

Abramovsky, Laura; Augsburg, Britta; Lührmann, Melanie; Oteiza, Francisco; Rud, Juan Pablo

Working Paper

Community matters: Heterogenous impacts of a sanitation intervention

IFS Working Papers, No. W18/28

Provided in Cooperation with:

The Institute for Fiscal Studies (IFS), London

Suggested Citation: Abramovsky, Laura; Augsburg, Britta; Lührmann, Melanie; Oteiza, Francisco; Rud, Juan Pablo (2018) : Community matters: Heterogenous impacts of a sanitation intervention, IFS Working Papers, No. W18/28, Institute for Fiscal Studies (IFS), London

This Version is available at:

<https://hdl.handle.net/10419/200317>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Community matters: heterogenous impacts of a sanitation intervention

IFS Working Paper W18/28

Laura Abramovsky

Britta Augsburg

Melanie Lührmann

Francisco Oteiza

Juan Pablo Rud

Community Matters: Heterogeneous Impacts of a Sanitation Intervention *

Laura Abramovsky[†] Britta Augsburg[†] Melanie Lührmann[‡] Francisco Oteiza[§]
Juan Pablo Rud[‡]

Abstract

We study the effectiveness of a community-level information and mobilization intervention to reduce open defecation (OD) and increase sanitation investments in Nigeria. The results of a cluster-randomized control trial in 246 communities, conducted between 2014 and 2018, suggest that average impacts are exiguous. However, these results hide important community heterogeneity, as the intervention has strong and lasting effects on OD habits in poorer communities. This result is robust across several measures of community socio-economic characteristics, and is not driven by baseline differences in toilet coverage. In poor communities, OD rates decreased by 9pp from a baseline level of 75%, while we find no effect in richer communities. The reduction in OD is achieved mainly through increased toilet ownership (+8pp from a baseline level of 24%). In terms of channels, the intervention appears to have raised the social status attached to toilet ownership among the poorer treated communities, and not in rich communities. Finally, we use data from our study and five other trials of similar interventions and show that estimated impacts on OD are stronger in poorer contexts, rationalizing the wide range of estimates in the literature and providing plausible external validity.

Keywords: External validity, Heterogeneous Treatment Effects, Sanitation, Information, Cluster-Randomized Control Trial.

*This research was carried as part of the Formal Research Component of WaterAid UK's Project "Sustainable Total Sanitation Nigeria -implementation, learning, research, and influence on practice and policy" (STS Nigeria), funded by the Bill and Melinda Gates Foundation. The authors would like to thank the participants of the EDePo Development Workshop. Corresponding author Francisco Oteiza: f.oteiza@ucl.ac.uk

[†]Centre for the Evaluation of Social Policies (EDePo), Institute for Fiscal Studies

[‡]Department of Economics, Royal Holloway, and IFS

[§]Institute of Education, University College London

1 Introduction

Countries in the developing world missed the United Nations' 2015 Millenium Development Goal target of halving the number of people without access to improved sanitation by almost 700 million people. In that year, close to 2.4 billion people lacked improved sanitation facilities and 1 billion still defecated in the open ([WHO/UNICEF, 2015](#)). While the costs of this practice in terms of child health, mortality and human capital accumulation are by now well understood among academics, practitioners and policy makers ([Adukia, 2017](#), [Alzua et al., 2017](#), [Prüss-Ustün et al., 2014](#), [Spears and Lamba, 2016](#), [United Nations, 2016](#)), lack of safe sanitation in poor areas remains a challenge.

This paper uses a community-level intervention in Nigeria to investigate the role of information in preventing the adoption of safe sanitation practices. If individuals are unaware of the health and environmental risks of current practices, then the perceived benefits of behavioral change, including sanitation investments, may be low. Existing evidence suggests that providing health related information to households may have a positive effect on household investments and health behaviour (see, for example, [Dupas \(2011\)](#)).¹ However, the success of information interventions aimed at changing sanitation perceptions has been mixed, ranging from large to very modest impacts ([Garn et al., 2017](#)). We explore the role of community characteristics in explaining disparate results in terms of program effectiveness.

The intervention, Community Led Total Sanitation (CLTS), aims at reducing open defecation (OD). While the construction of toilets may be instrumental in its eradication, the program does not promote specific types of toilets nor does it provide hardware or financial support such as credit or subsidies.² The stated objective of CLTS is to induce a cooperative approach that mobilizes the whole community to devise and engage in safer sanitation practices, rather than targeting individuals. This, and the fact that the intervention is delivered through community meetings, suggests that community characteristics may be important for CLTS effectiveness. Our study is conducted in Nigeria, a good testing ground as 34% of its population practices open defecation

¹ Information has been proven to play a role in other, non-health, investment decisions as well. For example, [Jensen \(2010\)](#) showed that by only providing information on market returns to education, average schooling increased by 0.20-0.35 additional years among a sample of students from the Dominican Republic.

² Indeed, CLTS was designed as a response to hardware subsidy programs in Bangladesh that saw increases in toilet construction but lagging improvements in terms of usage ([Kar, 2003](#)).

and toilet ownership rates are stagnating ([WHO/UNICEF, 2015](#)). We designed and conducted a cluster-randomized control trial in 246 communities in two states (Ekiti and Enugu), between 2014 and 2018. In each community, we interviewed 20 randomly selected households at baseline and in three follow-up surveys 8, 24 and 32 months after implementation.

We find that CLTS had exiguous average effects in reducing OD: households in treated communities were 3 percentage points (pp) less likely to declare to perform OD than control households. Effects are sustained until at least two years after CLTS, after which a fade-out is observed. However, average treatment effects hide meaningful heterogeneity: CLTS impacts are substantially stronger among, and restricted to, the poorer half of communities. In this sub-sample, OD fell by 9pp. Again, behavioural change is observed in the short-run (8 months post implementation), but maintained over time. Impact estimates of similar magnitude are found using three alternative measures of communities local economic and social conditions, i.e. geographic isolation, low population density and low average night light intensity. These effects cannot be ascribed to baseline differences in community levels of toilet ownership. Our findings identify community wealth as a crucial factor in mediating the effectiveness of the intervention.

We also find that the reduction in OD is mirrored by household investments: toilet construction in poorer communities increases by 10 percentage points. We find no evidence that better maintenance or sharing of existing toilets could explain changes in sanitation practices. We subsequently use our rich dataset to explore potential channels that may explain why households in poorer communities invest in toilet construction. Our results suggest that the intervention has raised the perceived benefits of sanitation, in particular the social status attached to toilet ownership. We find no evidence of this happening in rich communities. We also cannot ascribe changes to OD or toilet construction to other potential mechanisms, such as differences in delivery or implementation, social capital, social sanctions, trust or fragmentation.

Recent studies try to identify the underlying factors that drive heterogeneous impacts of interventions across and within studies (see, e.g., [Bandiera et al. \(2018\)](#), [Meager \(2018\)](#)). We investigate whether community wealth, the source of heterogeneity in our study, could rationalize the conflicting impact estimates across similar sanitation studies (namely, [Briceno et al. \(2015\)](#), [Cameron et al.](#)

(2013), Guiteras et al. (2015), Patil et al. (2014), Pickering et al. (2015)). Their results range from very large impacts - an increase in toilet ownership by 30pp, and an OD reduction by 23pp- in a trial in Mali (Pickering et al., 2015), to no evidence that an information campaign similar to CLTS improves toilet ownership in Bangladesh (Guiteras et al., 2015). We re-estimate study-specific impacts using a consistent method and order them by a proxy of local income, namely the average night light intensity observed during baseline in each study area. Our findings are consistent with an inverse relationship between area-level wealth and program effectiveness. By rationalizing the existing range of CLTS impact estimates across studies, we interpret that this result carries plausible external validity. The limited effectiveness of the CLTS approach in wealthier communities also suggests that governments using CLTS-type interventions as their national strategy may be wasting scarce resources.³ Wealthier areas, despite having high levels of open defecation, do not seem to respond to this type of interventions.

This paper is structured as follows. In the next section we describe the intervention and the experimental design. Section 3 discusses the empirical methods. In Section 4, we present impact estimates and discuss mechanisms of behavioural change regarding sanitation. Section 5 compares the results of our study to those of other CLTS interventions. Section 6 concludes.

2 Intervention, research design and data

2.1 The intervention: Community Led Total Sanitation

Community Led Total Sanitation (CLTS) is a village-level information and mobilization intervention aimed at reducing open-defecation (OD) and improving toilet coverage. The stated objective of the program is to ‘stimulate a collective sense of disgust and shame among community members as they confront the crude facts about mass open defecation and its negative impacts on the entire community (Kar, 2003).⁴ There are no monetary incentives attached to the reduction of OD, no subsidies or credit offered to finance toilet construction, nor is technical assistance or hardware provided. CLTS maps existing toilets, and links them to OD practices, but does not promote a

³ 24 Countries had adopted CLTS as their national strategy to combat open defecation in 2015 (Musyoki, 2016)

⁴ For a short overview of how CLTS was developed, please refer to appendix A.

particular toilet technology.

The first stage of the intervention consists of an advocacy and sensitization visit in which a team of facilitators meets with community leaders, village chiefs or other important local decision makers. In this meeting, the health implications of open defecation and the potential benefits of CLTS in achieving sustainable behavioural change are presented. Facilitators and civic leaders then arrange an appropriate date and time for the community meeting. Local leaders assume responsibility to mobilize the whole community.

The main component of the intervention is the community meeting, referred to as the triggering meeting. Facilitators engage attendees in a series of activities to inform and involve as many members of the community as possible. A triggering meeting typically starts with a mapping exercise of the village. Each attendee marks their household's location on a stylized village map, indicating whether they own a toilet or not. They next identify regular open defecation sites. With this information visualised on the map, facilitators trace the community's contamination paths of human feces into water supplies and food in a crude fashion. Graphic exercises illustrating the scale of the sanitation problem follow. In some cases, facilitators even rely on using fresh stool to contaminate a bottle or glass of drinking water to make the point as vivid as possible.

A key message is the importance of sanitation externalities and the impact of OD on individual health, i.e. that as long as a small number of people in the community continue to defecate in the open, all community members (and in particular children) are at risk of contracting sanitation-related diseases. When deemed necessary by facilitators, a followup so-called 'walk of shame' through the village may take place, where attendees of the triggering meeting walk through the village and point out visible feces in the environment, to evoke further disgust and shame.

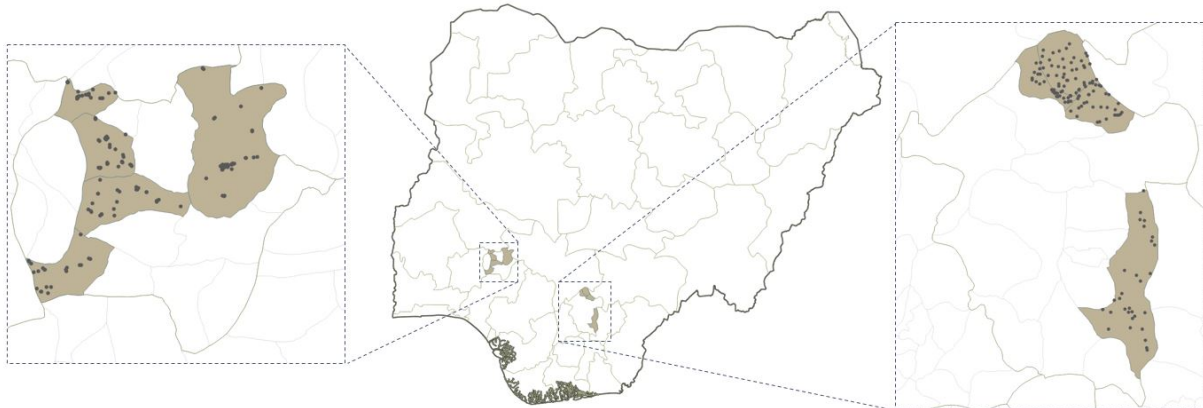
As a closing task, attendees are asked to draw up a community action plan to achieve open-defecation-free (ODF) status. This aspect of CLTS goes beyond simple information delivery and seeks to leverage the power of collaboration and social interactions. It includes discussions of how poor or vulnerable households can be supported by the community to achieve an open defecation free community status. The action plan is posted in a public spot. Volunteers (so-called 'natural leaders') are chosen to follow up regularly on each attendee's commitment towards implementing the

plan. After the triggering meeting, followup visits by the facilitators are organized. Eventually, the village might be certified for its achievements by the national Rural Water Supply and Sanitation Agency (RWASSA) and the National Task Group on Sanitation (NTGS).

2.2 Experimental design: Randomization and implementation

We study the effectiveness of CLTS in reducing open defecation through a cluster-randomized trial carried out in Nigeria.⁵ The intervention was implemented by WaterAid Nigeria, in collaboration with several local agencies in three Nigerian states, Enugu, Ekiti and Jigawa. Due to budget limitations and security concerns, the evaluation study was implemented in the states of Ekiti and Enugu only. Within each state, participating districts (denoted local government areas in Nigeria, or LGAs) were selected, making sure that communities had no relevant recent experience of CLTS, or CLTS-like interventions.⁶

Figure 1: Map of the study area in the Nigerian states of Ekiti and Enugu



Note: Approximate location of the study clusters in the selected local government areas (shaded) from the states of Ekiti (left) and Enugu (right).

We implemented randomized delivery of CLTS by dividing the study sample into ‘triggerable clusters’. The chosen approach reflects that CLTS is a community-level participatory intervention.

⁵ The study protocol was approved by the following IRBs: National Health Research Ethics Committee, Federal Ministry of Health, Nigeria (NHREC/01/01/2007-20-20/11/2014), University College London Ethics Committee (2168/009). The trial was registered at the ISRCTN registry (ISRCTN74165567)

⁶ In Enugu, the LGAs included were Igbo Eze North, Igbo Eze South, Nkanu East and Udenu. In Ekiti, the LGAs included were Ido Osi, Ikole, Moba, Irepodun Ifelodun and Ekiti South West.

Triggerable clusters were defined as groups of geographically close small villages, neighbourhoods or quarters (depending on the type of settlement they are located in).⁷ On average, clusters consist of 1.7 villages or quarters, and their size is similar in both states.⁸ A second criterion for triggerable clusters was that they be self-contained, i.e. that different clusters do not to share markets or large public areas with each other. If randomly selected, villages or quarters within a clusters are triggered at the same time. These steps were taken to reduce information spillover.

In total, 247 triggerable clusters were identified.⁹ Randomization of these clusters was performed after stratifying by LGA, in order to ensure balanced treatment and control samples at this level. Within each LGA, half of the clusters were randomly assigned to receive CLTS treatment and half to remain untreated as control clusters. The distribution is presented in Table 1. CLTS meetings were carried out once in each treatment cluster after approaching cluster officials, and all households from villages in this cluster were invited to attend.

While we collected baseline and followup data from all clusters and village officials were approached in each of them, not all clusters assigned to treatment were triggered. Actual delivery of CLTS triggering meetings took place in 78% of assigned clusters.¹⁰

Table 1: Number of triggerable clusters per study arm and state

	Control		CLTS		Total	
	Freq.	%	Freq.	%	Freq.	%
Ekiti	63	51.6	66	52.8	129	52.2
Enugu	59	48.4	59	47.2	118	47.8
Total	122	100.0	125	100.0	247	100.0

⁷ Clusters do not match Nigerian administrative units, but are smaller or equivalent to Settlements or Autonomous Communities.

⁸ The median and modal number of villages or quarters within a cluster is 1. The maximum number of villages in a cluster is 7, occurring only once.

⁹ One cluster from the original sampling of 247 clusters was subsequently dropped since data collection in any post-treatment survey wave was not possible due to civil unrest in the community.

¹⁰ Our main results are based on an intent-to-treat design based on treatment assignment rather than actual triggering. We discuss characteristics of non-triggered communities and conduct sensitivity analysis estimating average treatment effects in section 4.3.

2.3 Sampling and data collection

The study frame was established in October 2014 by first carrying out a household census in the nine participating LGAs from Ekiti and Enugu. The census collected basic household information from 50,333 households in the area (27,888 from Enugu and 22,445 from Ekiti).¹¹

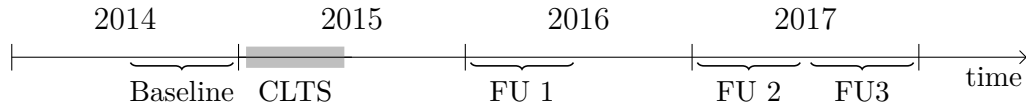
This listing formed the basis of our sampling frame. We randomly selected 20 households from each cluster for interview in our baseline and followup surveys. The pre-treatment baseline survey was carried out between December 2014 and January 2015. Our final sample consists of 4,671 households from 246 clusters, distributed evenly across the two states, or around 9% of the population in the area, according to our census.¹²

After baseline data collection was complete, CLTS was implemented in treatment areas between January and June 2015. Followup data was collected in three waves. The first wave (FU1) took place between December 2015 and February 2016, on average eight months after implementation. The second followup (FU2) was conducted in March-April 2017 and the third (FU3) between November 2017 and January 2018. FU2 and FU3 measured outcomes 24 months and 32 months after CLTS was implemented, on average. The three followup surveys allow us to study the dynamics of CLTS impacts over time and provide insight regarding the sustainability of program impacts. Figure 2 summarizes intervention and data collection timings.

¹¹ Based on program implementation planning, the initial estimate of the size of triggering clusters was around 150 households. It turned out that the clusters were larger on average. Budget limitations constrained to list a maximum of 180 households per cluster. To be able to list a representative sample of households in all clusters we adopted the following approach: For each triggerable cluster, we randomly ordered the villages/quarters and started the listing exercise in the first one on the list. This village/quarter was listed completely, independent of its size (i.e. going above 180 households threshold if needed). If the number of households was smaller than 150 within this village/quarter, and the TU consisted of more than one village/quarter, data collection would continue in the next village/quarter on the list, again listing every household in this village/quarter. This process continued until either all villages/quarters were listed in each TU, or until around 180 households were reached. This approach ensured that we have listing data from each of the study clusters, that our overall sample remains representative for the study area (since the ordering of village/quarter listing was randomly determined), and furthermore that whole villages and quarters were listed while remaining within the available budget.

