Sustainable Total Sanitation – Nigeria: Final Research Report

Laura Abramovsky
Britta Augsburg
Francisco Oteiza

Sustainable Total Sanitation – Nigeria: Final Research Report

Laura Abramovsky, Britta Augsburg and Francisco Oteiza;
in collaboration with Indepth Precision Consult, Nigeria and WaterAid

The Institute for Fiscal Studies

Published by

The Institute for Fiscal Studies

7 Ridgmount Street
London WC1E 7AE

Tel: +44 (0) 20-7291 4800
Fax: +44 (0) 20-7323 4780
Email: mailbox@ifs.org.uk
Website: http://www.ifs.org.uk

ISBN 978-1-912805-02-0

Preface

This report is part of the formal research component of WaterAid's 'Sustainable Total Sanitation in Nigeria' (STS Nigeria) project, funded by the Bill & Melinda Gates Foundation. Additional funding from the Economic and Social Research Council (ESRC) Centre for Microeconomic Analysis of Public Policy (CPP) is also gratefully acknowledged. The views expressed in this report are, however, those of the authors and do not necessarily reflect the views of the funders or of other individuals or institutions mentioned here, including the Institute for Fiscal Studies (IFS), which has no corporate view.

The impact evaluation was conducted in collaboration with Indepth Precision Consult (IPC), based in Abuja, Nigeria. All respondents agreed to participate in the surveys, and were assured of the confidentiality of any identifying information gathered. The University College London Ethics Review Board and the National Health Research Ethics Committee of Nigeria have approved this study (Project ID Number 2168-009).

The authors would like to thank Melanie Lührmann and Juan Pablo Rud for their helpful comments and intellectual advice throughout the duration of this impact evaluation, Orazio Attanasio for his intellectual guidance at the design stage of this evaluation, Sam Crossman for his excellent research assistance for this report and Richard Audoly and Tariya Yusuf for equally excellent research assistance at earlier stages of the project. Finally, the authors would like to thank WaterAid UK and WaterAid Nigeria for their cooperation while working on the implementation of their interventions and their help in the development of this impact evaluation. Any errors and all views expressed are those of the authors.

Contents

Abbreviations

ANCOVA – analysis of covariance
ATE – average treatment effect
CBP – concrete block producer
CLTS – Community-Led Total Sanitation
CR1 – cluster randomisation 1
CR2 – cluster randomisation 2
D2D – door-to-door
DHS – Demographic and Health Survey
DiD – difference-in-difference
ELF – ethnolinguistic fragmentation
FWR – familywise error rate
HH – household
IFS – Institute for Fiscal Studies
IR – individual randomisation
ITT – intention-to-treat
LGA – local government area
MICS – Multiple Indicator Cluster Survey
₦ – Nigerian naira
NGO – non-governmental organisation
NL – natural leader
OD – open defecation
ODF – open-defecation-free
OLS – ordinary least squares
PEC – personal entrepreneurial competency
pp – percentage point(s)
RCT – randomised controlled trial
SanMark – Sanitation Marketing
SATO – safe, affordable toilet pan
STS – Sustainable Total Sanitation
TU – triggerable unit
US$ – US dollars
VIP – ventilated improved pit
WASH – water, sanitation and hygiene
WET – water easy toilet

Executive summary

This report presents the final results of the impact evaluation of two of WaterAid Nigeria's main interventions within the programme 'Sustainable Total Sanitation Nigeria – implementation, learning, research, and influence on practice and policy' (STS Nigeria project for short). The impact evaluation was conducted in the states of Ekiti and Enugu. The two interventions under evaluation are Community-Led Total Sanitation (CLTS) and Sanitation Marketing (SanMark). Both interventions aim at increasing the level of improved toilet ownership and its sustained usage, with the final goal of eliminating community-wide open defecation (OD).

This study aims to contribute to a small but growing literature on experimental evaluations of CLTS and SanMark interventions. Our study contributes to the literature by providing evidence from two large-scale roll-outs of said interventions on their impact, studying possible interactions between them, and measuring outcomes over the course of 3 years. The Nigerian context is particularly suited for this, not least due to the large sanitation gap prevailing in the country. The 2014 National Nutrition and Health Survey revealed that just 29% of households in Enugu and 46% of households in Ekiti had access to improved sanitation. This resulted in rates of OD of 51% and 44% respectively, according to the same survey.

The research study was designed as a cluster-randomised controlled trial (RCT) in which communities were assigned to receive either CLTS, a component of SanMark, both of the interventions or neither. This design allowed the research team to compare sanitation outcomes at the household level, and to understand whether the interventions had successfully increased toilet ownership and reduced the prevalence of OD. Individual businesses were also randomly allocated to the SanMark intervention to study the determinants of technology adoption at the business level.

This report presents the final findings of this impact evaluation, which included a series of outputs documenting baseline and intermediate findings. The evaluation spanned 4 years, from 2014 to 2018. CLTS was rolled out during the first half of 2015. SanMark was implemented more gradually, introducing different sub-components consecutively rather than all at the same time, from late 2016 to late 2017. The most recent data presented in this report were collected between October and December 2017.

