		
	WTP for CT use	Perceived improvement	Demand for public intervention	Contribution in the PGG	Preference for maintenance
	(1)	(2)	(3)	(4)	(5)
Panel A					
Any treatment (T)	-0.037	0.027	0.053	0.001	-0.005
	(0.065)	(0.015)	(0.026)	(0.013)	(0.006)
	[0.57 ; 0.84]	[0.08 ; 0.30]	[0.04 ; 0.20]	[0.92 ; 0.93]	[0.39 ; 0.78]
Panel B					
Maintenance (T1)	0.047	0.030	0.055	-0.002	-0.009
	(0.083)	(0.018)	(0.032)	(0.015)	(0.007)
	[0.57 ; 0.84]	[0.10 ; 0.37]	[0.09 ; 0.42]	[0.89 ; 0.89]	[0.17 ; 0.44]
Maintenance + sensitization (T2)	-0.116	0.023	0.051	0.005	-0.001
	(0.070)	(0.019)	(0.032)	(0.016)	(0.007)
	[0.10 ; 0.45]	[0.21 ; 0.49]	[0.11 ; 0.39]	[0.77 ; 0.95]	[0.84 ; 0.84]
T1 = T2 (p-value)	0.052	0.738	0.911	0.665	0.197
Mean (control group)	1.146	0.154	0.096	0.174	0.212
Std. dev. (control group)	1.805	0.361	0.295	0.110	0.171
Observations	8808	4890	1580	1228	8808
Catchment areas	110	110	109	109	110
Follow-ups	1, 3, 5	1, 3, 5	3	5	1, 3, 5
Level of analysis	Respondent	Household	Household	Respondent	Respondent
Incentivized measurement	Yes	-	Yes	Yes	Yes

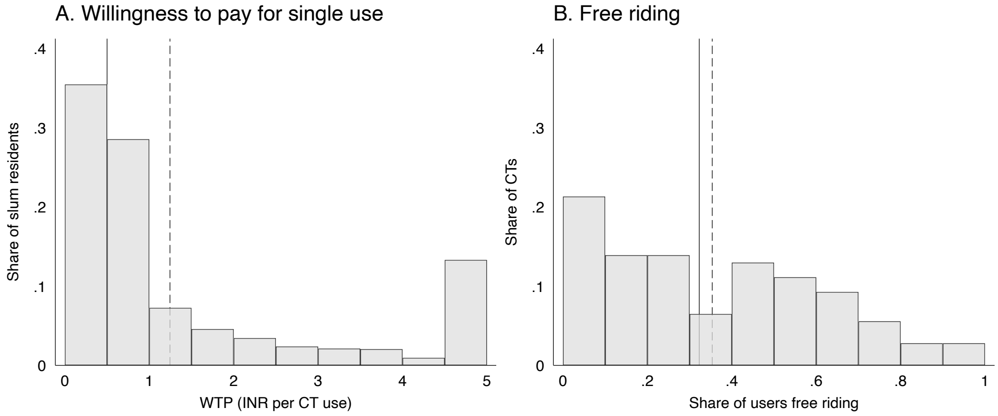
Note. Estimates based on respondent- and household-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B. Standard errors clustered by catchment area–round are reported in parentheses. P-values are presented in brackets, the first from individual testing, the second adjusting for testing that each treatment is jointly different from zero for all outcomes presented in the table (see Section 4 for details). Dependent variables by column: (1) *WTP for CT use*: incentivized willingness to pay for a single CT use (in Rupees), elicited for a bundle of ten tickets and divided by 10 to get at single use WTP; (2) *Perceived improvement*: indicator variable equal to 1 if the respondent perceived an improvement in the CT infrastructure, and 0 otherwise; (3) *Demand for public intervention*: indicator variable equal to 1 if the household asks for public intervention in the CTs O&M as incentivized through the voice-to-the-people initiative (Appendix C.8), and 0 otherwise; (4) *Contribution in the PGG*: share contributed in the public good game (Appendix C.7); (5) *Preference for maintenance*: share of the endowment that is donated by the respondent in the adapted dictator game (Appendix C.6). Columns (3) and (4) include only 109 catchment areas in the sample because the dependent variables were measured only at rounds 3 and 5, after a study slum was displaced. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. Specifications where the level of analysis is the respondent also include gender. Additional details about the variables are presented in Appendix A.

Table 4: Awareness of risks associated with unsafe sanitation and health

	Awareness of unsafe sanitation risks		**Morbidity and health expenditure**		
	Health and safety risks	Externalities	Morbidity	Expenditure (extensive)	Expenditure (intensive)
	(1)	(2)	(3)	(4)	(5)
Panel A					
Any treatment (T)	0.012	0.029	0.009	0.052	-43.860
	(0.005)	(0.018)	(0.019)	(0.023)	(187.709)
	[0.03 ; 0.12]	[0.10 ; 0.28]	[0.63 ; 0.87]	[0.03 ; 0.12]	[0.82 ; 0.80]
Panel B					
Maintenance (T1)	0.006	0.009	0.012	0.043	20.878
	(0.007)	(0.021)	(0.022)	(0.028)	(221.787)
	[0.35 ; 0.81]	[0.66 ; 0.89]	[0.58 ; 0.94]	[0.12 ; 0.49]	[0.93 ; 0.93]
Maintenance + sensitization (T2)	0.017	0.049	0.006	0.060	-105.831
	(0.006)	(0.020)	(0.021)	(0.024)	(220.743)
	[0.01 ; 0.03]	[0.02 ; 0.06]	[0.79 ; 0.78]	[0.01 ; 0.05]	[0.63 ; 0.87]
T1 = T2 (p-value)	0.106	0.063	0.761	0.494	0.590
Mean (control group)	0.837	0.661	0.400	0.635	1696.685
Std. dev. (control group)	0.120	0.474	0.490	0.482	5192.344
Observations	4757	4890	4890	3332	3332
Catchment areas	110	110	110	109	109
Follow-ups	1, 3, 5	1, 3, 5	1, 3, 5	3, 5	3, 5
Level of analysis	Household	Household	Household	Household	Household

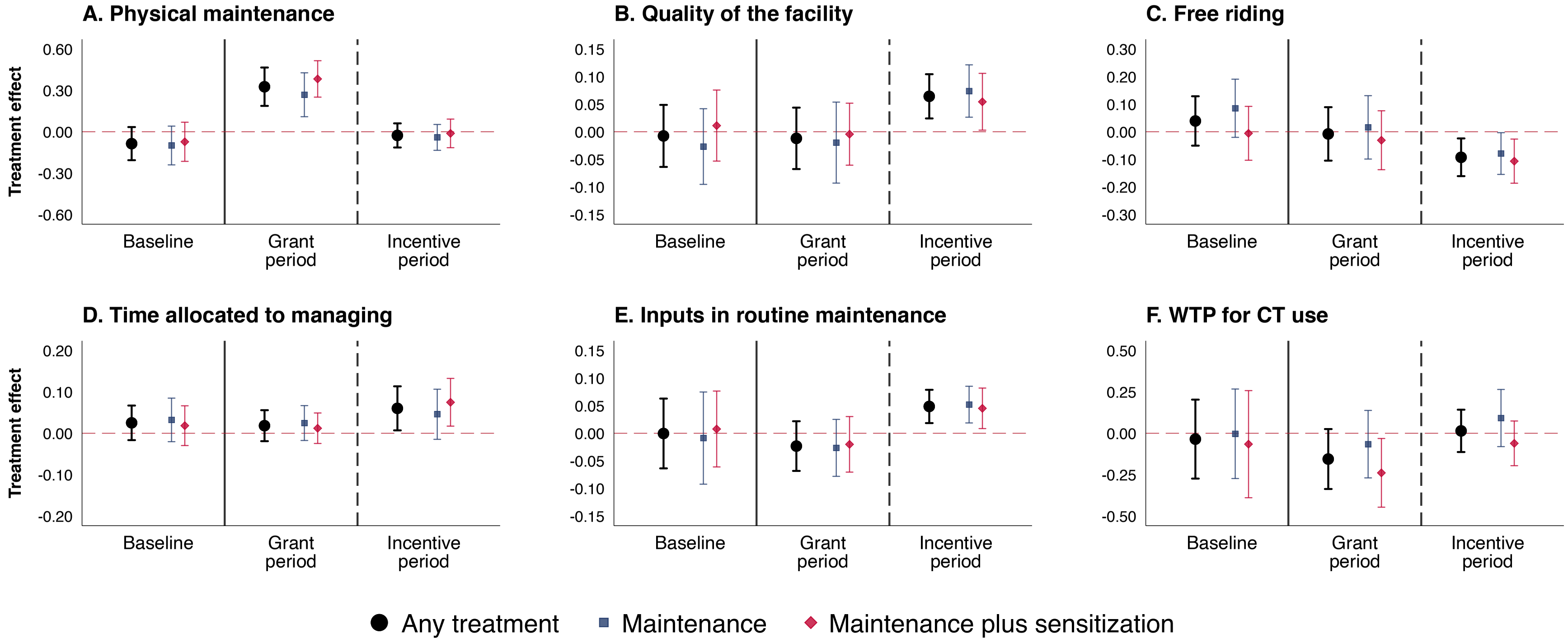
Note. Estimates based on household-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B. Standard errors clustered by catchment area–round are reported in parentheses. P-values are presented in brackets, the first from individual testing, the second adjusting for testing that each treatment is jointly different from zero for all outcomes presented in the table (see Section 4 for details). Dependent variables by column: (1) *Health and safety risks*: index aggregating continuous variables capturing the perception about health and safety risks from OD and from using dirty CTs (details about individual components are presented in Appendix C.10.1); (2) *Externalities*: indicator variable equal to 1 if the respondent reports that OD generates a health externality for their family, and 0 otherwise; (3) *Morbidity*: indicator variable equal to 1 if any household member had fever, diarrhea or cough during the two weeks previous to the interview, and 0 otherwise; (4) *Expenditure (extensive)*: indicator variable equal to 1 if the respondent had positive curative healthcare expenditures, and 0 otherwise; (5) *Expenditure (intensive)*: level of curative healthcare expenditures (in Rupees). Columns (4) and (5) include only 109 catchment areas in the sample because the dependent variables were measured only in rounds 3 and 5, after a study slum was displaced. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. Additional details about the variables are presented in Appendix A.

Figure 1: Baseline free riding and willingness to pay for CT use



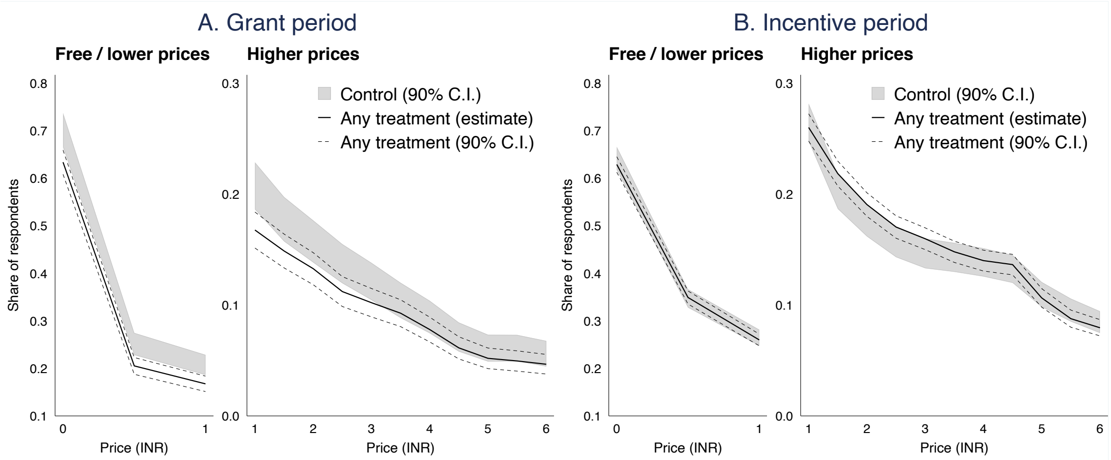
Note. Data collected at baseline. Panel A shows the distribution of the WTP for a single CT use among slum residents, measured using the incentivized elicitation of WTP (see Appendix C.5 for details about measurement). The distribution is censored at INR 5, the most common market price for a single CT use. Panel B reports the share of users who do not pay the fee for the use of the CT during 1 hour at dawn, measured by observers (see Appendix C.2 for details about measurement). The solid vertical lines represent the sample median, and the dashed vertical lines represent the sample mean. Additional details about the variables are presented in Appendix A.

Figure 2: Grant versus incentive periods: a comparison of treatment effects



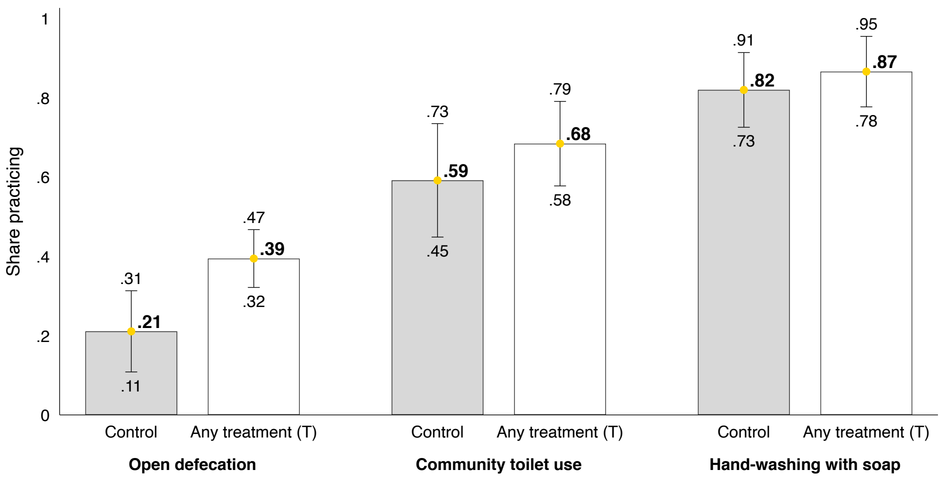
Note. Each panel presents estimates of treatment effects, and the mean of the control group at different points in time. Estimates of treatment effects are based on OLS regressions using equation (1) at the CT level or at the respondent level. Confidence intervals are built using statistical significance at the 10% level. *Baseline* includes the measurement at baseline, *Grant period* includes the measurement from follow-up 1, and *Incentive period* pools all subsequent follow-up measurements. See Section 1 for details about each intervention. When the regression is based on a single measurement period, robust standard errors are used for CT-level outcomes, and standard errors clustered at the catchment area are used for respondent level outcomes. When multiple measurement periods are pooled, standard errors are clustered at the catchment area for CT-level outcomes, and at the catchment area by collection round for respondent-level outcomes. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. For outcomes at the respondent-level, gender is also included as a control variable. Additional details about the variables are presented in Appendix A. Estimates pooling all post-baseline measurements are presented in Tables 1–4.

Figure 3: Inverse demand curve for CT use, by price level



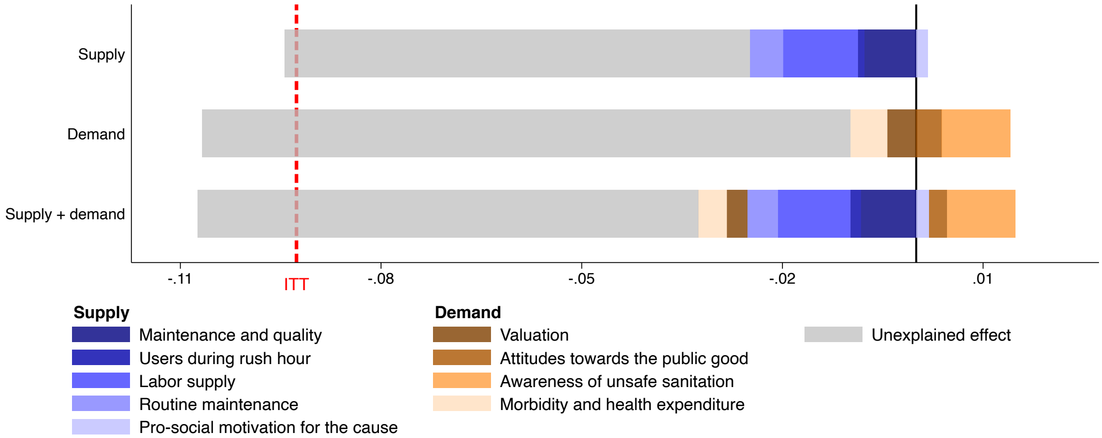
Note. Each curve indicates the share of respondents who prefer tickets for CT use versus cash at the corresponding price. This is elicited using a standard incentivized version of the multiple price list (or take-it-or-leave-it) methodology (Andersen et al., 2006). Details about the measurement are presented in Appendix C.5. Point-wise inference is computed using OLS regressions at prices ranging from INR 0 to 6 with increases of INR 0.5. For graphical representation, panels are split among free and lower prices (INR 0–1) and higher prices (INR 1–6). Panel A restricts the sample to the *Grant period*, which includes the measurement from follow-up 1, and Panel B restricts the sample to the *Incentive period*, which pools all subsequent follow-up measurements. See Section 1 for details about each intervention. When the regression is based on a single measurement period, standard errors are clustered at the catchment area. When multiple measurements are pooled, standard errors are clustered at the catchment area by collection round for respondent-level outcomes. Confidence intervals are built using statistical significance at the 10% level.

Figure 4: Sanitation- and hygiene-related behavior



Note. The figure shows the share of slum residents practicing each behavior in the day previous to the interview, estimating using a list randomization technique. Following this technique, shares are estimated as the difference in the number of items reported by respondents who faced the long list (which includes the sensitive behavior), and the respondents who faced the short list (which excludes the sensitive behavior). We compute this average separately in the control group, and in any treatment group. Confidence intervals are built using statistical significance at the 10% level and assuming errors are clustered at the level of the catchment area. Randomization of lists was performed at individual level, and data were collected during follow-up 5 only. Appendix C.4.1 provides additional details about the measurement.

Figure 5: A mediation analysis of the effect of any treatment on free riding



Note. The figure shows the decomposition of the ITT effect of any treatment on free riding. We use equation (1) and follow the procedure of Gelbach (2016) at the CT level and including all measurements in the incentive period (follow-ups 2–5). We include as mediators all outcome variables included in Tables 1–4, with the exception of free riding. We distinguish two groups of mediators: *supply* mediators, which include CT and caretaker outcomes, and *demand* mediators, which include the median value in the catchment area of slum resident outcomes. The contribution of each outcome variable to the ITT is aggregated in the group of variables presented in Tables 1–4. The decomposition is presented by including only *supply* mediators in the top bar, only *demand* mediators in the middle bar, and both groups of mediators in the bottom bar. The dashed vertical line indicates the ITT estimate from Table 1. The shaded gray areas represent the part of the ITT effect not explained by mediators. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. Additional details about the variables are presented in Appendix A.

ONLINE APPENDIX

Coordination and the Poor Maintenance Trap: an Experiment on Public Infrastructure in India

Alex Armand, Britta Augsburg, Antonella Bancalari

A Definition of variables

Variable	Description
CT and caretaker outcomes	
Physical maintenance	Indicator variable equal to 1 if the caretaker reports that the CT received physical maintenance (repairs and/or deep cleaning intervention) in the month previous to the visit or whether a physical maintenance activity was recorded as part of the maintenance intervention in the corresponding period, and 0 otherwise. The variable aggregates responses from the CT survey (Appendix C.2) and administrative data from the implementing team.
Quality	Index aggregating indicator variables about the physical quality of the facility, its cleanliness and the lack of bacteria. The index is re-scaled to be between zero and one. The variable aggregate survey responses from the CT survey (Appendix C.2), data from observers (Appendix C.2), and data from laboratory tests (Appendix C.3). Additional detail about the construction of the index and its components is provided in Appendix C.10.2.
Higher quality	Indicator variable equal to 1 if the quality index is above the 75th percentile, and 0 otherwise.
Users	Total number of users entering the CT (reported in logarithms). The variable uses data from observers during 1 hour at rush hour (Appendix C.2).
Free riding	Share of users who do not pay the entry fee. The variable uses data from observers during 1 hour at rush hour (Appendix C.2).
Hours worked	Number of hours worked by the caretaker. The variable is self-reported by the caretaker during each CT survey (Appendix C.2).
Time allocated to managing	Share of worked hours allocated by the caretaker to collecting fees and supervising cleaners. Alternative activities include conducting repairs, cleaning the facility, and spending time with the manager. The variable is self-reported by the caretaker during each CT survey (Appendix C.2).
Routine maintenance	*Awareness* is an indicator variable equal to 1 if the caretaker knows the recommended practices for cleaning routine and the need for deep cleaning, and 0 otherwise. The variable evaluates the correctness of questions about routine maintenance. These questions are asked during each CT survey (Appendix C.2). *Inputs* is the number of tools, equipment and cleaners used during the last routine maintenance for the CT. The variable aggregate survey responses from the CT survey (Appendix C.2). The number is normalized to be between 0, indicating that no tools reported in the questionnaire were used, and 1, indicating that all tools reported in the questionnaire were used. Tools include broom, mop, and safety equipment. Liquid tools include water, pressurized water and disinfectants. The baseline survey asks for information only on use of the broom, and disinfectants, while the full list is available for the following rounds of follow-ups. Details about individual components are presented in Appendix D.10.
Pro-social motivation for the cause	Share of the endowment donated by the caretaker in the adapted dictator game (Appendix C.6). The variable is incentivized and is measured for each caretaker.
Respondent-level outcomes	
WTP for CT use	Willingness to pay for a single CT use (in rupees). The variable is incentivized and elicited for a bundle of ten tickets, and is collected for both the household head and any partner separately in conjunction with the household survey (Appendix C.4). We divide the WTP for the bundle by 10 to get at measure of single use WTP. Appendix C.5 provides additional details about the measurement.