¹² Written consent was obtained from every household before interviews were carried out. An independent data collection company was in charge of all data collection rounds and was blinded to treatment status. Baseline questionnaires were carried out by pen and paper, while followup surveys were carried out using an electronic survey system.

Figure 2: Project timeline: implementation and data collection waves



2.4 Key outcomes and household characteristics

Our key outcome of interest is sanitation behaviour, in particular open defecation practices. We capture OD practices using two measures. The first dummy variable takes the value one if the survey respondent (typically the main woman of the household) states that at least one household member above the age of four years performs OD. The second variable takes the value one if the respondent herself declares to perform OD.

Among the menu of options to reduce OD, the most likely to come up in discussions in CLTS meetings, according to [Kar \(2003\)](#), is the ownership and use of private toilets. We therefore asked households whether they owned a toilet and about its type and status. We construct three main measures of interest: i) whether a household owns any toilet (i.e. functioning or not), ii) whether it owns any functioning toilet, and iii) whether it owns a functioning and improved toilet. While correlated, these three measures capture different dimensions of interest. The first keeps track of ownership. By ignoring functionality, our first measure is not able to capture toilet divestment through lack of maintenance. This dimension of changes in the intensity of maintenance efforts and investments into the existing stock of toilets is captured in our second measure. The third accounts for quality, satisfying the stricter criteria used by the WHO/UNICEF Joint Monitoring Program regarding improved sanitation.

Self-reported open defecation and toilet ownership may be subject to measurement error - more precisely, to over-reporting of toilet ownership and under-reporting of OD habits. To validate our toilet ownership measurements, interviewers asked households who reported toilet ownership whether they may look at the latrines at the end of each interview. Consent rates were high: At FU1, 89% of households who declared to own a toilet allowed interviewers to inspect them - 88% in CLTS and 89% among control households. If systematic measurement error was present, we would

expect it to be higher in later followup surveys. However, we find that only 3% of households who reported to own a toilet in FU2 refused inspection, and refusal rates are similar in treatment and control (3.5 versus 2.8%). Inspection did not yield discrepancies between self-reported and actual ownership levels. We interpret both - the similarity of consent rates for inspection in treatment and control (at baseline and in FU2) and the truthful reports of ownership by consenting households in both surveys- as evidence that there is little systematic measurement error in latrine ownership across groups. We cannot rule out measurement error with respect to our two sanitation practice measures. However, our results regarding the three measures of toilet ownership are qualitatively and quantitatively very similar, and CLTS impacts on toilet ownership are closely mirrored by reverse impacts on open defecation practice.

Outcome measures are complemented by rich data on household characteristics, which we use as control variables to increase precision, and to assess impact heterogeneity. Covariates used throughout the analysis capture characteristics of the household, such as size (i.e. number of members), whether the household’s main activity is farming, and asset wealth, and its head (age, gender, employment status (i.e. whether he/she worked for pay during the last week) and education (i.e. whether he/she completed primary school)). Household asset wealth is measured as the first factor of a principal component analysis based on a series of questions regarding asset ownership.¹³ Finally, we elicited expected benefits and costs of sanitation, as well as beliefs about social norms and awareness of health externalities, which we postulate can be mediators of intervention impacts. Appendix C.1 provides a detailed description of all household variables.

2.5 Community wealth and other characteristics

The CLTS Handbook, a practitioners’ guide drawing on experiences from 16 countries, suggests that successful implementation is more likely in i) small, ii) culturally and socially homogeneous settlements, iii) located in remote areas and iv) settlements with high prevalence of OD (Kar and Chambers, 2008). We posit that community wealth is a comprehensive proxy of the socio-economic conditions of a community. Using our survey, we construct an asset wealth score for each household

¹³ Details of its components and their factor loadings are provided in Table 13 in Appendix C.1.

and community and classify as ‘poor’ households and communities below the relevant median.¹⁴ As it is based on a detailed list of its inhabitants’ asset items, it may be subject to less measurement error than other community measurements that capture aspects of a community’s socio-economic status. Lists of household asset items are frequently elicited in household surveys in developing countries, thus facilitating using this index as a basis for policy making. As an illustration, we also construct a similar index - at the household and community level - using the 2013 Nigerian Demographic and Health Survey. More than 80% of the poor communities in our survey are below the median Nigerian community along the wealth distribution, while more than 85% of rich communities are above the national median (see Table 10 in Appendix F).¹⁵

We explore further how our measure of community wealth performs relative to other proxies of socio-economic conditions. These community measures are i) the pre-intervention nightlight intensity index, a proxy for local economic income¹⁶, ii) the community’s distance to the next LGA centre (in km), a measure of remoteness or isolation, and iii) the number of households within a 5km radius, a measure of population density. Table 2 shows pairwise correlations of these measures. While the correlations have the expected sign, it seems that each measure captures somewhat different aspects of local conditions. For instance, while population density and isolation appear to be highly correlated ($\rho = -0.58$), the correlation is lower between community asset wealth and average night light intensity ($\rho = 0.16$).¹⁷ We explore heterogeneous treatment impacts with respect to each of these in Section 4.1.

Table 2: Pairwise correlations of community level SES measures

SES measures	Asset wealth	Night lights	Density	Isolation
Asset wealth	1	-	-	-
Night lights	0.1579	1	-	-
Isolation	0.2524	0.4519	1	-
Density	-0.3434	-0.4871	-0.5762	1
Number of communities	247			

Notes: Pairwise correlations of our four alternative measures for community SES. Correlations for the discrete version of these variables can be found in Table 15 in Appendix C.

¹⁴ The aggregated index mean is close to zero and has a standard deviation of one.

¹⁵ We find similar results when looking at household-level wealth. See appendix F for more details.

¹⁶ Henderson et al. (2012) shows that changes to this measure are strongly correlated with local economic growth.

¹⁷ Note in this case that the inverse relationship between isolation and density reflects the intuitive result that higher population density is associated with shorter distances to the nearest LGA capital.

Additionally, we use variables that capture social cohesion and social capital at the community level as they are potentially important in improving the effectiveness of CLTS [Cameron et al. \(2015b\)](#). These are measures of i) trust, ii) social capital and iii) religious fragmentation. See Appendix C.2 for further details.

2.6 Summary statistics

Our baseline sample consists of 4,671 households from nine districts in the states of Ekiti and Enugu. We were able to track a large percentage of these initial study households over time, with very little attrition.¹⁸ Three years after the baseline survey, we successfully interviewed almost 90% of the households in our initial sample, and attrition is balanced across treatments. Hence, we focus our analysis on the sample of households that was successfully interviewed at each round of data collection, resulting in a balanced panel of 4,540 households. Table 3 shows summary statistics for our main outcomes and for the household covariates by treatment status, checking for balance in these variables. The upper panel displays on the main outcomes we study. We find that almost two thirds of study households had at least one member who defecated in the open, and typically, the main respondent was among those members. As discussed later in this paper, OD practices are closely linked with not owning a toilet in the study context. 37% of study households own some type of toilet, most of which are functioning (36% on average) and the majority is functioning and improved (33%). Toilet usage rates are similar to ownership rates, i.e around 37%. Among those who have a toilet, usage is very high at 93-94%.

The second panel displays characteristics of study households. The household head is predominantly male (64%), about 55 years of age, literate in that he/she finished primary school (68%) and is in employment (77%). The average study household consists of four members, 30% of households live with a child below the age of six years, and farming is the main activity of 45% of households.

The third panel of Table 3 shows balance in terms of the community-level characteristics we discussed above. Mean nightlight intensity is with values around 2 very low, given its range from 0 to 61. Communities have on average about 1,600 households within a 5 km radius, and the

¹⁸ Attrition rates are 2.56% at FU1, 9.27% at FU2 and 12.40% at FU3. Importantly, we see that treatment arms do not predict attrition in a statistically significant way, meaning that attrition was balanced across treatment groups.

distance to the nearest LGA capital is about 6.5 km¹⁹. On average 37% of households in the study communities have a toilet. Religious fragmentation is high²⁰, and people trust their neighbours somewhat.²¹

The last column of Table 3 indicates the p-value for equivalence in means between treatment and control. Statistically significant differences are observed only for household size which is slightly larger in the control group. We include it as a control variable throughout our analysis to ensure that it does not drive our results.²² Additionally, we ran an OLS regression taking treatment status as dependent variable, and included all variables in Table 3 as right-hand-side (RHS) variables. In an F-test of joint significance for the whole set of RHS variables, we find that we cannot reject the null hypothesis at the 1% level (p-value=0.038). Once we remove household size, the imbalanced variable, the explanatory power of the remaining variables falls significantly (p-value=0.27). This supports our claim that households assigned to treatment and control are on average identical, except in size. Finally, we find that attrition patterns between households in treatment and control communities are also balanced.

¹⁹ This is a walk of about an hour.

²⁰ 70% of countries recorded in Alesina's 2003 fractionalization dataset have lower fragmentation.

²¹ Mean social capital and community wealth are by construction close to zero.

²² The project's baseline report conducts balancedness checks on a wide range of additional variables.

Table 3: Balance between Treatment and Control groups at Baseline

	Control	Treatment	P-value
<i>Open Defecation and Toilet Ownership</i>			
At least 1 member (> 4 y/o) performs OD (%)	62.82	63.30	0.88
Main respondent performs OD (%)	62.48	62.80	0.92
Own a toilet (any condition, any type) (%)	37.09	36.48	0.84
Own a functioning toilet (any type) (%)	36.30	35.81	0.87
Own a functioning, improved toilet (%)	32.84	32.93	0.97
All members of the HH use a toilet (%)	37.18	36.70	0.88
All members of the HH use a toilet (cond. on ownership) (%)	93.00	93.94	0.52
<i>Household Characteristics</i>			
HH head male (%)	63.92	62.37	0.40
HH head age (years)	55.19	54.42	0.31
HH head employed (%)	76.98	76.54	0.81
HH head finished primary school (%)	67.65	68.08	0.84
Household size (SD = 2.45)	4.15	3.90	0.04**
Household has at least 1 child below 6 y/o (%)	29.99	29.53	0.80
HH primary activity is farming (%)	45.24	48.92	0.32
Relative asset wealth index score (SD = 2)	0.05	-0.01	0.66
<i>Community Characteristics</i>			
Community wealth (SD = 1)	-0.19	-0.30	0.43
Mean night light intensity within a 5km radius, 2013 (min = 0, max = 25)	2.131	2.089	0.91
Number of households within a 5km radius (SD = 989)	1,625	1,563	0.62
Distance to the nearest LGA HQ, in km (SD = 4.13)	6.52	6.98	0.39
Mean toilet ownership rate (%)	37.01	36.57	0.89
Mean trust in neighbours (0-None, 2-High, SD = 0.40)	0.89	0.90	0.88
Social capital (SD = 1)	-0.09	0.05	0.33
Religious fragmentation (0-Low, 1-High)	0.61	0.62	0.77
Attrition rate, endline (%)	11.14	12.04	0.51
Observations	4,540		
<i>p-value</i> F-test	0.0376		
<i>p-value</i> F-test (exc. HH size)	0.2686		

Notes: Mean values measured at baseline. Sample restricted to households interviewed at baseline and in all three followup survey waves. Statistically significant differences between CLTS and control households appear at the expected rate and are found only for household size. Improved toilets refer to toilets of the quality defined using the classification in [WHO/UNICEF \(2015\)](#). Household and community wealth indices are both mean zero, as is the social capital index. Nightlight intensity ranges from 0 to 25, with higher indices indicating more local economic activity. For a detailed description of household and community-level covariates, please see appendix C. ** $p < 0.05$.

3 Empirical method

In our analysis, we use an intent-to-treat (ITT) design based on cluster assignment to treatment, as shown in equation 1.²³ We present instrumented Average Treatment Effects estimates in Section 4.3, and compare the results to our ITT estimates.²⁴

We compare open defecation practices in household i living in community c in period t by treatment assignment. CLTS treatment status, assigned at the community level, is defined by T_c . To account for unobserved community differences and contemporaneous shocks, we include Local Government Area (LGA) and survey wave fixed effects, ω_g and δ_t . The parameter of interest, γ , captures the average impact of CLTS across all followup waves.

$$y_{ict} = \alpha + \gamma T_c + \omega_g + \delta_t + \epsilon_{ict} \quad (1)$$

Secondly, we use an ANCOVA specification (equation 2) by introducing the baseline value of the outcome variable, y_{ic0} . ANCOVA is more efficient than the alternative estimators, difference-in-difference and simple difference estimators, in experimental contexts, and always preferable if pre-treatment information is available and the outcome is strongly correlated over time, as shown in [McKenzie \(2012\)](#). In this specification, we further include baseline covariates (household head and household characteristics), X_{ic0} .

We use the specification:

$$y_{ict} = \alpha + \gamma T_c + X_{ic0}\beta + \theta y_{ic0} + \omega_g + \delta_t + \epsilon_{ict} \quad (2)$$

Subsequently, we explore how CLTS impacts evolve over time and estimate impacts by survey wave. We adopt this approach to understand when impacts unfold and whether they are sustained over time. This is particularly relevant due to possible divestment in toilet maintenance, as dis-

²³ There are two reasons for the intent-to-treat design: First, perfect take-up of interventions is rare, so the intent-to treat design informs us of the key parameter of interest, and likely carries more external validity. Second, due to successful randomisation, as shown in Table 3, intent-to-treat (ITT) designs avoid potentially biased estimates due to selective triggering.

²⁴ In the ATE estimates, we follow [Imbens and Angrist \(1994\)](#) and [Angrist and Imbens \(1995\)](#) and instrument treatment with treatment assignment.

cussed in section 2.4, or potentially short-term lived changes in OD that may not be sustained over time. Impacts are measured roughly eight (FU1), 24 (FU2), and 32 (FU3) months after CLTS implementation. We estimate three coefficients of interest, γ_1 , γ_2 and γ_3 , corresponding to impacts as measured at each followup survey wave, I_t :

$$y_{ict} = \alpha + \sum_{t=1}^3 \gamma_t (T_c \times I_t) + X_{ic0}\beta + \theta y_{ic0} + \omega_g + \delta_t + \epsilon_{ict} \quad (3)$$

Finally, we investigate heterogeneous program impacts by community characteristics. CLTS is a community level intervention, and its designers were keenly aware that its efficiency and impact may depend on local conditions (Kar and Chambers, 2008). We split our sample of communities along a margin of interest, CC_c , say wealth. The specification becomes:

$$y_{ict} = \alpha + \gamma_r T_c + \gamma_d (T_c \times CC_c) + \phi CC_c + X_{ic0}\beta + \theta y_{ic0} + \omega_g + \delta_t + \epsilon_{ict} \quad (4)$$

For robustness, and following the discussion above, we present estimates with and without the baseline value of the outcome variable and estimates with and without household and household head characteristics.

4 Results

In this section, we first present ITT estimates of CLTS' impact on the prevalence of OD, our main outcome of interest. In section 4.1, we explore heterogeneous impacts with regard to the community-level characteristics described in Section 2.5. We identify community wealth, in particular low levels of asset wealth, as an important factor in the effectiveness of CLTS. In Section 4.2, we link CLTS-induced changes in OD patterns to sanitation investments. Finally, we discuss several mechanisms how CLTS may bring about collective change in poor communities (Section 4.3).

Table 4 presents our first impact estimates: estimates of CLTS impacts on the prevalence of OD. The outcome variable takes the value 1 if at least one household member above the age of 4 is reported to perform OD in columns 1 to 3²⁵, while columns 4 to 6 report results for OD by the

²⁵ Note that this question was not asked at FU1.

main respondent. Columns 1, 2, 4 and 5 present difference-in-difference estimates (see equation 1), with and without controlling for household characteristics. Columns 3 and 6 present our preferred estimates, the ANCOVA specification described in equation 2, including household characteristics to maximize precision. All specifications include LGA and survey wave fixed effects.

Table 4: CLTS impacts on open defecation

Outcome =1 if:	OD by any member			OD by the main respondent		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A - Pooled impacts</i>						
CLTS (γ)	-0.03	-0.03	-0.03	-0.03	-0.04	-0.03
<i>p</i> -value	(0.26)	(0.16)	(0.06)	(0.22)	(0.13)	(0.04)
<i>Panel B - Dynamic impacts</i>						
CLTS x FU 1 (γ_1)				-0.04	-0.04	-0.04
<i>p</i> -value				(0.16)	(0.11)	(0.04)
CLTS x FU 2 (γ_2)	-0.04	-0.04	-0.04	-0.03	-0.04	-0.03
<i>p</i> -value	(0.18)	(0.11)	(0.06)	(0.23)	(0.14)	(0.09)
CLTS x FU 3 (γ_3)	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03
<i>p</i> -value	(0.41)	(0.32)	(0.16)	(0.41)	(0.33)	(0.22)
ANCOVA	No	No	Yes	No	No	Yes
Household controls	No	Yes	Yes	No	Yes	Yes
Control Mean (BL)	0.63	0.63	0.63	0.63	0.63	0.63
No. of communities	246	246	246	246	246	246
No. of observations	8,786	8,518	8,518	13,233	12,830	12,697

Notes: OD prevalence for any household member (columns 1-3) includes only members above the age of four, and was not measured in the first followup survey wave. **Household controls:** age, age squared, gender, education attainment level and employment status of the household head; household size, relative wealth asset score, and a dummy variable equal to one if farming is the household's main economic activity. Standard errors are clustered at the community level. *p*-values are shown in parenthesis.