Our analysis reveals no impacts of CLTS on toilet ownership and OD on average. However, when looking at the poorest half of the studied communities (defined based on an asset wealth index), we find that CLTS had strong and sustained impacts on toilet ownership and OD. In the last survey wave, conducted almost 3 years after the start of CLTS activities, households in poor treated clusters were 10 percentage points (pp) more likely to own a functioning toilet (of any kind), 7pp more likely to own an improved toilet, and 9–10pp less likely to report that the main respondent or any member of the household performs OD, than households in non-treated areas. No effects are detected among richer communities, which results in the lack of impacts over the whole sample. In terms of the margins of CLTS impact, we find that reductions in OD follow closely the improvements observed in toilet construction. In other words, reductions in OD are of a similar magnitude to increases in toilet ownership, and no reductions in OD are detected among households who already had toilets when the intervention began. This finding is likely to be primarily driven by high toilet usage rates, which seem to be specific to Nigeria.

We find no significant impacts of the SanMark intervention on ownership of toilets of any kind. Impacts on ownership of WET models in particular could not be estimated since only 13 households in our sample owned them. This is also true for areas exposed to both CLTS and SanMark activities, which, by the end of 2017, exhibit toilet ownership and OD rates statistically indistinguishable from those observed in similar areas that were not exposed to either intervention. Difficulties in successfully introducing the SanMark intervention and the relatively short implementation period of SanMark before the endline took place are likely to be important drivers of the lack of observed impact at the time of the endline survey.

The key findings are summarised in more detail in the table below.

This study has some important policy implications. First, it highlights the role of community characteristics in mediating the effectiveness of community-level sanitation interventions such as CLTS. Second, and related to the first point, researchers and policymakers in the sanitation sector should strive to better understand the key constraints preventing toilet ownership in each context before choosing suitable policy alternatives, which in turn should ideally be monitored and evaluated. CLTS had strong effects in communities with low average wealth, and no impact at all in the richer half of communities in our sample. In other words, half of the households in our sample experienced no detectable improvements in the outcomes measured in this study. This is likely related to the type of constraints and preferences of households in each type of community, and the community dynamics that CLTS aims to affect, which can vary by level of community wealth. Furthermore, our results suggest that while CLTS might have adjusted the expectations of households about the cost-benefit ratio of a good enough toilet in poor communities, the vast majority of non-toilet-owning households still report being financially constrained, and other policy alternatives may need to be considered to improve outcomes on this aspect. Better targeting of sanitation policies such as CLTS and SanMark should take into account the fact that there are not silver bullets and that these approaches may not be appropriate in all contexts.

Key findings and policy lessons

Reducing OD is intimately tied to increasing toilet ownership in Nigeria	• In Nigeria, almost 100% of the households who own toilets use them. • Reductions in OD are only achieved through increased ownership of functioning toilets. • At the time of study completion, by the end of 2017, 48% of the households in our sample did not own a functioning toilet (55% in poor communities).
CLTS improved sanitation and reduced OD in poor communities	• CLTS increased the ownership of functioning toilets by 10 percentage points (pp) and the ownership of improved toilets by 7pp in the poorest half of study communities. • In these poor communities, CLTS decreased OD by 9–10pp. • These impacts are sustained over time and detectable almost 3 years after CLTS triggering meetings took place. • Households in these poor communities are poor by Nigerian standards. Half of them belong to the poorest 20% of the country. • Evidence suggests that CLTS directed these households towards cheaper, more affordable toilet models, or that it corrected their perceptions of construction costs downwards. • Community-level wealth is a stronger predictor of CLTS impacts than other community characteristics such as baseline toilet ownership, social capital and religious fragmentation, and household-level wealth.
CLTS had no impact in rich communities	• No CLTS impacts are observed among the richest half of the communities, which explains why there is no effect on average over the whole sample. • Households in these richer communities are not all rich, but the communities exhibit a distribution of wealth moderately representative of the wider Nigerian context: that is they are evenly distributed over the five countrywide wealth quintiles. However, even though there are some poor households in these richer communities, the results suggest that CLTS would not lead to significant impacts in these communities. CLTS seems to unfold its potential to a measurable degree only in communities that have average wealth levels similar to the first quintile of the countrywide Nigerian wealth distribution.
Households with no toilets report financial constraints as the main barrier to toilet ownership	• The vast majority of households with no toilet report that the main reason they do not invest in a toilet is financial constraints (toilets being too expensive or not affordable). • This is true in both poor and rich communities within this study.
SanMark WET products are increasingly being sold by businesses, but sales remain low, leading	• One out of six businesses approached to participate in SanMark is selling WET products on a monthly basis. • Total sales were of the order of 400 units at the end of 2017. Less than 1% of households in our sample owned WET products by

to low WET ownership among households

December 2017.
- Monthly sales of WET products peaked 4–5 months after the roll-out of market-level activities and door-to-door (D2D) sales agents.
- SanMark and control businesses perform similarly in terms of revenues, costs and innovation.
- WET products are recognised by potential users as more affordable and attractive.
- Financial constraints are the main limiting factor behind investment in WET products.