(continued on next page)

Variable	Description
Perceived improvement	Indicator variable equal to 1 if the respondent perceived improvement in the CT infrastructure, and 0 otherwise. The variable is self-reported by the household head during the household survey (Appendix C.4).
Demand for public intervention	Indicator variable equal to 1 if the household asks for public intervention in the CTs O&M, and 0 otherwise. Since up to two participants per household could participate, the indicator variable is equal to 1 if any household member asks for public intervention in the CT's O&M. The information is incentivized and collected during the SCA voice-to-the-people initiative, in conjunction with follow-up 3. Appendix C.8 provides additional information about the measurement.
PGG contribution	Share contributed by the participant in the public good game. The variable is incentivized and is measured at the end of all the activities of the experiment. Appendix C.7 provides additional information about the measurement.
Preference for maintenance	Share of the endowment that is donated by the slum resident in the adapted dictator game. The variable is incentivized and is measured for both the household head and the partner separately in conjunction with the household survey (Appendix C.4). Appendix C.6 provides additional details about the measurement.
Health and safety risks	Index aggregating continuous variables capturing perception about health and safety risks from OD and from using dirty CTs. The variable aggregates responses from the household survey (Appendix C.4). Details about individual components are presented in Appendix C.10.1.
Externalities	Indicator variable equal to 1 if the respondent reports that OD generates a health externality for their family, and 0 otherwise. The variable is self-reported by the household head during the household survey (Appendix C.4).
Morbidity	Indicator variable equal to 1 if any household member had fever, diarrhea or cough during the two weeks previous to the interview, and 0 otherwise. The variable is self-reported by the household head during the household survey (Appendix C.4).
Health expenditure	*Extensive* (margin) is an indicator variable equal to 1 if the respondent had positive healthcare expenditures, and 0 otherwise. *Intensive* (margin) is the level of healthcare expenditures (in rupees). Healthcare expenditures include all expenses associated with the cure of illnesses (costs are associated with doctor visits when the person is ill, with the purchase of medicine, with hospitalization, and with x-rays, and include travel costs associated with these expenses). We do not include preventive expenditures, which include costs associated with regular doctor checks, vaccines, anti-worm tablets, bed-nets, and prenatal tests, and travel costs associated with these expenses). The variable is self-reported by the household head during the household survey (Appendix C.4), but is not collected during follow-up 1.
Open defecation	Aggregate share of slum residents who practiced open defecation the day before the interview. Data are obtained using the list randomization technique. Information is obtained from both the household head and any partner in conjunction with follow-up 5. Appendix C.4.1 provides details about the measurement.
CT use	Aggregate share of slum residents who used the CT the day before the interview. Data are obtained using the list randomization technique. Information is obtained from both the household head and any partner in conjunction with follow-up 5. Appendix C.4.1 provides details about the measurement.
Hand-washing with soap	Aggregate share of slum residents who washed their hands with soap after defecating on the day before the interview. Data are obtained from using the list randomization technique. Information is obtained from both the household head and any partner in conjunction with follow-up 5. Appendix C.4.1 provides details about the measurement.
Variables about implementation	

(continued on next page)

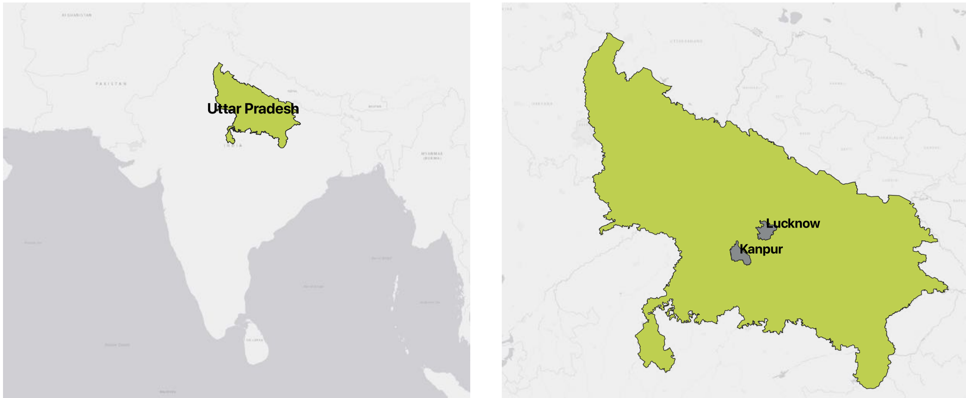
Variable	Description
Transfer to the CT	Transfer provided to the CT in the corresponding period as part of the intervention (in thousands of rupees). This includes the value of the initial grant to treated CTs, the subsidized use of tickets from the WTP game to both treated and control CTs, and the value of products bought with the transfer from slum residents as part of the adapted dictator game to both treated and control CTs. Information is based on administrative data from the implementing team.
Transfer to the caretaker	Transfer provided to the caretaker in the corresponding period as part of the intervention (in thousands of rupees). This includes the financial incentive provided in treated CTs and the amounts kept from each round of the adapted dictator game. Information is based on administrative data from the implementing team.
Any means	Indicator variable equal to 1 if the respondent is aware of any WASH campaign (including voice messages), and 0 otherwise. The variable is self-reported by the household head during the household survey (Appendix C.4).
Personal visits	Indicator variable equal to 1 if the respondent is aware of personal visits about WASH, and 0 otherwise. The variable is self-reported by the household head during the household survey (Appendix C.4).
Community activities	Indicator variable equal to 1 if the respondent is aware of any community activities about WASH, and 0 otherwise. The variable is self-reported by the household head during the household survey (Appendix C.4).
Posters	Indicator variable equal to 1 if the respondent is aware of any messages about WASH posted in the CT, and 0 otherwise. The variable is self-reported by the household head during the household survey (Appendix C.4).
Voice messages (recall)	Indicator variable equal to 1 if the respondent is aware of any voice message, and 0 otherwise. The variable is self-reported by the household head during the household survey (Appendix C.4).
Voice messages (exposure)	Proportion of voice messages about WASH and CTs listened to by the participant. The variable is built from administrative data derived from the implementation of voice messages as part of the sensitization campaign (Appendix B.2).
Variables about inputs	
Support from cleaners	Indicator variable equal to 1 if the caretaker is not the cleaner or the only cleaner in the CT, and 0 otherwise. The variable is self-reported by the caretaker during each CT survey (Appendix C.2).
Broom	Indicator variable equal to 1 if a broom was used during the last routine maintenance, and 0 otherwise. The variable is self-reported by the caretaker during each CT survey (Appendix C.2).
Mop	Indicator variable equal to 1 if a mop was used during the last routine maintenance, and 0 otherwise. The variable is self-reported by the caretaker during each CT survey (Appendix C.2).
Safety equipment	Indicator variable equal to 1 if safety equipment was used during the last routine maintenance, and 0 otherwise. The variable is self-reported by the caretaker during each CT survey (Appendix C.2).
Pressurized water machine	Indicator variable equal to 1 if a pressurized water machine was used during the last routine maintenance, and 0 otherwise. The variable is self-reported by the caretaker during each CT survey (Appendix C.2).
Bucket	Indicator variable equal to 1 if a bucket of water was used during the last routine maintenance, and 0 otherwise. The variable is self-reported by the caretaker during each CT survey (Appendix C.2).
Disinfectants	Indicator variable equal to 1 if disinfectants were used during the last routine maintenance, and 0 otherwise. The variable is self-reported by the caretaker during each CT survey (Appendix C.2).

B Details about the interventions

The intervention is implemented in the two largest urban agglomerations in the Indian state of Uttar Pradesh, Lucknow and Kanpur. Figure B1 shows the geographic location of the study.

Figure B1: Study location

Panel A. State | Panel B. Cities



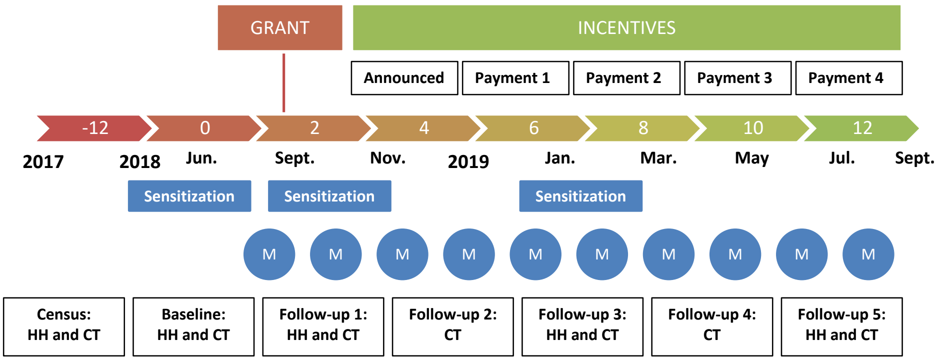
Note. Panel A shows the location of the state of Uttar Pradesh, while Panel B shows the location of Lucknow and Kanpur in the state. Basemap source: Esri (see Appendix C.9 for details and attributions).

These cities provide an ideal setting to study the role of infrastructure in developing cities as the results have external validity for a large number of contexts. First, similar to many expanding cities in South Asia and sub-Saharan Africa, Lucknow and Kanpur are characterized by rapid population growth and a relatively large prevalence of informal settlements. In 2015, among all urban agglomerations with more than 300,000 inhabitants, Lucknow and Kanpur were respectively the 129$^{\text{th}}$ and 141$^{\text{st}}$ largest cities worldwide (United Nations, 2018). In the period 2015–35, Lucknow is expected to grow from 3.2 to 5.2 million inhabitants (+59%), and Kanpur from 3.0 to 4.1 million inhabitants (+37%). These growth prospects are similar to those of similar-sized cities such as Accra (Ghana), Amman (Jordan), Jaipur (India), or Hyderabad (Pakistan) and of metropolises such as Karachi (Pakistan), Cairo (Egypt) or Manila (the Philippines). In terms of slum population, the share of inhabitants living in slums is comparable to that of other major cities in India (12.95% in Lucknow and 14.50% in Kanpur versus 14.66% in Delhi; Government of India, 2011). Second, the slum population of Lucknow and Kanpur is highly comparable to the average slum population of other cities in the state of Uttar Pradesh and the rest of the country (Appendix Table C1). Thirdly, Uttar Pradesh is a state in which access to basic services in urban settlements is highly salient. Out of 28 states and 8 union territories, it is the largest, the 4$^{\text{th}}$ most densely populated, and the 6$^{\text{th}}$ in terms of share of the population living in slums, which are home to more than 6 million people (Government of India, 2011). Finally, in terms of external validity at country level, in 2019, GDP per capita in India (2,104 USD) was comparable to that of sub-Saharan Africa (1,585 USD) and South Asia (1,959 USD). Population growth is also comparable, with 1.1% for India, 1.2% for Southeast Asia, and 2.7% for sub-Saharan Africa. In 2018, life expectancy at birth in India was the same as in South Asia

(69.4) and higher than in Sub-Saharan Africa (61.3) (World Bank, 2021).

The interventions are the results of a partnership between the research team and a wide number of institutions in Lucknow and Kanpur, including the owners and managers of CTs, Lucknow Municipal Corporation, Kanpur Municipal Corporation, Sulabh International, and a large number of local CT managers. The interventions were implementeed in partnership with FINISH Society, a Lucknow-based organization with more than 10 years of experience in promoting sustainable and equitable development of sanitation, hygiene and waste management in India. Figure B2 summarizes the timeline of activities.

Figure B2: Timeline of the intervention and data collection



Note. M indicates the delivery of voice messages (see Appendix B.2). HH indicates household surveys, and CT indicates Community Toilet surveys. Details about data collection activities are reported in Appendix C.

B.1 The maintenance intervention

The maintenance intervention consisted of two components provided one after the other: a grant and a financial reward. The **grant** was a one-off transfer to rehabilitate the infrastructure. The scheme offered three packages of similar monetary value from which the caretaker(s) could select one: (i) deep cleaning (i.e. septic tank sewage removal, unclogging latrines and sewerage pipes, and cleaning walls, floors and inside toilets), (ii) sanitation/water connection repairs and/or infrastructure refurbishment or (iii) cleaning tools and agents, and cleaning training. For CTs that selected repairs or deep cleaning, pictures of the CT area to be improved were taken before the work was done. Also, in this visit, a date was set for the works to be conducted. Based on this information, our partner FINISH arranged and supervised the work with an external contractor. To keep consistency in the works, we used the same contractor in all CTs. For CTs that selected training and cleaning tools, FINISH provided two sessions. First, the theory session covered why the facility should be frequently cleaned, at what frequency and hours it should be cleaned, and what tools and products should be used. Second, the practical session required caretakers and cleaners to accomplish cleaning tasks under the supervision of a trainer who provided feedback. In addition, FINISH endowed the CTs with several tools and cleaning products, including: four pairs of gloves, five floor cleaners, four toilet disinfectants, five liquid soaps, four toilet-cleaning brushes, two wipes, four nose masks, two brooms, two bucket and mop sets, three surf-ghadi (detergent), two hand-washing dispensers, two dustpans and two dustbins.

Figure B3: Examples of grant use

A. Pre-grant B. Post-grant

Note. Example of deep cleaning of walls and repair of locks in a CT in Lucknow. Panel A shows the status before the intervention, while Panel B shows the status after the deep cleaning. Source: Antonella Bancalari.

The **financial reward** was introduced in order to improve the quality of the service rendered by the caretaker. Two months after completion of the grant scheme works, we announced the financial reward scheme to caretakers in order to incentivize them to keep the CT clean. Caretakers could receive the following rewards: INR 500 conditional on soap availability in hand-washing facilities for both genders; INR 500 conditional on visible cleanliness of latrines, defined by whether cubicles were free from visible feces (both inside and outside the latrines); INR 1,000 conditional on bacteria counts being kept to a minimum standard (i.e., being below the median of the demeaned baseline distribution by city). We provided a greater pay-off for reducing bacteria because this is a harder task to achieve. The maximum amount a caretaker could receive was INR 2000, representing 40% of a monthly salary. Caretakers were informed that an external agent was going to return to measure each condition on a random day and time within the following two months, and that we would pay the financial reward depending on what the external agent measured. In CTs with more than one caretaker, the financial reward was split among them. After two months and with a bi-monthly frequency, we measured each of the conditions and paid accordingly. In each round, we reminded the caretaker(s) of the conditions to be awarded the financial reward. In each payment round, we informed caretakers of their past cleanliness performance to help them estimate the additional effort they needed to exert in order to achieve the conditions.

B.2 The sensitization campaign

The sensitization campaign *Awareness campaign to encourage CT use and maintenance in India* targeted all slum residents, in particular the heads of participant households and their spouses. The campaign was designed to provide information that was accessible to participants with low literacy levels. We provided key messages regarding the risks of unsafe sanitation behavior and the importance of paying the fee to fund operation and maintenance of the CTs through four different means. First, **door-to-door visits** used a flip chart with cartoons and messages targeted at all household members, especially household heads and spouses. This session covered the following sections: how open defecation affects your community; how open defecation affects your family; benefits of using CTs; what you and your family can do to make the CT better; your rights when you pay the fee for using the CT. The cartoons were made by a local graphic designer considering the context of urban slums in India (i.e. looks, clothes, infrastructure, environment). We piloted the visits before conducting the intervention in non-treated slums. Figure B4 shows the flip chart cover used for the campaign, and an example of delivery.

Figure B4: Door-to-door campaign

A. flip chart cover B. Delivery of the campaign

Note. Panel A shows the cover of the flip chart used to communicate key messages to residents in slums. It translates from Hindi as 'Awareness campaign to encourage CT use and maintenance in India". Panel B shows a moment of the sensitization campaign, in which a household head and spouse pay attention to the flip chart during a household visit in Lucknow. Source: Morsel.

Second, the main messages of the flip chart used in the door-to-door campaign were summarized into a four-page **leaflet** (Figure B5 distributed among slum residents. The key messages provided during the door-to-door visits were also summarized in a catchy fashion in a series of **posters** that were placed on CT walls. We placed three medium-sized and two large posters (Figure B6) in the entrance to CTs, in the area close to the hand-washing facilities and in each gender-specific area.

Third, monthly reminders in the form of **voice messages** were sent to participants' mobile

Figure B5: Leaflet

Note. The figure presents the leaflet circulated during the sensitization campaign. The first page from the left presents the 'benefits of CTs' and includes: (1) improved sanitation facilities; (2) operation and maintenance of infrastructure; (3) safety with doors, locks and lights; (4) hand-washing facilities; and (5) gender-specific areas. The second page presents 'duties of users' and includes: (1) paying the fee to use the CT; (2) not throwing trash into the latrines; (3) flushing after using; (4) not spitting; (5) helping the elderly in the family; (6) accompanying females in the family during darkness; and (7) keeping the facility clean. The third page presents the 'rights of users' and includes: (1) caretakers not allowing free riders; (2) regular cleaning; (3) repairs; (4) respecting opening hours; (5) functional doors, locks and lights; (6) keeping men out of female areas; and (7) respecting and giving priority to females with children and the elderly. The final page, the cover, is the same as the one provided in the flip chart, shown in Figure B4, and provides the title of the campaign.

Figure B6: Posters placed on CT walls

Poster 1

Poster 2

Poster 3

Poster 4

Poster 5

Note. The five posters placed on the walls of CTs read in Hindi: Poster 1, 'I choose to always defecate in CTs, I choose better health'; Poster 2, 'Health is happiness and cleanliness is godliness. Do your bit by using CTs'; Poster 3, 'We always pay and use CTs, do you? My family moved away from open defecation and now is healthier, safer and happier'; Poster 4, 'I value a clean and safe CT, that's why I pay the fee'; and Poster 5 replicates a famous Bollywood scene but replacing the words to make it relevant to CTs. The villain, depicted as a dirty man says 'I have buildings, properties, vehicles, what do you have?' and the hero replies 'I have my CT'.

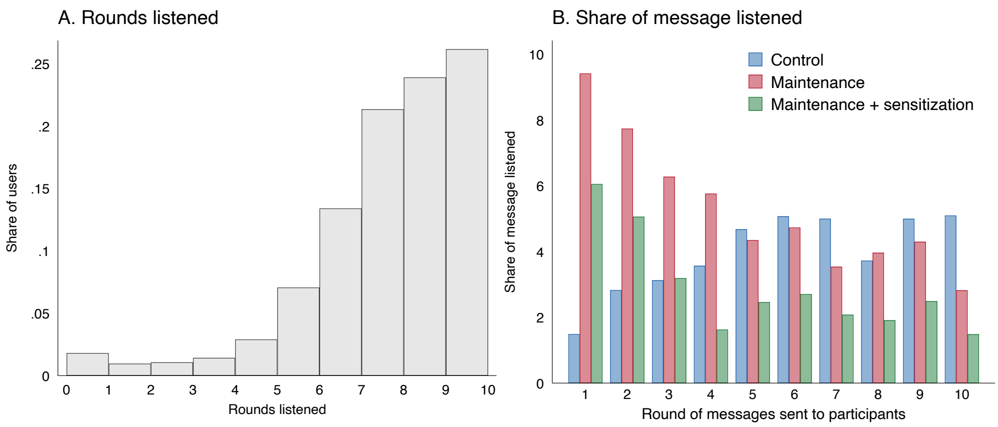
phones. We sent a total of 10 rounds of voice message between month 1 and 11 of the study. Table B1 reports the messages. This component was implemented using an purposely designed tracking app pre-populated with all mobile phone numbers. The tracking app records whether the number is valid, whether the voice call was answered, and the duration of the call. Households listened on average to 7 of the 10 monthly rounds of messages (Panel A in Figure B7), and listened to a good proportion of the message (Panel B). More than 20% of the information messages highlighting public and private health risks of unsafe sanitation, as well as supply-side messages, were heard.

Table B1: Voice messages

Treatment group	Message
Control	The community toilet is open from early morning until late evening.
Maintenance (T1)	Your community toilet has been granted aid to improve its quality. We hope you get to enjoy this better service.
Maintenance + sensitization (T2)	Do you know open defecation is one of the biggest causes of diarrhea which can even kill your children? Adopting good sanitation behavior will ensure a healthier future for your family!
	Open defecation is a big risk for your familys as well as your neighbors health. Use community toilets to defecate instead of polluting and contaminating your community with open defecation.
	Health is wealth! By not defecating in the open you are keeping your health safe and reducing expenses on medicines and treatment!
	Cleanliness is godliness! By using community toilets, you are contributing towards the cleanliness and health of your community.
	Do you know how unsafe it is for women and girls in your family to go for open defecation? Be the change and adopt the use of community toilets.
	Using community toilets ensures dignity of women in your community. Women should not feel ashamed of going to community toilets It is way better than open defecation!
	Using community toilets improves the health of your children and keeps medicines and doctors away!

Note. The table shows voice messages sent as part of the sensitization campaign. Households allocated to the sensitization campaign received the 'Maintenance + sensitization' messages in several rounds, plus the 'Maintenance' and 'Control' messages. Households allocated to the *maintenance* only received the 'Maintenance' and the 'Control' messages in every round. Households in the control received only the 'Control' message in every round.

Figure B7: Voice messages



Note. Panel A indicates the number of rounds in which at least one household member answered the phone. Panel B shows the average share of the message that is listened to by participants, distinguishing across treatment arms and rounds. We correct for the differential duration of messages per treatment arm by multiplying the mean share of the message listened to by the maximum length per treatment arm.

B.3 Implementation and effectiveness of interventions

We compute measures of exposure to interventions and report estimates of treatment effects in Table B2. For exposure to the maintenance intervention, in columns (1) and (2), we focus on the transfers (in thousands of INR) in the corresponding period to the CT and to the caretaker as part of the program. Throughout the study period, total transfers to the CT totaled on average INR 1,577 per CT in the control group, INR 24,777 per CT in the T1 treatment, and INR 25,791 per CT in the T2 treatment. Total transfers to the caretaker totaled on average INR 373 per CT in the control group, INR 4,133 per CT in the T1 treatment, and INR 4,227 per CT in the T2 treatment. For exposure to the sensitization campaign, columns (3)–(6) focus on the recall of WASH campaigns, while columns (7) and (8) focus on the recall of and exposure to voice messages. Concerning recall of both WASH campaigns and of voice messages, we asked respondents about their exposure to any WASH campaign, without direct indication of the campaign part of the experiment.

Table B2: Exposure to the interventions, by component

	Maintenance		**Sensitization campaign**					
	Transfer to the ...		Recall of WASH campaigns from ...				Voice messages	
	CT	Caretaker	Any means	Personal visits	Community activities	Posters	Recall	Exposure
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A								
Any treatment (T)	4.739	0.761	0.073	0.044	0.044	0.089	0.054	0.399
	(0.060)	(0.034)	(0.020)	(0.018)	(0.018)	(0.026)	(0.020)	(0.059)
	[0.00]	[0.00]	[0.00]	[0.02]	[0.01]	[0.00]	[0.01]	[0.00]
Panel B								
Maintenance (T1)	4.645	0.746	0.040	0.014	0.023	0.018	0.017	-0.041
	(0.081)	(0.045)	(0.024)	(0.022)	(0.020)	(0.030)	(0.024)	(0.047)
	[0.00]	[0.00]	[0.10]	[0.53]	[0.26]	[0.53]	[0.48]	[0.38]
Maintenance + sensitization (T2)	4.839	0.776	0.105	0.073	0.065	0.157	0.090	0.822
	(0.074)	(0.047)	(0.023)	(0.021)	(0.021)	(0.029)	(0.023)	(0.086)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
T1 = T2 (p-value)	0.063	0.636	0.010	0.014	0.051	0.000	0.003	0.000
Mean (control group)	0.315	0.063	0.676	0.245	0.321	0.326	0.572	0.188
Std. dev. (control group)	0	0	0	0	0	0	0	0
Observations	560	560	4890	4890	4890	3358	4890	4890
Catchment areas	110	110	328	328	328	218	328	328
Observation rounds	5	5	3	3	3	2	3	3

Note. In columns (1) and (2), estimates are based on CT-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B. Standard errors clustered by catchment area are reported in parentheses. Transfers are reported in thousands of INR. In columns (3)–(8), estimates are based on household-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B. Standard errors clustered by catchment area–round are reported in parentheses. P-values are presented in brackets, the first from individual testing, the second adjusting for jointly testing that each treatment is different from zero for all outcomes presented in the table. See Section 4 for details. Dependent variables are reported in the column header and are defined in Appendix A. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT.

B.4 Cost of intervention and quality scenarios

Table B3 presents a summary of the cost associated with each activity falling under the maintenance (Panel A) and sensitization interventions (Panel B). Apart from total intervention cost, we also provide the unitary cost at the CT level. Note that these are total costs throughout the project, while individual components have different timelines for implementation.

Based on input from our implementing partner FINISH Society, as well as Lucknow Municipal

Table B3: Cost of interventions

	Total expenditure		Cost per facility	
	INR	**US$**	**INR**	**US$**
A. Maintenance intervention				
Management	324,000	4,601	4,629	66
Implementation of grant scheme	1,688,500	23,678	24,121	343
Incentives for caretakers	267,000	3,792	3,814	54
Laboratory tests	210,000	2,982	3,000	42.60
Total	**2,489,500**	**35,352**	**35,564**	**505**
B. Sensitization intervention				
Management	81,000	1,150	2,314	32.86
Design and printing of material	50,000	710	1,429	20
Door-to-door campaign	440,770	6,259	12,593	179
Voice messages	21,662	308	619	8.79
Total	**593,432**	**8,427**	**16,955**	**241**

Note. For conversion of Indian rupees into US$, we assume an exchange rate of 70.42 INR/US$. The implementation of the grant component includes subcontracting, material for repairs, human resources, transportation and the overall management of the intervention. Door-to-door campaign includes transportation costs. Cost per facility is computed assuming 70 CTs in the maintenance intervention, and 35 in the sensitization intervention. Details about the interventions are provided in Appendices B.1 and B.2 .

Corporation, Table B4 provides information on O&M costs for the median CT in our study sample, defined by its age (20 years), size (four female WCs, six male WCs and two urinals), and number of daily users (average of 150). Cost items include salaries for a caretaker and cleaner(s), cleaning supplies, as well as electricity and costs for minor repairs.

Table B4: Monthly O&M costs and grant and incentive costs for median CT in study sample

	Maintenance level			
	Poor (status quo)		**Improved**	
	INR	US$	INR	US$
Panel A. O&M COSTS				
Salaries				
Caretaker (full-time)	5,000	71.00	12,000	170.41
Cleaner(s)	3,000	42.6	6,000	85.2
Supplies				
Cleaning agents	500	7.10	4,000	56.80
Cleaning equipment	200	2.84	2,200	31.24
Other				
Electricity	500	7.10	2,600	36.92
Minor repairs	1,000	14.20	2,000	28.40
Total	**10,200**	**144.85**	**28,800**	**408.97**
Total per eligible household	**300**	**4.26**	**847**	**12.03**
Panel B. INTERVENTION				
Maintenance grant				
Implementation	2,010	28.54		
Management	193	2.74		
Incentive scheme				
Amount paid to caretaker	477	6.77		
Management	289	4.11		
Laboratory tests	375	5.33		
Total	**3,344**	**47.49**		
Total per eligible household	**98**	**1.40**		
TOTAL (A + B)	**13,544**	**192.33**	**28,800**	**408.97**
TOTAL (A + B) per eligible household	**398**	**5.66**	**847**	**12.03**

Note. For conversion of INR into US$, we assume an exchange rate of 70.42 INR/US$. We assume that the grant is provided once a year and that incentives are provided on an ongoing basis every two months. We allocate 50% of total management cost to the maintenance grant implementation and 50% to the incentive scheme. To compute the total per eligible household, we consider the median number of households in the catchment area (34), and we assume no other household is using the CT.