Panel A pools observations across the three followup survey waves. It shows consistently across all specifications that CLTS reduced OD moderately. Estimates are similar across the two OD measures, suggesting that CLTS affects the main respondent's and their household's OD patterns. As expected, the estimated coefficients are identical across specifications, but precision is highest in the ANCOVA specification (columns 3 and 6). Using this as our preferred specification, we reject the null hypothesis of zero impact at the 10% level for both measures of OD: exposure to CLTS resulted in a reduction in OD by 3pp. In panel B, we allow impacts to vary over time, i.e. across post-intervention followup measurements (see equation 3). Point estimates are of similar

magnitude throughout the period of study, and are statistically significant for the first and second followup waves. These estimates point to short-run reductions in OD after eight months. These are sustained without fade-out until at least two years after CLTS.

4.1 Heterogeneous impacts across communities

CLTS is designed as a participatory intervention and implemented at the community-level, with the aim of bringing about collective change. This raises the question whether it has homogeneous impacts across communities. In spite of its current popularity as a low-cost sanitation policy, there is still scant experimental evidence as to where and how CLTS works.

In Section 2.5, we discussed community wealth as a widely available, comprehensive proxy for local economic and social conditions. In this section, we explore whether CLTS impacts vary by community wealth and by three alternative proxies of local socio-economic conditions: i) nightlight intensity, ii) population density, and iii) isolation.²⁶ We saw in Table 2, that poor communities are often remote, less densely populated and have lower nightlight activity. Yet, some pairwise correlations between the four characteristics are relatively low, suggesting that they each may emphasize a different aspect of socio-economic conditions.

Table 5 presents these heterogeneous impact estimates, expressed as percentage point changes in OD, using the pooled sample and specification 4. The outcome variable captures OD practice of the the main respondent, as it is measured in all three survey waves. Each column presents heterogeneous impacts by one of five measures of economic conditions at the community level, CC_c , which we discretize along the median cutoff. For example, poor communities are those with below median wealth.²⁷ We do the same for the other community measures. Our results are robust to an alternative linear specification (see Table 17 in Appendix D).

Two sets of p -values are presented. Naive or unadjusted p -values obtained from individual significant tests for each point estimate, calculated by taking 1,000 clustered bootstrapped samples, are shown in parentheses. Since we are testing multiple hypotheses simultaneously in this section

²⁶ Details on all four measurements are available in Appendix C.2.

²⁷ Specifically, we rank all communities according to their wealth score, and then define half of the communities in our sample as poor. CC_c becomes a dummy variable which is equal to one if community c has below median wealth, and zero otherwise.

(Tables 5 and 6), we report p -values that are adjusted for the family-wise error rate in brackets. We compute these using the methodology proposed by [Romano and Wolf \(2005\)](#) and the same cluster-bootstrapping procedure.

Table 5: Community economic conditions and CLTS impacts on OD

Community Characteristic (CC) at BL:	Dep. variable: main respondent performs OD				
	Asset wealth	Night lights	Density	Isolation	Toilet coverage
	(1)	(2)	(3)	(4)	(5)
CLTS x High (γ_r)	0.01	0.00	0.01	-0.07	-0.01
p -value (naive)	(0.50)	(0.86)	(0.73)	(0.01)	(0.50)
p -value (MHT robust)	[0.90]	[0.97]	[0.95]	[0.03]	[0.90]
CLTS x Low ($\gamma_r + \gamma_d$)	-0.09	-0.07	-0.08	0.00	-0.05
p -value (naive)	(0.01)	(0.01)	(0.01)	(0.98)	(0.09)
p -value (MHT robust)	[0.01]	[0.03]	[0.02]	[0.98]	[0.23]
Difference (γ_d)	-0.10	-0.07	-0.09	0.07	-0.04
p -value (naive)	(0.00)	(0.02)	(0.01)	(0.03)	(0.26)
p -value (MHT robust)	[0.02]	[0.12]	[0.06]	[0.12]	[0.66]
Control Mean (BL) - High	0.53	0.54	0.54	0.69	0.45
Control Mean (BL) - Low	0.75	0.71	0.73	0.57	0.85
No. of communities	246	246	246	246	246
No. of observations	12,697	12,697	12,697	12,697	12,697

Notes: All specifications control for the household and household head characteristics listed in Table 4. Errors are clustered at the community level. Naive (unadjusted) p -values shown in parenthesis. In brackets we present p -values adjusted by family wise error rate following [Romano and Wolf \(2005\)](#), using 1,000 cluster bootstrap samples and estimated jointly for the regressions presented in Tables 5 and 7.

Table 5 shows strikingly consistent heterogeneous CLTS impacts for four of our proxy measures for community wealth. CLTS reduced OD prevalence by 7-9pp in communities with low asset wealth, those that are far away from administrative capitals (note the scale reversal here), communities with low population density, and low night light intensity (see second row). In these communities, impact estimates more than double those found on average (see Table 4).

Yet, we find no evidence of any statistically significant CLTS impacts in richer communities (see first row), regardless of whether we proxy these through asset wealth, nightlight intensity, density or through low isolation, i.e. being in close proximity of the next LGA capital. In addition, the magnitude of all estimated coefficients is very close to zero.

The third row of Table 5 presents the point estimates for the difference in CLTS impacts between the two halves of the sample, i.e. the point estimates for the interaction term γ_d in specification 4. We can reject the hypothesis that CLTS had the same impact on communities below and above the

median of the respective characteristic if we find that γ_d is statistically significantly different from zero. We reject the hypothesis for wealth, and find that the difference in CLTS impacts between poor and rich communities was 10pp (see column 1). Similarly, we reject it for all wealth measures according to naive p-values, at the 5% level at least. Using multiple hypothesis testing, we reject it for wealth and density. These results suggest that CLTS was indeed effective in a sub-sample of communities that shared underlying characteristics of low socio-economic status.

One possible explanation for our findings could be that poorer communities have lower toilet coverage (and higher OD rates) at baseline, hence leaving ‘more room for improved sanitation practices following the intervention. We define toilet coverage as the mean toilet ownership rate at baseline and explore this hypothesis in column 5 of Table 5. We find no evidence that CLTS is more effective in communities with lower toilet coverage at baseline.

As a further robustness check, we run a horserace between the four correlated community characteristics. If community wealth captures heterogeneous impacts well, the estimates should be robust to including interactions with the other three dimensions of community socio-economic conditions. Columns 1 to 3 in Table 6 show estimates allowing for pairwise heterogeneity (i.e. by community wealth and one of the other three variables), while column 4 displays the estimates of including all four characteristics (and their interaction terms) simultaneously. Due to their relatively low correlations with community wealth (see Table 2), we find heterogeneous impacts with respect to all three alternative characteristics in the pairwise specifications. This suggests that each indicator captures a dimension of community which matters for the successful implementation of CLTS. Yet, the most striking result is this: We detect strongly statistically significant and economically identical community wealth-specific CLTS impacts (that are also of equal magnitude than those estimated in Table 5) in all specifications. In fact, when we include all interaction terms simultaneously, only the community wealth-specific impacts are statistically significant at the 5% level. These results suggest that community wealth is a good proxy for communities’ socio-economic conditions, and that CLTS impacts vary by communities’ socio-economic status in a robust fashion.

An additional concern may be that community wealth is an aggregated measure of house-

Table 6: Performance of different proxies for community-level economic conditions

Asset wealth compared to:	Night lights (1)	Density (2)	Isolation (3)	All (4)
CLTS	0.04 (0.11)	0.05 (0.07)	0.03 (0.15)	0.05 (0.05)
Low asset wealth	0.08 (0.00)	0.08 (0.00)	0.08 (0.00)	0.08 (0.00)
CLTS x Low asset wealth	-0.09 (0.01)	-0.09 (0.01)	-0.09 (0.01)	-0.09 (0.01)
Low night lights	0.07 (0.00)			0.04 (0.12)
CLTS x Low night lights	-0.05 (0.10)			-0.03 (0.55)
Low density		0.08 (0.00)		0.06 (0.03)
CLTS x Low density		-0.08 (0.01)		-0.07 (0.07)
High isolation			0.07 (0.00)	0.01 (0.78)
CLTS x High isolation			-0.05 (0.09)	0.01 (0.91)
No. of observations	12,697	12,697	12,697	12,697

Notes: All specifications control for the household and household head characteristics listed in Table 4. *p*-values are shown in parenthesis. Standard errors are clustered at the community level.

hold wealth. Richer (poorer) communities are indeed on average composed of richer (poorer) households,²⁸ so our estimates may simply proxy household rather than capture community-level heterogeneity. So is CLTS is simply (more) effective among poorer households? We test this by accounting for household-level wealth. Our findings suggest that treatment heterogeneity seems predominantly driven by community rather than household wealth. Yet, aggregating wealth to the community level may reduce measurement error in household wealth, and provide a more robust measure of households' economic conditions, or long term wealth. To address this concern, we use a proxy for household wealth: whether the household head had completed primary education at baseline. Again, we see no evidence of heterogeneity in treatment impacts along this dimension

²⁸ 69% of the households living in poor TUs have below-median asset wealth, while 66% of the households living in rich TUs have above-median wealth.

(see Table 19 in appendix D.2). We also do not find evidence that household characteristics that are often associated with higher levels of investment in sanitation and health technologies in the literature (see, e.g. [Miller and Mobarak \(2013\)](#)), such as having a female head of household and the presence of children in the household, mediate the impact of CLTS. A detailed discussion of these results can be found in Appendix D.2.

Following [Kar and Chambers \(2008\)](#) and [Cameron et al. \(2015a\)](#), we further test whether CLTS implementation in our study context had heterogeneous impacts along i) community-level trust, ii) social capital and iii) religious fragmentation. Trust is constructed as the average community score for a measure of the degree to which households trust their neighbors, following [Alesina and La Ferrara \(2002\)](#). Social capital captures households' participation in community life, social contact and provision of services to the community, as in [Cameron et al. \(2015a\)](#). We extract a community-level index, using principal component analysis. Fragmentation is based on religious fragmentation, following the measures designed by literature on ethnolinguistic fragmentation (ELF).²⁹ We find no evidence that CLTS had heterogeneous impacts across communities with varying levels of neighbourhood trust, social capital or religious fragmentation.

In summary, our results suggest that average impact estimates mask large heterogeneity in CLTS impacts, and that the intervention has strong and lasting effects on OD habits in communities of low socio-economic status. In poor communities, OD rates decreased by 9pp from a baseline level of 75% - more than twice as much as in the average impact estimates.

4.2 CLTS, toilet construction and open defecation

CLTS's ultimate aim is to reduce open defecation and as we have seen thus far, it successfully reduced it by 10pp in poor communities. How was this reduction achieved? We propose four possible channels that could have contributed. First, CLTS could have promoted investment in new toilets, increasing toilet coverage and its usage. Second, it could have persuaded households who owned toilets at baseline to invest more heavily into their maintenance and upkeep. This would

²⁹ We use religious rather than ethnic fragmentation to study 'homogeneity', because our study sample is homogeneous along ethnic lines but very diverse in terms of religion. Table 5 in the appendix shows the distribution of these measures.

Table 7: CLTS impacts on OD by other community characteristics

Community Characteristic (CC) at BL:	Dep. variable: main respondent performs OD		
	Trust (1)	Social capital (2)	Fragmentation (3)
CLTS x High (γ_r)	-0.03	-0.05	-0.02
<i>p</i> -value (naive)	(0.17)	(0.04)	(0.30)
<i>p</i> -value (MHT robust)	[0.57]	[0.20]	[0.79]
CLTS x Low ($\gamma_r + \gamma_d$)	-0.04	-0.01	-0.04
<i>p</i> -value (naive)	(0.11)	(0.49)	(0.08)
<i>p</i> -value (MHT robust)	[0.23]	[0.49]	[0.23]
Difference (γ_d)	-0.02	0.04	-0.02
<i>p</i> -value (naive)	(0.60)	(0.25)	(0.56)
<i>p</i> -value (MHT robust)	[0.79]	[0.66]	[0.79]
Control Mean (BL) - High	0.63	0.66	0.62
Control Mean (BL) - Low	0.62	0.59	0.63
No. of communities	246	246	246
No. of observations	12,697	12,697	12,697

Notes: All specifications control for the household and household head characteristics listed in Table 4. Errors are clustered at the community level. Naive (unadjusted) p-values shown in parenthesis. In brackets we present p-values adjusted by family wise error rate following [Romano and Wolf \(2005\)](#), using 1,000 cluster bootstrap samples and estimated jointly for the regressions presented in Tables 5 and 7.

increase the stock of *functioning* toilets, by reducing de-accumulation. In addition to investment decisions, CLTS may have influenced the sanitation behavior of treated households, increasing toilet *usage* among toilet-owning households. Finally, CLTS may have stimulated shared toilet usage, e.g. by neighbors.

We make use of the detailed toilet stock, flow and quality measures, described in section 2.4, to determine the importance of these channels. Table 8 presents the results of re-estimating Tables 4 and column 1 in Table 5 using these measures. As an additional check, we include a measure of shared toilet use equal to one if the main respondent declares to use a shared toilet (e.g. owned by a neighbour, public toilet or toilet at school or work). The reduction by 9pp in OD observed in poor communities was almost identically matched by an increase in toilet ownership by 8pp (column 1), pointing to an increase in the stock of toilets. Ownership of functioning toilets, i.e. maintained stock, increased by 10pp, slightly more than the reduction in OD reported by the main respondent in poor communities (see column 2). In contrast, usage of existing toilets (column 3) and of shared facilities in poor communities increased by much less or not at all, relative to the two

Table 8: Channels of OD reduction

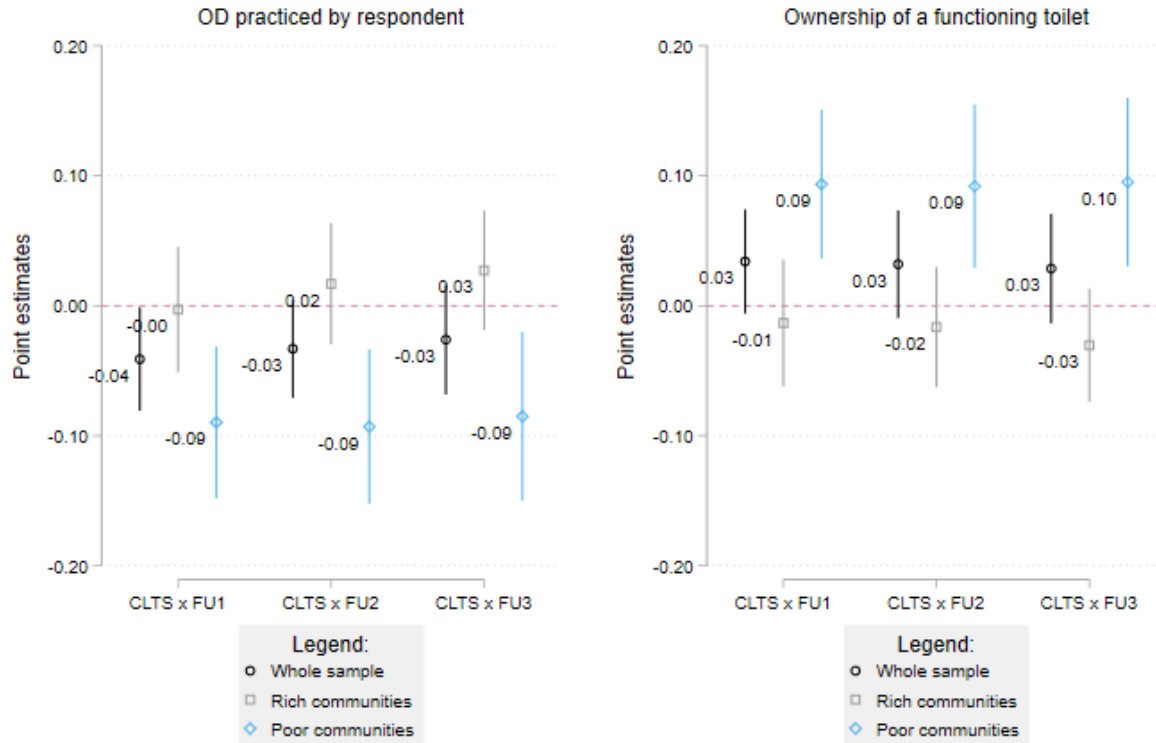
Outcome =1 if:	Owns toilet (1)	Owns functioning toilet (2)	Usage (if functioning) (3)	Shared use (4)
<i>Panel A - Pooled impacts</i>				
CLTS (γ)	0.02	0.03	0.00	0.01
<i>p</i> -value (naive)	(0.25)	(0.07)	(0.83)	(0.25)
<i>p</i> -value (MHT robust)	[0.55]	[0.21]	[0.83]	[0.55]
<i>Panel B - By community wealth group</i>				
CLTS x Rich (γ_r)	-0.03	-0.02	-0.01	0.01
<i>p</i> -value (naive)	(0.10)	(0.25)	(0.41)	(0.02)
<i>p</i> -value (MHT robust)	[0.23]	[0.42]	[0.42]	[0.07]
CLTS x Poor ($\gamma_r + \gamma_d$)	0.08	0.10	0.04	-0.00
<i>p</i> -value (naive)	0.01	0.01	0.16	0.99
<i>p</i> -value (MHT robust)	[0.01]	[0.01]	[0.31]	[0.99]
Difference (γ_d)	0.12	0.12	0.05	-0.01
<i>p</i> -value (naive)	(0.00)	(0.00)	(0.11)	(0.37)
<i>p</i> -value (MHT robust)	[0.01]	[0.01]	[0.20]	[0.37]
Control Mean (BL) - Rich	0.47	0.45	0.94	0.04
Control Mean (BL) - Poor	0.24	0.24	0.91	0.03
No. of communities	246	246	245	246
No. of observations	12,497	12,497	7,113	12,697

Notes: All specifications control for the household and household head characteristics listed in Table 4. *p*-values are shown in parenthesis. Standard errors are clustered at the community level and are adjusted for family-wise error rate following Romano and Wolf (2005), using 1,000 cluster bootstrap samples.

investment mechanisms. In both of these cases, the differential impact of CLTS across community wealth levels is not statistically significant. Appendix D.3 shows that results are robust when we estimate heterogeneous impacts using the three alternative measures of communities socio-economic conditions. In sum, we interpret the results from Table 8 as supporting the hypothesis that OD reductions brought about by CLTS were due almost exclusively to increases in toilet ownership, with a similar likelihood of these toilet being used as before CLTS was implemented.