Door-to-door sales agents are important

- Sales agents appear to play an important role in facilitating WET sales, being involved in every second WET product sale.
- They are primarily involved in the sale of WET products to households who previously did not own private sanitation facilities of any kind and to households with unimproved toilets that want to upgrade their facilities.

Lessons for policy

- Targeting CLTS interventions based on community characteristics (in particular their relative wealth status) can increase policy impacts.
- CLTS increased toilet ownership among households in poor areas without actually removing financial constraints, but these constraints remain important for households with no toilet.
- SanMark is still a young intervention, and it is difficult to assess its effectiveness at addressing the sanitation gap at this stage. Policymakers should monitor and continue to evaluate the cost-effectiveness of this intervention further before considering a SanMark scale-up.
- Policymakers should consider alternative policies that address financial constraints in both poor and richer areas, such as targeted subsidies or credit lines. These policies could complement the efforts of both CLTS and SanMark by alleviating households' main constraints.
- In poorer areas, a combination of CLTS with targeted subsidies or credit might prove effective.

1. Introduction

This report presents the final results of the impact evaluation of two of WaterAid Nigeria's main interventions within the programme 'Sustainable Total Sanitation Nigeria – implementation, learning, research, and influence on practice and policy' (STS Nigeria project for short). The impact evaluation is part of the STS project's formal research component and focuses on two of the three states in which STS was implemented: Ekiti and Enugu. The two interventions under evaluation are Community-Led Total Sanitation (CLTS) and Sanitation Marketing (SanMark). Both interventions aim at increasing the level of improved toilet ownership and sustained usage in rural communities, with the final goal of eliminating community-wide open defecation (OD). These objectives shape the choice of unit of analysis and main outcomes for the impact evaluation.

The STS project has wider objectives, including to better understand learning processes and the impact of advocacy activities conducted by both WaterAid Nigeria and its implementing partners [both interventions were implemented by a number of local government areas (LGAs) and non-governmental organisations (NGOs)]. It is worth noting that the evaluation is not focused on these aspects.

This is the final report in a series of studies documenting baseline, intermediate and endline findings as part of the impact evaluation.[1] The evaluation spanned 4 years, from 2014 to 2018, although researchers from the Institute for Fiscal Studies (IFS) and staff from WaterAid and its implementing partners conducted exploratory work back in late 2013. Implementers, evaluators and data collectors worked closely together to ensure that data collection and the implementation of the interventions were carefully planned and coordinated in pursuit of maximising the success of both the interventions and the impact evaluation, while minimising impacts on intervention implementation.

CLTS was implemented in the study areas during the first half of 2015. Since then, and up until the end of 2017, follow-up activities were carried out in these areas, which consisted mainly of monitoring progress (described in Chapter 3). SanMark was implemented more gradually, introducing each of its sub-components at a time, from late 2016 to mid 2017. With regards to data collection, baseline data were collected between December 2014 and January 2015. Two intermediate follow-up survey rounds were conducted in December 2015 and March 2017. Endline data were collected between November and December 2017. Researchers from IFS have carried out data collection, data analysis and the compilation of this report in collaboration with Indepth Precision Consult, Nigeria, and with input from WaterAid UK and WaterAid Nigeria.

The analysis focuses on household- and business-level outcomes. First, we present the impacts of CLTS and SanMark on sanitation uptake and OD at the household level. Second, we estimate the rates of sanitation technology adoption of SanMark businesses and of non-selected businesses; and discuss emerging evidence on the take-up of SanMark sanitation products by businesses in our study areas.

1 Two baseline reports documented average household- and firm-level characteristics and carried out balance tests (Abramovsky, Augsburg and Oteiza, 2015 and 2016b). Three intermediate reports documented the progress observed in short follow-up surveys carried out between 2016 and 2017 (Abramovsky, Augsburg and Oteiza, 2016a, 2017a and 2017b).

This report is structured as follows. Chapter 2 provides background information on the context in which STS has been implemented and on the nature of the sanitation problem addressed by the interventions. Chapter 3 describes the two interventions in detail. Chapter 4 discusses the research design for the impact evaluation, Chapter 5 describes the data collection and survey instruments, and Chapter 6 presents balance tests and survey attrition results. Chapters 7 to 9 present the results from our impact evaluation for both CLTS and SanMark. Chapter 10 considers some limitations of our study and Chapter 11 concludes.

2. Background

Health benefits from improved sanitation and the elimination of OD are thought to be substantial (Prüss et al., 2002). The presence of excreta in the environment can affect human health by polluting drinking water, by entering the food chain and by providing breeding sites for flies and insects that spread disease. OD poses an important threat to health, human capital accumulation and economic growth more generally.

Faecal-oral infectious diseases can result in acute symptoms such as diarrhoea, vomiting, fever and sight impairment or be subclinical.[2] Symptomatic diseases could result in death in extreme cases or temporary incapacitation of adults and children that prevents them from working or attending school and also affects caregivers, who may have to give up their regular activities to look after the sickly members of the household.[3] Some subclinical and clinical infections can also reduce the ability of the gut to absorb nutrients and generate malnutrition, stunting and cognitive deficits in children, having a long-term impact on human capital accumulation.[4]

Benefits from better sanitation go beyond health, and include convenience and comfort, privacy and safety, avoidance of sexual harassment and assault for females, less embarrassment (usually an issue when friends and family visit) and dignity. Because of the multiple channels through which pathogens can affect humans, OD generates negative externalities affecting individuals beyond those practising OD. In economic terms, pay-offs from better sanitation practices are both private and public. The benefits from using a latrine for an individual household will depend on the sanitation decisions of other households in their community.[5] This lends credence to the argument that sanitation interventions should be carried out at the community, instead of the household, level.