The monthly maintenance cost for the current scenario (which we term as 'status quo') is INR 10,200 (US$ 144.85). Under the current scenario, salaries represent 78% of the total budget, and cover the costs for a full-time caretaker and for one cleaner performing a daily routine clean. We consider one alternative cost scenario that was deemed to support an 'improved' maintenance level. Under this scenario, we assume that the number of users remains constant. The scenario introduces a higher salary for the caretaker (which allows hiring a more experienced caretaker), higher input costs, and a yearly investment into cleaning machinery, such as a pressurized water cleaner, which costs about INR 20,000 (US$ 284.01). This scenario leads to a total of INR 28,800 (US$ 408.97) per month, with salaries representing 63% of the total. It is important to note that we do not claim that this scenario is optimal, and it can be improved further. The table also shows cost per eligible household (see Appendix C for eligibility and proximity criteria), of which there are 34 in the median CT. In Panel B of Table B4 we convert the total intervention expenditures of the supply intervention (Table B3) into monthly expenditures. Adding these costs, the total monthly costs become INR 13,544 (US$ 192.33) per CT.

B.5 Treatment effects on revenues

Table B5 provides estimates of treatment effects on monthly revenues during the rush hour. Revenues are imputed using information from observers about the number of people using the CT and the share of them who are is paying the fee (assuming a standard fee of INR 5). We compute revenues based on the actual number of users observed, and on a constant number of users (assumed to be at the baseline level).

Table B5: Revenues during rush hour

	Monthly revenues during rush hour							
	All periods				Incentive period			
	Actual users		Constant users		Actual users		Constant users	
	Positive	Level	Positive	Level	Positive	Level	Positive	Level
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A								
Any treatment (T)	0.033	267.073	0.026	267.073	0.035	294.981	0.027	294.981
	(0.033)	(254.664)	(0.033)	(254.664)	(0.032)	(265.241)	(0.031)	(265.241)
	[0.32]	[0.29]	[0.43]	[0.29]	[0.27]	[0.27]	[0.39]	[0.27]
Panel B								
Maintenance (T1)	0.020	127.399	0.011	127.399	0.027	197.439	0.016	197.439
	(0.041)	(293.467)	(0.043)	(293.467)	(0.039)	(297.060)	(0.042)	(297.060)
	[0.62]	[0.66]	[0.80]	[0.66]	[0.50]	[0.51]	[0.70]	[0.51]
Maintenance + sensitization (T2)	0.045	407.935	0.040	407.935	0.044	394.460	0.038	394.460
	(0.029)	(290.198)	(0.029)	(290.198)	(0.029)	(304.695)	(0.028)	(304.695)
	[0.13]	[0.16]	[0.17]	[0.16]	[0.13]	[0.20]	[0.18]	[0.20]
T1 = T2 (p-value)	0.379	0.330	0.363	0.330	0.534	0.490	0.496	0.490
Mean (control group)	0.948	3027.202	0.953	3151.252	0.948	2840.260	0.954	3023.497
Std. dev. (control group)	0.222	1870.703	0.212	1750.153	0.223	1779.927	0.210	1651.547
Observations	542	542	528	542	434	434	421	434
Catchment areas	110	110	108	110	110	110	108	110
Follow-ups	1–5	1–5	1–5	1–5	2–5	2–5	2–5	2–5

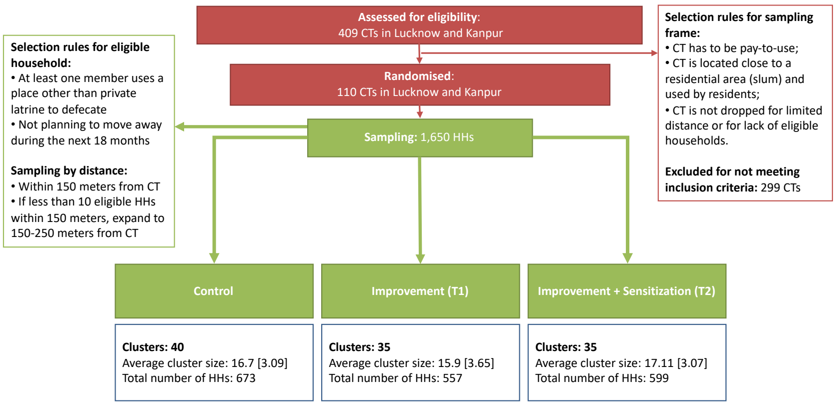
Note. In columns (1), (3), (5) and (7), estimates are based on CT-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B. In columns (2), (4), (6) and (8), estimates are based on CT-level tobit regressions using equation (1) in Panel A, and equation (2) in Panel B, and imposing censoring at zero. Standard errors clustered by catchment area are reported in parentheses. P-values are presented in brackets. See Section 4 for details. Dependent variables are reported in columns. *Positive* is an indicator variable equal to 1 if the revenues are larger than zero, and 0 otherwise. *Level* is the revenues reported in levels. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. Additional details about the variables are presented in Appendix A.

C Sampling, data collection and measurement

C.1 CT census, slum resident census, and sampling strategy

Figure C1 summarizes the sampling procedure for CTs and households. In order to obtain the sampling frame, during the first half of 2017, we performed mapping of slums and a census of all CTs in both study cities.[2] The census questionnaire was administered to caretakers and/or supervisors, and collected information on the geolocation and the main characteristics of CTs, such as main users, building characteristics, ownership, management structure, and payment system. A total of 409 CTs (201 in Lucknow and 208 Kanpur) were identified. Out of these, we dropped CTs free to use, CTs located outside slum areas, and CTs used by non-residents (generally located in slums near market areas). In addition, to avoid cases in which residents can choose between different CTs, we drop clusters of CTs within a slum area. Specifically, we drop CTs that are closer than 300 meters to each other, and CTs that have two other CTs closer than 350 meters. This resulted in a total of 110 CTs, which were all selected for the study.[3]

Figure C1: Sampling frame definition and sampling procedure



Note. The flowchart summarizes the procedure followed for the selection of CTs and the sampling of households within their catchment areas. Details of the procedure are discussed in Appendix C.

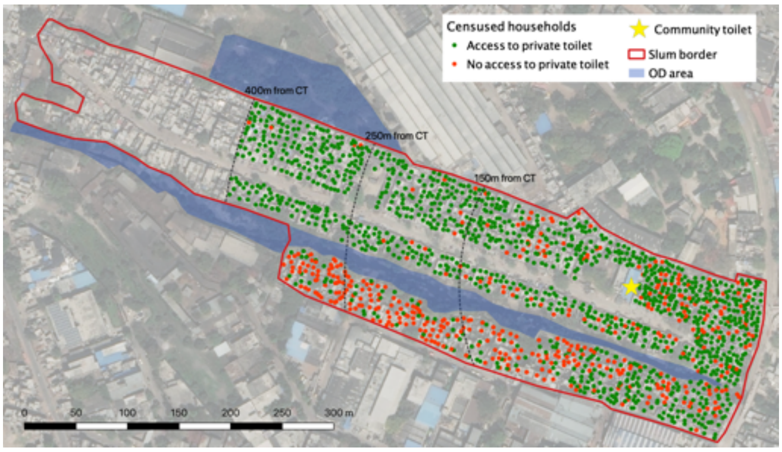
The second level of the sampling frame is characterized by all households living in proximity to the CT without access to a private toilet. To identify these households, during the second half of 2017, we performed a census of all households living within slum borders and within 400 meters of the selected CTs. The distance bound was selected based on qualitative evidence about the maximum distance one person would walk to opt for CT use versus open defecation (Armand et al., 2020a). The questionnaire was administered to household heads. The census

[2]The definition of *slums* follows official criteria used in the Indian Census. According to Government of India (2011), a *slum* is defined as a 'residential area where dwellings are unfit for human habitation by reasons of dilapidation, overcrowding, faulty arrangements and design of such buildings, narrowness or faulty arrangement of street, lack of ventilation, light, or sanitation facilities or any combination of these factors which are detrimental to the safety and health'. We make use of the definition of *identified* slum, 'a compact area of at least 300 population or about 60-70 households of poorly built congested tenements, in unhygienic environment usually with inadequate infrastructure and lacking in proper sanitary and drinking water facilities.'

[3]Following the household census discussed in the next paragraph, we also dropped very small complexes, i.e. CTs in whose catchment areas are living fewer than eight eligible households.

gathered information about demographic and dwelling characteristics (including geolocation), and sanitation-related behavior for more than 30,000 households. To identify potential users of CTs, we defined a household to be *eligible* for the study if all these conditions are met: the household lives in the catchment area of a selected CT, defined by the area within the slum and within 150 or 250 meters in straight distance from the CT building (see Appendix E.1 for a discussion about the definition of catchment area); at least one household member reports using a CT or a shared toilet (i.e. neighbors, makeshift, work, school), or practicing open defecation; the household does not intend to migrate during the 18 months following the census interview. Figure C2 shows an example of this selection process. Figure C3 provides the spatial distribution of CTs selected for the study, and their allocation to different treatment groups.

Figure C2: Definition of sampling frame: an example



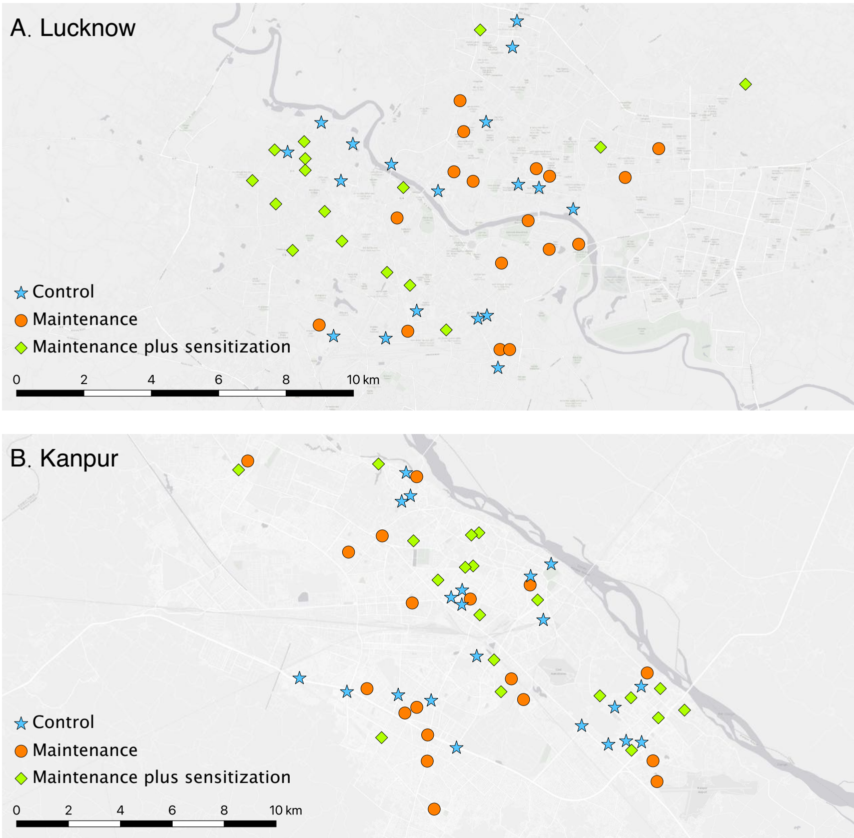
Note. The figure shows an example of the selection process for constructing the sampling frame using a hypothetical slum. Each dot represents a censused household. The area within the slum border but more than 400 meters from the CT was not covered by the census. Distance bounds are computed as straight distance from the CT. Basemap source: Esri (see Appendix C for details and attributions).

Within each of the 110 catchment areas, we sampled up to 17 eligible households. For catchment areas with fewer than 10 eligible households available within 150 meters, we selected all households within this bound, and randomly selected the remaining ones (up to 17 households) from the households living between 150 and 250 meters from the CT. In total, we obtained a sample of 1,650 households living in 110 catchment areas. Table C1 provides a comparison between sampled households and the average characteristics of slum residents across all states of India and in Uttar Pradesh.

C.2 CT surveys and observation

During regular unannounced visits to the CTs, we administered a questionnaire to the caretaker to collect data on cleaning practices, CT management and time allocation to different tasks. In addition to self-reported information from the caretakers, we also gathered information about the condition and cleanliness of CTs using observers. To provide uniform reports from observers, the data collection manual defined conditions for the visual evaluation of the state of the CT.

Figure C3: Geographical distribution of CTs, by city and treatment group



Note. Panel A shows the geographical distribution of CTs selected for the study in the city of Lucknow. Panel B shows the geographical distribution of CTs selected for the study in the city of Kanpur. Details about the procedure to select CTs is provided in Appendix C. Basemap source: Esri (see Appendix C.9 for details and attributions).

Table C1: Descriptive statistics of slum populations

	2011 Census of India		Study sample
	India	Uttar Pradesh	Lucknow and Kanpur
	(1)	(2)	(3)
A. Share of the population			
Male	0.52	0.53	0.53
Female	0.48	0.47	0.47
Children (0-6 y.o.)	0.12	0.14	0.09
Scheduled caste	0.20	0.22	0.45
B. Other characteristics			
Sex ratio (female to male)	1.08	1.12	1.12
Literacy rate	0.78	0.69	0.46

Note. The table provides descriptive statistics for the slum population in India in Column (1), for the slum population in Uttar Pradesh in Column (2), and for the study sample in Column (3). The source for Columns (1) and (2) is the 2011 Indian Slum Population Census (Government of India, 2011).

This measure was supplemented with bacterial counts derived from lab analyses (see Appendix C.3). Observers also recorded the number of users and the share of users who pay the fee for the duration of 1 hour using manual counters. This observation was done at dawn, when congestion at the CT is highest. We collected CT-level data in a sequence of six waves (refer to Figure B2 for the timing and label of each wave).

Table C2: Selected CTs and households, by treatment arm and city

	Control		T1		T2		Total	
	N	%	N	%	N	%	N	%
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
A. CTs								
Lucknow	19	36.5	17	32.7	16	30.8	52	100
Kanpur	21	36.2	18	31.0	19	32.8	58	100
Total	40	36.4	35	31.8	35	31.8	110	100
B. Households								
Lucknow	255	35.5	225	31.3	239	33.2	719	100
Kanpur	321	37.5	262	30.6	273	31.9	856	100
Total	576	36.6	487	30.9	512	32.5	1,575	100

Note. The table presents the distribution of selected CTs (Panel A) and households (Panel B) by treatment arm and city. *T1* indicates the 'maintenance' treatment group, and *T2* indicates the'maintenance + sensitization' treatment group.

We collected data for all of the 110 selected CTs at the baseline, but only for 108 in follow-up 1, 109 in follow-up 2, 107 in follow-up 3, 105 in follow-up 4 and 106 in follow-up 5, given that some CTs closed temporarily/permanently for refurbishment, and one slum was completely displaced after follow-up 2. In some cases, we were able to collect observations and bacteria swabs, while not being able to survey caretakers. In 92% of CTs, we surveyed caretakers in all five follow-up rounds. In addition, two new CTs opened very close to the study toilets, both allocated to the maintenance (T1) treatment arm. Because households in the catchment area also used these new CTs, we collected data from these toilets during follow-ups 2 to 5. These new CTs did not increase the number of clusters, since we consider them part of the same cluster as the old CTs given their close proximity. Table C3 shows that the number of CT observations and caretaker surveys and the addition of CTs are orthogonal to treatment allocation.

Table C3: CT observations and caretaker surveys balanced across treatment arms

	Observations collected FU rounds	**Caretaker surveyed** FU rounds	**New CT in catchment area**
	(1)	(2)	(3)
Panel A			
Any treatment (T)	0.100	0.125	0.014
	(0.100)	(0.135)	(0.010)
	[0.32]	[0.36]	[0.16]
Panel B			
Maintenance (T1)	0.100	0.052	0.028
	(0.100)	(0.170)	(0.020)
	[0.32]	[0.76]	[0.15]
Maintenance + sensitization (T2)	0.100	0.198	0.001
	(0.100)	(0.124)	(0.002)
	[0.32]	[0.11]	[0.82]
T1 = T2 (p-value)	0.942	0.238	0.153
Mean (control group)	4.900	4.775	0.000
Catchment areas	110	110	110

Note. Estimates based on CT-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B. Robust standard errors in parentheses. P-values are presented in brackets. Dependent variables by column: (1) *Observations collected*: number of follow-up surveys where CT observation were collected; (2) *Caretaker surveyed*: number of follow-up surveys where the CT caretaker was surveyed; (3) *New CT in catchment area*: indicator variable equal to 1 if an additional CT opened in the same catchment area and hence was included in the study later on, and 0 otherwise. All specifications include strata indicators for city and the provider of the CT.

C.3 Laboratory tests

In order to measure the presence of health hazards at the CT-level, we collected data about bacteria and mold presence using samples analyzed in the laboratory. We first focus on the presence of the species *Escherichia coli* (*E. coli*) of genus *Escherichia*, an indicator of fecal contamination (Sclar et al., 2016). For E. coli, because it is present in almost all samples, we focus on the bacteria count (CFU per cm^2). To compute counts, we follow Benke and Hamilton (2008) and World Health Organisation (2017) and use the arithmetic mean among the samples collected in the CT during each measurement round. We supplement E. coli counts with tests for the presence of potentially harmful bacteria of the genus *Bacillus*, genus *Staphylococcus*, genus *Klebsiella*, and genus *Salmonella*. For further information on the effect of bacteria on human health, refer, for instance, to Jenkins and Maddocks (2019). In addition, we test for the presence of mold, which can cause allergic reactions and respiratory problems (Gent et al., 2002).

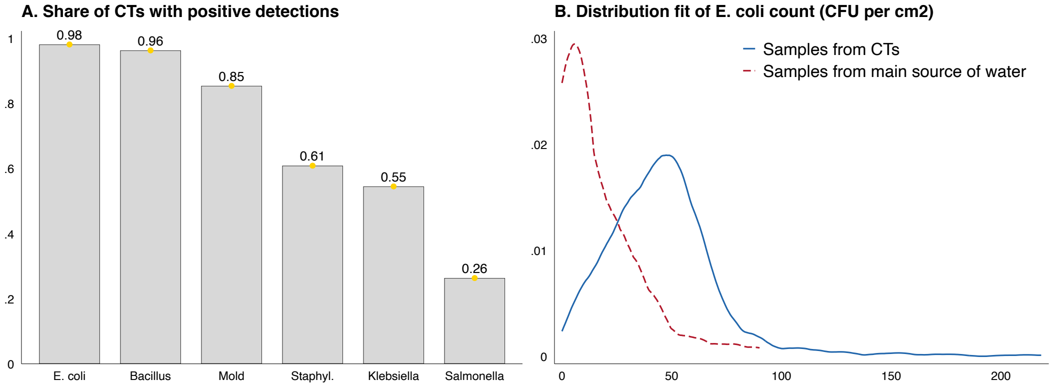
To implementat these measurements, we prepared a protocol in conjunction with a laboratory based in Lucknow, which analyzed the samples (the full protocol is available upon request to the authors). For each CT and during each survey round, three samples were collected using swabs in specific locations of the CT based on evidence about the microbial bio-geography in public toilets (Flores et al., 2011; McGinnis et al., 2019). CTs were first randomized into two groups: a *male* group, in which the swabs were collected in the male area of the CT throughout the study, and a *female* group, in which the same was performed in the female area of the CT. During each visit, the enumerator collected three samples from different parts of the CT. The first two samples were collected from the floor of the cubicles at the mid-point between the entrance wall and the latrine/water. Cubicles were randomly selected in each round to avoid the caretaker focusing on the cleanliness of a specific space in the CT. The randomly selected cubicles were provided to the enumerator by the research team in the form of a number, and the enumerator was instructed to select the cubicle according to the random number counting from right to left. A third sample, aimed at collecting information about the area where most people walk, was collected from the floor where one would take one's first step to enter the cubicle hallway.

At baseline, we also collected information about access to clean water. We collected and analyzed two samples of water for each catchment area. During the baseline survey, we asked households about their main source of water, and we then randomly selected two households and collected water samples from their indicated source. Whenever the source of water is shared by all eligible households in the slum, such as the case of a public tap, we collected only one water sample. Figure C4 shows descriptive statistics at baseline for these measurements.

C.4 Slum resident survey

We collected household-level data in a sequence of four waves (refer to Figure B2 for the timing and label of each wave). This was a standard household survey, collecting information on demographic and socio-economic characteristics, such as household composition, dwelling

Figure C4: Bacteria and mold detection at baseline



Note. Panel A presents the share of CTs where each bacteria type or mold was detected in at least one of the three samples. Panel B shows the distribution of the E. coli count from CT and water samples. The distribution fits are estimated non-parametrically using kernel density estimation assuming an Epanechnikov kernel function. Bandwidths are estimated by Silvermans rule of thumb (Silverman, 1986).

characteristics, assets, income and expenditure. This information was supplemented by a section on health and sanitation behavior, including attitudes and expectations associated with different sanitation practices. At baseline, we further collected information on child health for children under the age of 6, and on intra-household dynamics between spouses. The instrument, including all modules, has an average duration of one hour. All follow-up surveys aimed to recollect some of the same information collected at baseline. Some information collected at baseline was not collected during the follow-up surveys, such as detailed information on child health and childcare, and intra-household decision-making. Some new information was collected after the baseline survey, such as information related to exposure to the interventions.

The respondent is the main decision maker in the household. We select the respondent using the following rules: if the household head is present, then the respondent is the household head; if the household head is absent, then the respondent is the spouse of the household head; if the household head and spouse are both absent, the household is revisited; if the household head and spouse are both absent during the revisit, then the respondent is the most senior member (over 18 years old) who is actively participates in the households decision-making. At baseline, we also interviewed the spouse of the household head to gather information about intra-household decision-making, and the primary caregiver to collect information on child health for children aged 5 years or younger.

In total, we interviewed 1,575 households at baseline (an average of 12 households per cluster), 1,532 households during follow-up 1, 1,578 households at follow-up 3, and 1,772 households in the follow-up 5. Each baseline household was interviewed in 2.6 out of 3 follow-up measurements, with only 2% of baseline households that was never re-interviewed at follow-up. In terms of attrition from baseline to a specific follow-up surveys, the rate is 9% for follow-up 1, 19% for follow-up 3, and 14% for follow-up 5.[4] Columns (1)–(5) in Table C4 estimates the

[4]The attrition rate was the highest at follow-up 3 because the survey coincided with school vacations, a period

probability of attrition for each of these indicators as a function of the treatment status. Attrition does not differ between treatment and control groups for any of the attrition indicators. In order to maintain a comparable sample size in all follow-up surveys, we handled attrition with replacements at random using the sampling frame used for the baseline sampling. Column (6) tests whether the replacement was introduced differently across treatment arms. In total, 16% of follow-up observations are replacements, with no statistical difference across treatment arms. Appendix D.3 present results using ANCOVA and IPW specifications.

Table C4: Attrition and replacements across treatment groups

	Follow-up interviews per baseline household	**Interviewed at baseline and not re-interviewed in ...**				**Replacements**
		Any follow-up	Follow-up 1	Follow-up 3	Follow-up 5	Household is replacement
	(1)	(2)	(3)	(4)	(5)	(6)
Maintenance (T1)	0.029	0.004	0.013	-0.026	-0.016	0.008
	(0.072)	(0.011)	(0.022)	(0.037)	(0.035)	(0.015)
	[0.69]	[0.73]	[0.57]	[0.48]	[0.65]	[0.60]
Maintenance + sensitization (T2)	0.013	0.008	0.003	-0.014	-0.002	-0.000
	(0.078)	(0.014)	(0.021)	(0.041)	(0.034)	(0.014)
	[0.87]	[0.54]	[0.87]	[0.73]	[0.96]	[0.99]
T1 = T2 (p-value)	0.807	0.754	0.678	0.706	0.656	0.594
Attrition rate	2.575	0.025	0.090	0.194	0.142	0.161
Observations	1575	1575	1575	1575	1575	6711

Note. Figure B2 provides the timing of each follow-up survey. Dependent variables by column: (1) indicator variable equal to 1 if the household was interviewed at baseline and was not re-interviewed in any of the follow-ups, and zero otherwise; (2) indicator variable equal to 1 if the household was interviewed at baseline and was not re-interviewed in two out of three follow-ups, and 0 otherwise; (3)–(5) indicator variable equal to 1 if the household was interviewed at baseline and was not re-interviewed at follow-up 1 or follow-up 2 or follow-up 3, and 0 otherwise; (6) indicator variable equal to 1 if the household is part of the replacement sample (it was interviewed in any of the follow-ups, but it was not interviewed at baseline), and 0 otherwise. In columns (1)–(5), the sample is restricted to baseline observations, while in column (6) the sample is restricted to follow-up observations. All specifications include strata indicators for city and the provider of the CT. Standard errors clustered by catchment area are presented in parenthesis in columns (1)–(5). Standard errors clustered by catchment area and follow-up round are presented in parenthesis in column (6).