As shown in Section 4.1, CLTS reduced OD in the short-run in poorer communities, and impacts were sustained over time. We re-examine these dynamic results to provide further evidence that CLTS reduces open defecation through increases in toilet ownership (see Figure 3). The estimated short- and long-run impacts of CLTS on both OD (see left panel) and toilet ownership (right panel) across the three followup periods are remarkably constant over time among poor communities. In consequence, our interpretation that CLTS-induced OD reductions are realized via increased in

Figure 3: CLTS treatment effects on OD and toilet ownership over time



Note: Graphs plot the point estimates for CLTS impacts by follow up survey wave and community level wealth. Results for the whole sample from the leftmost plot are equivalent to those presented in column 6 of Table 4. All specifications control for the household and household head characteristics listed in Table 4. Robust standard errors are clustered at the community level.

toilet ownership holds also in the dynamic context. This suggests that CLTS has had an instantaneous and persistent effect in poor communities on OD through toilet uptake, akin to a one-shot policy.

4.3 Mechanisms of impact in poor communities

So far, we have established the degree to which CLTS was successful in reducing OD in Nigeria. OD fell by 9pp in poor communities, and this was tied to an increase in the construction of toilets in these areas of about the same magnitude. In this section, we explore why CLTS, an intervention with fairly standardized design and content, that provides no-subsidy, no-credit intervention, led

to increased toilet construction in poor communities.

The heterogeneous impacts by community wealth can arise due to ‘supply side effects’, i.e. heterogeneity across communities in the delivery of the intervention, or due to differential ‘demand-side responses’. We will discuss these briefly. Our study is informative about some aspects of delivery quality and demand side factors which have not received much attention in the literature.

CLTS is a fairly standardized intervention, so any delivery differences are likely to arise from differential triggering and the effectiveness or skill of the delivery agent. In recent work based on the same RCT, [Abramovsky et al. \(2016\)](#) show that CLTS triggering meetings are more likely to fail, and not be carried out at all, in areas with high population density which is positively correlated with community wealth. Using community wealth directly, we find that 71% (83%) of rich (poor) communities were triggered. Yet, we find no evidence of selective triggering by community wealth (p-value=0.138).

A plausible hypothesis would nevertheless be that CLTS is more effective in communities in which the CLTS meeting actually took place, and these communities tend to be poorer. Our ITT estimates do not capture this. Hence, we replace treatment assignment, T (as used in our ITT design), by actual triggering, D , and re-estimate CLTS impacts, using treatment assignment as instrumental variable (see left panel of Table 9). We find an overall decrease in OD by 4pp (see column 1), which is very similar to our ITT results from Table 4. When we estimate heterogeneous impacts (columns 2 and 3), the decline in OD in by 10pp in poor communities is again very similar to the ITT results from Table 5. Hence, we conclude that the likelihood of arranging and conducting the CLTS meeting is only slightly higher in poor communities, and differential CLTS impacts by community wealth are not due to differences in triggering.

Supply side differences in the quality of delivery may have differed across states, as the teams organizing and delivering the treatment in Enugu and Ekiti were not the same. While LGA fixed effects may deal with these issues, there may be systematic differences between delivery teams in rich and poor areas across states. Results in columns (4) to (7) in Table 9 show that this is not the case, as poorer communities in both states show similar reductions in open defecation.

On the demand side, effects arise if the information provided by CLTS changes households’

Table 9: Other possible channels of CLTS impacts

Sample:	Triggering			Implementation			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. var.: OD	All	Rich	Poor	Enugu	Enugu	Ekiti	Ekiti
CLTS triggered (1=yes, 0=no)	-0.04 (0.04)	0.02 (0.55)	-0.10 (0.00)				
CLTS				-0.02 (0.35)	0.06 (0.07)	-0.05 (0.04)	-0.02 (0.44)
CLTS \times Poor TU					-0.13 (0.00)		-0.13 (0.07)
Estimation method	IV	IV	IV	OLS	OLS	OLS	OLS
No. of communities	246	123	123	118	118	128	128
No. of observations	12,697	6,515	6,182	6,523	6,523	6,174	6,174

Notes: Dependent variable: whether main respondent performs OD. The independent variable in columns 1-3 is an indicator of whether the CLTS triggering meeting was successfully conducted, and is instrumented by selection into the treatment (CLTS) group. All specifications control for the household and household head characteristics listed in Table 4. p -values in parenthesis. Errors are clustered at the community level.

perceptions of sanitation (more strongly) in poor communities. At the point of investment, a household's utility from owning a toilet depends on the net perceived benefit of this investment, NB_{ic} . These are a function of the expected private benefits received from sanitation, B_{ic} , the expected costs of construction, C_{ic} , and household characteristics η_{ic} , such that $NB_{ic} = g(B_{ic}, C_{ic}, \eta_{ic})$. A first channel through which CLTS may increase sanitation investment is by raising households' perception of the expected benefits of owning a private toilet. It does so by stressing the negative private and social impacts of OD. While facilitators leave the course of action in the decision of the community, open discussions in the CLTS meetings usually highlight private toilets. As discussed in Section 2, CLTS is also designed to instill a sense of pride and collective contribution of toilet ownership. If successful, this would imply that the expected benefit, B_{ic} , changes through treatment exposure: $B_{ic} = f(T_c, \epsilon_{ic})$. Hence, we test i) whether expected benefits increase in treated communities $E[B|X, C, T = 1] \geq E[B|X, C, T = 0]$ generally, and ii) whether they increase by more among poor treated communities (in comparison to poor control communities). The second channel through which CLTS may increase sanitation investment is by decreasing households' perception of the expected costs of the investment. While construction costs are not discussed in the meeting, households might gather information post triggering that correct their previously held

cost expectations.

In the household surveys, we asked several questions about the expected benefits of sanitation (see details in Appendix C.1), the policy lever targeted by CLTS, as well as the expected costs of toilet construction. We use this data to construct the following measures. First, we develop two measures of expected benefits: first, we create an index of awareness about the costs of OD in terms of health and physical risks. Given that OD is the natural outside option for toilet ownership, and we found OD practices and lack of private toilet ownership to be closely linked in the last section, we interpret the perceived costs of OD as a measure of expected benefits from toilet ownership. Second, we asked households that did not own a toilet what they believe their social standing in their community would be if they were to own a toilet. Finally, we asked households how much they think it would cost to build a toilet, weighing the response estimates of four different toilet models.

We estimate CLTS impacts on expected benefits and costs, otherwise keeping the same specification (see equation 4). Instead of the baseline level of the dependent variable, y_{ic0} , we control for baseline costs C_{ic} to account for initial perceptions that may persist over time. CLTS may be more successful in correcting misperceptions about the benefits of sanitation in poor communities, as people in poor communities may lack information about the benefits of sanitation before the intervention. However, we find no statistically significant differences at baseline between treatment and control groups in expected benefits, that would lend support to this hypothesis. We also do not find pre-intervention differences in the social status derived from toilets between poor and rich communities. Expected construction costs for households in rich communities were 0.3 standard deviations higher than for those in poor communities.

Our findings suggest that the expected costs of construction have been adjusted downwards in poorer communities after the intervention (see column 3 in Table 10). At the same time, CLTS raises awareness of OD costs (column 1) and the social status of toilet ownership (column 2) in poor communities, but not in rich ones (see row 2 in Table 10). Yet, the impact estimates are statistically significant at the 10% level (using naive standard errors or a multiple hypothesis testing adjustment)

Table 10: CLTS impacts on expected benefits and costs from toilet ownership

Components of net benefits (NB):	Expected benefits from toilet ownership (B)		Expected costs of toilet construction (C)
	Awareness of OD costs	Social status if toilet	Weighted by type in logs
	PCA Index	1-Low, 10-High	
	(1)	(2)	(3)
CLTS x Rich (γ_r)	-0.11	-0.24	0.01
<i>p</i> -value (naive)	(0.42)	(0.26)	(0.87)
<i>p</i> -value (MHT robust)	[0.65]	[0.60]	[0.87]
CLTS x Poor ($\gamma_r + \gamma_d$)	0.21	0.38	-0.08
<i>p</i> -value (naive)	(0.16)	(0.04)	(0.05)
<i>p</i> -value (MHT robust)	[0.16]	[0.12]	[0.12]
Difference (γ_d)	0.32	0.63	-0.09
<i>p</i> -value (naive)	(0.12)	(0.03)	(0.13)
<i>p</i> -value (MHT robust)	[0.21]	[0.07]	[0.21]
Mean dep. variable (BL) - Rich	-0.07		-0.05
Mean dep. variable (BL) - Poor	0.04		0.03
No. of communities	230	212	244
No. of observations	3,850	1,188	3,103

Notes: Restricted to observations from FU3. The dependent variable used in column 1 is an index of awareness of the costs of OD built using answers to four questions. In column 2, we use the answers to a question asking what the social status of the household would be, if they were to own a toilet. This was asked only to households with no toilets. In column 3 we use the natural logarithm of the expected cost of construction of a toilet, which was asked to all households at baseline and in FU3. See appendix C for more details on the construction of the outcome variables. All specifications control for the household and household head characteristics listed in Table 4. Errors are clustered at the community level and adjusted by family wise error rate following [Romano and Wolf \(2005\)](#).

only for our measure of toilets as a good that delivers social status.³⁰ The third row of Table 10 shows differential impacts between rich and poor communities. It confirms that CLTS increases the expected status derived from toilet ownership in poor communities relative to richer communities. There may be several explanations for this. First, this is consistent with CLTS' aim to instill a sense of pride and collective contribution from toilet ownership. This intervention component may be more effective in triggering sanitation investment, as toilets may be more visible, and, due to lower levels of wealth, more of a luxury good in poor communities than in richer ones.

Additionally, another demand-side channel may arise if social interactions in poorer communities are stronger. In Appendix G we consider several additional aspects of social interactions, such as

³⁰ We measure expected benefits and costs of toilet ownership in the last followup survey wave, but include averages at baseline when available. Our estimates therefore have lower power, particularly after performing multiple hypothesis adjustments to the *p*-values.

sanctions, private or public help or better diffusion of information about externalities, and find no evidence suggesting that these channels are at work.

5 Comparison across CLTS studies

To the best of our knowledge, five other studies have experimentally estimated the impact of CLTS-like interventions in developing countries. The World Bank’s Water and Sanitation Programme (WSP) conducted three of these studies, located in Tanzania ([Briceno et al., 2015](#)), the Indian state of Madhya Pradesh ([Patil et al., 2014](#)) and East Java, Indonesia ([Cameron et al., 2013](#)). [Pickering et al. \(2015\)](#) conducted a similar evaluation in rural Mali. Finally, [Guiteras et al. \(2015\)](#) carried out a cluster-randomized trial in Tanore, Bangladesh, in which they evaluated the impact of three different policy approaches, one of which one a CLTS-style intervention.³¹ These studies have produced inconclusive results on the impact of CLTS. Results range from increases in toilet ownership by 30pp in Mali to no statistically-detectable impacts in the studies conducted in Bangladesh and Indonesia.

In this section, we argue that community-level wealth might be the underlying factor explaining these diverging results. We have documented the strong heterogeneity of CLTS impacts by community wealth in our field experiment in Nigeria. If these results are externally valid, CLTS interventions in richer areas will generally have lower (or no detectable) impact than those administered in poorer areas. To this end, we exploit the variation in average characteristics of each study sample and, abstracting from sampling and implementation differences across studies, analyze how CLTS effectiveness varied along these characteristics.

We faced two main challenges in this comparison exercise. The first one was to arrive at consistent estimates for all studies. While our preferred specification has been an ANCOVA with household level controls, other studies have relied on simple-differences or first-differences estimations. A first step in our analysis was therefore to download the datasets behind each of the studies and estimate CLTS treatment effects in a consistent manner.³² Datasets were not pooled: the

³¹ The other two were supply side technical assistance and subsidy provision.

³² The datasets for the studies conducted in Bangladesh, India, Indonesia and Tanzania were made publicly available by the authors of the studies. They can be found at: [Guiteras et al. \(2017\)](#) (Bangladesh), [Water and](#)

objective was to calculate study-specific treatment effects only. Data from the Mali intervention were not accessible at the time of writing this paper, so the treatment effects reported in the paper were used instead.

The only specification all studies had enough data for was a simple-differences approach. So for each study, including our own, we estimated CLTS impacts using the following regression model:

$$y_{icd} = \alpha + \gamma T_{cd} + \omega_d + \epsilon_{icd} \quad (5)$$

The outcome variables y_{icd} , were ownership of a private, functioning toilet and whether any member of the household performed OD. T_{cd} is a dummy variable equal to one if cluster c from district d was assigned to CLTS, and zero otherwise.³³ We did not include baseline values for the outcome variables, since no baseline data was collected in the case of Tanzania. Our study was the only one to include multiple post-treatment survey waves, so we only used observations from the last wave (FU3), which is the closest in terms of time-since-treatment to the other studies. We included geographical fixed effects (ω_d) where appropriate, to account for stratified randomization.³⁴ Standard errors were clustered at the unit of randomization. The data used in [Pickering et al. \(2015\)](#) was not publicly available, so we use the estimates presented in their paper, also obtained by a means comparison of outcomes between treatment and control households at endline. The OD outcome selected for this exercise was whether any adult female performed OD.

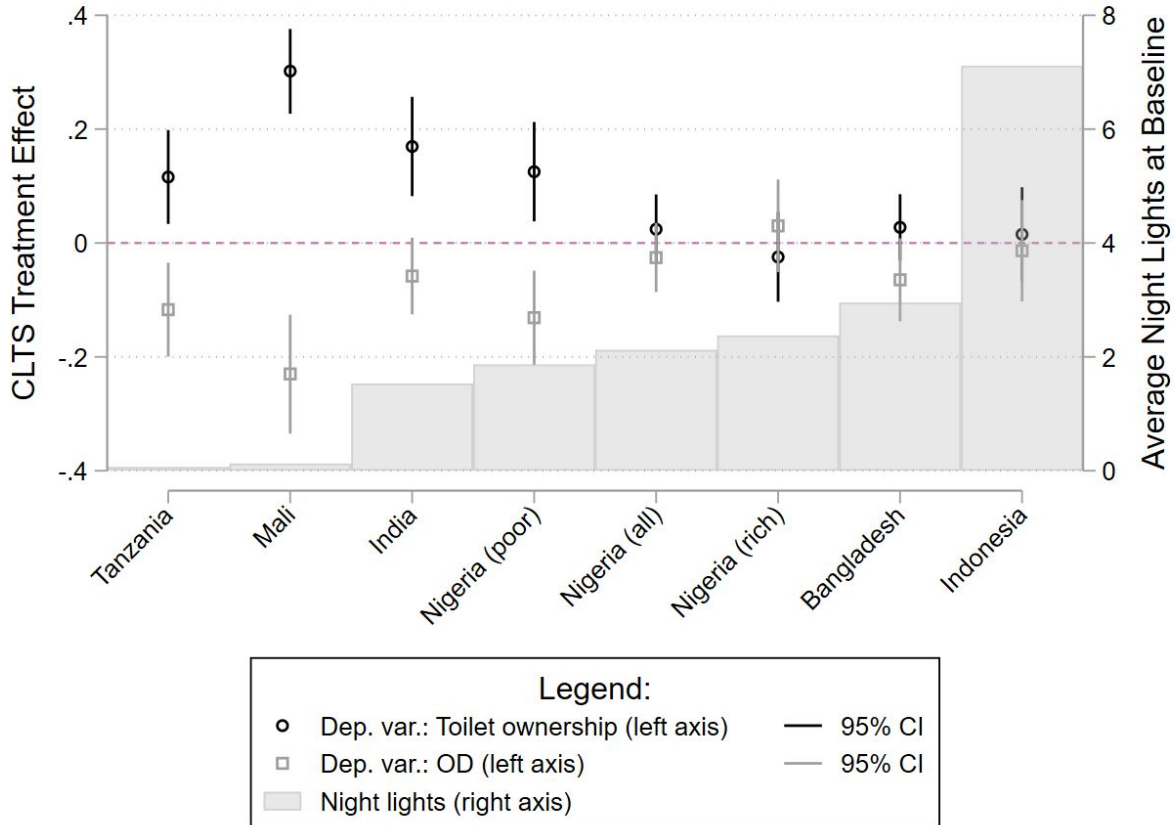
The second hurdle in the exercise was to obtain a comparable measure of the average wealth of each study area. Asset wealth indices as the one used in our study are highly context-specific ([Filmer and Pritchett, 2001](#)). In addition, information on the ownership of consumer durables was not available for all studies. The same limitations applied to two other measures used in Table 7: distance to the nearest district headquarters and population density. The exact location of study sites was not available for any of the cited studies for data confidentiality reasons. We therefore used

Sanitation Program and Network for Engineering and Economics Research and Management (2009) (India), [Water and Sanitation Program \(2008\)](#) (Indonesia), and [Briceno et al. \(2012\)](#) (Tanzania).

³³ In cases where multiple treatment arms existed, we included only the CLTS arm and the control group.

³⁴ District level fixed effects were used to estimate impacts for the case of Nigeria, Tanzania and Indonesia. No geographical fixed effects were included in the cases of Bangladesh and India, since both studies were conducted in a single district.