In the past three decades, most sanitation interventions in rural areas of the developing world have focused on increasing individual demand for sanitation goods by raising awareness, sometimes combined with subsidies for toilet construction.[6] But progress has been slow in countries such as Nigeria, which led policymakers in the sector to consider new approaches.

In this context, WaterAid Nigeria implemented the STS Nigeria project, which is based on a complementary two-pronged intervention design that combines CLTS and SanMark. CLTS was chosen due to its community-led approach, assuming that the main constraints

2 See, for instance, WHO/UNICEF (2008) and Clasen et al. (2014). Diseases include cholera, typhoid, infectious hepatitis, polio, cryptosporidiosis, trachoma and ascariasis as well as environmental enteropathy.

3 The World Health Organisation estimates that poor sanitation costs Nigeria US$3 billion (or US$20 per person or 1.3% of the national GDP) per year (Water and Sanitation Program, 2012).

4 Recent evidence shows that chronic but subclinical environmental enteropathy, which is a disorder caused by faecal contamination which decreases nutrient absorption while increasing the permeability of the small intestine to pathogens, can generate malnutrition, stunting and cognitive deficits without necessarily resulting in diarrhoea episodes (Petri et al., 2008; Mondal et al., 2011).

5 Recent evidence on the impact of sanitation interventions on health in India is consistent with the presence of externalities (Spears, 2012; Hammer and Spears, 2013).

6 Cairncross (2004), for example, describes how, in many developing countries, subsidies for toilet construction have been particularly ineffective. While they increased ownership of toilets, they failed to drive the necessary change in usage behaviour; hence, many constructed toilets were abandoned and not maintained.

affecting the demand side – i.e. toilet construction and usage in rural Nigeria – were related to information and coordination failures, and because CLTS is believed to be more cost-effective for the government or implementing agencies since it does not involve direct subsidies or other financial incentives. CLTS is also the nationally agreed approach for rural sanitation promotion.

SanMark was chosen due to the understanding that frictions existing in the supply of desirable sanitation materials were considerable and could be addressed. Both interventions are described in the next chapter. STS Nigeria aims to improve the effectiveness, efficiency, inclusion and sustainability of total sanitation approaches in rural communities of three states of Nigeria – Ekiti, Enugu and Jigawa. STS Nigeria also aims to contribute to wider regional, national and international good practice, by providing evidence on the effectiveness of CLTS and SanMark as implemented in the study areas within Ekiti and Enugu.

3. The interventions

In this report, we evaluate the impact of two interventions carried out as part of the STS Nigeria project: Community-Led Total Sanitation and SanMark. The motivation behind the combined implementation of these two policies was to tackle the demand for and the supply of desirable, affordable and safe sanitation products. We describe these interventions in detail in Sections 3.1 and 3.2 respectively. We give details of the intended timing of both interventions in Section 3.3 and review the most recent evidence on the effectiveness of CLTS and SanMark in Section 3.4.

3.1 Community-Led Total Sanitation (CLTS)

CLTS aims at changing social norms and increasing the demand for sanitation. It promotes a collective sense of disgust and shame around the practice of open defecation (making evident that 'open shit goes to open mouth') and a change in social norms regarding sanitation practices. Community-level activities conducted under CLTS are typically referred to as 'triggering activities'. CLTS conveys the message that so long as a small number of people in the community continue to open defecate, all community members are at risk of contracting diseases related to unsafe sanitation. This understanding is expected to lead to community members coming up with a coordinated solution to increase the ownership and sustainable usage of toilets in their community.

CLTS was first developed in Bangladesh and has since been introduced in more than 60 countries in South Asia (e.g. India, Nepal and Pakistan), Southeast Asia (e.g. Cambodia, Indonesia and Vietnam) and sub-Saharan Africa, among others, with many governments having adopted it as their national policy to fight open defecation.[7] Nigeria declared CLTS as its main approach for rural sanitation in its 2008 Strategy for Scaling Up Rural Sanitation, and has since then gradually implemented it in rural communities across most states.

The CLTS approach implemented by WaterAid in our study areas can be broken down into four distinct phases, in line with the recommendations from the Handbook on CLTS (Kar and Chambers, 2008):

1. *Planning*. Organising phases 2 and 3 – mobilisation and triggering – is a desk-based activity that can take about four hours.

2. *Mobilisation*. The CLTS triggering team visits communities to be triggered and talks with community leaders. The aim of this visit is to engage the leaders and to agree on a date and a time for triggering activities to take place. The date should be chosen to be suitable for the majority of community members to attend. For large communities, a single date will be set for the triggering of multiple clusters concurrently or consecutively. Sometimes two or three mobilisation visits are needed to set a date for triggering. Each visit takes between one and two hours, excluding travel time.