C.4.1 List randomization for reporting of sensitive behavior

The questionnaire for follow-up 5 was supplemented with a list randomization technique (see, i.e., Karlan and Zinman, 2012; Treibich and Lépine, 2019). Following this methodology, respondents were randomly allocated to one of four groups. Depending on the group, respondents received a different list of statements, and were asked to report how many of them were true. Table C5 provides the list of statements. Group A received only a list of statements related to general behavior. Groups B–D received the same list and one extra statement capturing sensitive behavior (OD, use of CT, or hand-washing). In aggregate, the difference between the mean number of true statements in the first group and in the other groups allows estimation of the proportion of the sample that engaged in each sensitive behavior.

when a share of slum residents goes back to their native villages.

Table C5: Statements used for list randomization

Group A	Group B	Group C	Group D
- I cooked yesterday - I bought milk yesterday - I watched TV yesterday	- I cooked yesterday - I bought milk yesterday - I watched TV yesterday - I defecated in the open yesterday	- I cooked yesterday - I bought milk yesterday - I watched TV yesterday - I used the CT to defecate yesterday	- I cooked yesterday - I bought milk yesterday - I watched TV yesterday - I washed my hands with soap yesterday

Note. Group A reports a list of statements related to general behavior. Groups B–D provide the same list, but adding one extra statement capturing sensitive behavior (OD, use of CT, or hand-washing).

C.5 Lab-in-the-field experiment: WTP for CT use

WTP for CT use is elicited to the respondent of the household survey and the spouse (up to two respondents per household). We measure WTP four times during the study in conjunction with the household survey. WTP is elicited using a standard incentivized version of the multiple price list (or take-it-or-leave-it) methodology (Andersen et al., 2006). Participants were prompted to choose between different amounts of cash (ranging from INR 0 to 60 with increases of INR 5) and a bundle of 10 tickets to use the CT in the catchment area where they live.[5] In total, participants face 13 combinations. After all choices are made, one of the options is then randomly selected by drawing a numbered ball from a bag, and the decisions are realized. Following the realization of the game, in the case of the bundle of tickets being assigned, the respondent could allocate the 10 tickets or some of them to either male or female use. Before participating in the game, the participant was introduced to a practice round of the game using a bar of soap to facilitate familiarity with the rules. The exact explanation of the game read by the enumerator to the participant was as follows:

> *Now let us do the prize draw for 10 tickets to use the [CT name]. These tickets are being officially provided by [CT name] as a promotion to encourage people to use the CT. They can be used at any time in the next 2 months. You will be given the choice later to decide how many of the 10 tickets you would like to be for men and boys, and how many you would like to be for women and girls. We are going to ask you to make a series of choices between either receiving these 10 tickets or instead receiving amounts of cash. At the end of all of the choices, you will draw a ball from a bag to determine which one of these choices will be randomly selected for your lucky draw – you will get the tickets or the money, depending on what you chose. This means that any one of the choices that you make could be selected at the end. Therefore it is in your best interest just to answer your honest opinion about which option you would prefer in every single choice.*

In conjunction with the incentivized version of the WTP elicitation, we also collected information from participants about the price that female and male residents face to use the CT, and we asked about WTP for the use of a hypothetical higher-quality CT in a non-incentivized setting. For the latter, participants were asked directly whether they would purchase a ticket for different amounts of money, ranging from INR 0 to 10, with increases of INR 1. The exact question reads as follows:

[5] During follow-up 5, following the introduction at the end of the study of monthly passes for families in a limited number of CTs, we also elicited the WTP for a monthly family pass for up to five members.

Now I want to tell you an imaginary story. Imagine that starting from tomorrow, the owners of the nearest CT decided to change the price for using the defecation cubicles. At the same time, they would improve the quality of the CT to the highest standard, ensuring it was very clean, had good hand-washing facilities, and that all the cubicles had a light and a lock. Would you be willing to buy a ticket to use the defecation cubicles of the community toilet, if the price was...

C.6 Lab-in-the-field experiment: adapted dictator game

To measure preference for maintenance among slum residents, we played an adapted dictator game in which participants are endowed with INR 50 and are given the option to donate all or part of it to a fund to purchase cleaning products for the CT. This component was administered to the respondent of the household survey and the spouse (up to two respondents per household), and measured in conjunction with each household survey. Having collected all the contributions to the cleanliness of the CT within each slum, the total amount was used to purchase cleaning products, which were then delivered to the caretaker. The design of this game is similar to the one proposed by Ashraf et al. (2014b), which has proven effective at measuring pro-social motivation. The exact setting reads as follows:

I would like to inform you that as an additional thank-you for participating in this study, you will receive an extra INR 50 in cash. We are asking all participants to choose between keeping some or all of this INR 50 for themselves, and donating some or all of this INR 50 for a special fund for cleaning products that we will deliver to the CT. How would you like to split the INR 50 between cash for yourself, and donation to the cleaning product fund for your CT?

Similarly, to measure pro-social motivation for the cause among caretakers, we implemented an adapted dictator game in which the caretaker is endowed with INR 50 and is given the option to donate all or part of it to fund a sanitation project implemented by our partner, FINISH Society. Pro-social motivation among caretakers was measured during each CT survey. Having collected the contributions from all caretakers, the total amount was donated to the FINISH Society project. The exact setting faced by the caretakers reads as follows:

I would like to inform you that as a thank-you for participating in this study, you will receive INR 100 in cash. You can keep the full amount for yourself or you have the opportunity to donate some or all of it to FINISH Society to help with improving water access, sanitation and hygiene in disadvantaged areas of India. How would you like to split the INR 100 between cash for yourself and donation to charity?

C.7 Lab-in-the-field experiment: public goods game (PGG)

To measure willingness to cooperate among the slum residents, we implemented a standard public goods game with the experiment participants. The game is based on the voluntary contribution mechanism, in which participants receive an endowment of INR 100, and they have to decide whether to keep the endowment or to invest part or all of it in a public pot. The contributions in the group are increased by a multiplier and shared equally among participants. The multiplier is randomly varied at catchment-area level to either double or triple the contributions. The game is designed so that while the total return to the investment in the pots is higher than the return from keeping the endowment, there is no incentive to invest in the former because of the higher individual pay-off that can be obtained from keeping the endowment. The dominant strategy is therefore to not contribute at all, while the social optimum is to invest in the pot. We played simultaneously with three groups of equal size (ranging from four to six participants) in each community. Participants also received INR 20 as show-up fee. The instructions given to groups of six participants are the following (note that x is either 2 or 3):

> *In this game, each player receives an endowment of INR 100 and you can choose to contribute (C) to the shared pot or keep (K) it. Out of the INR 100, you can decide how much to contribute and how much to keep. Secretly, you will put your donation amount in the pink envelope and the amount you want to keep in the blue envelope. All contributions will be summed and we will increase the total contribution by [x]. The final pot will be split equally among players. Let's look at some examples. If all 6 players contribute the INR 100, their individual payoffs would be equal to INR [$600 \cdot x/6$]; if one player contributes and other players keep the endowment, then the payoff of each player contributing is equal to INR [$pot \cdot x/6$], and the payoff of the player keeping is equal to INR [$100 + pot \cdot x/6$]; if all players keep, then their individual payoffs are INR 100.*

C.8 SCA: voice-to-the-people initiative

Figure C5 shows the card distributed to participants as part of this initiative asking about the most pressing issue in their community. Participants were invited to circle only one option among the following: children are frequently ill, water availability is limited, the community is dirty, the quality of roads is poor, there is no waste collection, the CT is dirty, jobs are missing, access to healthcare is limited, and lighting at night is poor. We used visual representations to facilitate selection of the issue among illiterate participants. Individual anonymized codes allowed the returned cards to be matched with the households in the sample. Two letters with the results for the corresponding cities were sent to the heads of sanitation and environment at the municipal corporations of Lucknow and Kanpur in October 2019. Descriptive statistics and treatment effect estimates related to this activity are reported in Appendix D.11. The specific instruction reads as follows:

> *We are collecting anonymous reports about the most pressing issue in your community. We will communicate this to the district municipal corporation to raise awareness among administrators. If you want to tell us your opinion, fill in the card and return to one of our team members. Your voice is important! In my community, I am most concerned about: [circle only one option]*

Figure C5: Card distributed for the voice-to-the-people initiative

Note. The figure shows the card distributed to participants as part of the voice-to-the-people initiative. The options read in Hindi: children frequently ill (A); limited water availability (B); community is dirty (C); poor quality of roads (D); no trash collection (E); Community toilet dirty (F); no jobs (G); limited access to healthcare (H); poor lighting at night (I). In each household, up to two participants had the option to mark one of the issues.

C.9 Additional data sources

Table C6 presents a description of additional data sources used in the paper.

Table C6: Additional data sources

Data (source)	Description
Basemaps (Esri)	Basemaps throughout the paper were created using ArcGIS® software by Esri®. Basemaps are used in line with the Esri Master License Agreement, specifically for the inclusion of screen captures in academic publications. We use the *World Light Gray Base* (sources: Esri, HERE, Garmin, ®OpenStreetMap contributors, and the GIS User Community).

C.10 Construction of indices

C.10.1 Awareness of health and safety risks

We construct an index capturing awareness of health and safety risks from open defecation or the use of a dirty community toilet. We use variables capturing the extent to which respondents would expect their community to be clean, healthy and safe if everybody practiced OD, or used a dirty or high-quality CT. In particular, we use principal components analysis, keeping the principal component for each family of outcomes and first eigenvalue loading. The index is standardized to range between 0 and 1 for the baseline and for the follow-up data collection. Table C7 report the list of variables contained in the index, how they were measured and the loading.

C.10.2 Quality of the CT

To construct a measure capturing the overall quality of the CT, we use all indicators from the observation of CTs (Appendix C.2) related to the facility's physical quality and cleanliness, and to laboratory tests that capture lack of bacteria (Appendix C.3). Since the overall quality of the CT is multidimensional and varies over time, we build an index using item response theory

Table C7: Indicators used for the construction of the awareness of health and safety risks index

Variables	**Measurement**	**Factor loading**
Disagrees with *Adult members in family would be healthier with OD*	Likert scale	0.1677
Disagrees with *Children in family would be healthier with OD*	Likert scale	0.1688
Agrees with *Family would have higher expenditure on healthcare with OD*	Likert scale	0.0959
Disagrees with *Community would be cleaner with OD*	Likert scale	0.1750
Disagrees with *Children in community would be healthier with OD*	Likert scale	0.1814
Agrees with *Water in community would be dirtier with OD*	Likert scale	0.0970
Disagrees with *Adults in community would be healthier with OD*	Likert scale	0.1788
Agrees with *Adult members in family would be healthier with high-quality CT*	Likert scale	0.1882
Agrees with *Children in family would be healthier with high-quality CT*	Likert scale	0.2023
Agrees with *Community would be cleaner with high-quality CT*	Likert scale	0.1952
Agrees with *Children in community would be healthier with high-quality CT*	Likert scale	0.1999
Agrees with *Adult in community would be healthier with high-quality CT*	Likert scale	0.1727
Adult women that would get sick if OD	0-10	0.1599
Adult men that would get sick if OD	0-10	0.1601
Girls that would get sick if OD	0-10	0.2076
Boys that would get sick if OD	0-10	0.2040
Adult women that would get sick if use CT and not OD	0-10	0.2020
Adult men that would get sick if use CT and not OD	0-10	0.2026
Girls that would get sick if use CT and not OD	0-10	0.2104
Boys that would get sick if use CT and not OD	0-10	0.2108
Adult women that would be unsafe if OD	0-10	0.0970
Adult women that would be unsafe if use CT and not OD	0-10	0.0820
Adult women that would get sick if use dirty CT	0-10	0.1961
Adult men that would get sick if use dirty CT	0-10	0.1940
Girls that would get sick if use dirty CT	0-10	0.2237
Boys that would get sick if use dirty CT	0-10	0.2228
Adult women that would get sick if use clean CT	0-10	0.1942
Adult men that would get sick if use clean CT	0-10	0.1917
Girls that would get sick if use clean CT	0-10	0.1922
Boys that would get sick if use clean CT	0-10	0.1878

(IRT) or latent trait analysis, a technique used to describe the relationship between individual responses to questionnaire items and an unobserved latent trait, generally a concept. As compared with other statistical models such as principal component analysis, it gives the advantage of providing information on the reliability of each individual item in the index (Gordon et al., 2012; Kline, 2014). We build the index of CT quality using a two parameter IRT model with the two parameters being an ability score, which could be used as a weight in constructing the index, and a discrimination score, which measures how well the indicator differentiates between low- and high-quality CTs. The index is re-scaled to be between 0 and 1, with 1 indicating the highest quality in the sample.[6] Table C8 provides the list of all indicators included, while Figure C6 shows the distribution of the resulting index by control/treatment group.

We build three separate indices using IRT to measure the physical quality of the facility, the cleanliness of the CT, and the lack of bacteria. For each component, Table C9 shows treatment effects on the indices and on indicator variables equal to 1 if the CT is in the top 25 percentiles of the sample distribution of each index. Figure C7 shows the effect on CT quality by component.

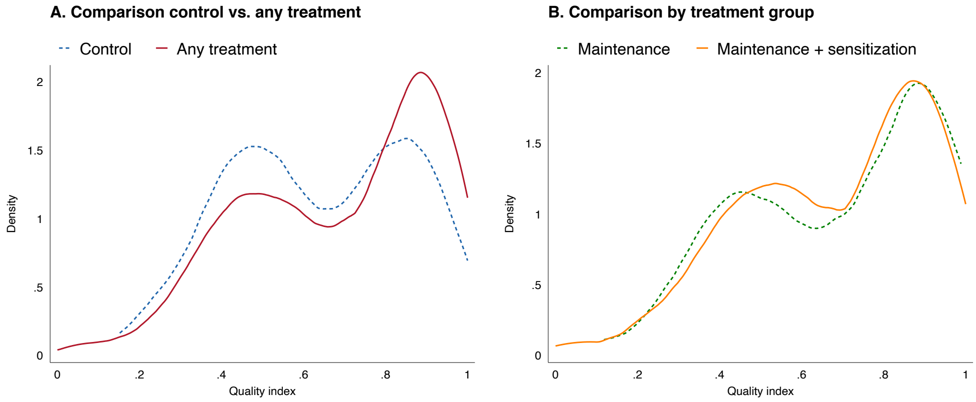
[6] We compute the index separately for baseline and for all follow-up measurements due to the fact that the baseline survey includes a lower number of indicators. At baseline, due to convergence, we adopt a one parameter IRT model.

Table C8: Indicators used for the construction of the quality index

Indicator variables	Ability score (1)	Discrimination (2)
Physical quality		
All cubicle doors are functioning	1.971	0.247
All locks are functioning	-0.603	0.435
Compound has finished walls	2.259	0.412
Internal walls are in good condition	3.156	0.294
Soap is available and visible for both genders	1.731	0.572
Hand-washing facility available for both genders	1.667	0.811
Female area has lighting	1.842	1.002
Male area has lighting	1.751	1.059
Common area has lighting	2.960	0.762
Cleanliness		
Toilets in female area are not dirty	0.699	3.705
Toilets in female area do not stink	0.640	4.121
Flies not present in the female area	0.837	3.904
Toilets in male area are not dirty	0.570	4.843
Toilets in male area do not stink	0.771	3.431
Flies not present in the male area	0.525	5.990
Feces not visible inside the latrine in the female area	1.009	5.186
Feces not visible outside the latrine in the female area	1.200	4.523
Feces not visible inside the latrine in the male area	0.987	3.699
Feces not visible outside the latrine in the male area	1.192	3.134
Common area is not dirty	1.276	2.924
Common area does not stink	1.254	3.254
Flies not present in the common area	1.272	2.764
No visible sewage leaks inside the compound	2.449	2.235
Lack of bacteria		
Bacteria count of E. coli is low	-0.379	-0.196
Bacteria of bacillus are not detected	2.148	-3.145
Bacteria of staphylococcus are not detected	-25.405	-0.097
Bacteria of salmonella are not detected	38.091	0.025
Bacteria of klebsiella are not detected	10.820	-0.123
Mold is not detected	3.537	-0.455

Note. All indicator variables are equal to 1 if the condition is true, and 0 otherwise. The table reports the main parameters in the index build using IRT, with the ability score reported in column (1) and the discrimination reported in column (2). Observations are restricted to follow-ups 1–5 for computing the index. The manual for observers defines the rules for the visual evaluation of CTs (Appendix C.2). *Finished walls* are defined as built in cement, and bricks, with no cracks or crumbles on the paintwork or tiles. *Dirt* is reported as the presence of mud, mold, red spitting, urine or feces on floors or walls. *Stink* is reported as the presence of an unpleasant smell from urine or feces. *Sewage leaks* are identified by fecally contaminated black waters leaking from a septic tank, pit/cesspool or pipes.

Figure C6: Distribution of the quality index at follow-up



The p-value of a Kolmogorov-Smirnov test of equality of distributions is equal to .007 for Panel A and .531 for Panel B.

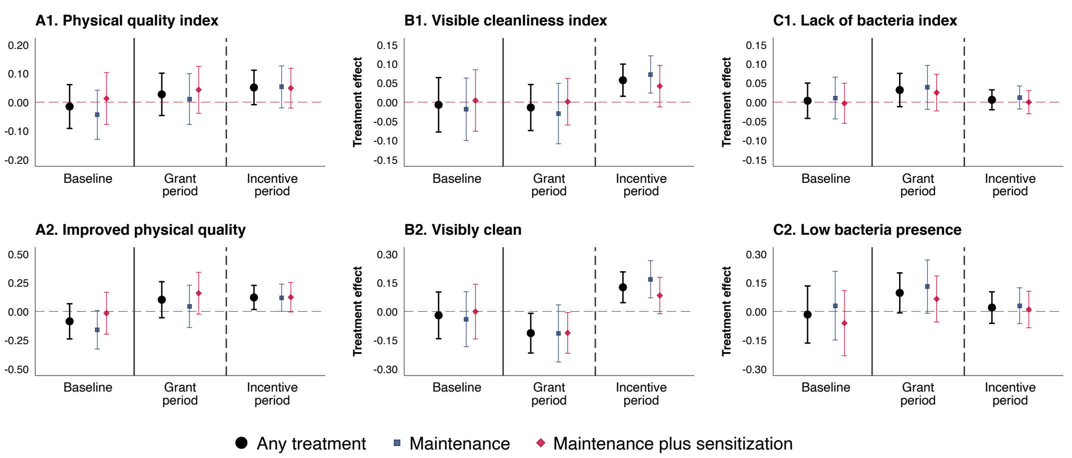
Note. The distributions include all follow-up measurements. The CT quality index is built using a two parameter IRT model (see Appendix C.10.2 for details about the construction). The distribution fits are estimated non-parametrically using kernel density estimation assuming an Epanechnikov kernel function. Bandwidths are estimated by Silvermans rule of thumb (Silverman, 1986). Panel A shows the comparison between the control group and any treatment group. Panel B shows the comparison between the two treatment groups individually (maintenance and maintenance plus sensitization). The two-sample Kolmogorov-Smirnov test of equality of distributions is performed using exact p-values and is equal to 0.007 for Panel A and 0.531 for Panel B.

Table C9: Effect on CT quality by indicator

	Physical quality		**Visible cleanliness**		**Lack of bacteria**	
	Index	Higher quality	Index	Higher quality	Index	Higher quality
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Any treatment (T)	0.046	0.117	0.044	0.080	0.011	0.035
	(0.036)	(0.062)	(0.023)	(0.044)	(0.014)	(0.045)
	[0.20]	[0.06]	[0.06]	[0.07]	[0.42]	[0.43]
Panel B						
Maintenance (T1)	0.045	0.103	0.055	0.116	0.017	0.049
	(0.043)	(0.072)	(0.029)	(0.056)	(0.016)	(0.050)
	[0.31]	[0.15]	[0.06]	[0.04]	[0.28]	[0.33]
Maintenance + sensitization (T2)	0.048	0.130	0.034	0.043	0.005	0.021
	(0.041)	(0.073)	(0.029)	(0.051)	(0.016)	(0.052)
	[0.24]	[0.08]	[0.25]	[0.40]	[0.75]	[0.68]
T1 = T2 (p-value)	0.943	0.717	0.525	0.239	0.468	0.577
Mean (control group)	0.649	0.309	0.751	0.487	0.427	0.492
Std. Dev. (control group)	0.242	0.463	0.265	0.501	0.172	0.501
Observations	542	542	542	542	542	542
Catchment areas	110	110	110	110	110	110

Note. Estimates based on CT-level OLS regressions using equation (1). Standard errors clustered by catchment area are reported in parentheses. *P*-values corresponding to testing that the individual coefficient is different from zero are presented in brackets. Dependent variables by column: (1) *Physical quality (index)*: index aggregating indicator variables about the physical quality of the facility; (2) *Physical quality (higher quality)*: indicator variable equal to 1 if the physical quality index is above the 75th percentile, and 0 otherwise; (3) *Visible cleanliness (index)*: index aggregating indicator variables about the visible cleanliness of the CT; (4) *Visible cleanliness (higher quality)*: indicator variable equal to 1 if the visible cleanliness index is above the 75th percentile, and 0 otherwise; (5) *Lack of bacteria (index)*: index aggregating indicator variables about the lack of bacteria; (6) *Lack of bacteria (higher quality)*: indicator variable equal to 1 if the lack of bacteria index is above the 75th percentile, and 0 otherwise. Individual indicators used for the creation of indices are presented in Appendix Table C8. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. Additional details about the dependent variables are presented in Appendix A.

Figure C7: Effect on CT quality by component: grant versus incentive period



Note. Each panel presents estimates of treatment effects based on OLS regressions using equation (1) at the CT level. Confidence intervals are built using statistical significance at the 10% level. *Baseline* includes the measurement at baseline, *Grant period* includes the measurement from follow-up 1, and *Incentive period* pools all subsequent follow-up measurements. See Section 1 for details about each intervention. When the regression is based on a single measurement period, robust standard errors are used. When multiple measurement periods are pooled, standard errors are clustered at the catchment area. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. Additional details about the variables are presented in Appendix C.10.2 and Appendix A.

D Additional analysis

D.1 Balance in observable characteristics

Tables D1 and D2 present the balance test for CT, caretaker and household characteristics.