Figure 4: CLTS impacts on OD and toilet ownership by average study area night light index



Note: Study-specific point estimates from a simple differences regression of CLTS impacts on OD prevalence (gray squares) and toilet ownership (black dots). Gray bars show the average night light intensity recorded by NASA's Operational Line Scan (OLS) system, on the year of each study's baseline survey. In the case of Tanzania, no baseline survey was conducted, so the base year for night light measures was set at 2010, the year during which the intervention was implemented. Data from the Mali study was unavailable, so results from Table 2 from [Pickering et al. \(2015\)](#) were used instead (in the case of OD, results for adult women were used).

each study area's average night light index, which we showed to be a robust measure of community wealth effects in Table 7. First level administrative divisions were used in Bangladesh, Indonesia, Mali and Tanzania. This is the lowest level for which we could match the observations of all six studies. In India, we calculated average night lighting by district (second level) instead, due to the large size of the state where the intervention was conducted (Madhya Pradesh). As a result, we obtained an average night light intensity value for each study.

Figure 4 plots the point estimates for CLTS treatment effects by study (left axis), together with the average night light index over the areas of study (right axis). For Nigeria, we included our

estimates for the whole sample, as well as those for the two halves (poor and rich), split along the lines of community wealth. Gray squares present the point estimates for CLTS treatment effects on OD, while black dots do the same for ownership of a functioning toilet.

The four samples with high average night light index (i.e. Nigeria (all), Nigeria (rich), Bangladesh and Indonesia) all reported statistically insignificant reductions in OD. The same is true when we look at impacts on toilet ownership. Confidence intervals for both outcomes include zero in all four samples. In contrast, reductions in OD were detected at standard levels of significance in three of the four samples with low night light index averages (Tanzania, Mali and Nigeria (poor)), and toilet ownership increased in all four of these studies. Thus, the heterogeneity in CLTS impacts across studies can be rationalised by differences in the average area wealth (measured through nightlight) in which they were conducted.³⁵

Our comparison analysis suggests that the key result of our field experiment - that CLTS impacts are higher (or non-zero) in poorer areas - extends to other contexts as well. In other words, community characteristics matter. This has three important implications: First, the source of heterogeneity observed in the Nigerian CLTS experience might hold the key to understanding previous contradicting results from studies conducted elsewhere. Second, our findings suggest that interventions like CLTS may have a higher return when targeted at poorer communities. We find no evidence of CLTS being effective at reducing OD or increasing toilet ownership in areas with higher or equal night light intensity, neither within our Nigerian field experiment nor when we compare across studies. In poorer areas (i.e those with lower night light intensity), however, we find that CLTS is effective. This finding strengthens our hypothesis about the importance of community level characteristics as key mediators of policy effectiveness. Finally, our results provide an additional reason for why scale-up of interventions is not trivial. The literature has suggested that spillover

³⁵ In a separate exercise, we pooled the five studies for which data was available (i.e. we excluded Mali, where stronger effects have been found) and extracted average night light intensity at the second administrative level (districts). We included an interaction term between T and the measure of nightlight at the district level in equation 5, and found that CLTS increases toilet ownership (coefficient 0.092, p-value 0.001) and reduces open defecation (coefficient -0.080, p-value 0.007). As suggested in Figure 4, the treatment effect is slightly lower in areas with higher night light intensity, as observed in the coefficient on the interaction term, for both toilet ownership (coefficient -0.015, p-value 0.072) and open defecation decreases (coefficient 0.010, p-value 0.220). These results are only suggestive, but agree with our findings above. More precise estimations would require access to the exact locations of the study areas, in order to estimate more accurately the degree of night lighting in each community, but this is currently unavailable for confidentiality reasons.

and moderating general equilibrium effects may lead to a lowering of the return, when interventions conducted in areas with specific characteristics are being rolled out universally, e.g. in richer areas.

6 Conclusions

Sanitation remains an urgent concern for policy makers in the developing world. Effective policy design requires a nuanced understanding of what drives households investment choices regarding health and sanitation, and associated behavioral change. This paper contributes to this effort using findings from a cluster randomized experiment carried out in the Nigerian states of Ekiti and Enugu.

We demonstrate that CLTS, an information and mobilization intervention, designed to curb open defecation levels with no monetary incentives involved, had positive but moderate effects on open defecation and toilet construction. However, these results hide important community heterogeneity, as the intervention has strong and lasting effects on OD habits in poorer communities, and increased sanitation investments. This result is robust across several measures of community socio-economic characteristics, and is not driven by baseline differences in toilet coverage. In poor communities, OD rates decreased by 9pp from a baseline level of 75%, while we find no effect in richer communities. The reduction in OD is achieved mainly through increased toilet ownership (+8pp from a baseline level of 24%). These statistics highlight an important caveat: While CLTS was effective in these poor communities, it was far from successful in achieving open defecation *free* communities, its main aim. In terms of channels, CLTS is designed to activate pride in toilet ownership, by mapping out existing toilets and by highlighting the value of sanitation for the community. Using subjective reports of households, we find that the intervention raised the social status that households in treated communities attach to toilet ownership, but only in poor communities.

Finally, we use data from our study and five other RCTs of similar interventions and show that estimated impacts on OD are generally stronger and significant only in poorer contexts. More specifically, we find an inverse relationship between area-level wealth, measured by nightlight intensity, and program effectiveness. This points to an important role of community characteristics in

mediating the effectiveness of sanitation interventions, and rationalizes the wide range of estimates in the literature.

This has important implications for policy and for evaluations of interventions. First, several arguments have been made why the impacts of interventions may not scale-up successfully in a national roll-out. These comprise general equilibrium and spillover effects. We highlight that scale-up may yield differential impacts if intervention impacts differ by local community characteristics. We show that CLTS-type sanitation interventions are (more) effective in poor communities. Second, our findings are important for the targeting of interventions. Governments with restricted funds may achieve larger improvements in sanitation if they target this type of policies to less wealthy communities first.

References

- Abramovsky, Laura, Britta Augsburg, Erin Flynn, and Francisco Oteiza**, “Improving CLTS targeting: Evidence from Nigeria,” Technical Report IFS Briefing Note BN183, Institute for Fiscal Studies 2016.
- Adukia, Anjali**, “Sanitation and Education,” *American Economic Journal: Applied Economics*, April 2017, 9 (2), 23–59.
- Alesina, Alberto and Eliana La Ferrara**, “Who trusts others?,” *Journal of Public Economics*, 2002, 85 (2), 207–234.
- , **Reza Baqir, and William Easterly**, “Public goods and ethnic divisions,” *Quarterly Journal of Economics*, 1999, 114 (4), 1243–1284.
- Alzua, Maria Laura, Habiba Djebbari, and Amy J Pickering**, “A collective approach promotes private supply of sanitation,” 2017.
- Angrist, Joshua D. and Guido W. Imbens**, “Two-stage least squares estimation of average causal effects in models with variable treatment intensity,” *Journal of the American Statistical Association*, 1995.
- Bandiera, Oriana, Robin Burgess, Erika Deserrano, Imran Rasul, and Munshi Sulaiman**, “Social Ties, Identity and the Delivery of Public Services,” 2018.
- Briceno, B, Aidan Coville, and Sebastian Martinez**, “Promoting handwashing and sanitation: evidence from a large-scale randomized trial in rural Tanzania.,” *Policy Research Working Paper - World Bank*, 2015, 7164 (WPS7164), 58–pp.
- Briceno, Bertha, Aidan Coville, and Sebastian Martinez**, “Impact Evaluation of Scaling-up Handwashing and Rural Sanitation Behavior Projects in Tanzania 2012, Endline Survey (SHRSBIE-EL). Ref. TZA_2012_SHRSBIE-EL_v01_M_v01_A_PUF,” 2012.
- Cameron, Lisa A., Susan Olivia, and Manisha Shah**, “Initial Conditions Matter: Social Capital and Participatory Development,” 2015.

- Cameron, Lisa, Manisha Shah, and Susan Olivia**, “Impact Evaluation of a Large-Scale Rural Sanitation Project in Indonesia,” *Policy Research Working Papers*, 2013, 6360 (WPS6360), 1–8.
- , **Susan Olivia, and Manisha Shah**, “Initial Conditions Matter: Social Capital and Participatory Development,” 2015.
- Dupas, Pascaline**, “Health Behavior in Developing Countries,” *Annual Review of Economics*, 2011, 3 (1), 425–449.
- Easterly, William and Ross Levine**, “Africa’s growth tragedy: Policies and ethnic divisions,” *Quarterly Journal of Economics*, 1997, 112 (4), 1203–1250.
- Filmer, Deon and Lant Pritchett**, “Estimating Wealth Effects Without Expenditure Data—Or Tears: An Application to Educational Enrollments in States of India,” *Demography*, 2001, 38 (1), 115–132.
- Garn, Joshua V., Gloria D. Sclar, Matthew C. Freeman, Gauthami Penakalapati, Kelly T. Alexander, Patrick Brooks, Eva A. Rehfuss, Sophie Boisson, Kate O. Medlicott, and Thomas F. Clasen**, “The impact of sanitation interventions on latrine coverage and latrine use: A systematic review and meta-analysis,” *International Journal of Hygiene and Environmental Health*, 4 2017, 220 (2), 329–340.
- Guiteras, Raymond, James Levinsohn, and Ahmed Mushfiq Mobarak**, “Sanitation subsidies. Encouraging sanitation investment in the developing world: a cluster-randomized trial.,” *Science*, 2015, 348 (6237), 903–906.
- , – , **and** – , “Encouraging Sanitation Investment in the Developing World: A Cluster-Randomized Trial,” 2017.
- Henderson, J. Vernon, Adam Storeygard, and David N. Weil**, “Measuring economic growth from outer space,” *American Economic Review*, 2012, 102 (2), 994–1028.
- Imbens, Guido W. and Joshua D. Angrist**, “Identification and Estimation of Local Average Treatment Effects,” *Econometrica*, 1994.

- Jensen, Robert**, “The (Perceived) returns to education and the demand for schooling,” *Quarterly Journal of Economics*, 2010, 125 (2), 515–548.
- Kar, Kamal**, “Subsidy or self-respect? Participatory total community sanitation in Bangladesh,” 9 2003.
- **and Robert Chambers**, *Handbook on Community-Led Total Sanitation*, Vol. 44, Institute of Development Studies, 2008.
- Mauro, P.**, “Corruption and Growth,” *The Quarterly Journal of Economics*, 1995, 110 (3), 681–712.
- McKenzie, David**, “Beyond baseline and follow-up: The case for more T in experiments,” *Journal of Development Economics*, 2012, 99 (2), 210–221.
- McKenzie, David J.**, “Measuring inequality with asset indicators,” *Journal of Population Economics*, 2005, 18 (2), 229–260.
- Meager, Rachael**, “Understanding the average impact of microcredit expansions: A Bayesian hierarchical analysis of seven randomized experiments,” *American Economic Journal: Applied Economics*, 2018, *forthcomin.*
- Miller, Grant and A. Mushfiq Mobarak**, “Gender Differences in Preferences, Intra-Household Externalities, and Low Demand for Improved Cookstoves,” Technical Report, National Bureau of Economic Research 2013.
- Musyoki, Samuel Musembi**, “CHAPTER 10: Roles and responsibilities for post-ODF engagement: building an enabling institutional environment for CLTS sustainability,” 7 2016.
- Patil, Sumeet R, Benjamin F Arnold, Alicia L Salvatore, Bertha Briceño, Sandipan Ganguly, John M Colford, and Paul J Gertler**, “The effect of India’s Total Sanitation Campaign on defecation behaviors and child health in rural Madhya Pradesh: a cluster randomized controlled trial,” *PLOS Medicine*, 2014, 11 (8), e1001709.

- Pickering, Amy J, Habiba Djebbari, Carolina Lopez, Massa Coulibaly, and Maria Laura Alzua**, “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*, 2015, 3 (11), e701–e711.
- Prüss-Ustün, Annette, Jamie Bartram, Thomas Clasen, John M Colford, Valerie Curtis, Sophie Bonjour, Oliver Cumming, Alan D Dangour, Lorna Fewtrell, Jennifer De France, Matthew Freeman, Bruce Gordon, Paul Hunter, Richard B Johnston, Kate Medlicott, Daniel Mäusezahl, Colin Mathers, Maria Neira, Meredith Stocks, Jenyfer Wolf, and Sandy Cairncross**, “Burden of diarrhoeal disease from water, sanitation and hygiene in low resource settings: a retrospective analysis of data from 147 countries,” *Trop Med Int Health (in submission)*, 2014, 19 (8), 894–905.
- Romano, Joseph P and Michael Wolf**, “Stepwise multiple testing as formalized data snooping,” *Econometrica*, 2005, 73 (4), 1237–1282.
- Spears, Dean and Sneha Lamba**, “Effects of Early-Life Exposure to Sanitation on Childhood Cognitive Skills: Evidence from India’s Total Sanitation Campaign,” *Journal of Human Resources*, 2016, 51 (2), 298–327.
- United Nations**, “Clean Water and Sanitation: Why It Matters,” *Sustainable Development Goals Briefs*, 2016.
- Water and Sanitation Program**, “Indonesia WSP Global Scaling Up Rural Sanitation Access (TSSM) Impact Evaluation, Baseline and Endline Surveys 2009-2011. Ref. IDN_2009_2011_WSP-IE_v01_M_v01_A_PUF,” 2008.
- and Network for Engineering and Economics Research and Management**, “WSP Global Scaling Up Rural Sanitation Access (TSSM) Impact Evaluation, Madhya Pradesh State, Baseline and Endline Surveys 2009-2011,” 2009.
- WHO/UNICEF**, “Progress on Sanitation and Drinking Water: 2015 Update and MDG Assessment,” Technical Report 2015.

Appendix

A History of CLTS

The concept of Community Led Total Sanitation (CLTS) was first developed by Kamal Kar and the Village Education Resource Centre (VERC) in Bangladesh, in 2000.³⁶ While carrying out an impact assessment of WaterAid’s decade-old water and sanitation strategy in Bangladesh, they noticed that the existing strategies, heavily reliant on subsidies for toilet construction, fell short of their objectives. Though toilet uptake had increased, new construction was mostly concentrated among middle and high income households. Additionally, open defecation remained common practice, even among households with toilets (Kar, 2003). To tackle this problem, they developed the new ‘no subsidy community empowerment approach’ (Kar, 2003). It focused on asking every member of the community to first consider the sanitation situation in the village, and then agree on a collective action plan to change it. Since its first trials in Bangladesh in 2000, CLTS has been rolled out to several Asian and African countries. It has been the sanitation approach of choice for the Nigerian Government’s Strategy for Scaling Up Sanitation and Hygiene since 2007.

WaterAid conducted piloting activities along with UNICEF and local government authorities (LGAs) before scale-up of CLTS in Nigeria commenced in 2008. Within the states in our study, Wateraid Nigeria has been implementing CLTS in selected, mostly rural communities since 2012, and has tailored the intervention to the local context. In the main text, we describe the intervention as it was implemented in Nigeria since late 2014.

³⁶See <http://www.communityledtotalsanitation.org/page/clts-approach> for more details.

B Household attrition

While random attrition would only reduce the precision of our estimates, non-random attrition of households in treatment and control communities could undo the successful randomization demonstrated in Table 3. Therefore, as a precursor to any impact analysis, it is important to check whether our sample is still balanced and whether attrition is correlated with treatment assignment.

First, our study showed remarkably low attrition rates: 2.56% in FU1, 9.23% in FU2 and 12.22% in FU3 (see Table 11). Three years after the baseline survey, we successfully interviewed almost 90% of the households in our initial sample.

Table 11: Attrition rates by survey wave and treatment status

Completed interviews as % of total attempted	Survey wave		
	FU1 (1)	FU2 (2)	FU3 (3)
Whole sample	2.53	8.81	11.58
<i>By treatment status:</i>			
CLTS	2.51	8.61	12.04
Control	2.55	9.01	11.14
<i>By community level wealth:</i>			
Rich communities	3.21	9.85	14.09
Poor communities	1.80	7.69	8.98

Notes: Unconditional attrition rates, as a share of the total attempted interviews in each survey wave.

Secondly, attrition rates were similar across treatment arms. In Table 12 we formally test whether treatment status can predict attrition conditional on baseline characteristics. Columns 1, 4 and 7 show the results of a regression of attrition on the treatment indicator and district fixed effects, for FU1, FU2, and FU3 respectively. We see that assignment to CLTS does not predict attrition, meaning that attrition was balanced across treatment groups. This is also robust to the inclusion of household-level controls (columns 2, 5 and 8) and community wealth group (columns 3, 6 and 9), suggesting that attrition was balanced across treatment arms throughout.

Table 12 additionally tests whether attrition was different across community level wealth groups (see columns 3, 6 and 9). While this is not a threat to the identification of treatment effects, imbalanced attrition in this dimension will question the representativeness of the sample we are

Table 12: Attrition rates by survey wave and treatment status

Survey wave:	FU1			FU2			FU3		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment Status	0.00 (0.90)	-0.00 (0.96)	0.00 (0.90)	0.01 (0.60)	-0.01 (0.49)	-0.01 (0.48)	0.01 (0.22)	0.00 (0.71)	0.01 (0.55)
Poor community			-0.01 (0.06)			0.00 (0.74)			-0.03 (0.02)
Household controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Number of observations	4,647	4,505	4,505	4,647	4,505	4,505	4,722	4,546	4,546

Notes: Estimation results for regressions of attrition, by wave, on treatment status, household and community level characteristics. Errors are clustered at the community level and p -values shown in parenthesis.

working with. We find that once we condition on treatment status, household characteristics and districts, attrition rates were not significantly different across rich and poor communities in FU1 and FU2. In FU3, however, these were higher among rich communities than poor ones, by 3pp. Recall from the previous table that in FU3 the unconditional attrition rate in poor communities was 8.98%.

C Variable definitions

In this appendix we provide additional details on a series of measures we use to construct household and community-level characteristics. These are based on our household surveys and other auxiliary datasets, which were used as part of our impact evaluation and analysis of mechanisms.