[7] Source: http://www.communityledtotalsanitation.org/page/clts-approach.

3. *Triggering*. On the agreed date for triggering, a facilitating team of at least four staff members – comprised of LGA water, sanitation and hygiene (WASH) unit staff, NGO staff and sometimes WaterAid Nigeria staff – go to the community at the agreed location, where community members are supposed to be already present.[8] Facilitators run the triggering, with the help of the community leaders. They engage attendees with a series of activities aimed to inform but also involve as many members of the community as possible. These activities might include some or all of the following activities:

    a. A mapping of the community in which each attendee marks their household, and then their regular OD site, if any.

    b. A transect walk in which facilitators then trace the contamination paths of human faeces into water supplies and food in a crude fashion.

    c. Graphic exercises involving faeces and food or water. For example, fresh stool might be used to contaminate a bottle of pure drinking water, to make the point as graphic as possible.

    d. Community-level calculations. Attendees may be encouraged to calculate the total amount of faeces dropped around the village on a weekly, monthly and yearly basis. Calculations of the total medical expenditures incurred by households that are associated with lack of appropriate sanitation may also follow.

    e. Action plan. Attendees are asked to draw a community action plan, which is determined using the contributions of as many members as possible. The plan's objective is for the community to achieve open-defecation-free (ODF) status, and the plan is then posted in a public spot that will make it visible to all community members.

    These different activities are meant to 'trigger' communities into action, and facilitators are instructed to implement activities at their own discretion. The village mapping of OD sites (a) and the drawing of the community action plan (e) are the only activities that all triggering meetings must carry out. It is up to the facilitators to assess which exercises need to be conducted before moving on to the action plan. Table 3.1 shows the significant amount of heterogeneity observed in CLTS triggering events in our sample. For example, the graphic exercise involving faeces and drinking water was conducted in every community for which there is valid information about this in Ekiti, but in none in Enugu (conditional on triggering activities having been conducted). During the meeting, natural leaders

[8] If the team of facilitators sees that not enough people have turned up at the set time, they try to gather more people, with the support of community leaders, by going to people's houses or busy areas. Between 45 minutes and one hour is spent trying to gather more people. If attempts to gather people fail (i.e. the team agrees that an insufficient proportion of the community are present) after an hour, the triggering is cancelled. The team apologises to the people who have turned up, and requests that they mobilise more people in the future. This means that at least four people spend at least four hours (two hours in the community and two hours travelling to the community) in this phase.

(NLs) emerge who cooperate with facilitators in the subsequent steps of the implementation.

4. *Follow-up*. Weekly visits to triggered communities are conducted by WASH LGA staff for up to 6 months. After that, visits occur on a monthly basis. In these monitoring visits, WASH LGA staff assess and monitor progress towards ODF status, movement up the sanitation ladder and use of facilities. Importantly, in contrast to other CLTS experiences in other contexts such as Mali, these visits do not necessarily include any additional community-wide activities or reinforcements of the CLTS message.[9] The purpose of follow-up activities is to monitor and record toilet construction.

Table 3.1. Activities conducted at CLTS triggering events among CLTS communities, by state

	Enugu	Ekiti	Total
Defecation mapping (%)	97% [106 of 109]	90% [52 of 58]	95% [158 of 167]
Transect walk (%)	0% [0 of 109]	3% [2 of 58]	1% [2 of 167]
Faeces & food/water graphic exercises (%)	0% [0 of 109]	41% [24 of 58]	14% [24 of 167]
Faeces & medical expenses calculations (%)	6% [6 of 109]	10% [6 of 58]	7% [12 of 167]
Action plan (%)	100% [109 of 109]	100% [58 of 58]	100% [167 of 167]

Note: Table indicates the share of triggering events in which each of the CLTS activities was conducted (conditional on triggering activities taking place). Monitoring information from CLTS triggering meetings is incomplete, where information on some activities was missing; we assumed that that activity did not take place. Only TUs for which no triggering data is available are excluded from the above table.

In the study areas, CLTS implementation started in January 2015, just after the completion of the baseline survey (data collection exercises are described in more detail in Chapter 5). The last triggering meeting was conducted in June 2015. All targeted communities were approached by the local implementing partners initiating the planning and mobilisation stages just described. However, the triggering meeting could not be conducted in all cases.

In Enugu, implementation was successful in all 109 study communities up to and including the third phase (‘triggering’), i.e. all communities were triggered. However, in Ekiti, a different reality was observed. The team was always able to talk to and engage the community leader as part of the second phase. However, the third phase, which is reliant on the mobilisation of enough community members, failed in 33 out of 91 study communities in Ekiti state.[10] This prevented the delivery of all subsequent activities. In a policy report written as part of this impact evaluation, we identified a series of community

9 For details on the Mali CLTS trial, see Pickering et al. (2015).
10 Additionally, on six occasions, two communities were triggered together.

characteristics that are associated with higher likelihood of successful triggering, as well as those that are associated with lower likelihood of said success (Abramovsky et al., 2016). In line with the reasons stated by programme staff for failed triggering in these 33 study communities, the reasons identified related mostly to the communities' more urban nature.