Table D1: CT characteristics at baseline, by treatment group

	Descriptive statistics		**Differences from control group, by treatment group**			
	All	Control	Any treatment	Maintenance	Maintenance + sensiti-zation	P-value joint test (4)-(5)
	(1)	(2)	(3)	(4)	(5)	(6)
Year of construction	1996.98	1995.26	2.78	2.34	3.23	0.32
	[8.85]	[9.29]	(1.88)	(2.11)	(2.19)	
Distance to closest CT	0.54	0.58	-0.06	-0.04	-0.07	0.76
	[0.44]	[0.66]	(0.11)	(0.11)	(0.11)	
Surrounding market	0.33	0.35	-0.04	-0.01	-0.06	0.82
	[0.47]	[0.48]	(0.10)	(0.11)	(0.11)	
Surrounding road	0.84	0.88	-0.06	-0.05	-0.08	0.67
	[0.37]	[0.33]	(0.07)	(0.09)	(0.09)	
Surrounding government office	0.25	0.20	0.07	0.08	0.06	0.69
	[0.43]	[0.41]	(0.08)	(0.10)	(0.10)	
Only residents use CT	0.12	0.07	0.07	0.07	0.07	0.53
	[0.32]	[0.27]	(0.06)	(0.07)	(0.07)	
Single caretaker	0.80	0.82	-0.04	0.03	-0.11	0.28
	[0.40]	[0.39]	(0.07)	(0.08)	(0.09)	
Share of female caretakers	0.18	0.22	-0.06	-0.02	-0.10	0.42
	[0.37]	[0.39]	(0.07)	(0.08)	(0.08)	
Caretaker is also cleaner	0.27	0.28	-0.02	-0.02	-0.03	0.96
	[0.45]	[0.46]	(0.09)	(0.10)	(0.10)	
Caretaker is from local community	0.44	0.49	-0.07	-0.11	-0.02	0.60
	[0.50]	[0.51]	(0.10)	(0.12)	(0.12)	
Caretaker's experience (months)	125.28	129.91	-5.43	1.37	-11.53	0.86
	[103.45]	[109.34]	(22.81)	(26.60)	(25.96)	
CT is cleaned frequently	0.86	0.87	-0.02	-0.02	-0.02	0.97
	[0.35]	[0.34]	(0.07)	(0.08)	(0.08)	
Time allocated to managing	0.68	0.66	0.03	0.03	0.02	0.58
	[0.14]	[0.11]	(0.03)	(0.03)	(0.03)	
Capacity	13.00	13.21	-0.32	-0.46	-0.17	0.94
	[5.57]	[5.52]	(1.11)	(1.27)	(1.34)	
Daily opening hours	17.76	17.88	-0.19	-0.35	-0.02	0.53
	[1.49]	[1.59]	(0.28)	(0.36)	(0.27)	
Share of functioning toilets	0.90	0.88	0.03	0.05	0.01	0.47
	[0.22]	[0.23]	(0.04)	(0.04)	(0.05)	
WTP (avg. catchment area)	1.41	1.44	-0.05	-0.03	-0.06	0.95
	[0.83]	[0.65]	(0.15)	(0.17)	(0.20)	
Distance from CT (avg. catchment area)	128.71	128.77	-0.01	-2.22	2.21	0.94
	[49.56]	[43.87]	(9.26)	(10.21)	(12.25)	

Note. Columns (1) and (2) report sample mean with standard deviation in brackets for the whole sample and for the control group, respectively. Column (3) reports the difference from the control group with any treatment group. Columns (4) and (5) report the difference from the control group for each treatment group. Differences in columns (3)–(5) are estimated using OLS and controlling for strata indicators for city and the provider of the CT. Robust standard errors are reported in parentheses. Column (6) presents a joint test of significance of the coefficients for each treatment dummy. Statistical significance denoted by *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table D2: Household characteristics at baseline, by treatment group

	Descriptive statistics		Differences from control group, by treatment group			
	All	Control	Any treatment	Maintenance	Maintenance + sensitization	P-value joint test (4)-(5)
	(1)	(2)	(3)	(4)	(5)	(6)
Household head is male	0.75	0.73	0.03	0.05	0.01	0.28
	[0.43]	[0.44]	(0.02)	(0.03)	(0.03)	
Household head is married	0.77	0.76	0.01	0.01	0.01	0.89
	[0.42]	[0.43]	(0.03)	(0.03)	(0.03)	
Age of household head	45.43	46.02	-0.84	-0.87	-0.82	0.57
	[12.82]	[13.42]	(0.79)	(0.97)	(0.86)	
Age of spouse	39.13	39.38	-0.31	-0.74	0.10	0.60
	[11.39]	[11.99]	(0.76)	(0.94)	(0.78)	
Household head has no education	0.54	0.55	-0.02	-0.07	0.03	0.04
	[0.50]	[0.50]	(0.04)	(0.05)	(0.04)	
Spouse has no education	0.45	0.45	-0.00	0.01	-0.01	0.92
	[0.50]	[0.50]	(0.03)	(0.04)	(0.04)	
Household members	4.94	4.94	0.00	0.01	0.00	1.00
	[1.99]	[2.08]	(0.13)	(0.15)	(0.14)	
Household members (0-5 y.o.)	0.47	0.50	-0.05	-0.05	-0.06	0.64
	[0.77]	[0.82]	(0.06)	(0.06)	(0.07)	
Household members (older than 5 y.o.)	4.47	4.44	0.06	0.04	0.07	0.86
	[1.83]	[1.92]	(0.11)	(0.13)	(0.12)	
Muslim	0.17	0.12	0.08**	0.11*	0.06	0.13
	[0.37]	[0.32]	(0.04)	(0.06)	(0.04)	
Spent on religious items	0.94	0.94	-0.01	-0.01	-0.00	0.83
	[0.25]	[0.24]	(0.01)	(0.02)	(0.02)	
General caste	0.07	0.05	0.03	0.03	0.02	0.29
	[0.26]	[0.23]	(0.02)	(0.02)	(0.02)	
Asset index	0.53	0.53	0.00	0.01	-0.00	0.78
	[0.15]	[0.16]	(0.02)	(0.02)	(0.02)	
Household members per room	3.99	3.90	0.13	0.05	0.21	0.31
	[1.86]	[1.93]	(0.14)	(0.16)	(0.15)	
Access to piped water	0.71	0.70	0.01	-0.02	0.04	0.68
	[0.45]	[0.46]	(0.05)	(0.06)	(0.06)	
Access to private toilet	0.08	0.07	0.01	0.01	0.02	0.68
	[0.27]	[0.26]	(0.02)	(0.02)	(0.02)	
Expenditure on CT use (INR)	180.53	173.42	10.92	-2.71	23.88	0.65
	[244.52]	[221.41]	(23.01)	(22.99)	(30.61)	
Prevalence of diarrhea (last 15 days)	0.08	0.07	0.02	0.01	0.03	0.25
	[0.28]	[0.26]	(0.02)	(0.02)	(0.02)	
Prevalence of fever (last 15 days)	0.18	0.18	-0.01	-0.02	-0.01	0.85
	[0.38]	[0.39]	(0.02)	(0.03)	(0.03)	
Distance to CT (meters)	126.13	126.26	-1.05	-2.12	-0.04	0.97
	[79.89]	[80.43]	(8.73)	(9.63)	(11.54)	

Note. Columns (1) and (2) report sample mean with standard deviation in brackets for the whole sample and for the control group, respectively. Column (3) reports the difference from the control group with any treatment group. Columns (4) and (5) report the difference from the control group for each treatment group. Differences in columns (3)–(5) are estimated using OLS and controlling for strata indicators for city and the provider of the CT. Standard errors clustered at slum level are reported in parentheses. Column (6) presents a joint test of significance of the coefficients for each treatment dummy. Statistical significance denoted by *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

D.2 Self-reported sanitation and hygiene behavior

Self-reported sanitation behavior was measured by asking survey respondents where each demographic group went to defecate the last two times. To prevent under-reporting of open defecation due to social stigma, we included the following prelude: ‘I’ve been to many similar communities and I’ve seen that even people owning latrines and having nearby community toilets defecate in the open.’ We also asked about the sanitation behavior of the most intimate neighbor in order to identify the extent of response bias when asked to disclose the behavior of their own household (Yeatman and Trinitapoli, 2011). Table D3 shows estimates of treatment effects on open defecation in Columns (1)–(6), and on hand-washing with soap in Column (7).

Table D3: Self-reported sanitation and hygiene behavior, by demographic group

	Open defecation						**Soap**
	Respondent	Spouse	Male > 14y	Female > 14y	Male 6–14y	Female 6–14y	Respondent
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A							
Any treatment (T)	0.015	-0.013	-0.011	-0.010	0.004	-0.001	-0.004
	(0.007)	(0.025)	(0.022)	(0.015)	(0.013)	(0.012)	(0.014)
	[0.04]	[0.59]	[0.60]	[0.49]	[0.79]	[0.91]	[0.78]
Panel B							
Maintenance (T1)	0.018	0.017	0.021	0.006	0.022	-0.001	0.006
	(0.007)	(0.032)	(0.028)	(0.020)	(0.018)	(0.014)	(0.019)
	[0.02]	[0.59]	[0.45]	[0.75]	[0.23]	[0.94]	[0.75]
Maintenance + sensitization (T2)	0.012	-0.043	-0.042	-0.026	-0.014	-0.001	-0.013
	(0.008)	(0.027)	(0.023)	(0.017)	(0.014)	(0.014)	(0.016)
	[0.14]	[0.11]	[0.07]	[0.11]	[0.33]	[0.92]	[0.39]
T1 = T2 (p-value)	0.290	0.075	0.025	0.114	0.063	0.984	0.331
Mean (control group)	0.967	0.147	0.116	0.091	0.059	0.058	0.089
Std. Dev. (control group)	0.179	0.354	0.320	0.287	0.235	0.233	0.284
Observations	9780	9780	8052	9780	9780	9780	9780
Catchment areas	110	110	110	110	110	110	110

Note. Estimates based on household-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B. Standard errors clustered by catchment area–round are reported in parentheses and p-values in brackets. Dependent variables by column: (1)–(6) *Open defecation*: is an indicator variable equal to 1 if the household member (by demographic group) is reported to have practiced open defecation the last time they defecated, and 0 otherwise; (7) *Soap*: is an indicator variable equal to 1 if the respondent reports washing her/his hands with soap, and 0 otherwise. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. Additional details about the variables are presented in Appendix A.

D.3 Estimates of treatment effects using ANCOVA and IPW specifications

Tables D4–D7 present estimates of treatment effects using equations (1) and (2) adding the value at baseline of the dependent variable as a control variable (ANCOVA specification). The organization of the results and the order of the variables are the same as in Tables 1–4 in the main text. Tables D8 and D9 present estimates of treatment effects using equations (1) and (2) weighting observations by inverse probability weights (Wooldridge, 2002). Weights are estimated at baseline using a probit regression on indicator variables for attrition at different follow-ups on observable characteristics of the household and of the catchment area where the household resides.

Table D4: CT-level outcomes: ANCOVA specification

	Maintenance and quality of the facility			Traffic during rush hour	
	Physical maintenance	Quality	Higher quality	Users	Free riding
	(1)	(2)	(3)	(4)	(5)
Panel A					
Any treatment (T)	0.059	0.045	0.109	-0.051	-0.081
	(0.049)	(0.022)	(0.043)	(0.049)	(0.038)
	[0.23]	[0.04]	[0.01]	[0.30]	[0.04]
Panel B					
Maintenance (T1)	0.044	0.053	0.128	-0.085	-0.074
	(0.052)	(0.027)	(0.055)	(0.060)	(0.043)
	[0.40]	[0.05]	[0.02]	[0.16]	[0.09]
Maintenance + sensitization (T2)	0.074	0.038	0.090	-0.018	-0.087
	(0.056)	(0.026)	(0.049)	(0.055)	(0.045)
	[0.19]	[0.15]	[0.07]	[0.74]	[0.05]
T1 = T2 (p-value)	0.507	0.599	0.521	0.272	0.754
Observations	536	536	536	536	536
ANCOVA specification	Yes	Yes	Yes	Yes	Yes
Follow-ups	1–5	1–5	1–5	1–5	1–5

Note. Estimates based on CT-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B, controlling for the baseline value of the dependent variable if available (see *ANCOVA specification* row). Standard errors clustered by catchment area are reported in parentheses. P-values are presented in brackets. The dependent variables are indicated in column headers and are defined in Appendix A. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT.

Table D5: Caretaker's behavior: ANCOVA specification

	Labor supply		Routine maintenance		Motivation
	Hours worked	Time allocated to managing	Awareness	Inputs	Pro-social motivation for the cause
	(1)	(2)	(3)	(4)	(5)
Panel A					
Any treatment (T)	0.190	0.046	0.089	0.033	-0.032
	(0.318)	(0.027)	(0.035)	(0.014)	(0.023)
	[0.55]	[0.09]	[0.01]	[0.02]	[0.17]
Panel B					
Maintenance (T1)	-0.020	0.034	0.076	0.036	-0.033
	(0.393)	(0.031)	(0.043)	(0.018)	(0.029)
	[0.96]	[0.29]	[0.08]	[0.04]	[0.25]
Maintenance + sensitization (T2)	0.396	0.059	0.102	0.030	-0.031
	(0.341)	(0.029)	(0.049)	(0.015)	(0.028)
	[0.25]	[0.04]	[0.04]	[0.05]	[0.26]
T1 = T2 (p-value)	0.265	0.309	0.663	0.712	0.959
Observations	536	536	542	536	536
ANCOVA specification	Yes	Yes	-	Yes	Yes
Follow-ups	1–5	1–5	1–5	1– 5	1–5
incentivized measurement	-	-	-	-	Yes

Note. Estimates based on CT-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B, controlling for the baseline value of the dependent variable if available (see *ANCOVA specification* row). Standard errors clustered by catchment area are reported in parentheses. P-values are presented in brackets. The dependent variables are indicated in column headers and are defined in Appendix A. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT.

Table D6: Valuation and attitudes towards the public good: ANCOVA specification

	Valuation		**Attitudes toward the public good**		
	WTP for CT use	Perceived improvement	Demand for public intervention	Contribution in the PGG	Preference for maintenance
	(1)	(2)	(3)	(4)	(5)
Panel A					
Any treatment (T)	-0.035	0.027	0.053	0.001	-0.004
	(0.064)	(0.015)	(0.026)	(0.013)	(0.006)
	[0.58]	[0.08]	[0.04]	[0.92]	[0.52]
Panel B					
Maintenance (T1)	0.041	0.031	0.055	-0.002	-0.007
	(0.082)	(0.018)	(0.032)	(0.015)	(0.006)
	[0.62]	[0.09]	[0.09]	[0.89]	[0.29]
Maintenance + sensitization (T2)	-0.109	0.023	0.051	0.005	-0.001
	(0.068)	(0.019)	(0.032)	(0.016)	(0.007)
	[0.11]	[0.22]	[0.11]	[0.77]	[0.90]
T1 = T2 (p-value)	0.067	0.717	0.911	0.665	0.313
Observations	8808	4890	1580	1228	8808
ANCOVA specification	Yes	Yes	-	-	Yes
Follow-ups	1, 3, 5	1, 3, 5	3	5	1, 3, 5
Level of analysis	Respondent	Household	Household	Respondent	Respondent
incentivized measurement	Yes	-	Yes	Yes	Yes

Note. Estimates based on respondent- and household-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B, controlling for the baseline value of the dependent variable if available (see *ANCOVA specification* row). When the baseline value is missing, we impute it with the baseline average of the dependent variable in the corresponding catchment area. Standard errors clustered by catchment area–round are reported in parentheses. P-values are presented in brackets. The dependent variables are indicated in column headers and are defined in Appendix A. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. Specifications where the level of analysis is the respondent also include gender.

Table D7: Awareness of risks associated with unsafe sanitation, and health: ANCOVA specification

	Awareness of unsafe sanitation risks		**Morbidity and health expenditure**		
	Health and safety risks	Externalities	Morbidity	Expenditure (extensive)	Expenditure (intensive)
	(1)	(2)	(3)	(4)	(5)
Panel A					
Any treatment (T)	0.011	0.029	0.009	0.051	-43.026
	(0.005)	(0.018)	(0.019)	(0.023)	(187.352)
	[0.03]	[0.10]	[0.65]	[0.03]	[0.82]
Panel B					
Maintenance (T1)	0.006	0.009	0.013	0.043	18.040
	(0.007)	(0.021)	(0.022)	(0.028)	(222.175)
	[0.35]	[0.67]	[0.55]	[0.12]	[0.94]
Maintenance + sensitization (T2)	0.016	0.048	0.004	0.059	-101.596
	(0.006)	(0.020)	(0.021)	(0.024)	(221.133)
	[0.01]	[0.02]	[0.85]	[0.02]	[0.65]
T1 = T2 (p-value)	0.122	0.065	0.657	0.507	0.615
Observations	4757	4890	4890	3332	3332
ANCOVA specification	Yes	Yes	Yes	Yes	Yes
Follow-ups	1, 3, 5	1, 3, 5	1, 3, 5	3, 5	3, 5
Level of analysis	Household	Household	Household	Household	Household

Note. Estimates based on household-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B, controlling for the baseline value of the dependent variable if available (see *ANCOVA specification* row). Standard errors clustered by catchment area–round are reported in parentheses. P-values are presented in brackets. The dependent variables are indicated in column headers and are defined in Appendix A. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT.

Table D8: Valuation and attitudes towards the public good: IPW specification

	Valuation		**Attitudes toward the public good**		
	WTP for CT use	Perceived improvement	Demand for public intervention	Contribution in the PGG	Preference for maintenance
	(1)	(2)	(3)	(4)	(5)
Panel A					
Any treatment (T)	-0.046	0.027	0.054	0.007	-0.007
	(0.066)	(0.015)	(0.026)	(0.014)	(0.006)
	[0.49]	[0.08]	[0.04]	[0.61]	[0.27]
Panel B					
Maintenance (T1)	0.044	0.032	0.054	0.003	-0.010
	(0.085)	(0.018)	(0.032)	(0.015)	(0.007)
	[0.60]	[0.08]	[0.09]	[0.82]	[0.13]
Maintenance + sensitization (T2)	-0.132	0.022	0.053	0.010	-0.003
	(0.071)	(0.019)	(0.033)	(0.016)	(0.007)
	[0.06]	[0.24]	[0.11]	[0.53]	[0.62]
T1 = T2 (p-value)	0.038	0.626	0.992	0.679	0.239
Observations	8808	4890	1580	1228	8808
Follow-ups	1, 3, 5	1, 3, 5	3	5	1, 3, 5
Level of analysis	Respondent	Household	Household	Respondent	Respondent
incentivized measurement	Yes	-	Yes	Yes	Yes

Note. Estimates based on respondent- and household-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B, weighting observations by inverse probability weights. Standard errors clustered by catchment area–round are reported in parentheses. P-values are presented in brackets. The dependent variables are indicated in column headers and are defined in Appendix A. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. Specifications where the level of analysis is the respondent also include gender.

Table D9: Awareness of risks associated with unsafe sanitation and health: IPW specification

	Awareness of unsafe sanitation risks		**Morbidity and health expenditure**		
	Health and safety risks	Externalities	Morbidity	Expenditure (extensive)	Expenditure (intensive)
	(1)	(2)	(3)	(4)	(5)
Panel A					
Any treatment (T)	0.011	0.028	0.008	0.050	-68.495
	(0.005)	(0.018)	(0.019)	(0.023)	(196.104)
	[0.04]	[0.11]	[0.68]	[0.03]	[0.73]
Panel B					
Maintenance (T1)	0.005	0.008	0.012	0.042	23.770
	(0.007)	(0.021)	(0.022)	(0.028)	(230.495)
	[0.42]	[0.71]	[0.59]	[0.14]	[0.92]
Maintenance + sensitization (T2)	0.016	0.048	0.004	0.058	-157.068
	(0.006)	(0.020)	(0.021)	(0.024)	(230.877)
	[0.01]	[0.02]	[0.86]	[0.02]	[0.50]
T1 = T2 (p-value)	0.093	0.065	0.710	0.485	0.460
Observations	4757	4890	4890	3332	3332
Follow-ups	1, 3, 5	1, 3, 5	1, 3, 5	3, 5	3, 5
Level of analysis	Household	Household	Household	Household	Household

Note. Estimates based on household-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B, weighting observations by inverse probability weights. Standard errors clustered by catchment area–round are reported in parentheses. P-values are presented in brackets. The dependent variables are indicated in column headers and are defined in Appendix A. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT.

D.4 Robustness to the inclusion of control variables

Table D10 presents estimates of the effect of any treatment (T) using equation (1) in Columns (1)–(3), and the post-double selection LASSO (PDSL) procedure (Tibshirani, 1996; Belloni et al., 2013) in columns (4)–(6). The PDSL procedure provides a method for model selection in the presence of a large number of control variables. To build the set of potential control variables, we include the following observable characteristics in the procedure (all continuous variables are also included in their squared term and are standardized):

- **CT characteristics**: variables describing the facility at baseline included in Table D1;
- **Caretaker characteristics**: variables related to caretakers at baseline included in Table D1;
- **Catchment area characteristics**: for CT- and caretaker-level outcomes, we include the catchment-area average at baseline for the household head's gender, education, marital status, religion and caste, WTP for CT use, trust of the community, bacteria contamination in water sources, share practicing OD, and distance from the CT.
- **Individual characteristics**: for household- and respondent-level outcomes, we include the baseline characteristics of the household and of the respondent included in Table D2;
- **Outcome variables**: for CT- and caretaker-level estimates, we include the baseline value of outcomes presented in Tables 1–2, while for household- and respondent-level outcomes, the baseline values of outcomes are presented in Tables 3–4.

Table D10: Effect of any treatment: comparison between main estimates and PDSL

	No control variables			Post-double selection LASSO			
	β	se	p-value	β	se	p-value	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CT-/caretaker-level outcomes							
Physical maintenance	0.04	0.05	0.36	0.04	0.05	0.45	542
Quality	0.05	0.02	0.03	0.04	0.02	0.07	542
Higher quality	0.12	0.04	0.01	0.09	0.04	0.02	542
Users during rush hour	-0.06	0.05	0.24	-0.06	0.05	0.21	542
Free riding	-0.08	0.04	0.07	-0.07	0.04	0.10	542
Hours worked	0.17	0.36	0.62	0.14	0.36	0.69	542
Time allocated to managing	0.05	0.03	0.07	0.05	0.02	0.05	542
Awareness of routine maintenance	0.09	0.04	0.01	0.09	0.03	0.01	542
Inputs in routine maintenance	0.03	0.02	0.05	0.03	0.01	0.03	542
Pro-social motivation for the cause	-0.02	0.02	0.33	-0.03	0.02	0.29	542
Household-/respondent-level outcomes							
WTP for CT use	-0.04	0.06	0.57	0.04	0.07	0.61	8808
Perceived improvement in the facility	0.03	0.02	0.08	0.04	0.02	0.01	4890
Demand for public intervention	0.05	0.03	0.04	0.05	0.03	0.04	1580
PGG contribution	0.00	0.01	0.92	0.01	0.01	0.57	1228
Preference for maintenance	-0.01	0.01	0.39	-0.00	0.01	0.68	8808
Awareness of health and safety risks	0.01	0.01	0.03	0.01	0.01	0.03	9514
Awareness of externalities	0.03	0.02	0.10	0.03	0.02	0.11	9780
Morbidity	0.01	0.02	0.63	0.01	0.02	0.60	9780
Health expenditure (extensive margin)	0.05	0.02	0.03	0.05	0.02	0.06	6664
Health expenditure (intensive margin)	-43.86	187.67	0.82	-28.51	208.33	0.89	6664

Note. Columns (1)–(3) show estimates using equation (1), while columns (4)–(6) show estimates using the PDSL procedure (Tibshirani, 1996; Belloni et al., 2013), with selection over a large number of baseline-level control variables. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. N indicates the sample size. In order to have the same sample size of estimates as in the main tables, missing values are replaced by the value 0 and an indicator variable equal to 1 if the observation had a missing value is introduced for all variables. Additional information about outcome variables is provided in Appendix A.

D.5 Robustness to estimation of treatment effects via causal forest

To verify the robustness of estimates to potential selection into treatment and to check the presence of heterogeneous impacts, we present estimates of ATE of any treatment on all outcome variables using the causal forest procedure of Wager and Athey (2018) and Athey et al. (2019). Specifically, we follow the cluster-robust procedure of Basu et al. (2018) and Athey and Wager (2019). In the procedure, we use the set of variables from Appendix D.4. Table D11 presents the results. Columns (1)–(3) present estimates of the ATE and the p-value of a two-sided test for the ATE being different from 0. Results are in line with those presented in the main text. In addition, to verify the overall presence of heterogeneity in the impacts, Columns (4)–(5) implement a calibration test based on the best linear predictor method of Chernozhukov et al. (2017). Column (4) presents the p-value for the equality to 1 of the coefficient on the mean forest prediction, with 1 indicating that the mean forest prediction is correct. Column (5) presents the p-value for the equality to 1 of the coefficient on the quality of the estimates of treatment heterogeneity, with 1 indicating that the forest has captured heterogeneity in the underlying signal. While the model is performing relatively well in predicting outcome variables, we do not highlight heterogeneity in treatment effects for most outcomes.