C.1 Household characteristics

C.1.1 Relative asset wealth

Our survey asked households to report their income over the past year but response rates for this question were relatively low: 27% of the households interviewed reported no income at all, or refused to answer. Instead, we constructed an index of relative asset wealth following [Filmer and Pritchett \(2001\)](#), using a set of questions regarding the ownership of consumer durables which had higher response rates. This index was constructed using principal components analysis and is often used as a proxy for the long-term asset wealth of a household. Asset wealth indices of this

kind are included in the Demographic and Health Surveys (DHS) conducted by USAID, and have been used in the past as a way of targeting the PROGRESA conditional cash-transfer programme (McKenzie, 2005). Table 13 lists the set of assets elicited in the household survey, and shows their factor loadings which we use to construct the household wealth indicator.

Table 13: Questions used to construct the household relative asset wealth index

Survey question	Factor loading
<i>Ownership of the following durable assets: (1=Yes, 0=No)</i>	
Motorcycle/scooter/tricycle	0.1302
Furniture: chairs	0.1561
Furniture: tables	0.1823
Furniture: beds	0.1075
Refrigerator	0.2998
Washing machine	0.1826
Microwave oven	0.1914
Gas cooker	0.2507
Plasma/flat screen TV	0.2173
Other TV	0.2867
Satellite dish (monthly subscription)	0.2272
Other satellite dish (DSTV, etc)	0.2391
Radio/CD/DVD Player	0.2241
Smart phones	0.1265
Other Telephone / phones	0.0886
Computer	0.2195
Air conditioner	0.1061
Power generator	0.2777
Sewing machine	0.1323
Electric iron	0.3172
Pressure cooker	0.1557
Electric fans	0.3162
Number of households included (N=4,722)	4,622

Notes: Questions were coded equal to one if the household reported to own at least one of each of the items listed in each category. The relative wealth index was then constructed using principal component analysis. Households with missing data for at least one of the categories were excluded.

C.1.2 Expected benefits from toilet ownership

In our exploration of possible mechanisms for CLTS impacts, we use two measures of expected benefits from toilet ownership. The first one is a measure of awareness of the costs of OD. As a community level participatory intervention, CLTS aims to change the social norms surrounding OD and toilet ownership, and therefore this first measure is based on what the households think the community's beliefs are. The questions included, and the factor loadings used in the construction

of the index by principal components analysis are shown in Table 14. The answers ranged from strongly agree (scoring 5 points) down to strongly disagree (scoring 1 point).

Table 14: Questions used to construct the index of awareness of OD costs

Survey question	Factor loading
<i>Most people in your community believe that open defecation is...</i>	
... unhealthy	0.4666
... dangerous	0.5188
... bad for childrens' health	0.4822
... attended a political meeting or rally?	0.5297
Number of households included (N=4,722)	4,139

Notes: Respondents were asked whether they agreed with these statements, and the pre-specified answers ranged from ‘Strongly agree’ to ‘Strongly disagree’ (a five-point scale). For the purposes of constructing the index, these responses were standardized, before conducting the principal component analysis. Households with missing data for at least one of the categories were excluded. This module was included only in the FU3 household survey wave.

Our second measure of expected benefits from toilet ownership is related to social status. Households with no toilets were presented with a 10 rung ladder, read the following statement:

“In the next set of questions, I am going to ask about how you see your status within your own community. Think of this ladder as representing where people stand in your community. At the top of the ladder are people who have the highest standing in your community. At the bottom are people who have the lowest standing in your community.”

Next, households that did not own a toilet in FU3 were asked the following: ‘Where would you place yourself, if you were to own a toilet of any kind?’. Possible answers ranged from 1 (lowest status) to 10 (highest status). The responses to this question were used as our second measure of the expected benefits from toilet ownership.

C.1.3 Expected costs of construction of a toilet

Expected construction costs are a weighted average of the expected costs of four different toilet models. Respondents were presented with four images showing standard toilet models used in the area (a simple pit latrine, a ventilated improved pit latrine, a pour squat latrine and a flush toilet with a ceramic pan), and asked to estimate what it would cost them to construct one of these. All households were asked this question at BL and FU3, regardless of whether they owned a toilet.

The weights used are the shares of existing toilets of each type in the respondent’s district at BL, before CLTS was implemented.

C.2 Community characteristics

In this subsection we provide a detailed description of the measures used to analyze heterogeneous CLTS impacts by community level characteristics as shown in section 4.1. Community denotes the cluster to which a household belongs. On average, clusters are composed of 220 households, of which 20 were randomly selected to be part of our survey, as described in Section 2.2.

C.2.1 Community wealth

Asset wealth at the community level is estimated as the median household’s relative asset wealth score. Our household survey targeted 20 households per community, so we chose the median, instead of the mean, to limit possible distortions due to outliers (i.e. households with extremely high or low asset wealth).

For our discrete measure, communities were ordered by asset wealth, and two sub-samples were created. Rich communities were those with above or equal asset wealth than the median community. Poor communities were those with lower asset wealth than the median community. This resulted in two (approximately) identically sized samples.

Table 7 studies CLTS heterogeneous treatment effects along different measures that proxy for community SES in our sample. Our preferred measure is community asset wealth, which was described above, but we propose a series of alternatives, population density, isolation and a nightlight index to test the robustness of our results to the use of other outcomes variables.

C.2.2 Nightlight index

The first alternative measure is the average night light index recorded in 2013, before the intervention began, over each community’s buffer area (as defined above). The intensity of nighttime lighting has been shown to be correlated with economic growth, and proposed as a tool for inferring growth rates at the sub-national level ([Henderson et al., 2012](#)). We used nighttime lights data made

available by the U.S. National Oceanographic and Atmospheric Administration (NOAA). The observations on which the data is assembled are made by the Operational Linescan System (OLS) flown on the Defense Meteorological Satellite Program (DMSP) satellites.

C.2.3 Isolation

The second alternative is a measure of geographical isolation. Using household GPS coordinates, we calculate the geographical centroid of each community. Next, we use GIS software to calculate the distance, on a straight line, from each community centroid to the nearest administrative capital.³⁷ As in the case for asset wealth, we split the communities in our sample in two groups along the median distance to create to (approximately) identically sized sub-samples.

C.2.4 Population density

Third, we draw a 5km radius circle around the centroid of each community. We compute the number of households living within the 5km radius buffer area around each community centroid. We rely on a census of households living in the area during mid 2014 conducted in preparation for the intervention. Since the area of the buffers is equal for all communities ($\pi \times 5^2 = 79km^2$), this acts as a measure of population density.

C.2.5 Correlations between community-level wealth, nightlight index, isolation, and population density

As a result of these four measures of communities' socio-economic status, communities were classified according to whether they had above or below median asset wealth, nightlight intensity, isolation, and density. A value of 1 was assigned to each of these variables if the community scored below the median value in each of these measures (i.e. below-median asset wealth, below-median night light intensity, below-median population density and below-median distance to district capitals), and zero otherwise. Note that while the first three measures could be interpreted as proxies

³⁷Nigerian states are divided into local government areas (LGAs). Each LGA has its own administrative headquarters, which we use as the relevant administrative capital in this context. Results are unchanged if using the distance to the nearest state capital instead.

of low community SES, the last is an indicator of being relatively near to district capitals, and thus a proxy for *high* SES. Table 15 shows the pairwise correlations for these community level variables.

Table 15: Pairwise correlations of community level SES measures

High SES by measure	Asset wealth	Night lights	Density	Isolation
Asset wealth	1	-	-	-
Night lights	0.2308	1	-	-
Density	0.1417	0.4251	1	-
Isolation	-0.2146	-0.6437	-0.5709	1
Number of communities	247			

Notes: Pairwise correlations of our four alternative measures for community SES. One observation per community, as measured at baseline. Each of these variables takes a value of 1 if the community scores below the median in each measure, i.e. low asset wealth, low night light index, low population density, low isolation (e.g. short distance to the nearest LGA HQ).

Table 15 shows that while these alternative measures are all significantly correlated, these correlations are far from perfect. While isolation and night light intensity are very highly correlated ($\rho=-0.64$), asset wealth and population density are much less so ($\rho=0.14$).

C.2.6 Toilet coverage

Toilet coverage (column 2 of Table 5, is simply the share of households in the community that owned toilets at baseline.

C.2.7 Religious fragmentation

Fragmentation at the community level is measured using a religious diversity index that estimates the probability that two randomly selected households from our sample are of different religions. Indices of this sort, first used by [Mauro \(1995\)](#), are often used to study ethnic diversity, and its impact on economic growth ([Easterly and Levine, 1997](#)) or public goods provision ([Alesina et al., 1999](#)). We choose to study religious instead of ethnic fragmentation as a way of proxying for higher or lower levels of ‘homogeneity’, because our study sample is extremely homogeneous along ethnic lines but has considerable levels of religious diversity.

C.2.8 Trust

In a seminal paper on trust, [Alesina and La Ferrara \(2002\)](#) use a question from the US General Social Survey that asked respondents if they thought that ‘most people can be trusted’. While this measure may be a useful proxy for social trust, we instead rely on a question that asks respondents the following: ‘Generally speaking, would you say that you trust the people in your neighborhood a lot, only a little, or not at all?’. Responses were coded with 2 (a lot), 1 (a little) or 0 (not at all). Our choice is motivated by the fact that we want to uncover the levels of trust *within* the community. We use mean values of household responses within a community as community-level aggregate.

C.2.9 Social capital

The index was constructed by principal components analysis using questions related to participation in community life, social contact and provision of services to the community. [Cameron et al. \(2015a\)](#) use a similar index to measure social capital in their study of CLTS impacts in the province of East Java, Indonesia. Table 16 shows the set of survey questions used to construct the household-level social capital index, together with each item’s factor loadings. Again, we aggregate household responses to a community-level measure by taking the community-specific mean.

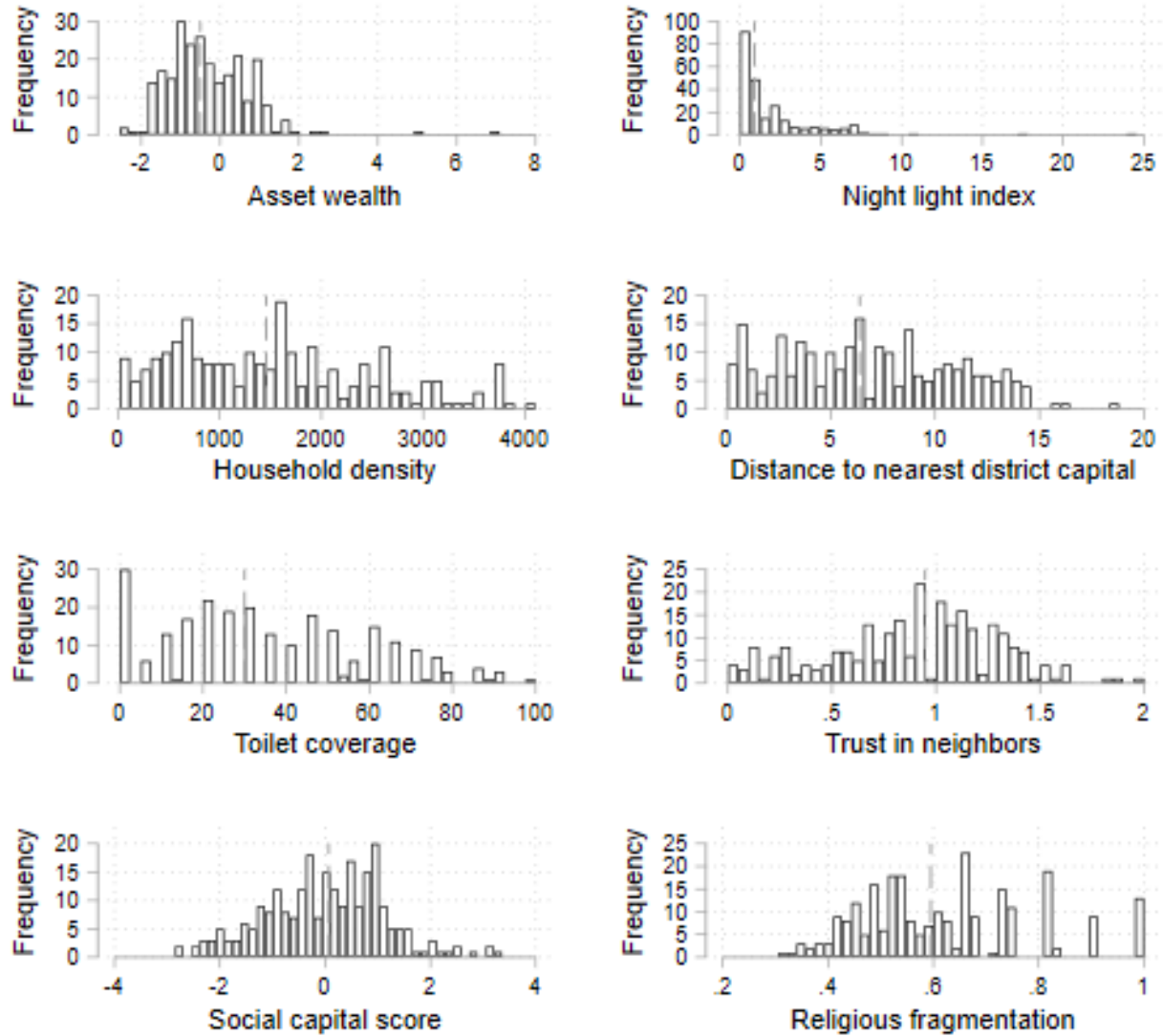
Figure 5 shows the distributions of each community-level characteristic. These characteristics show significant variation, except for the case of night lights, in the top right panel. This skewed distribution in night light intensity across communities is one of the reasons why in our preferred specification, we analyze heterogeneous treatment effects in a discrete manner (high vs low) as opposed to a continuous one. The Figure shows clearly that a specification that interacts treatment status with a continuous measure of night light intensity does not represent the underlying distribution of night light intensity values appropriately.

Table 16: Questions used to construct the household-level social capital index

Survey question	Factor loading
<i>How many times in the past 12 months have you ...</i>	
... donated blood?	0.0573
...worked on a community project?	0.3129
...attended any public meeting in which there was discussion of town or school affairs?	0.2939
...attended a political meeting or rally?	0.2387
...attended any club or organizational meeting (not including meetings for work)?	0.3204
...had friends over to your home?	0.3662
...been in the home of/invited a friend of a different race?	0.3166
...been in the home of/invited someone of a different neighbourhood?	0.3573
...been in the home of/invited someone you consider to be a community leader?	0.3699
...volunteered?	0.3185
...served as an official or served on a committee of a club or community association?	0.2259
Not including weddings and funerals, how often do you attend religious services?	0.0382
Number of households included (N=4,722)	4,227

Notes: These were multiple choice questions in which the (pre-specified) answers ranged from ‘Never did this’ to ‘More than once a week’. For the purposes of constructing the social capital index, these responses were standardized, before conducting the principal component analysis. Households with missing data for at least one of the categories were excluded.

Figure 5: Community characteristics at baseline



Note: Histograms plot the distribution of communities according to the eight characteristics studied in Tables 5 and 7. Dashed lines indicate median values, along which the sample was split to arrive at groups with high and low scores for each trait.

D Additional evidence on heterogeneous impacts on OD and sanitation investments

D.1 Results using continuous community characteristics

Table 17 presents CLTS impact estimates by community type, as in Table 5. Instead of discretizing measures along their median cutoffs, we use continuous measures of the four community characteristics here. The results look very similar. Poor treated communities whose wealth is one standard deviation below the mean of zero display a reduction in OD by 10%, while the two coefficients level out to a close to zero impact for rich CLTS communities. Similar results are found for density and isolation. As with the discrete measures, the interaction term is statistically insignificant once we correct for family wise errors.

In contrast to estimates using the discrete measure, we find no heterogeneous impact when we use the continuous measure of nightlight intensity. Figure 5 shows why. About 90 out of 247 communities in our sample are strongly bunched around a close to zero night light intensity. Hence, a discrete specification seems more suited to capture community variation in this variable.

Table 17: **New** CLTS impacts on OD using four measures of community SES

Community Characteristic (CC) at BL:	Dep. variable: main respondent performs OD				
	Asset wealth (1)	Night lights (2)	Density (3)	Isolation (4)	Toilet coverage (5)
CLTS	-0.04	-0.03	-0.04	-0.04	-0.03
<i>p</i> -value (naive)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
<i>p</i> -value (MHT robust)	[0.07]	[0.07]	[0.07]	[0.07]	[0.07]
CLTS x CC	0.06	0.00	0.04	-0.04	0.03
<i>p</i> -value (naive)	(0.00)	(0.98)	(0.02)	(0.05)	(0.15)
<i>p</i> -value (MHT robust)	[0.01]	[0.98]	[0.09]	[0.22]	[0.53]
No. of communities	246	246	246	246	246
No. of observations	12,697	12,697	12,697	12,697	12,697

Notes: All specifications control for the household and household head characteristics listed in Table 4. Errors are clustered at the community level. Naive (unadjusted) *p*-values shown in parenthesis. In brackets we present *p*-values adjusted by family wise error rate following Romano and Wolf (2005), using 1,000 cluster bootstrap samples and estimated jointly for the regressions presented in Tables 17 and 18.

Table 18 analogously shows estimates based on the continuous measures of fragmentation, social capital and trust. The results are again very similar.

Table 18: **New** CLTS impacts on OD by community characteristics

Community Characteristic (CC) at BL:	Dep. variable: main respondent performs OD		
	Trust (1)	Social capital (2)	Fragmentation (3)
CLTS (γ)	-0.03	-0.03	-0.03
<i>p</i> -value (naive)	(0.05)	(0.05)	(0.05)
<i>p</i> -value (MHT robust)	[0.07]	[0.07]	[0.07]
CLTS x CC	0.00	-0.02	0.01
<i>p</i> -value (naive)	(0.82)	(0.41)	(0.43)
<i>p</i> -value (MHT robust)	[0.98]	[0.89]	[0.89]
No. of communities	246	246	246
No. of observations	12,697	12,697	12,697

Notes: All specifications control for the household and household head characteristics listed in Table 4. Errors are clustered at the community level. Naive (unadjusted) *p*-values shown in parenthesis. In brackets we present *p*-values adjusted by family wise error rate following [Romano and Wolf \(2005\)](#), using 1,000 cluster bootstrap samples and estimated jointly for the regressions presented in Tables 5 and 7.