The fact that not all study communities went through the full set of intervention activities has implications for the interpretation of the findings of this evaluation. We discuss these implications in Section 4.4.

3.2 Sanitation Marketing (SanMark)

SanMark is a relatively new market-based approach that aims to sustainably increase improved sanitation access by addressing demand and supply simultaneously. In its current form, it was developed by WaterAid in Nigeria and inspired by strategies that have been shown to be successful in other parts of the world; see, for example, Mara et al. (2010). The hypothesis behind the design of the SanMark intervention is that a supply of affordable and safe sanitation products and services should be ready and accessible at the time of stimulating demand among households, to maximise the probability that households invest in toilets. This should be combined with a series of marketing activities designed to continue driving demand for a better / more desirable product than those constructed by households themselves.

A key component of SanMark, as implemented by WaterAid in Nigeria,[11] is the design and market introduction of a set of three new toilet models that are safe, affordable and water-efficient. Their development went through a number of stages in order to tailor the model to the local market (Abramovsky, Augsburg and Oteiza, 2015). First, an abridged formative research, or 'deep-dive', exercise was conducted, to identify the sanitation product desired by the targeted market, to gather insights into the drivers of demand for sanitation and current blockages to sanitation adoption, and to achieve a greater understanding of what communities desire and aspire to, when moving from OD to using a private toilet. The final versions of the three water easy toilet (WET) products developed for Nigeria are a direct pit toilet, an offset toilet (which includes an additional height offset) and a dual set that combines both while sharing a single pit. The models can be seen in Figure A.1 in the appendix.

As part of the SanMark programme, participating businesses are granted free access to a metal mould for the casting of WET components, and can purchase the plastic pans, a key component of WET products, from WaterAid (Figure A.2 in the appendix). During the period of our study, WaterAid Nigeria was the sole importer of plastic toilet pans into the country. The cost of each pan for suppliers at the time of writing this report was ₦1,000, around US$3.[12] The metal mould, on the other hand, could be reproduced by other

[11] Henceforth when referring to 'SanMark', we refer to the approach designed and implemented by WaterAid in Nigeria.

[12] This price was set by WaterAid Nigeria staff and was based on a realistic estimate of what the safe, affordable toilet pan (SATO) would cost if available in Nigeria. It used benchmarks from other markets, such as Bangladesh and Cambodia, and was meant to simulate a 'real market price', so as not to distort the future market. In parallel to this, WaterAid facilitated the entry into the market of the manufacturer (Lixil) and, at the time of writing this report, an agreement had been signed with a local manufacturing company. Note that this

businesses, at an estimated cost of US$400. These toilet models were then introduced into the market in two distinct phases. A first phase involved working with businesses operating in the construction sector (e.g. concrete block producers) which supply inputs necessary for the construction of toilets. A second phase consisted of market- and household-level activities, directed at consumers.

SanMark phase 1 – businesses: September 2016 to December 2017

The first phase is composed of the following steps:

- *Entry point*. As is the case for CLTS, WaterAid staff and the LGA WASH coordinator carry out a familiarisation visit to each SanMark community, to introduce the intervention to the community leaders. This introduction serves the purpose of gaining buy-in, discussing the planned intervention strategy within the community leadership and appealing for their support for the programme. The team also explains the randomised controlled trial (RCT) design used in the impact evaluation and its implication in terms of targeting, ensuring that stakeholders understand that not all eligible suppliers in each community will be approached by WaterAid as part of the SanMark intervention. Using a panel survey especially designed and collected for this project and its evaluation, a sample of 135 businesses that regularly built and sold concrete blocks was selected as eligible for the first phase of the SanMark intervention.[13]

- *Approach*. Out of all eligible businesses, a group of randomly selected businesses are approached by the SanMark team and the LGA WASH coordinator and invited to a group pitching session. During this session, they receive a personal entrepreneurial competency (PEC) test, which is their private information (they do not need to reveal their score to the organisers). The SanMark programme is then described in detail and businesses are invited to the subsequent training sessions.

- *Product demonstration session*. A master trainer describes the quantities and qualities of materials needed in the construction of each WET component that makes up the dual set model.[14] The steps required for installation of each of the models are discussed, as well as the advantages that WET models offer over other options in the market. Hands-on training sessions are carried out, during which participating businesses go through these steps themselves and construct the three WET products. Integrity tests are also conducted after a casting period of products as a process of quality assurance of the products. These sessions take place in the premises of one of the participating businesses, chosen so as to minimise travel time for all other participants.

---

pan was only used in the WET product design because it tested so well during the deep-dive market research process.

[13] Detailed information about how eligible businesses have been identified and selected for SanMark can be found in Chapter 4 on Research Design (methodology), in particular in section 4.2 on businesses-level randomisation.

[14] A master trainer is a skilled building expert (e.g. an engineer) with good knowledge of concrete mixtures and consistency, quantity, cost and quality of materials required to produce WET products. They are recruited by WaterAid Nigeria to participate in product demonstration sessions. They are familiar with the design and installation process for the WET product line. The first master trainer witnessed the initial product development process and was supported by an international expert.