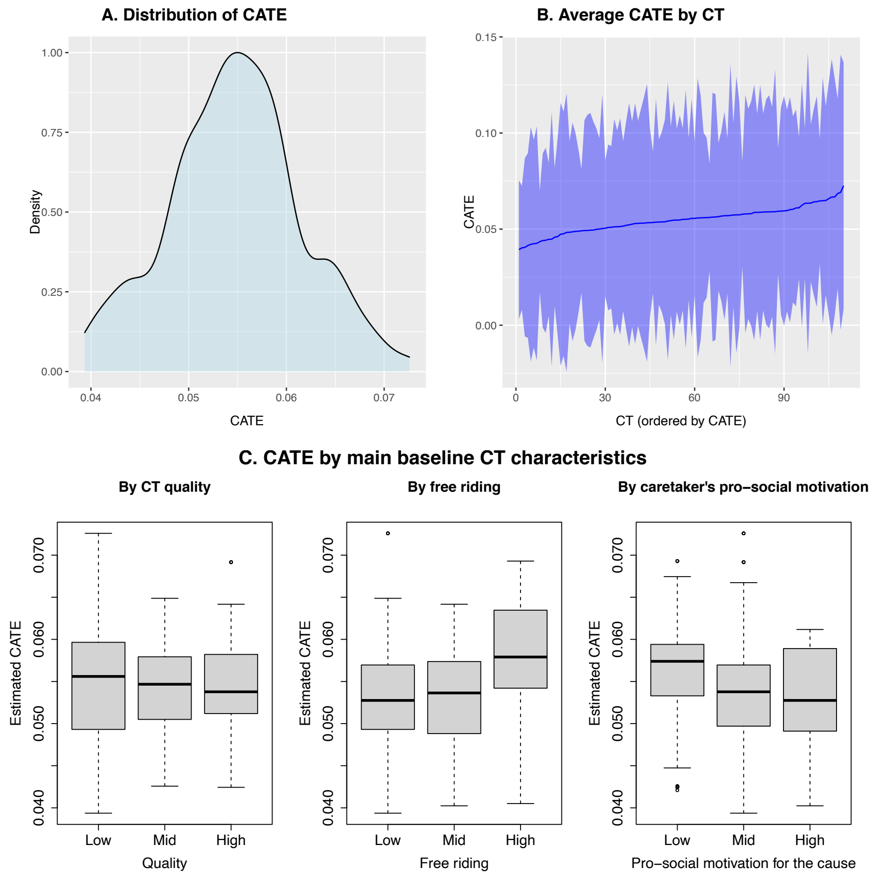
Table D11: Treatment effects of any treatment using cluster-robust causal forest procedure

	ATE via causal forest procedure			**Calibration test**	
	β	se	p-value	Mean prediction (p-value)	Heterogeneity (p-value)
	(1)	(2)	(3)	(4)	(5)
CT-/caretaker-level outcomes					
Physical maintenance	0.039	0.050	0.435	0.103	1.000
Quality	0.054	0.027	0.042	0.010	1.000
Higher quality	0.107	0.044	0.014	0.000	1.000
Users during rush hour	-0.046	0.050	0.356	0.086	1.000
Free riding	-0.094	0.043	0.029	0.002	1.000
Hours worked	0.442	0.456	0.333	0.182	0.945
Time allocated to managing	0.044	0.029	0.129	0.065	1.000
Awareness of routine maintenance	0.097	0.037	0.008	0.002	0.999
Inputs in routine maintenance	0.036	0.016	0.027	0.000	1.000
Pro-social motivation for the cause	-0.024	0.025	0.343	0.127	1.000
Household-/respondent-level outcomes					
WTP for CT use	-0.031	0.070	0.662	0.350	0.887
Perceived improvement in the facility	0.041	0.017	0.013	0.008	1.000
Demand for public intervention	0.049	0.029	0.087	0.041	0.192
PGG contribution	0.006	0.014	0.658	0.305	0.895
Preference for maintenance	-0.006	0.006	0.300	0.214	0.047
Awareness of health and safety risks	0.016	0.006	0.006	0.001	1.000
Awareness of externalities	0.026	0.020	0.205	0.190	1.000
Morbidity	0.010	0.020	0.620	0.352	0.784
Health expenditure (extensive margin)	0.05	0.025	0.046	0.035	0.981
Health expenditure (intensive margin)	-28.849	187.107	0.877	0.366	1.000

Note. Estimates of the ATE of any treatment on the outcome variables presented in the first column based on the cluster-robust causal forest procedure of Basu et al. (2018) and Athey and Wager (2019). We use the set of variables used in Appendix D.4, and we maintain the same assumptions about clustering implemented in Tables 1–4. Columns (1)–(3) present estimates of the ATE and the p-value of a two-sided test for the ATE being different from zero. Columns (4)–(5) implement a calibration test based on the best linear predictor method of Chernozhukov et al. (2017). Column (4) presents the p-value for the equality to 1 of the coefficient on the mean forest prediction, with 1 indicating that the mean forest prediction is correct. Column (5) presents the p-value for the equality to 1 of the coefficient on the quality of the estimates of treatment heterogeneity, with 1 indicating that the forest has captured heterogeneity in the underlying signal. Additional information about outcome variables is provided in Appendix A.

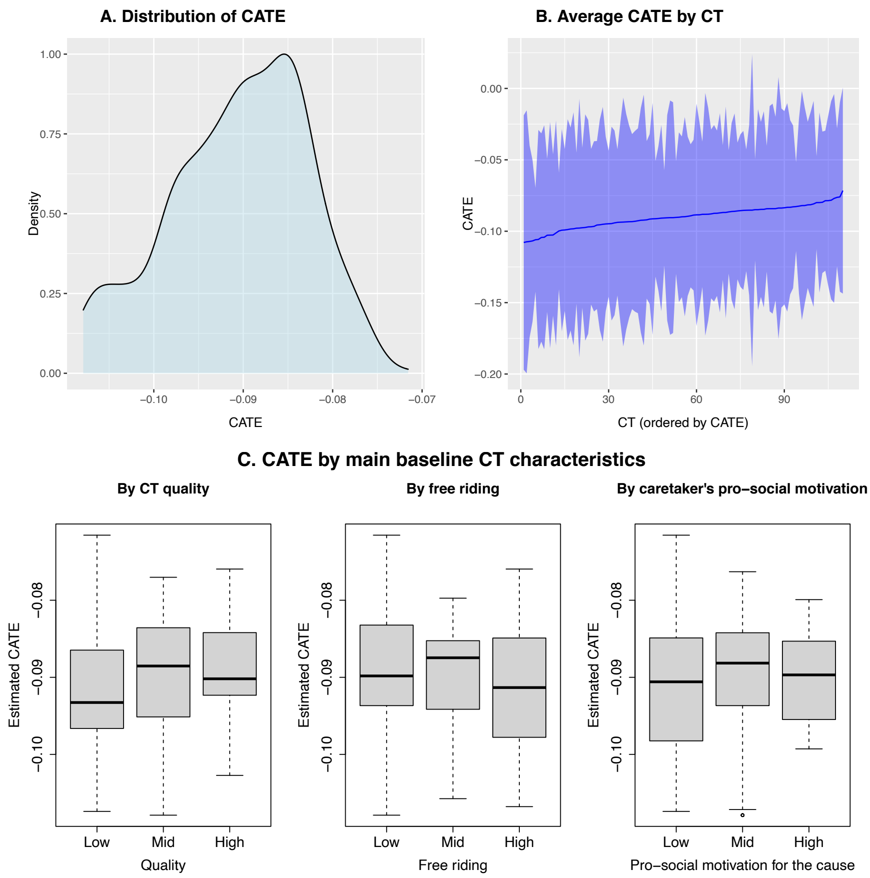
Figures D1 and D2 summarizes the causal forest results on heterogeneity of the effect on the quality of the facility and on free riding. Panel A shows the distribution of the Conditional ATE (CATE), while Panel B averages the CATE at CT level and includes the 90% confidence interval. Panel C shows instead how CATE estimates vary according to three baseline characteristics of the CT: quality, free riding, and caretaker's pro-social motivation for the cause (see Appendix A for the definition of these variables). Results show the relatively homogeneous impact of the interventions.

Figure D1: Conditional ATE of any treatment on quality



Note. Panel A shows the distribution of the Conditional ATE (CATE) of any treatment on quality computed using the cluster-robust causal forest procedure of Basu et al. (2018) and Athey and Wager (2019). Panel B shows the average CATE at CT level with the 90% confidence interval. Panel C shows variation of the CATE by baseline characteristics of the facility using a box plot. Low, mid and high indicates the first, second and third terciles in the distribution of the characteristic. In the box plot, each rectangle represents the inter-quartile range, with the top indicating the upper quartile, the bottom the lower quartile, and the middle line the median. The vertical line indicates the whiskers, i.e., the smallest value greater than the lower quartile minus 1.5 times the inter-quartile range, and the largest value less than the upper quartile plus 1.5 times the inter-quartile range. Additional information about the variables is provided in Appendix A.

Figure D2: Conditional ATE of any treatment on free riding



Note. Panel A shows the distribution of the Conditional ATE (CATE) of any treatment on free riding computed using the cluster-robust causal forest procedure of Basu et al. (2018) and Athey and Wager (2019). Panel B shows the average CATE at CT level with the 90% confidence interval. Panel C shows variation of the CATE by baseline characteristics of the facility using a box plot. Low, mid and high indicates the first, second and third terciles in the distribution of the characteristic. In the box plot, each rectangle represents the inter-quartile range, with the top indicating the upper quartile, the bottom the lower quartile, and the middle line the median. The vertical line indicates the whiskers, i.e., the smallest value greater than the lower quartile minus 1.5 times the inter-quartile range, and the largest value less than the upper quartile plus 1.5 times the inter-quartile range. Additional information about the variables is provided in Appendix A.

D.6 Spillover analysis

Table D12 shows a test for contagion or spillover effects by estimating heterogeneous treatment effects according to the average distance of a CT or catchment area to another treated CT or catchment area. We define a catchment area to be close to (far from) another treated catchment area if the distance is below or equal to (above) the sample median. Among all outcome variables, we do not observe any heterogeneous effect, suggesting the absence of spillover effects.

Table D12: Contagion and spillover effects

	Effect of any treatment, by average distance to another treatment						**Het. test**
	β	se	N	β	se	N	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Close to T			**Far from T**			
Physical maintenance	0.02	0.05	266	0.07	0.09	276	0.62
Quality	0.04	0.03	266	0.06	0.04	276	0.61
Higher quality	0.12*	0.06	266	0.12*	0.06	276	0.97
Users during rush hour	-0.04	0.07	265	-0.06	0.07	277	0.85
Free riding	-0.08*	0.04	265	-0.07	0.07	277	0.89
Hours worked	-0.25	0.40	266	0.66	0.59	276	0.20
Time allocated to managing	0.01	0.02	266	0.10*	0.05	276	0.13
Awareness of routine maintenance	0.08***	0.03	266	0.09	0.06	276	0.95
Inputs in routine maintenance	0.03	0.02	266	0.04	0.03	276	0.86
Pro-social motivation for the cause	-0.02	0.04	266	-0.03	0.03	276	0.80
WTP for CT use	-0.07	0.09	4406	-0.00	0.08	4402	0.69
Perceived improvement in the facility	0.03	0.02	2482	0.03	0.02	2408	0.94
Demand for public intervention	0.05	0.04	814	0.05	0.03	766	0.86
PGG contribution	-0.00	0.02	604	0.00	0.02	624	0.87
Preference for maintenance	-0.00	0.01	4406	-0.00	0.01	4402	0.99

Note. Close to (*far from*) indicates whether the average distance is below or equal to (above) the sample median. Variables referring to catchment areas are averages of the corresponding variable within the catchment area. In columns (1)–(6), estimates are based on CT-, respondent- or household-level OLS regressions using equation (1) separately for each category. Column (7) presents a heterogeneity test based on CT-level OLS regressions using equation (1) and adding an interaction term between the treatment indicator T and an indicator variable for the first category. Standard errors are clustered by catchment area for CT-level outcomes and by catchment-area–round for respondent- and household-level outcomes. The dependent variables are indicated in the rows and are defined in Appendix A. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. Statistical significance is denoted by *** p<0.01, ** p<0.05, * p<0.1.

D.7 Heterogeneity of treatment effects on pre-specified dimensions

This section presents estimates of heterogeneous effects by a series of variables identified in the pre-analysis plan (Armand et al., 2018). Table D13 presents an analysis of heterogeneity for CT- and caretaker-level outcomes. Tables D14 and D15 refer instead to respondent- and household-level outcomes.

Table D13: Heterogeneity by catchment area or CT characteristics

	Effect of any treatment, by category						Het. test
	β	se	N	β	se	N	p-value
Outcome variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Lower WTP in catchment area			**Higher WTP in catchment area**			
Physical maintenance	0.08	0.06	272	0.01	0.07	270	0.39
Quality	0.06*	0.03	272	0.05*	0.03	270	0.85
Higher quality	0.15**	0.06	272	0.09	0.06	270	0.48
Users during rush hour	0.05	0.07	273	-0.13*	0.07	269	0.07
Free riding	-0.08	0.06	273	-0.07	0.05	269	0.90
Hours worked	0.28	0.54	272	0.08	0.45	270	0.77
Time allocated to managing	0.07**	0.03	272	0.03	0.04	270	0.38
Awareness of routine maintenance	0.08	0.06	272	0.03	0.03	270	0.39
Inputs in routine maintenance	0.08***	0.03	272	0.01	0.02	270	0.05
Pro-social motivation for the cause	-0.03	0.04	272	-0.01	0.04	270	0.81
	Lower quality of the CT			**Higher quality of the CT**			
Physical maintenance	0.13**	0.06	294	-0.06	0.07	248	0.05
Quality	0.05	0.03	294	0.04	0.03	248	0.70
Higher quality	0.11*	0.06	294	0.11*	0.06	248	0.98
Users during rush hour	-0.05	0.06	294	-0.07	0.08	248	0.85
Free riding	-0.07	0.06	294	-0.08	0.06	248	0.91
Hours worked	-0.16	0.56	294	0.46	0.40	248	0.37
Time allocated to managing	0.05	0.05	294	0.06**	0.03	248	0.87
Awareness of routine maintenance	0.13**	0.05	294	0.04	0.04	248	0.16
Inputs in routine maintenance	0.02	0.02	294	0.04*	0.02	248	0.48
Pro-social motivation for the cause	-0.01	0.03	294	-0.05	0.04	248	0.36
	Lower traffic during rush hour			**Higher traffic during rush hour**			
Physical maintenance	0.06	0.08	204	0.05	0.06	338	0.98
Quality	0.00	0.03	204	0.08***	0.03	338	0.09
Higher quality	0.08	0.07	204	0.13**	0.05	338	0.56
Users during rush hour	-0.21**	0.10	203	0.03	0.05	339	0.05
Free riding	-0.08	0.06	203	-0.07	0.05	339	0.92
Hours worked	-0.01	0.54	204	0.33	0.45	338	0.59
Time allocated to managing	0.03	0.03	204	0.06	0.04	338	0.71
Awareness of routine maintenance	0.16**	0.06	204	0.05	0.04	338	0.12
Inputs in routine maintenance	0.01	0.03	204	0.05**	0.02	338	0.29
Pro-social motivation for the cause	-0.12***	0.04	204	0.02	0.03	338	0.01
	Lower free-riding			**Higher free-riding**			
Physical maintenance	0.04	0.07	270	0.06	0.08	272	0.88
Quality	0.03	0.03	270	0.08**	0.03	272	0.26
Higher quality	0.11	0.07	270	0.14***	0.04	272	0.74
Users during rush hour	-0.08	0.07	270	-0.04	0.07	272	0.62
Free riding	-0.07	0.05	270	-0.10*	0.06	272	0.65
Hours worked	-0.33	0.36	270	0.75	0.59	272	0.12
Time allocated to managing	0.03	0.02	270	0.07	0.05	272	0.50
Awareness of routine maintenance	0.15***	0.05	270	0.01	0.05	272	0.04
Inputs in routine maintenance	0.02	0.02	270	0.05*	0.03	272	0.56
Pro-social motivation for the cause	-0.03	0.04	270	-0.02	0.03	272	0.93
	Lower pro-social motivation			**Higher pro-social motivation**			
Physical maintenance	0.16**	0.08	264	-0.08	0.06	278	0.01
Quality	0.06*	0.03	264	0.03	0.04	278	0.54
Higher quality	0.11**	0.05	264	0.09	0.07	278	0.72
Users during rush hour	0.04	0.08	265	-0.14**	0.06	277	0.07
Free riding	-0.08	0.07	265	-0.06	0.05	277	0.77
Hours worked	0.60	0.60	264	-0.25	0.40	278	0.23
Time allocated to managing	0.04	0.05	264	0.05*	0.03	278	0.95
Awareness of routine maintenance	0.03	0.05	264	0.15***	0.05	278	0.08
Inputs in routine maintenance	0.02	0.03	264	0.04	0.02	278	0.71
Pro-social motivation for the cause	-0.05	0.03	264	-0.00	0.04	278	0.37

Note. Categories for heterogeneity analysis are defined at baseline, with *lower* (*higher*) indicating whether the variable is smaller than or equal to (larger than) the sample median. Variables referring to catchment areas are averages of the corresponding variable within the catchment area. In columns (1)–(6), estimates are based on CT- or caretaker-level OLS regressions using equation (1) separately for each category. Column (7) presents a heterogeneity test based on CT- or caretaker-level OLS regressions using equation (1) and adding an interaction term between the treatment indicator T and an indicator variable for the first category. The p-value is relative to the significance of the coefficient on the interaction term. Standard errors clustered by catchment area. The dependent variables are indicated in the rows and are defined in Appendix A. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. Statistical significance is denoted by *** p<0.01, ** p<0.05, * p<0.1.

Table D14: Heterogeneity by individual characteristics: respondent-level outcomes

	Effect of any treatment, by category						**Het. test**
	β	se	N	β	se	N	p-value
Outcome variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Lower WTP for CT use			**Higher WTP for CT use**			
WTP for CT use	-0.06	0.07	4130	-0.01	0.08	4678	0.59
Perceived improvement in the facility	0.05**	0.02	2357	0.01	0.02	2533	0.06
Demand for public intervention	0.07**	0.03	733	0.04	0.03	847	0.46
PGG contribution	-0.01	0.02	564	0.01	0.01	664	0.05
Preference for maintenance	-0.00	0.01	4130	-0.01	0.01	4678	0.68
Awareness of health and safety risks	-0.00	0.01	2301	0.01	0.01	2456	0.20
Awareness of externalities	-0.03	0.03	2357	0.04*	0.02	2533	0.06
Morbidity	0.00	0.03	2357	0.02	0.03	2533	0.66
Health expenditure (extensive margin)	0.05	0.04	1516	0.04	0.03	1816	0.78
Health expenditure (intensive margin)	288.33	321.02	1516	-214.11	291.00	1816	0.23
	Lower awareness of health and safety risks			**Higher awareness of health and safety risks**			
WTP for CT use	0.02	0.10	2562	-0.07	0.07	6246	0.47
Perceived improvement in the facility	0.02	0.02	1408	0.03*	0.02	3482	0.60
Demand for public intervention	0.04	0.04	448	0.06**	0.03	1132	0.65
PGG contribution	-0.02	0.02	349	0.01	0.01	879	0.11
Preference for maintenance	-0.00	0.01	2562	-0.01	0.01	6246	0.61
Awareness of health and safety risks	0.00	0.01	1361	0.01	0.01	3396	0.63
Awareness of externalities	-0.03	0.03	1408	0.03	0.02	3482	0.08
Morbidity	0.04	0.03	1408	0.00	0.03	3482	0.49
Health expenditure (extensive margin)	0.06*	0.04	911	0.03	0.03	2421	0.58
Health expenditure (intensive margin)	345.95	425.72	911	-120.75	239.08	2421	0.30
	Lower trust in community to keep CT clean			**Higher trust in community to keep CT clean**			
WTP for CT use	-0.05	0.07	7192	0.01	0.12	1616	0.54
Perceived improvement in the facility	0.03	0.02	3983	0.04	0.03	907	0.72
Demand for public intervention	0.08***	0.03	1245	-0.03	0.04	335	0.01
PGG contribution	-0.00	0.01	966	0.02	0.02	262	0.25
Preference for maintenance	-0.00	0.01	7192	-0.01	0.01	1616	0.58
Awareness of health and safety risks	0.01	0.01	3878	0.01	0.01	879	0.84
Awareness of externalities	-0.00	0.02	3983	0.05	0.04	907	0.20
Morbidity	0.01	0.02	3983	0.05	0.04	907	0.40
Health expenditure (extensive margin)	0.04	0.03	2560	0.05	0.05	772	0.94
Health expenditure (intensive margin)	-29.83	252.60	2560	224.02	460.72	772	0.64
	Shorter distance to CT			**Longer distance to CT**			
WTP for CT use	-0.14	0.09	4344	0.06	0.08	4464	0.09
Perceived improvement in the facility	0.03	0.02	2417	0.03	0.02	2473	0.98
Demand for public intervention	0.07*	0.04	788	0.03	0.03	792	0.33
PGG contribution	-0.01	0.02	599	0.01	0.01	629	0.50
Preference for maintenance	-0.01	0.01	4344	0.00	0.01	4464	0.33
Awareness of health and safety risks	0.00	0.01	2360	0.01	0.01	2397	0.34
Awareness of externalities	0.01	0.03	2417	0.01	0.03	2473	0.95
Morbidity	0.02	0.03	2417	-0.00	0.03	2473	0.52
Health expenditure (extensive margin)	0.04	0.04	1641	0.04	0.03	1691	1.00
Health expenditure (intensive margin)	-394.47	338.83	1641	410.45	275.24	1691	0.06

Note. Categories for heterogeneity analysis are defined at baseline, with *lower* (*higher*) indicating whether the variable is smaller than or equal to (larger than) the sample median. In columns (1)–(6), estimates are based on respondent- and household-level OLS regressions using equation (1) separately for each category. Column (7) presents a heterogeneity test based on CT-level OLS regressions using equation (1) and adding an interaction term between the treatment indicator T and an indicator variable for the first category. The p-value is relative to the significance of the coefficient on the interaction term. Standard errors are clustered by catchment-area–round of observation. The dependent variables are indicated in the rows and are defined in Appendix A. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. Specifications where the level of analysis is the respondent also include gender. Statistical significance denoted by *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table D15: Heterogeneity by catchment area or CT characteristics: respondent-level outcomes

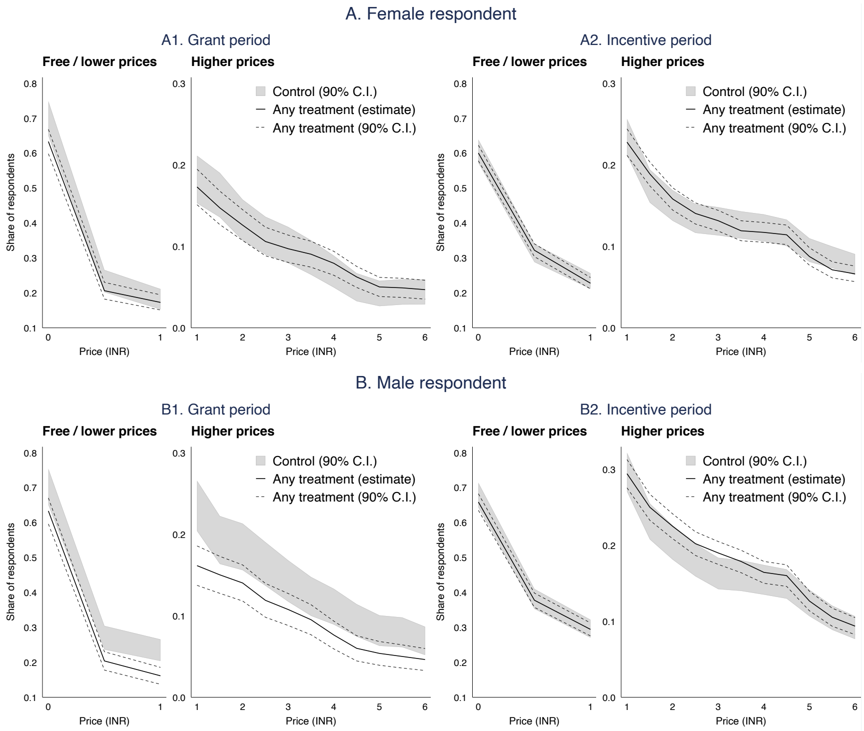
	Effect of any treatment, by category						Het. test
	β	se	N	β	se	N	p-value
Outcome variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Lower water quality			**Higher water quality**			
WTP for CT use	-0.02	0.08	4338	-0.05	0.10	4470	0.85
Perceived improvement in the facility	0.03	0.02	2404	0.02	0.02	2486	0.68
Demand for public intervention	0.02	0.04	769	0.08**	0.03	811	0.32
PGG contribution	0.04**	0.02	619	-0.03*	0.02	609	0.00
Preference for maintenance	0.00	0.01	4338	-0.02**	0.01	4470	0.16
Awareness of health and safety risks	0.01	0.01	2335	-0.00	0.01	2422	0.26
Awareness of externalities	-0.03	0.03	2404	0.03	0.03	2486	0.15
Morbidity	0.05	0.03	2404	-0.02	0.03	2486	0.12
Health expenditure (extensive margin)	0.06*	0.04	1625	0.02	0.04	1707	0.40
Health expenditure (intensive margin)	582.04*	304.56	1625	-456.99	301.54	1707	0.02
	Lower CT quality			**Higher CT quality**			
WTP for CT use	-0.19**	0.09	4660	0.10	0.10	4148	0.06
Perceived improvement in the facility	0.03	0.02	2564	0.01	0.02	2326	0.43
Demand for public intervention	0.08*	0.04	834	0.04	0.03	746	0.41
PGG contribution	-0.01	0.02	647	0.01	0.02	581	0.55
Preference for maintenance	-0.01	0.01	4660	0.00	0.01	4148	0.42
Awareness of health and safety risks	0.00	0.01	2494	0.01	0.01	2263	0.69
Awareness of externalities	0.02	0.03	2564	0.00	0.03	2326	0.64
Morbidity	0.03	0.03	2564	-0.02	0.04	2326	0.29
Health expenditure (extensive margin)	0.01	0.04	1737	0.08*	0.04	1595	0.26
Health expenditure (intensive margin)	-420.15	266.32	1737	462.34	336.49	1595	0.04
	Lower free riding			**Higher free riding**			
WTP for CT use	-0.03	0.09	4273	-0.05	0.10	4535	0.87
Perceived improvement in the facility	0.03	0.02	2382	0.03	0.02	2508	0.93
Demand for public intervention	0.05	0.04	773	0.05	0.04	807	0.87
PGG contribution	0.00	0.02	600	-0.00	0.02	628	0.84
Preference for maintenance	-0.01	0.01	4273	-0.00	0.01	4535	0.76
Awareness of health and safety risks	0.00	0.01	2310	0.01	0.01	2447	0.67
Awareness of externalities	0.04	0.03	2382	-0.01	0.03	2508	0.25
Morbidity	0.07**	0.03	2382	-0.04	0.03	2508	0.03
Health expenditure (extensive margin)	0.04	0.05	1640	0.04	0.04	1692	0.95
Health expenditure (intensive margin)	-165.32	322.34	1640	231.14	305.11	1692	0.37
	Lower preferences for maintenance			**Higher preferences for maintenance**			
WTP for CT use	0.05	0.09	1516	-0.05	0.07	7292	0.34
Perceived improvement in the facility	0.03	0.03	869	0.03*	0.02	4021	0.82
Demand for public intervention	0.17***	0.06	268	0.03	0.02	1312	0.01
PGG contribution	-0.00	0.02	221	0.00	0.01	1007	0.83
Preference for maintenance	0.01*	0.01	1516	-0.01	0.01	7292	0.04
Awareness of health and safety risks	-0.01	0.01	849	0.01	0.01	3908	0.11
Awareness of externalities	-0.03	0.04	869	0.02	0.02	4021	0.30
Morbidity	0.02	0.04	869	0.01	0.02	4021	0.68
Health expenditure (extensive margin)	0.08	0.06	559	0.04	0.03	2773	0.43
Health expenditure (intensive margin)	-43.09	448.50	559	30.33	240.91	2773	0.86
	Caretaker is less motivated			**Caretaker is more motivated**			
WTP for CT use	0.12	0.08	4210	-0.19*	0.10	4598	0.03
Perceived improvement in the facility	0.01	0.02	2337	0.04*	0.02	2553	0.32
Demand for public intervention	0.03	0.04	745	0.09**	0.04	835	0.33
PGG contribution	-0.00	0.02	595	-0.00	0.02	633	0.88
Preference for maintenance	0.00	0.01	4210	-0.01	0.01	4598	0.33
Awareness of health and safety risks	0.02*	0.01	2270	0.00	0.01	2487	0.22
Awareness of externalities	0.01	0.03	2337	0.02	0.03	2553	0.80
Morbidity	0.01	0.03	2337	0.02	0.03	2553	0.81
Health expenditure (extensive margin)	0.03	0.04	1586	0.08**	0.04	1746	0.46
Health expenditure (intensive margin)	131.17	341.40	1586	-120.24	285.11	1746	0.58

Note. Categories for heterogeneity analysis are defined at baseline, with *lower* (*higher*) indicating whether the variable is smaller than or equal to (larger than) the sample median. In columns (1)–(6), estimates are based on respondent- and household-level OLS regressions using equation (1) separately for each category. Column (7) presents a heterogeneity test based on CT-level OLS regressions using equation (1) and adding an interaction term between the treatment indicator T and an indicator variable for the first category. The p-value is relative to the significance of the coefficient on the interaction term. Standard errors are clustered by catchment area–round of observation. The dependent variables are indicated in the rows and are defined in Appendix A. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. Specifications where the level of analysis is the respondent also include gender. Statistical significance is denoted by *** p<0.01, ** p<0.05, * p<0.1.