D.2 Are Community-specific impacts capturing household-level variation in the effectiveness of CLTS?

An additional concern may be that community wealth is an aggregated measure of household wealth. Richer (poorer) communities are indeed on average composed of richer (poorer) households,³⁸ so our estimates may simply proxy household rather than capture community-level heterogeneity. So is CLTS simply (more) effective among poorer households? We test this by splitting communities into rich and poor as before. Then we split households into two groups. More accurately, a household in a rich community is classified as having low (high) asset wealth if its asset wealth score is below (above or equal to) the median of households in rich communities, and likewise for households in poor communities. As a result, our sample is split into four almost equally sized groups along the lines of community and household wealth.

If CLTS impacts are heterogeneous across households rather than communities, we would expect to find stronger CLTS impacts (in effect size and statistical significance) among poor households in either community type. Table 19 shows that while no impacts are detected among the richest households in our sample (first row), a significant 5pp reduction of OD is observed among the poorest (second row). However, we do not reject the hypothesis of no difference between these two

³⁸ 69% of the households living in poor TUs have below-median asset wealth, while 66% of the households living in rich TUs have above-median wealth.

estimates (third row). Thus, treatment heterogeneity seems predominantly driven by community-level wealth, and not household characteristics.

Table 19: CLTS impacts on OD by household characteristics

Household characteristic (at BL):	Dep. variable: main respondent performs OD			
	Low asset wealth (1)	Low ed. HoH (2)	Female HoH (3)	Any children (4)
CLTS x No (γ_r)	-0.03	-0.02	-0.03	-0.03
<i>p</i> -value (naive)	(0.12)	(0.21)	(0.15)	(0.14)
<i>p</i> -value (MHT robust)	[0.26]	[0.26]	[0.26]	[0.26]
CLTS x Yes ($\gamma_r + \gamma_d$)	-0.04	-0.06	-0.04	-0.05
<i>p</i> -value (naive)	(0.08)	(0.04)	(0.04)	(0.03)
<i>p</i> -value (MHT robust)	[0.08]	[0.08]	[0.08]	[0.08]
Difference (γ_d)	-0.01	-0.03	-0.02	-0.02
<i>p</i> -value (naive)	(0.57)	(0.13)	(0.46)	(0.38)
<i>p</i> -value (MHT robust)	[0.78]	[0.43]	[0.78]	[0.78]
Control Mean (BL) - No	0.53	0.59	0.61	0.63
Control Mean (BL) - Yes	0.73	0.70	0.65	0.61
No. of communities	246	246	246	246
No. of observations	12,709	12,868	12,880	12,880

Notes: All specifications control for age and employment status of the household head, as well as household size. Errors are clustered at the community level and adjusted by family wise error rate following [Romano and Wolf \(2005\)](#).

Yet, aggregating to community level wealth may reduce measurement error in household wealth, and thus provide a more robust measure of households' SES, or long term wealth. In this case, household-level characteristics could be the driver behind the higher CLTS impacts observed in poor communities. To address this concern, we use a proxy for household wealth (see column 2 of Table 19): whether the household head had completed primary education at baseline. Again, we see no evidence of heterogeneity in treatment impacts along this dimension, as indicated by the statistically insignificant interaction term shown in the third row. Finally, in columns 3 and 4, we test for two additional household characteristics that have been often cited as correlated with higher levels of investment in sanitation and health technologies in general: having a female head of household and the presence of children in the household.³⁹ No difference in CLTS impacts is

³⁹With respect to CLTS, [Kar \(2003\)](#) poses that women are 'one of the greatest internal forces for mobilisation and promotional activities in the villages'. Women may enjoy larger returns from private sanitation in terms of personal safety and privacy. Evidence from other health enhancing investments suggests the existence gender-specific preferences in certain domains, such as health and children's welfare ([Miller and Mobarak, 2013](#)).

observed along these dimensions either.

D.3 Community-specific impacts on sanitation investments

In this appendix, we re-estimate Table 8, replacing community wealth by the other three measures of socio-economic conditions. We find very similar results: CLTS increases (functioning and general) toilet ownership only in communities with low nightlight intensity (Table 20), low density (Table 21) and high isolation (Table 22), with little change in usage of existing toilets, or shared usage. We find some indication that shared toilet use increases in areas with high nightlight intensity, higher population density and those that are close to the next LGA capital. Yet, the differential impact of CLTS across community groups is never statistically significant.

Table 20: Channels of OD reduction - by night light measures

Outcome =1 if:	Owns toilet (1)	Owns functioning toilet (2)	Usage (if functioning) (3)	Shared use (4)
CLTS x High (γ_r)	-0.01	-0.02	-0.02	0.02
<i>p</i> -value (naive)	(0.70)	(0.45)	(0.31)	(0.02)
<i>p</i> -value (MHT robust)	[0.70]	[0.57]	[0.57]	[0.05]
CLTS x Low ($\gamma_r + \gamma_d$)	0.05	0.08	0.03	0.00
<i>p</i> -value (naive)	0.05	0.01	0.13	0.97
<i>p</i> -value (MHT robust)	[0.18]	[0.01]	[0.25]	[0.97]
Difference (γ_d)	0.06	0.09	0.05	-0.01
<i>p</i> -value (naive)	(0.10)	(0.01)	(0.08)	(0.36)
<i>p</i> -value (MHT robust)	[0.19]	[0.02]	[0.19]	[0.36]
Control Mean (BL) - High	0.45	0.44	0.93	0.03
Control Mean (BL) - Low	0.29	0.28	0.93	0.03
Number of communities	246	246	245	246
Number of observations	12,497	12,497	7,113	12,697

Notes: All specifications control for the household and household head characteristics listed in Table 4. *p*-values are shown in parenthesis. Standard errors are clustered at the community level and are adjusted for family-wise error rate following Romano and Wolf (2005), using 1,000 cluster bootstrap samples.

Table 21: Channels of OD reduction - by household density

Outcome =1 if:	Owns toilet (1)	Owns functioning toilet (2)	Usage (if functioning) (3)	Shared use (4)
CLTS x High (γ_r)	-0.02	-0.01	-0.00	0.01
<i>p</i> -value (naive)	(0.30)	(0.54)	(0.80)	(0.03)
<i>p</i> -value (MHT robust)	[0.56]	[0.78]	[0.80]	[0.09]
CLTS x Low ($\gamma_r + \gamma_d$)	0.07	0.09	0.02	0.00
<i>p</i> -value (naive)	0.01	0.00	0.43	0.81
<i>p</i> -value (MHT robust)	[0.03]	[0.01]	[0.69]	[0.81]
Difference (γ_d)	0.09	0.10	0.02	-0.01
<i>p</i> -value (naive)	(0.01)	(0.01)	(0.43)	(0.63)
<i>p</i> -value (MHT robust)	[0.03]	[0.02]	[0.68]	[0.68]
Control Mean (BL) - High	0.45	0.45	0.93	0.03
Control Mean (BL) - Low	0.26	0.25	0.93	0.03
Number of communities	246	246	245	246
Number of observations	12,497	12,497	7,113	12,697

Notes: All specifications control for the household and household head characteristics listed in Table 4. *p*-values are shown in parenthesis. Standard errors are clustered at the community level and are adjusted for family-wise error rate following [Romano and Wolf \(2005\)](#), using 1,000 cluster bootstrap samples.

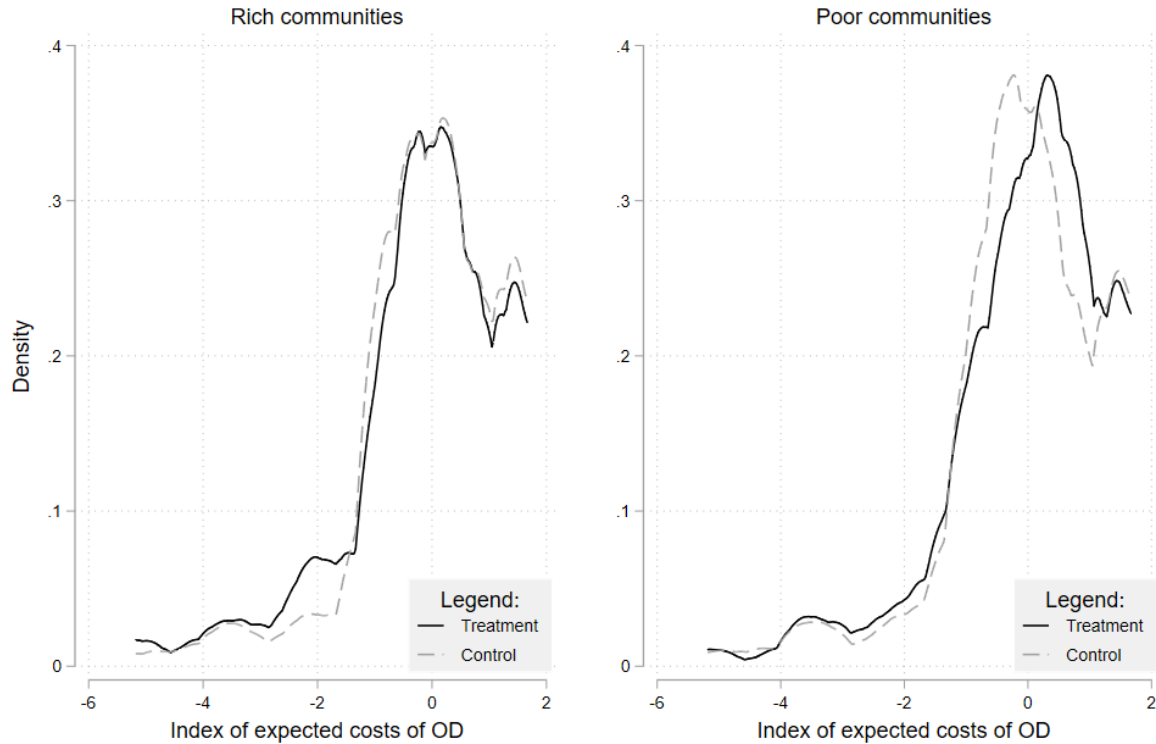
Table 22: Channels of OD reduction - by degree of isolation

Outcome =1 if:	Owns toilet (1)	Owns functioning toilet (2)	Usage (if functioning) (3)	Shared use (4)
CLTS x High (γ_r)	0.05	0.08	0.02	0.00
<i>p</i> -value (naive)	(0.05)	(0.01)	(0.34)	(0.77)
<i>p</i> -value (MHT robust)	[0.12]	[0.01]	[0.59]	[0.77]
CLTS x Low ($\gamma_r + \gamma_d$)	-0.01	-0.01	-0.01	0.01
<i>p</i> -value (naive)	0.79	0.68	0.64	0.08
<i>p</i> -value (MHT robust)	[0.90]	[0.90]	[0.90]	[0.23]
Difference (γ_d)	-0.06	-0.09	-0.03	0.01
<i>p</i> -value (naive)	(0.08)	(0.01)	(0.32)	(0.68)
<i>p</i> -value (MHT robust)	[0.22]	[0.03]	[0.54]	[0.68]
Control Mean (BL) - High	0.30	0.30	0.93	0.03
Control Mean (BL) - Low	0.42	0.42	0.93	0.03
Number of communities	246	246	245	246
Number of observations	12,497	12,497	7,113	12,697

Notes: All specifications control for the household and household head characteristics listed in Table 4. *p*-values are shown in parenthesis. Standard errors are clustered at the community level and are adjusted for family-wise error rate following [Romano and Wolf \(2005\)](#), using 1,000 cluster bootstrap samples.

E Expected benefits and costs of sanitation

Figure 6: Expected benefits from toilet ownership at baseline



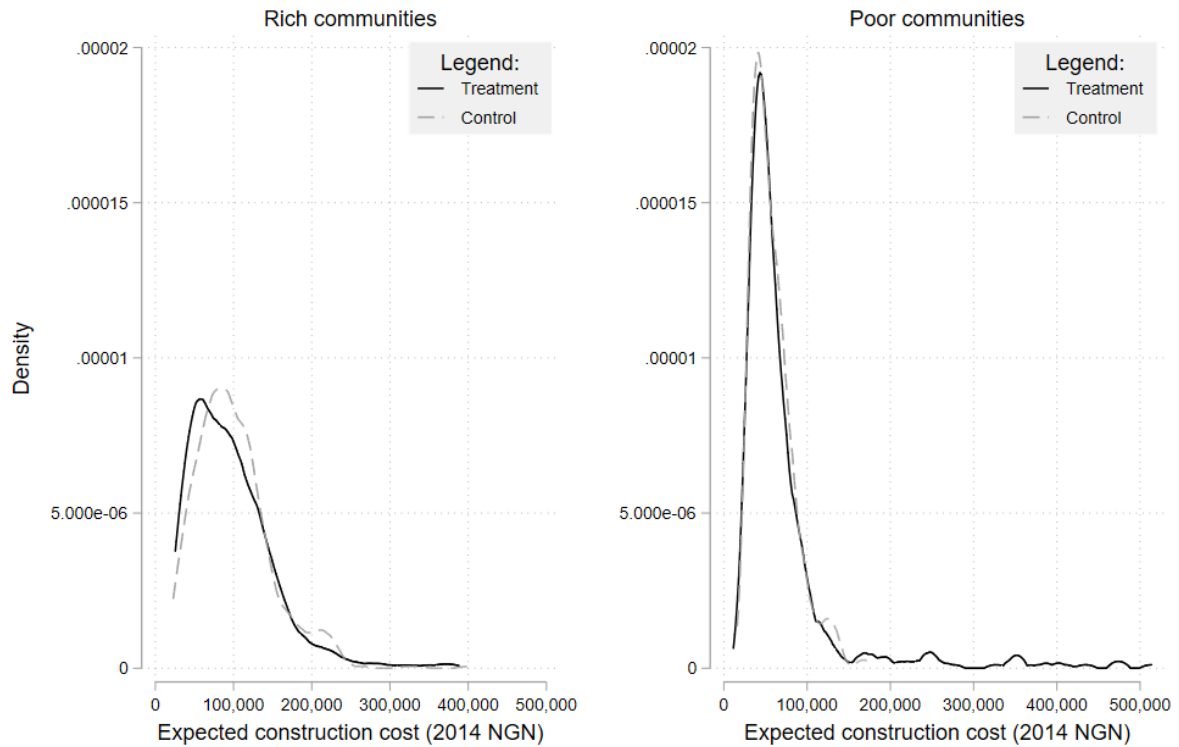
Note: Kernel density estimates of the index of expected benefits of toilet ownership reported by households during the baseline survey. This index was constructed by principal component analysis. At baseline, households were asked how the life of a hypothetical neighbor who did not own a toilet would change if they were invest in one. The following three questions were used to construct the index: ‘The neighbour’s family will be healthier because of the toilet/latrine’, ‘The neighbour’s family will have a higher status in society’ and ‘The neighbour’s family will feel that women in the family will be safer with this toilet/latrine’. Answers, coded as 1 for ‘Strongly disagree’ to 5 ‘Strongly agree’, were then standardized and used to construct the index. The Epanechnikov kernel was used in all cases.

Table 23: CLTS impacts on actual construction costs

LHS: Log(actual cost in NGN)	All toilets built 2015-2017		
	Unimproved (1)	Improved (2)	Any kind (3)
CLTS x Rich (γ_r)	0.18	-0.35	-0.33
<i>p</i> -value (naive)	(0.70)	(0.07)	(0.09)
<i>p</i> -value (MHT robust)	[0.72]	[0.15]	[0.17]
CLTS x Poor community ($\gamma_r + \gamma_d$)	-0.85	-0.11	-0.16
<i>p</i> -value (naive)	(0.03)	(0.59)	(0.40)
<i>p</i> -value (MHT robust)	[0.06]	[0.60]	[0.47]
Difference (γ_d)	-1.02	0.24	0.17
<i>p</i> -value (naive)	(0.16)	(0.37)	(0.51)
<i>p</i> -value (MHT robust)	[0.29]	[0.41]	[0.51]
No. of communities	26	125	129
No. of observations	32	249	282

Notes: Sample includes all toilets constructed since the intervention ended for which we have data on construction costs and year of construction. All columns include district and year of construction fixed effects to account for changes in prices over time. Improved and unimproved toilets classified using WHO/UNICEF criteria. Errors are clustered at the community level and, in square brackets, are adjusted by family wise error rate following [Romano and Wolf \(2005\)](#).

Figure 7: Expected costs of construction at baseline, weighed by toilet model



Note: Kernel density estimates of the expected costs of construction reported by households during the baseline survey. Households were asked what was their expected cost of constructing four different toilet models: a simple pit latrine with no slab, a ventilated improved pit latrine, a squat pour flush toilet, and a ceramic sitting flush toilet. A single expected cost variable was constructed by taking a weighted average of these expected costs. The weights used were the share of existing toilets of each type in each household’s district and community wealth group. Amounts are expressed in Nigerian Naira of the time of the survey (late 2014), when the average exchange rate was 170 NGN/USD. The Epanechnikov kernel was used in all cases.

F Community wealth of the study sample in the Nigerian context

In this section, we compare the wealth of our study households and communities to the distribution of these measures in the Nigerian population. We do this by bench-marking our sample with that of the nationally representative 2013 Nigerian Demographic and Health Survey (DHS).

F.1 Constructing a common asset wealth index across samples

Given the lack of accurate income data in our study, to carry out this comparison we must first construct a measure that can be applied to households in our study sample and in the DHS sample. We construct a new asset wealth index (the ‘DHS index’), using the sub-set of durable asset ownership questions that overlap between our survey and the DHS survey.⁴⁰ To ensure representativeness, we construct the index by performing a principal component analysis of the questions on the DHS sample only, and then create the resulting index for both sets of households by applying the estimated factor loadings on their responses.