- *Business development training*. In a second set of training sessions, business development consultants talk to the businesses about basic business management skills, such as costing, pricing and record keeping. Participating businesses also receive training in using data that can help determine demand in surrounding communities. A particular feature of this context is that the implementing agency had access to ownership rate figures for nearby communities, from our household surveys. Participating businesses are therefore provided with a list of nearby communities and their toilet ownership rates. In addition, both CLTS and, as we will see below, SanMark marketing activities at the community level were hypothesised to spur demand for improved sanitation in targeted communities, so SanMark businesses are 'nudged' towards these communities.

- *Product support*. After participating in the training and being introduced to the WET product, SanMark businesses are offered metal moulds for the production of the WET offset model. These moulds remain the property of WaterAid and can be retrieved if the business is not using them. During this research phase, SanMark businesses can also purchase plastic toilet pans (the plastic component of WET products) from WaterAid Nigeria staff at a cost of ₦1000 (equivalent to US$4), who are the only providers of the pans in the area as local production of the pans is only planned to start in the second half of 2018.

- *Continued support to SanMark businesses*. Training sessions are followed by monthly visits by WaterAid staff, partners and business consultants to provide on-the-spot and basic business problem diagnosis and to recommend solutions as well as validate reported sales.

SanMark phase 2 – marketing: March 2017 to December 2017

The second phase of the intervention is directed at stimulating demand for WET products among households. Door-to-door (D2D) sales agents are recruited from within communities. They work on commission for WET producers and visit households and organise community-level marketing activities to raise awareness about the products and their benefits.[15] The details of these components are as follows:

- *Recruitment and training of D2D agents*. D2D sales agents are recruited from SanMark communities by local government WASH unit staff, following a set of criteria established by WaterAid. Applicants are informed about the commercial nature of the job and that no payment will be received from WaterAid. Their compensation comes from making sales on behalf of participating businesses and is negotiated directly with them. SanMark support to businesses includes guidelines about compensation schemes for D2D agents, but the size of the commissions paid is negotiated privately among the parties (i.e. without support or influence from either WaterAid or WASH unit staff). Applicants are invited to marketing and sales training sessions. The training introduces

[15] With support from 17 Triggers, a marketing design agency, WaterAid developed promotional and marketing materials to advertise the WET products and to test the emerging business model for sanitation marketing in the project communities. Below-the-line marketing was identified as appropriate for the context, focusing on interpersonal communications, direct consumer contact, roadshows and market activities in public spaces. Above-the-line marketing, proven to be less effective in other contexts, was also restricted due to the RCT design.

the product line and the product components, a suggested pricing formula, pitching techniques, the use of promotional materials, and documentation of pitching and outcomes. During training, D2D agents practise both individual and group pitching of the WET product. After the training, D2D agents make household visits at the intensity and frequency of their choice, with the objective of performing WET sales.

- *Community-level marketing activities*. Community-level marketing activities are carried out in public areas, mostly local markets, to raise the profile of WET products and to promote their advantages, highlighting both health benefits and aspirational drivers. Initially, while the D2D agents are being mentored, these events are planned by the LGA WASH unit with guidance from WaterAid and conducted by D2D sales agents and LGA unit staff. Once mentoring is over, businesses and D2D agents are in charge of conducting the activities, and LGA staff may attend. Generally, marketplace events are conducted with the businesses displaying the WET range accompanied by side events such as dramas and dancing to attract interest. Attendees see the WET products and hear their advantages. These activities are an opportunity for D2D agents to meet as many prospective buyers as possible.

The recruitment and training of D2D agents were conducted between February and April 2017. The first community-level marketing events run by these agents occurred between the end of March and the middle of April 2017. D2D agents' sales visits to individual households and marketing events were conducted until December 2017.

It is worth highlighting how SanMark differs from CLTS in terms of the mechanisms used to increase demand for sanitation. While CLTS discussed the dangers involved in the practice of OD, it did not promote the construction of any particular toilet model. SanMark, on the other hand, was entirely focused on advertising a particular set of toilets developed for this purpose, the WET product line. SanMark's strategy of WET promotion relied on aspirational drivers and the health benefits associated with improved sanitation access. This is different from the CLTS approach, which associates feelings of disgust and shame with OD, and pride and social status with toilet ownership. The hypothesis underlying the combined roll-out of CLTS and SanMark is that the interventions may complement each other: by using different ways of increasing demand and also by addressing demand and supply simultaneously.

3.3 Timing of the CLTS and SanMark approaches

The initial plan of the STS project aimed to roll out the SanMark intervention straight after the last CLTS triggering meeting took place. This immediate follow-up of one intervention by another was designed to allow households freshly exposed to the demand-creating influence of CLTS to decide between self-building their toilet and buying an affordable, safe option from the market.