D.8 Treatment heterogeneity on WTP and sanitation-related behavior, by gender

Figure D3 shows estimates of the inverse demand curve for CT use separately for the control and treatment groups during the grant period, while Figure D4 shows estimates of the effect of interventions on hygiene- and sanitation-related behavior by gender using the list randomization technique (see Appendix C.4.1 for details about measurement).

Figure D3: Inverse demand curve for single CT use, by gender and price level



Note. Each curve indicates the share of respondents who prefer tickets for CT use to cash at the corresponding price. Panels A1 and A2 restrict the sample to female respondents only, while Panels B1 and B2 restrict the sample to male respondents only. The inverse demand curve is elicited using a standard incentivized version of the multiple price list (or take-it-or-leave-it) methodology (Andersen et al., 2006). Details about the measurement are presented in Appendix C.5. Confidence intervals are built using statistical significance at the 10% level.

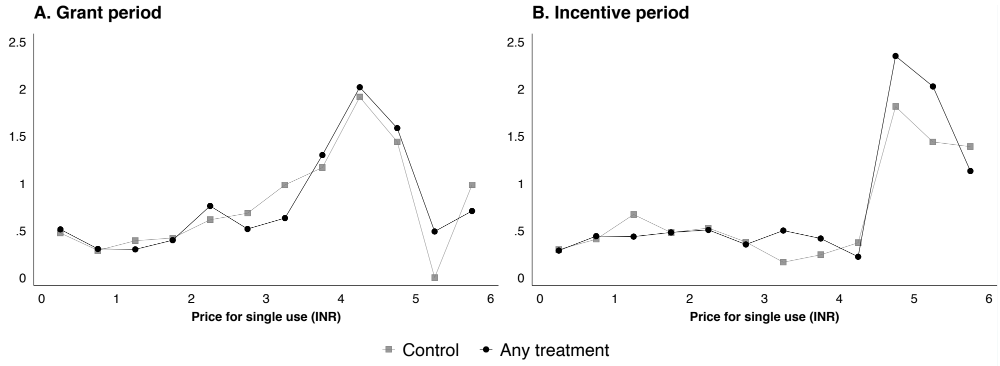
Figure D4: Sanitation- and hygiene-related behavior, by gender

Note. The figure shows the share of slum residents practicing each behavior in the day previous to the interview, estimating using a list randomization technique. Panel A restricts the sample to female respondents, while Panel B restricts the sample to male respondents. Randomization of lists was performed at individual level, and data were collected during follow-up 5 only. Appendix C.4.1 provides additional details about the measurement.

D.9 Price elasticity of the demand for CT use

Figure D5 shows the price elasticity of the demand for CT use. This derived from the incentivized WTP (see Appendix C.5). In both the control and the treatment group, the demand at low prices is relatively inelastic, while it becomes more elastic for prices closer to the market price (INR 5).

Figure D5: Price elasticity of the demand for CT use, by intervention period



Note. The vertical axis is the price elasticity computed as an arc elasticity and calculated between each point and plotted at the midpoint of each segment. Appendix C.5 provides additional details about the measurement.

D.10 Effect on inputs used during routine maintenance

Table D16 presents estimates of treatment effects on the use of inputs during routine maintenance.

Table D16: Effect on the use of individual inputs

	Support from cleaners (1)	Broom (2)	Mop (3)	Disinfectants (4)	Bucket (5)	Safety equipment (6)	Pressurized water
Panel A							
Any treatment (T)	0.133	-0.001	0.050	0.005	0.026	0.016	0.012
	(0.075)	(0.009)	(0.028)	(0.020)	(0.035)	(0.026)	(0.027)
	[0.08]	[0.91]	[0.08]	[0.80]	[0.46]	[0.54]	[0.65]
Panel B							
Maintenance (T1)	0.112	-0.001	0.035	0.026	0.040	0.034	0.013
	(0.090)	(0.011)	(0.032)	(0.020)	(0.039)	(0.033)	(0.031)
	[0.22]	[0.92]	[0.28]	[0.19]	[0.30]	[0.30]	[0.67]
Maintenance + sensitization (T2)	0.153	-0.001	0.065	-0.016	0.012	-0.001	0.012
	(0.085)	(0.011)	(0.036)	(0.023)	(0.044)	(0.032)	(0.033)
	[0.07]	[0.93]	[0.07]	[0.49]	[0.79]	[0.97]	[0.72]
T1 = T2 (p-value)	0.649	0.988	0.442	0.021	0.516	0.343	0.973
Mean (control group)	0.586	0.990	0.759	0.958	0.738	0.110	0.115
Std. Dev. (control group)	0.494	0.102	0.429	0.201	0.441	0.314	0.320
Observations	542	542	542	542	542	542	542

Note. Estimates based on CT-level OLS regressions using equation (1) in Panel A and equation 2 in Panel B. Standard errors clustered by catchment area are reported in parentheses. P-values corresponding to testing that the individual coefficient is different from zero are presented in brackets. The dependent variables are indicated in column headers and are defined in Appendix A. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT.

D.11 Effect on issues reported in the voice-to-the-people initiative

Table D17 shows demand for public interventions for other categories included in the voice-to-the-people initiative (see Appendix C.8 for details).

Table D17: Effect on the reporting of main issues in the community

	Healthcare	Water availability	Dirtiness	Roads and waste	Jobs	Lighting
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Any treatment (T)	0.008	-0.036	-0.073	0.017	-0.001	0.001
	(0.026)	(0.032)	(0.036)	(0.029)	(0.028)	(0.010)
	[0.77]	[0.26]	[0.05]	[0.56]	[0.98]	[0.89]
Panel B						
Maintenance (T1)	-0.007	-0.024	-0.078	0.001	0.006	-0.011
	(0.032)	(0.039)	(0.039)	(0.038)	(0.032)	(0.009)
	[0.83]	[0.55]	[0.05]	[0.97]	[0.86]	[0.23]
Maintenance + sensitization (T2)	0.022	-0.047	-0.069	0.032	-0.007	0.014
	(0.030)	(0.035)	(0.042)	(0.032)	(0.032)	(0.012)
	[0.48]	[0.18]	[0.10]	[0.33]	[0.83]	[0.27]
T1 = T2 (p-value)	0.398	0.556	0.806	0.454	0.683	0.016
Mean (control group)	0.182	0.280	0.432	0.372	0.216	0.024
Std. dev. (control group)	0.386	0.449	0.496	0.484	0.412	0.155
Observations	1580	1582	1582	1580	1581	1580

Note. Estimates based on respondent- and household-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B. Standard errors clustered by catchment area are reported in parentheses. P-values are presented in brackets (see Section 4 for details). Dependent variables are indicator variables equal to 1 if the household reported the corresponding issue in the voice-to-the-people initiative (Appendix C.8), and 0 otherwise. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. Additional details about the variables are presented in Appendix A. *Healthcare* includes the reporting of children's illnesses or poor healthcare as main issues. *Roads and waste* includes the reporting of the poor quality of roads or the lack of waste collection as main issues.

D.12 Refusals of entry and free riding

Table D18 shows estimates of treatment effects on the reporting of refusals of entry to the CT and on free riding, by the level of free riding in the CT at baseline.

Table D18: Refusals of entry and free riding, by level of free riding at baseline

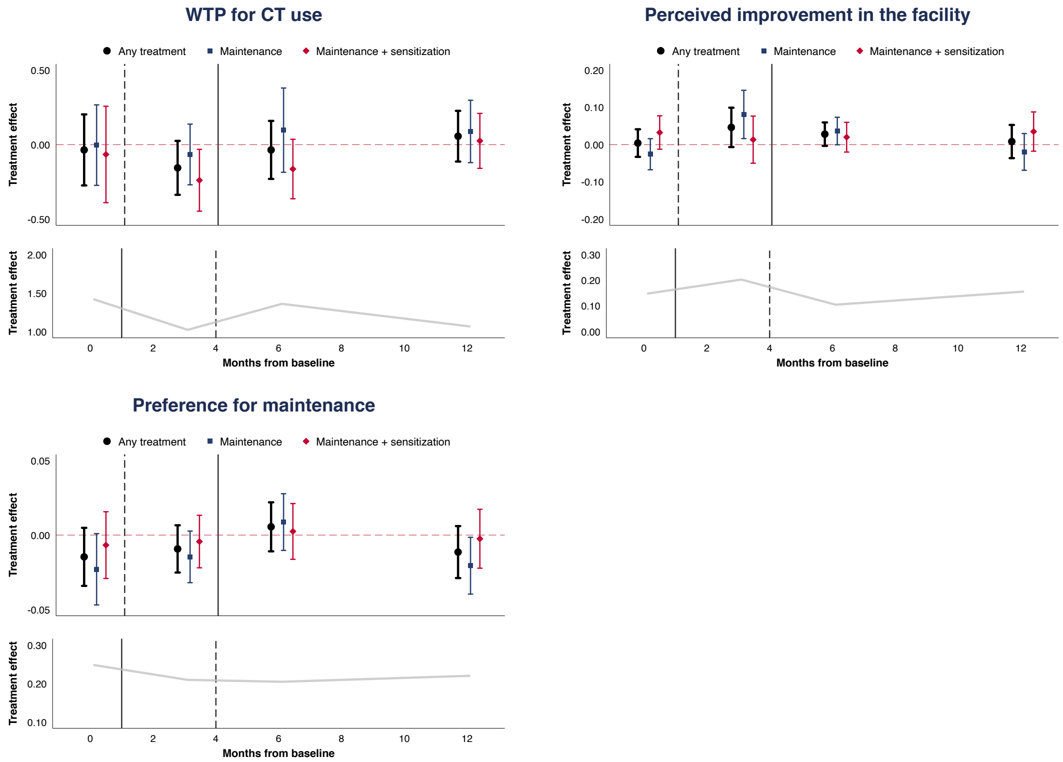
	Caretaker refused entry to the facility			**Refused entry for not paying the entry fee**			**Free riding**	
Level of baseline free riding	Any	Low	High	Any	Low	High	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A								
Any treatment (T)	0.016	-0.017	0.050	0.006	-0.030	0.047	-0.068	-0.104
	(0.020)	(0.030)	(0.023)	(0.020)	(0.031)	(0.022)	(0.052)	(0.058)
	[0.42]	[0.58]	[0.03]	[0.75]	[0.34]	[0.04]	[0.19]	[0.08]
Panel B								
Maintenance (T1)	0.009	-0.001	0.026	-0.006	-0.026	0.022	-0.024	-0.116
	(0.024)	(0.042)	(0.021)	(0.023)	(0.039)	(0.021)	(0.063)	(0.064)
	[0.72]	[0.98]	[0.23]	[0.78]	[0.50]	[0.30]	[0.70]	[0.07]
Maintenance + sensitization (T2)	0.023	-0.031	0.075	0.018	-0.033	0.073	-0.105	-0.090
	(0.024)	(0.031)	(0.034)	(0.024)	(0.036)	(0.033)	(0.056)	(0.069)
	[0.35]	[0.32]	[0.03]	[0.46]	[0.36]	[0.03]	[0.06]	[0.20]
T1 = T2 (p-value)	0.614	0.484	0.172	0.346	0.869	0.146	0.177	0.689
Mean (control group)	0.075	0.104	0.042	0.073	0.104	0.039	0.353	0.500
Std. dev. (control group)	0.263	0.305	0.202	0.261	0.305	0.194	0.285	0.278
Observations	1661	817	844	1661	817	844	266	276
Follow-ups	5	5	5	5	5	5	1–5	1–5

Note. In columns (1)–(6), estimates are based on household-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B. In columns (7) and (8), estimates are based on CT-level OLS regressions using equation (1) in Panel A, and equation (2) in Panel B. Standard errors clustered by catchment area are reported in parentheses. P-values are presented in brackets (see Section 4 for details). Dependent variables reported in column: (1)–(3) *Caretaker refused entry to the facility* is an indicator variable equal to 1 if the respondent reports that he/she observed the caretaker refusing entry to the CT to someone, and 0 otherwise; (4)–(6) *Refused entry for not paying the entry fee* is an indicator variable equal to 1 if the respondent reports that he/she has been refused entry to the CT for not paying the entry fee, and 0 otherwise; (7) and (8) *Free riding*: share of users who do not pay the entry fee observed during 1 hour at rush hour. All specifications include indicator variables for data collection rounds, and strata indicators for city and the provider of the CT. Additional details about the variables are presented in Appendix A. Low (high) baseline free riding indicates CT with free riding at baseline below (above or equal to) the sample median (32% of users free riding).

D.13 Estimates of treatment effects by survey

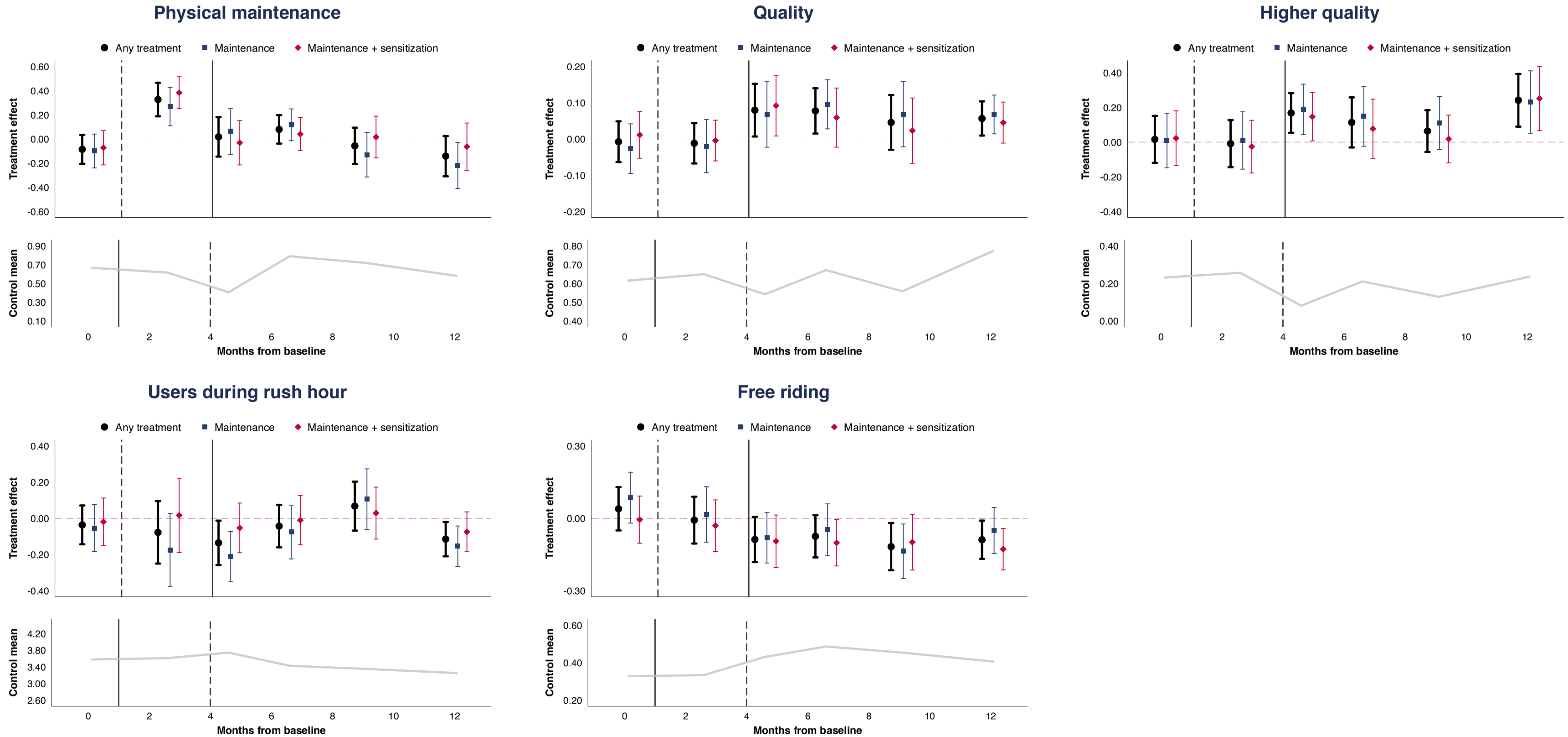
For the outcomes presented in Tables 1–4, this section presents estimates of equation (1) and equation (2) separately for each survey. Estimates are presented in Figures D6–D9. The upper part of each panel presents estimates of treatment effects on the corresponding variable, while the lower part reports the evolution over time of the average of the corresponding variable in the control group. Figures D6–D9 do not report variables that were measured only once.

Figure D6: Timing of effects: WTP and preference for maintenance



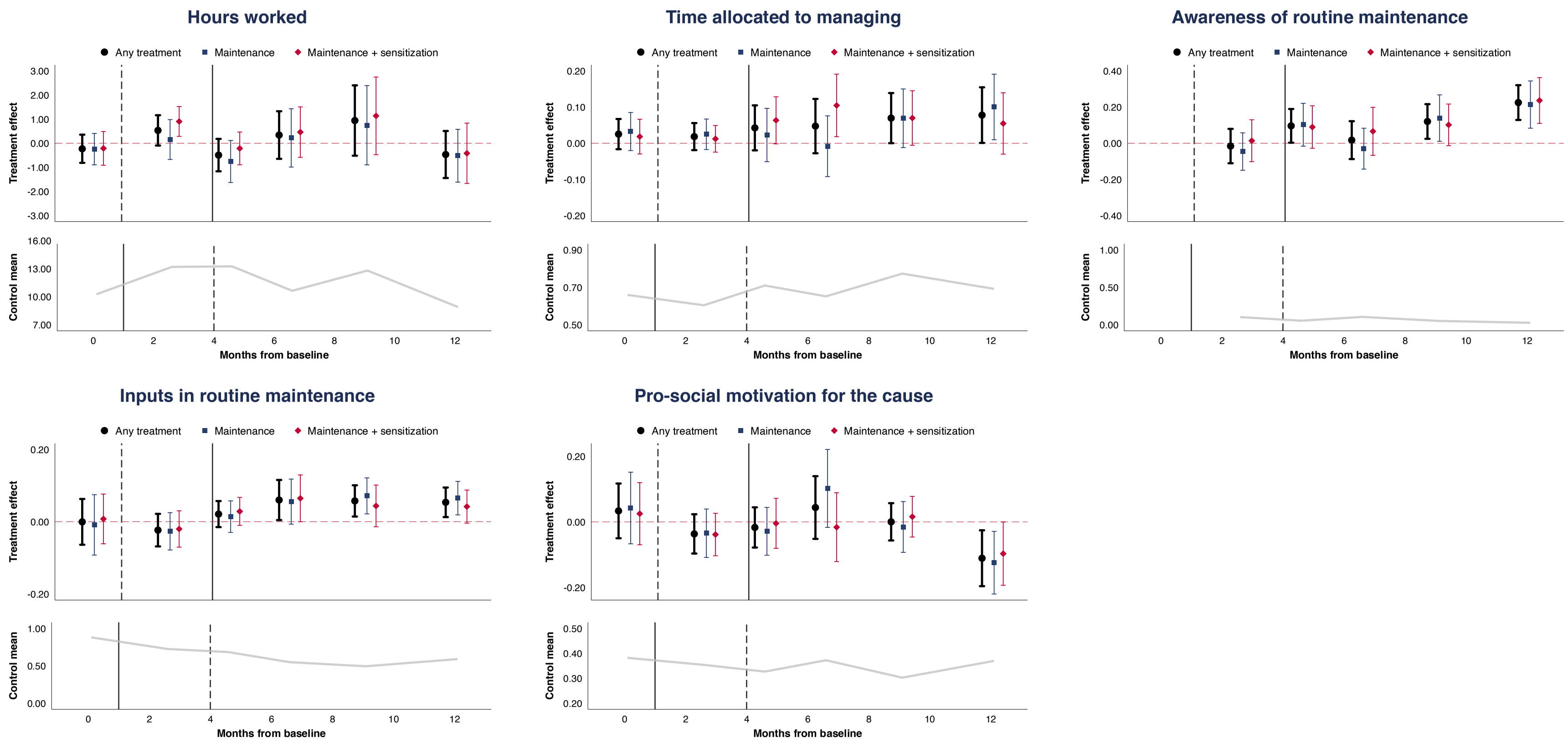
Notes. Estimates based on respondent- and household-level OLS regressions using equation (1) and equation (2) separately for each data collection period. Confidence intervals are computed at the 10% level of confidence using standard errors clustered at the catchment area. Outcome variables are defined in Appendix A. All specifications include strata indicators for city and the provider of the CT. Respondent-level regressions include a control for the gender of the respondent.