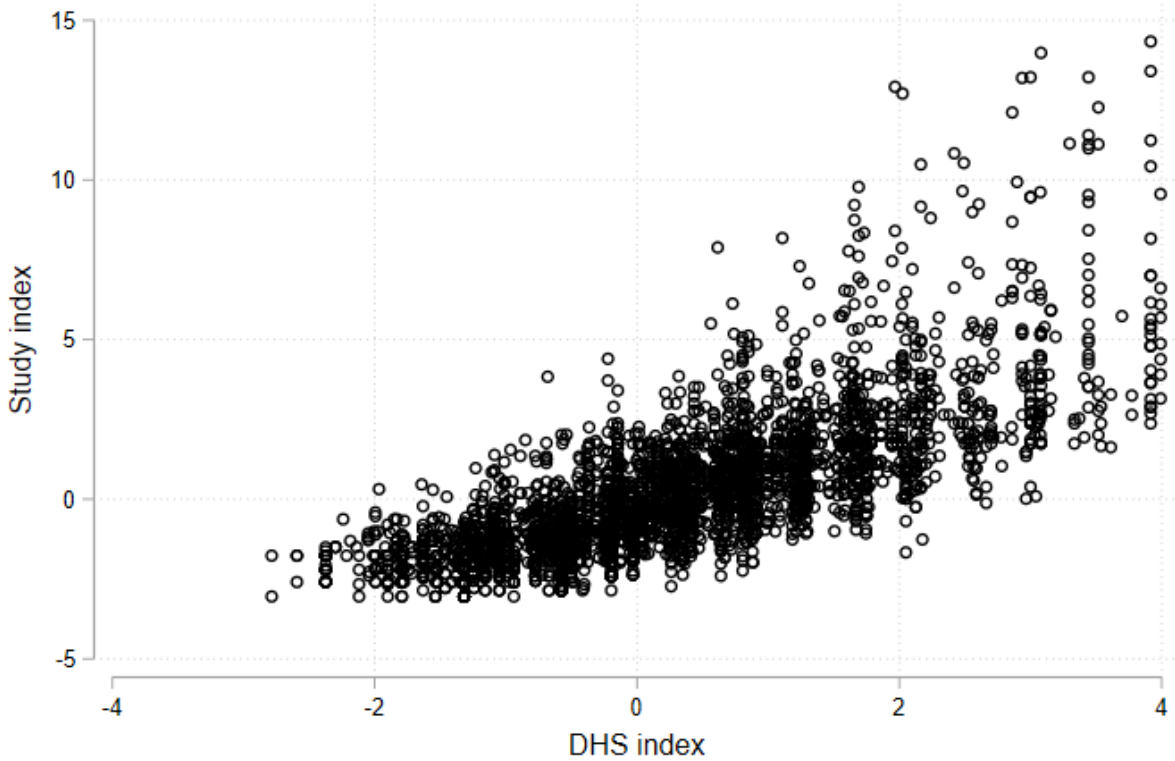
The DHS index and our previously used index (henceforth: the ‘Study index’), based on a more comprehensive list of asset items (see Table 13), are highly positively correlated in our study sample ($\rho=0.77$, significant at the 1% level). Figure 8 plots the new and original household wealth indices for the households in our study sample. The figure shows the positive correlation between the asset wealth indices. The newly created DHS index explains up to 58% of the variation in the more comprehensive Study index.

F.2 Putting our sample into context

Our objective is to compare the wealth of the households and communities in our sample to the broader Nigerian context. In terms of households, Figure 9 plots the distribution of DHS index scores in the DHS sample (black line), in poor communities in our study (blue line) and in rich communities in our study (gray line). The plot reveals important insights. First, from the leftmost

⁴⁰Asset items elicited in both surveys are: bicycles, motorcycles (scooters, tricycles, etc), cars and trucks, refrigerators, radio, TV, bank account, telephone (mobile or fixed), improved water, improved sanitation and livestock (cattle, goats and sheep, pigs, poultry). The inclusion of livestock is motivated by the purpose of capturing farming households. Indeed, the loading factors for these four elements are negative.

Figure 8: Household asset wealth indices from our study data are highly consistent

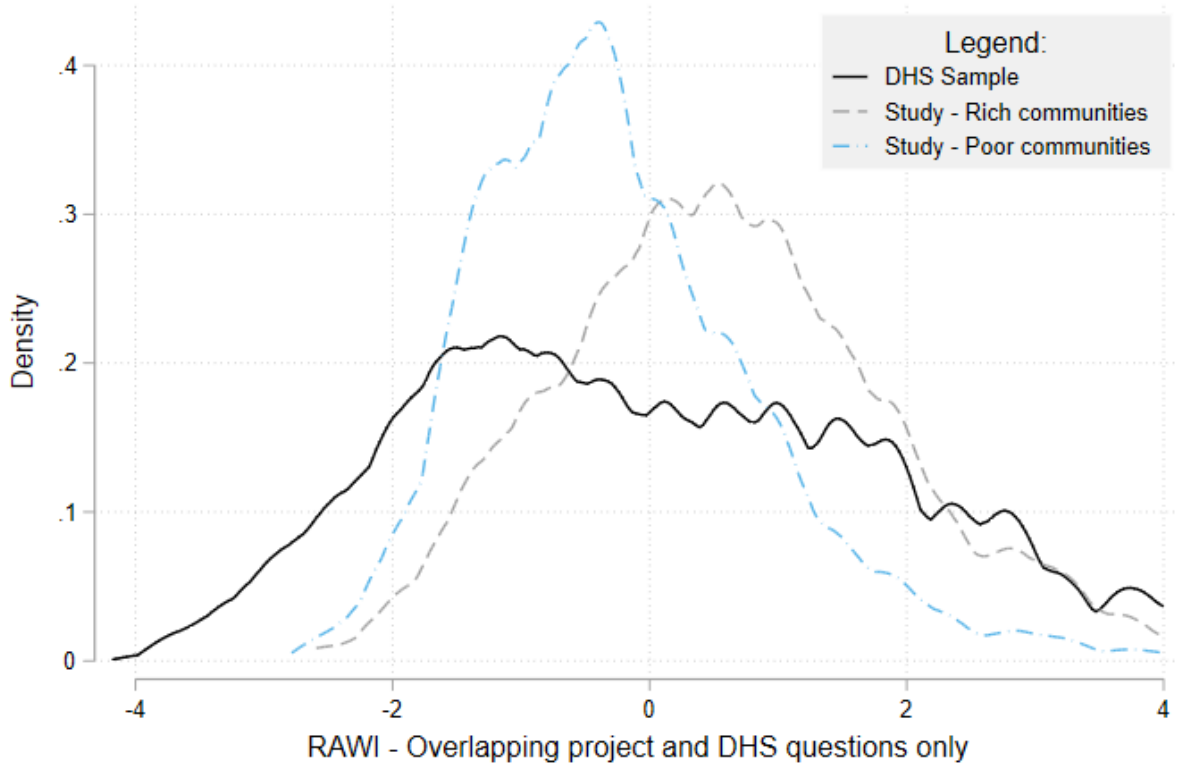


Note: Scatter plot showing the asset wealth scores obtained for each household in our study sample. Each dot represents one household. The y-axis shows the value for the asset wealth index used in our study, built using asset ownership questions from our household survey (see Table 13). The x-axis shows the score obtained in the asset wealth index constructed using only those questions that were included in both our household survey and in the DHS, for the same sample of households. See text for more details.

quarter of the graph we see that our study sample does not include any households with the lowest levels of asset wealth in the country. This is not surprising since Ekiti and Enugu are two states located in the relatively wealthier Centre-South of the country, with the poorest states located in the North. Secondly, the households in our study from poor and rich communities appear to be mostly located just to the left and just to the right of the median Nigerian household, respectively.

Yet, our main focus is on community rather than household wealth. So where do the communities in this study fall within the Nigerian distribution of community level wealth? To answer this question, we construct cluster level asset wealth indicators, as described in section C.2. We

Figure 9: Distribution of the DHS wealth index by sample

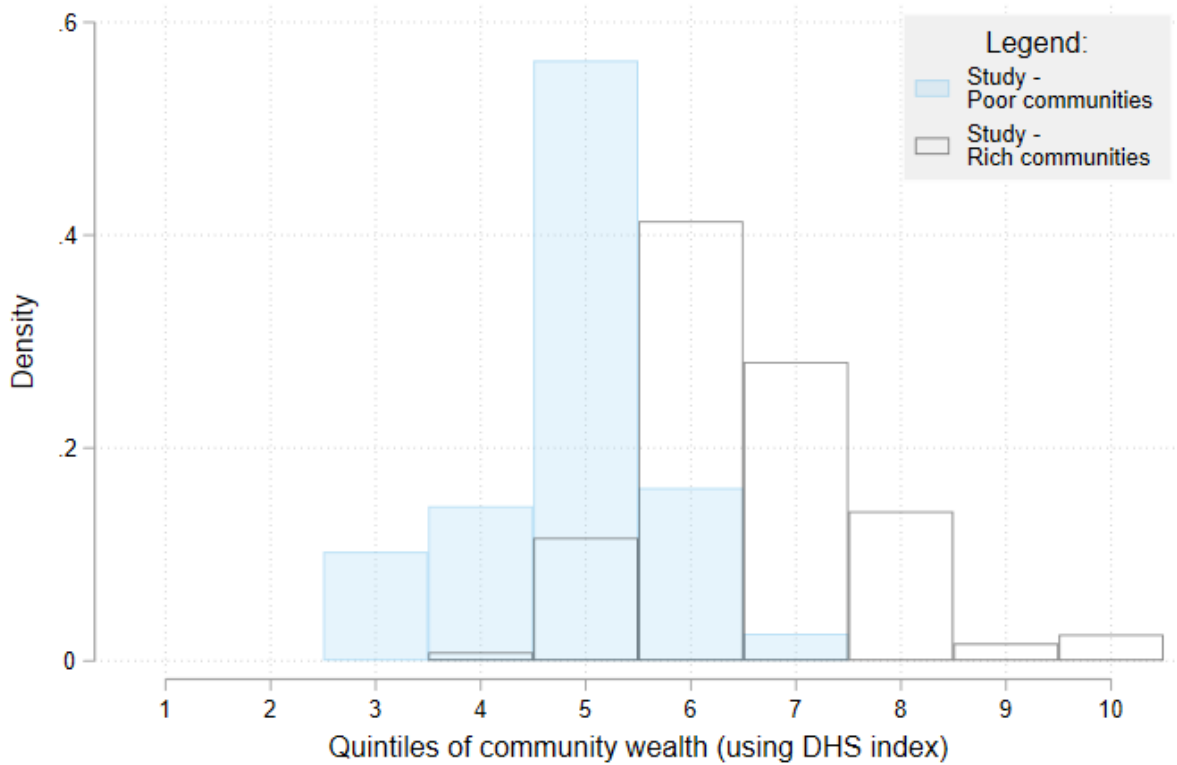


Note: Kernel density plot showing the distribution of DHS index scores among the DHS sample (black), rich (gray) and poor (blue) communities in our study sample. The DHS index is the asset wealth index constructed using only those questions that were included in both our household survey and the DHS.

calculate the median DHS index value within each DHS cluster, and within each study community. Then, using only the DHS sample, we constructed community wealth deciles representative of the Nigerian distribution of cluster level wealth. Finally, using the decile cutoff values from the previous step, we assigned the communities from our survey to the corresponding wealth deciles.

Figure 10 shows the distribution of poor (blue) and rich (white) communities in our sample along the country-wide distribution of (community-level) wealth deciles. Community level wealth in this Figure is determined using the DHS index, to ensure comparability with the ten Nigerian-wide wealth deciles. We retain, however, the classification of rich and poor communities used in the remainder of the paper, which is defined using the more comprehensive Study wealth index, and the

Figure 10: Distribution of study communities in DHS community wealth deciles for Nigeria



Note: Distribution of the communities in our study along community wealth deciles for the whole of Nigeria, estimated using the 2013 DHS. The wealth index used for this comparison was constructed using a set of questions that was included in both our survey and the DHS household questionnaire.

wealth variation within our sample. Thus, some communities that we have classified as poor in our study fall into the higher deciles of the Nigerian community wealth distribution. Overall, however, 80% of the poor communities in our sample are in deciles 3 to 5 of the Nigerian distribution, while 88% of the richest communities in our sample are in the upper deciles of the Nigerian distribution, i.e. deciles 6 to 10. Hence, poor communities in our sample are generally below the median of the national distribution. On the other hand, the richest 50% of communities in our study are richer than the median community in Nigeria.

G Other possible mechanisms

So far, we investigated households' expected costs and benefits from toilet adoption as our main mechanisms of impact. Yet, other possible channels should be considered, in particular if we consider two important traits of the intervention and the framework. First, CLTS puts a strong emphasis on the collective aspect of sanitation. By highlighting the impact of every single member's actions on the health and cleanliness of the rest of the community, negative externalities of OD are emphasised. This attention to the public nature of the sanitation problem is reinforced by a recommendation that community members establish a joint plan to work towards eliminating OD from their community. The CLTS approach is therefore designed as a trigger for collective action solutions to a public good provision problem (i.e. how to keep the community clean).

In consequence, we provide additional tests of two mechanisms: First, we consider whether poorer treated communities become more aware of sanitation externalities. Awareness of externalities is measured through a series of questions in aimed at capturing the social norms of the community. We construct an index using principal component analysis. The questions and their factor loadings are shown in Table 24.

Table 24: Questions used to construct the index of awareness of OD externalities

Survey question	Factor loading
<i>Most people in my community believe that if your neighbour's family uses a toilet...</i>	
...the environment will be cleaner	0.4522
...the water in the community will be cleaner	0.4886
...the food in the community will be cleaner	0.4895
...the children in the community will be healthier	0.4788
...the rest of the community will be cleaner	0.2964
Number of households included (N=4,722)	4,139

Notes: Respondents were asked whether they agreed with these statements, and the pre-specified answers ranged from 'Strongly agree' to 'Strongly disagree' (a five-point scale). For the purposes of constructing the index, these responses were standardized, before conducting the principal component analysis. Households with missing data for at least one of the categories were excluded. This module was included only in the FU3 household survey wave.

Second, we consider whether people in poorer communities are more likely to help each other to construct toilets. We asked households who owned a toilet at FU3 whether they received any help in its construction (financial or otherwise) from family, neighbors, other community members, friends or family from outside the community, members of the same church, or members of an

NGO. 46% of the households declared to have received help from at least one of these informal sources. The results in column 2 in Table 25 are based on this indicator variable. Similarly, we define a second indicator variable that takes the value one if respondents declared to have received help from village officials, traditional village leaders, district officials, or other government officials. Just 2% of the households asked had received some type of official help. We show estimates based on this variable in column 3 of Table 25.

Columns 1 to 3 of Table 25 shows impact estimates of CLTS on each of these potential mechanisms. Results suggest neither an overall increase in construction help or awareness of externalities, nor differential impacts between poor and rich treatment clusters, although impacts might be too small for our study to detect them.

Table 25: Other possible channels of CLTS impacts

Mechanism:	Externalities	Help from others		Sanctions		
	Index (1)	Informal (2)	Official (3)	Shame (4)	Shunning (5)	Fines (6)
CLTS x Rich community (γ_r)	-0.00	0.03	0.01	0.04	0.08	-0.01
<i>p</i> -value (naive)	(0.99)	(0.55)	(0.26)	(0.53)	(0.40)	(0.77)
<i>p</i> -value (MHT robust)	[0.99]	[0.91]	[0.78]	[0.91]	[0.87]	[0.94]
CLTS x Poor community ($\gamma_r + \gamma_d$)	0.04	0.04	-0.01	0.03	-0.13	0.05
<i>p</i> -value (naive)	(0.79)	(0.45)	(0.79)	(0.65)	(0.22)	(0.11)
<i>p</i> -value (MHT robust)	[0.96]	[0.90]	[0.96]	[0.96]	[0.69]	[0.46]
Difference (γ_d)	0.04	0.02	-0.02	-0.01	-0.21	0.06
<i>p</i> -value (naive)	(0.85)	(0.83)	(0.57)	(0.89)	(0.14)	(0.25)
<i>p</i> -value (MHT robust)	[0.99]	[0.99]	[0.96]	[0.99]	[0.56]	[0.71]
No. of communities	245	231	231	245	245	245
No. of observations	4,013	1,946	1,943	4,013	4,013	3,860

Notes: Restricted to observations from FU3 only. The dependent variables used in columns 2 and 3 are responses to questions asked only to households who owned a toilet. All specifications control for the household and household head characteristics listed in Table 4. Errors are clustered at the community level and adjusted by family wise error rate following [Romano and Wolf \(2005\)](#).

A second important aspect for the analysis is that, unlike other health enhancing actions such as hand-washing, toilet ownership and OD practice are public actions. This is especially true in the context of small rural communities where such household actions are usually visible, and where OD is performed pre-established areas within and around the community. Hence, we consider two further mechanisms: social and institutional sanctions to achieve collective change.

We measure changes in the prevalence of *social* sanctions using two statements questions from the FU3 household survey. The first one is a measure of social shaming, where respondents were presented with the following statement: ‘Most people would feel embarrassed to defecate in the open’. They were then asked to rank their agreement with this statement using a 5-point Likert scale, with 5 denoting ‘Strongly agree’ to 1, signifying ‘Strongly disagree’. This question did failed to capture a significant amount of variation: 87% of the respondents said they either agreed or strongly agreed with it. The second alternative measures the degree of shunning that communities practice towards member that defecate in the open. This second statement read: ‘Most people avoid people who do not use a toilet/latrine’, and was coded identically. In this case, 62% of respondents disagreed or disagreed strongly with the statement, while 33% agreed. Finally, to test for the existence of *institutional* sanctions, we asked households whether village leaders punish OD by imposing fines (column 6). Responses were either ‘Yes’ (coded 1) or ‘No’ (coded 0), with 20% of respondents answering affirmatively. The results, shown in columns 4 to 6 in Table 25, do not suggest a large role to these factors.