Figure D7: Timing of effects: CT-level outcomes



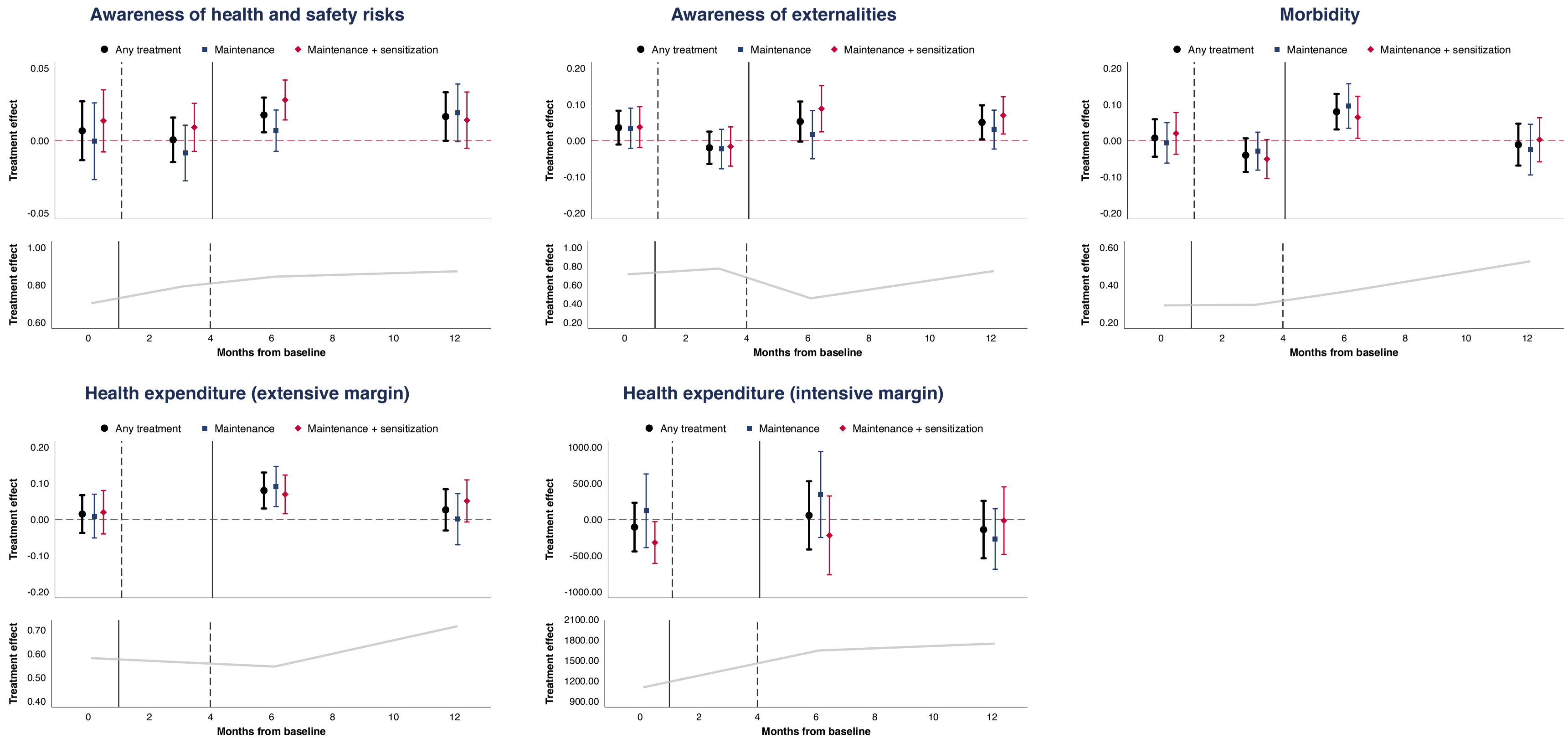
Notes. Estimates based on CT-level OLS regressions using equation (1) and equation (2) separately for each data collection period. Confidence intervals are computed at the 10% level of confidence using robust standard errors. Outcome variables are defined in Appendix A. All specifications include strata indicators for city and the provider of the CT.

Figure D8: Timing of effects: caretaker's behavior



Notes. Estimates based on CT-level OLS regressions using equation (1) and equation (2) separately for each data collection period. Confidence intervals are computed at the 10% level of confidence using robust standard errors. Outcome variables are defined in Appendix A. All specifications include strata indicators for city and the provider of the CT.

Figure D9: Timing of treatment effects: awareness of risks associated with unsafe sanitation and health



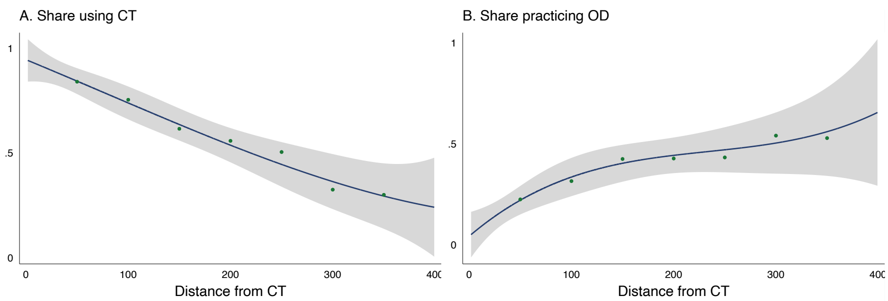
Notes. Estimates based on household-level OLS regressions using equation (1) and equation (2) separately for each data collection period. Confidence intervals are computed at the 10% level of confidence using standard errors clustered at the catchment area. Outcome variables are defined in Appendix A. All specifications include strata indicators for city and the provider of the CT. Respondent-level regressions include a control for the gender of the respondent.

E An analysis of CTs in the study area

E.1 Sanitation behavior in the slum

The location of a CT in the slum is a strong predictor of sanitation-related behavior among eligible households. Using self-reported data from the census of slum residents (see Appendix C), we can study how distance from a facility affects the use of the CT and the practice of OD among eligible households (i.e. those without access to private toilets; see Appendix C for details). Information from the census of slum residents includes self-reported behavior. Figure E1 presents cubic fits for the relationships between CT use and distance from the facility (Panel A), and between the practice of OD and distance from CT (Panel B). The use of a CT reduces rapidly with distance from the facility. At 200 meters from a CT, only 50% of eligible households report using the CT, and more than 40% report practicing OD. At 400 meters from a CT, very few households report making use of the CT, while over 50% of respondents report practicing OD. Notice that these statistics might suffer from reporting bias since information is self-reported by the respondent (see Appendix C.4.1). While we cannot draw any conclusion in terms of causality, it is important to note that either CTs are placed endogenously where slum residents have a higher chance of using them, or the location of the CT is indeed shaping the behavior of slum residents.

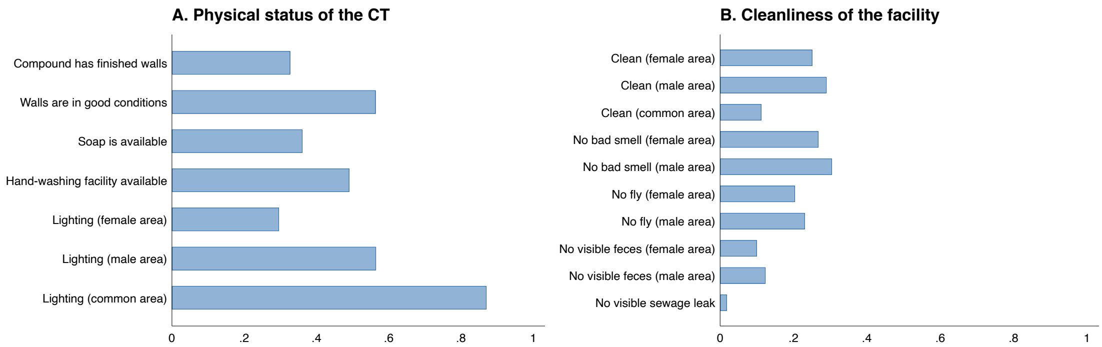
Figure E1: Sanitation behavior in the slum, by distance from a facility



Note. Data source is the slum resident census (see Appendix C). The figures present cubic fits of the share using CT (Panel A) and of the share practicing OD (Panel B) on distance from the closest CT. Dots show the average of the variable indicated on the vertical axis for equally-spaced intervals on the variable indicated on the horizontal axis. The shaded area presents the 90% confidence intervals, assuming standard errors are clustered at the slum level. The sample includes all households in the census that are considered eligible for the study (see Appendix C for details).

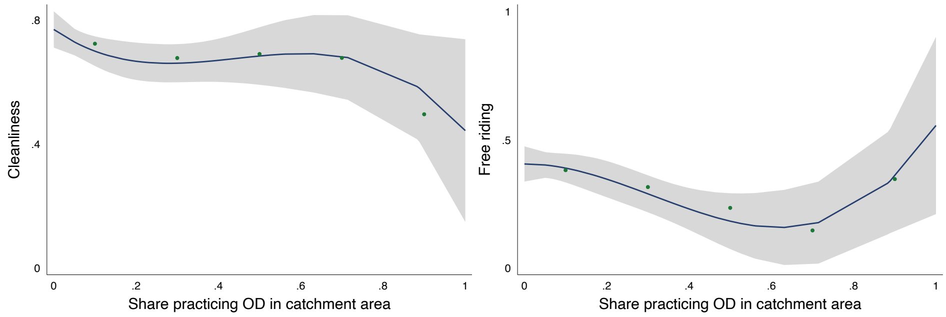
In the urban slums of Uttar Pradesh, CTs are often found in poor condition. Figure E2 summarizes the average status of the facilities as collected by observers at baseline (Appendix C.2). Panel A refers to the physical status of the facility, while Panel B refers to the cleanliness of the facility. CTs are characterized by the poor quality of the construction, by the lack of hand-washing facilities, and by a general lack of cleanliness. Figure E3 shows the baseline relationship between the level of OD among eligible households in the catchment area and the cleanliness of the CT (left panel), and between the level of OD among eligible households in the catchment area and free riding among CT users (right panel). Cleanliness decreases as the share practicing OD increases, with larger drops at very high levels of OD. For free riding, we observe a U-shaped relationship. At low levels of OD, increases in OD are associated with reductions in free riding. On the contrary, at very high levels of OD, free riding is much higher. This potentially indicates that for CTs where payment enforcement is stricter, more residents have to practice OD as they cannot access the CT.

Figure E2: Status of facilities at baseline



Note. Share of CTs that have or have access to the corresponding characteristic. Information is measured at baseline by observers. Appendix C.2 provides further details about the measurement.

Figure E3: CT characteristics by average OD in catchment area



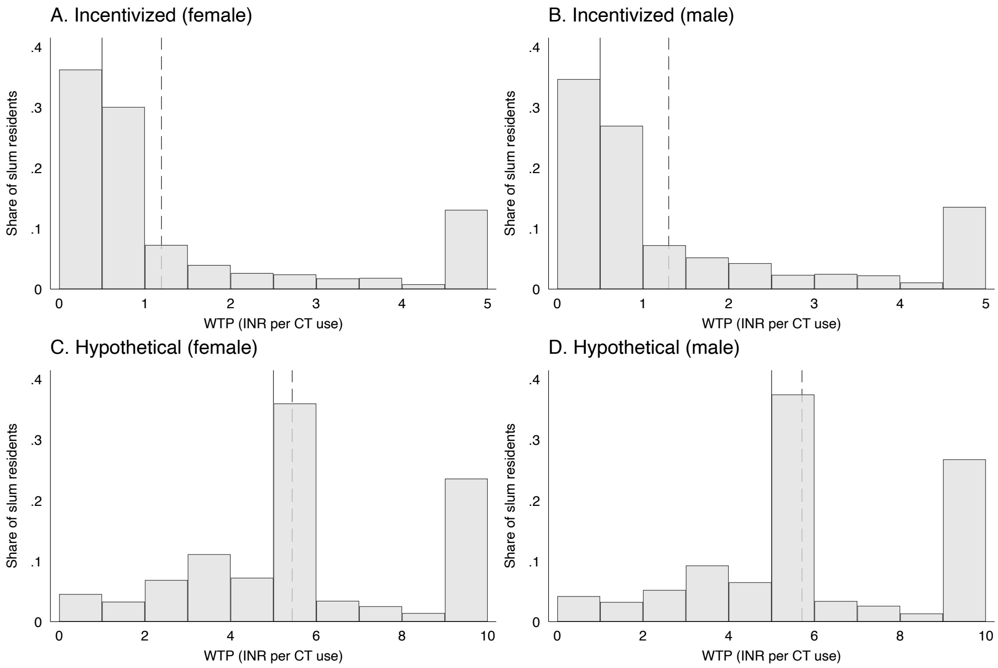
Note. Each panel shows a cubic fit for the relationship between the variables indicated on the horizontal and vertical axis. Dots show the average of the variable indicated on the vertical axis for equally spaced intervals of the variable indicated on the horizontal axis. The shaded area presents the 90% confidence intervals. Information for the variables indicated in the vertical axis is restricted to the baseline CT survey (Appendix C.2), while information for the variables indicated on the horizontal axis is restricted to the baseline household survey (Appendix C.4). Information about OD in the catchment area is computed as the average of the self-reported practice of OD among respondents (after controlling for catchment area fixed effects to deal with reporting bias). Additional details about the variable are presented in Appendix A.

E.2 The role of slum residents

Low quality is generally perceived by slum residents, with less than half of households reporting that they like the services offered in the local CT, 36% reporting the CT is clean, 15% reporting that they like the facility, and 28% reporting they considered it safe. **WTP for using the CT** is also particularly low among potential users. Figure E4 shows the distribution of WTP among male and female respondents, measured using the incentivized elicitation of WTP (Panel A and B) and a non-incentivized elicitation for a hypothetical higher-quality CT (see Appendix C.5). On average, slum residents are willing to pay INR 1.40 to use the CT, corresponding to just 28% of the official market rate of INR 5 per ticket. WTP is slightly higher for male respondents

(INR 1.46 versus 1.36 for female respondents). The share of respondents with zero WTP is 35% and is highly comparable across male and female respondents (34% versus 36%). WTP is slightly higher in households that always use the CT (INR 1.53) than in households that do not always use the CT (INR 1.33). For a hypothetical higher-quality CT, both women and men are, on average, willing to pay above the market price of INR 5. For a hypothetical CT with a high level of cleanliness, good hand-washing facilities, and well-lit and locked cubicles, the average WTP is INR 5.57, slightly higher for men (INR 5.71 versus INR 5.44 for women).

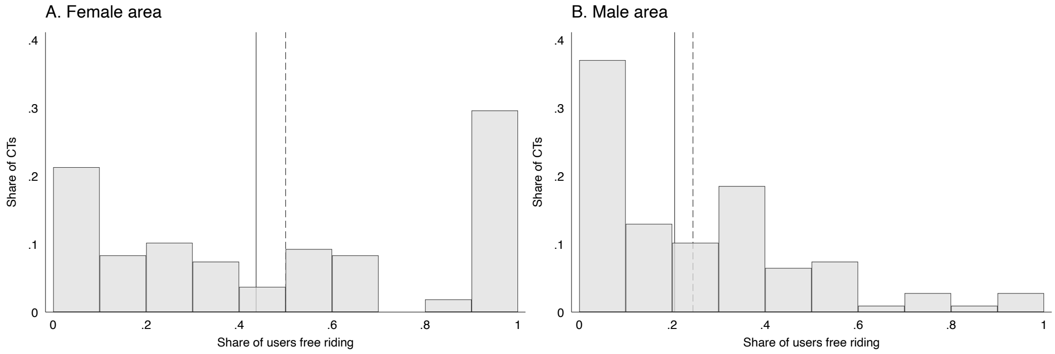
Figure E4: WTP for CT use at baseline, by gender



Note. Data collected at baseline. Panel A (Panel B) shows the distribution of the WTP for a single CT use among female (male) slum residents, measured using the incentivized elicitation of WTP. Panel C (Panel D) shows the distribution of the WTP for a single CT use among female (male) slum residents, measured in a non-incentivized setting for a hypothetical higher-quality CT. Appendix C.5 provides further details about measurement. The solid vertical lines represent the sample median, and the dashed vertical lines represent the sample mean. Additional details about the variable are presented in Appendix A.

In the absence of free riding, an average household of four adult members would spend INR 600 per month to use the CT daily, around 8% of their average household income and less than the amount they spend on intoxicants (INR 817). However, while the maintenance of CTs relies primarily on users fees, **free riding** is often found to be rampant. Figure E5 documents the distribution of the share of women and men who use the CT without paying during the rush hour. On average, only 66% of users pay the CT fee. This is mainly driven by women, among whom 50% do not pay the fee, as compared with 24% among men. There is also wide heterogeneity at the extremes of the distribution. For female areas, 30% of CTs receive no payment, while in 21% everybody pays the fee. For male areas, in 34% of CTs everybody pays the fee, while in only 3% everybody pays the fee. Payments are only partly enforced by caretakers At follow-up 5, we observe that 8% of slum residents report having been prevented from using the facility for not paying the fee during the month previous to the interview.

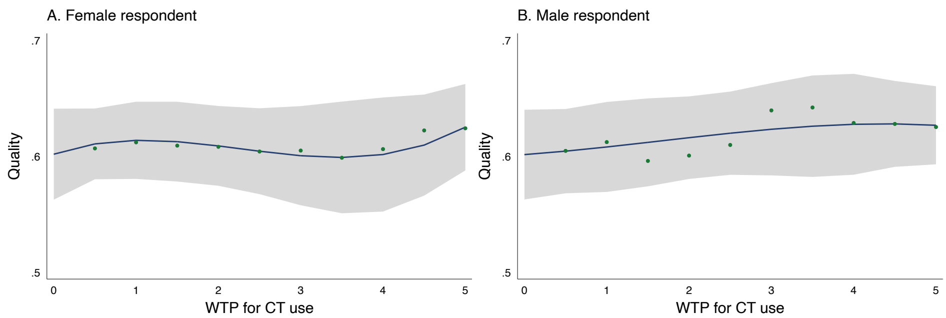
Figure E5: Free riding at baseline, by gender-specific area



Note. Data collected at baseline. Panel A (Panel B) reports the share of female (male) users who do not pay the fee for the use of the female (male) area of the CT during 1 hour at dawn (rush hour), measured by observers. Appendix C.2 provides further details about the measurement. The solid vertical lines represent the sample median, and the dashed vertical lines represent the sample mean. Additional details about the variable are presented in Appendix A.

Figure E6 looks at how the WTP for CT use relates to the quality of the facility at baseline. For both female and male respondents, the relationship is relatively flat; if anything the slope is only slightly positive. This indicates that quality is not driving variation in WTP among potential users at baseline.

Figure E6: WTP in the catchment area and CT quality

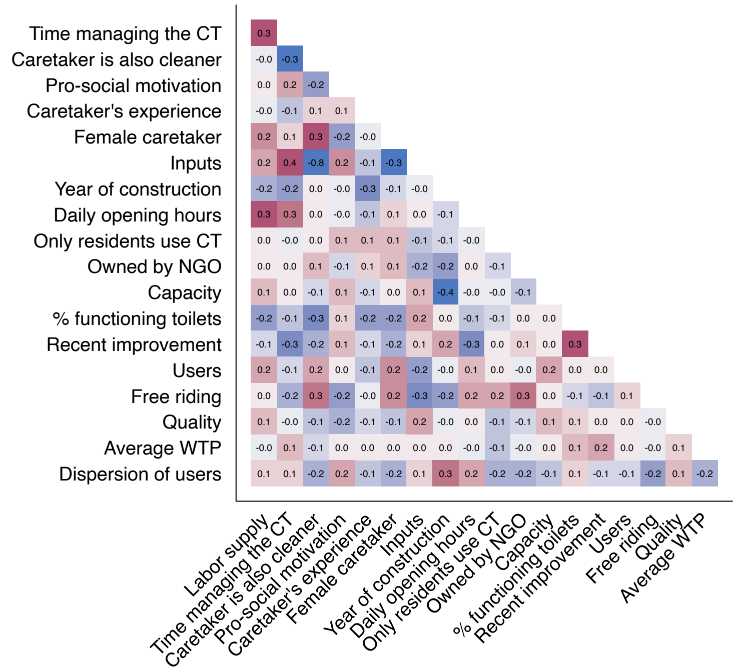


Note. Panel A is restricted to female respondents, while Panel B is restricted to male respondents. Each panel shows a cubic fit for the relationship between quality of the CT and WTP, measured using the incentivized elicitation of WTP. Appendix C.5 provides further details about measurement. Dots show the average of the variable indicated on the vertical axis for equally spaced intervals of the variable indicated on the horizontal axis. The shaded area presents the 90% confidence intervals, assuming standard errors are clustered at the catchment-area level.

E.3 The role of the caretaker

To understand the role of the caretaker in maintaining the facility, Figure E7 presents an unconditional correlation matrix measured at baseline for a variety of indicators associated with the CT, including caretakers' characteristics and average characteristics of slum residents in the catchment area. Caretakers' labor supply is positively related to the opening hours of the CT and the share of time allocated to managing the CT. Caretakers who are also working as cleaners tend to be female, have lower pro-social motivation and work in worse CTs (as measured by the share of functioning toilets). Recent improvements positively correlate with the share of functioning toilets, which is higher where caretakers have higher experience. Free riding is especially concentrated in CTs owned by an NGO, and in CTs used only by residents. In terms of caretakers' characteristics, free riding is lower where caretakers are male, they are more pro-socially motivated, they spend a higher share of time managing the CT, they do not work as cleaners, and they use better inputs. Overall, these statistics indicate the importance of the caretaker in determining outcomes related to the CT.

Figure E7: Correlation matrix for caretaker and CT characteristics



Note. The figure presents unconditional correlations between CT indicators. Variables are measured at baseline. Refer to Appendix A for variable definitions. *Caretaker's experience* is the number of months the caretaker has worked in the CT. *Dispersion of users* is the square of the average distance of potential users to a CT within the catchment area.

Appendix Bibliography

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, 9(4):383–405.

Armand, A., Augsburg, B., and Bancalari, A. (2018). Community Toilet Use in Slums: Willingness to Pay and the Role of Informational and Supply Side Constraints. AEA RCT Registry. June 20. https://doi.org/10.1257/rct.3087-1.0.

Armand, A., Augsburg, B., Bancalari, A., and Trivedi, B. (2020). Community toilet use in slums: Willingness-to-pay and the role of informational and supply side constraints. *3ie Impact Evaluation Report 113*.

Ashraf, N., Bandiera, O., and Lee, S. S. (2014). Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior and Organization*, 100:44–63.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *Annals of Statistics*, 47(2):1179–1203.

Athey, S. and Wager, S. (2019). Estimating Treatment Effects with Causal Forests: An Application. *Observational Studies*, 5.

Basu, S., Kumbier, K., Brown, J. B., and Yu, B. (2018). Iterative random forests to discover predictive and stable high-order interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 115(8):1943–1948.

Belloni, A., Chernozhukov, V., and Hansen, C. (2013). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2):608–650.

Benke, K. K. and Hamilton, A. J. (2008). Quantitative microbial risk assessment: Uncertainty and measures of central tendency for skewed distributions. *Stochastic Environmental Research and Risk Assessment*, 22(4):533–539.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2017). Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments. Technical report, National Bureau of Economic Research.

Flores, G. E., Bates, S. T., Knights, D., Lauber, C. L., Stombaugh, J., Knight, R., and Fierer, N. (2011). Microbial biogeography of public restroom surfaces. *PLoS ONE*, 6(11):1–7.

Gelbach, J. B. (2016). When do covariates matter? And which ones, and how much? *Journal of Labor Economics*, 34(2):509–543.

Gent, J. F., Ren, P., Belanger, K., Triche, E., Bracken, M. B., Holford, T. R., and Leaderer, B. P. (2002). Levels of household mold associated with respiratory symptoms in the first year of life in a cohort at risk for asthma. *Environmental Health Perspectives*, 110(12):A781–A786.

Gordon, D., Howe, L. D., Galobardes, B., Matijasevich, A., Johnston, D., Onwujekwe, O., Patel, R., Webb, E. A., Lawlor, D. A., and Hargreaves, J. R. (2012). Authors' response to: Alternatives to principal components analysis to derive asset-based indices to measure socio-economic position in low- and middle-income countries: The case for multiple correspondence analysis. *International Journal of Epidemiology*, 41(4):1209–1210.

Government of India (2011). Census of India 2011. Office of the Registrar General and Census Commissioner, Ministry of Home Affairs, Government of India, New Delhi, India.

Jenkins, R. and Maddocks, S. (2019). *Bacteriology methods for the study of infectious diseases*. Academic Press.

Karlan, D. S. and Zinman, J. (2012). List randomization for sensitive behavior: An application for measuring use of loan proceeds. *Journal of Development Economics*, 98(1):71–75.

Kline, T. (2014). *Psychological Testing: A Practical Approach to Design and Evaluation*. Sage Publications.

McGinnis, S., Marini, D., Amatya, P., and Murphy, H. M. (2019). Bacterial contamination on latrine surfaces in community and household latrines in Kathmandu, Nepal. *International Journal of Environmental Research and Public Health*, 16(2):257.

Sclar, G. D., Penakalapati, G., Amato, H. K., Garn, J. V., Alexander, K., Freeman, M. C., Boisson, S., Medlicott, K. O., and Clasen, T. (2016). Assessing the impact of sanitation on indicators of fecal exposure along principal transmission pathways: A systematic review. *International Journal of Hygiene and Environmental Health*, 219(8):709–723.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, volume 26. CRC press.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Treibich, C. and Lépine, A. (2019). Estimating misreporting in condom use and its determinants among sex workers: Evidence from the list randomisation method. *Health Economics*, 28(1):144–160.

United Nations (2018). World Urbanization Prospects: The 2018 Revision. New York, US: United Nations Department of Economic and Social Affairs, Population Division.

Wager, S. and Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

World Bank (2021). World Development Indicators. Washington D.C.: The World Bank.

World Health Organisation (2017). Guidelines for drinking-water quality, 4th edition: 1st addendum. Technical Report April 2017, Guidelines Review Committee – Water, Sanitation, Hygiene and Health, World Health Organization.

Yeatman, S. and Trinitapoli, J. (2011). Best-friend reports: A tool for measuring the prevalence of sensitive behaviors. *American Journal of Public Health*, 101(9):1666–1667.