Unfortunately, the implementing team faced several obstacles along the way that delayed SanMark's arrival to the study area. First, problems in the design of the WET product line delayed the start of SanMark's first phase. The last CLTS triggering meeting was held in June 2015 and the first SanMark phase was only rolled out to SanMark businesses outside the piloting LGA in September 2016, 15 months later. At this point, a second set of obstacles appeared. The initial plan was to implement both of SanMark's phases simultaneously, but the implementing team soon realised that the initial recruiting and training of D2D agents had failed. Community health workers and volunteers had been targeted, and thus the pool of candidates had little sales experience. The team realised a new recruitment and training process had to be put in place, this time emphasising the importance of a profit motive and that D2D agents would work on commission for WET-producing businesses. This second round was more successful, and provided important learning points for further fine-tuning of the SanMark programme, as we discuss in Chapter 9. However, this delayed the roll-out of SanMark's second phase to March 2017, 6 months after the start of its first phase.

This resulted in a 'watered-down' version of the SanMark intervention during the evaluation period, and undermined the objective of assessing the combined effect of SanMark with CLTS. It is therefore difficult to test the effectiveness of a comprehensive policy package addressing both the supply and demand sides of the sanitation market using evidence from this case. Instead, in their actual form, CLTS and SanMark were rolled out at significantly separate times and, importantly, SanMark was completed much closer to the endline survey than originally planned. This implies that SanMark had little time to reach its potential, if any, on its own as well as in combination with CLTS, which our experiment was originally designed to assess.

3.4 Past evidence on the impact of CLTS and SanMark

This study aims to contribute to a small but growing literature on experimental evaluations of CLTS and SanMark interventions. The Nigerian context is particularly suited for this, not least due to the large sanitation gap prevailing in the country. At the same time, the interventions under study were implemented in coordination with, and with the active participation of, local authorities. We discuss the literature briefly below.

CLTS

In spite of its popularity as a rural sanitation approach, rigorous evidence on the impacts of CLTS remains scarce. A recent systematic literature review of CLTS's effectiveness found a number of RCT studies evaluating impacts in six different countries (Venkataramanan et al., 2018). Most of these studies support the notion that CLTS can be effective in improving toilet ownership and reducing OD. Often, however, the approach evaluated in these studies is not a pure CLTS approach, but instead includes construction subsidies, training of local masons or other supply-side components. For example, Clasen et al. (2014) found an increase of 28 percentage points (pp) in ownership of functioning toilets, up to 2 years after a rural sanitation campaign was implemented in Odisha, India, which included CLTS-like promotion activities combined with the delivery of toilet construction subsidies. An evaluation of the same programme using households from the state of Madhya Pradesh estimated that 21 months after the intervention was delivered, toilet ownership increased by 19pp and OD fell by 10pp (Patil et al., 2014). In an evaluation of a large-scale programme conducted in East Java, Indonesia, Cameron, Shah and Olivia (2013) found more moderate impacts: increases in toilet ownership of 3pp, and reductions in OD of 6pp, 24 months after the intervention. In their case, CLTS was combined with training local masons and a social marketing campaign. A similar sanitation campaign run in Tanzania was found to have resulted in an increase in toilet ownership (among non-toilet-owners) of 13pp, and an equivalent reduction of OD, 23 months after the campaign was conducted

(Gertler et al., 2016). In all these studies, CLTS-like programmes were accompanied by other supporting policies, making it hard to disentangle the individual impact of CLTS.

A few studies have estimated the impact of CLTS implemented in the absence of other policies. Among these, results vary widely according to the context. Pickering et al. (2015) found that 18 months after a CLTS campaign was conducted in Mali, toilet ownership increased by 30pp and adult OD fell by 23pp. On the other hand, evidence from a CLTS-inspired intervention conducted in Bangladesh, and implemented by the same NGOs who developed CLTS a decade and a half earlier, found no significant impacts on toilet ownership or OD 18 months after the intervention, unless it was paired with subsidy provision (Guiteras, Levinsohn and Mobarak, 2015).

Considering the significant number of countries that adopted CLTS as their main strategy to fight OD, this sparsity of evidence on the effectiveness of CLTS-only approaches, together with the lack of evidence from the Nigerian context, calls for more rigorous evaluation of experimental designs to build a more solid evidence base from which to design sanitation policy. Early qualitative work from Nigerian CLTS experiences carried out by WaterAid suggested that the country presented unique challenges, and was one of the reasons for the implementation of the STS Nigeria project (Robinson, 2009).

SanMark

Much less evidence is available regarding SanMark’s impacts and its possible interactions with CLTS, particularly in Nigeria. A review of peer-reviewed studies on the effectiveness of social marketing campaigns for water and sanitation products found mixed results in terms of behavioural change (Evans et al., 2014). Guiteras, Levinsohn and Mobarak (2015), whose study looked at CLTS and is cited in the previous subsection, also tested the effectiveness of a supply-side policy that hired local masons to provide technical assistance to households who invested in toilets. The assistants helped households identify where to purchase the materials needed, verify the quality of the products offered and maintain their toilets. The authors found that this intervention had no effect on toilet ownership or usage rates, either on its own or when combined with a CLTS-like demand-side policy. Experimental evidence on larger-scale supply-side approaches and strategies that develop local markets to introduce new, cheaper, desirable and higher-quality products and financing mechanisms, such as SanMark, is still to the best of our knowledge unavailable. This report should contribute to filling that knowledge gap.

4. Research design

Our chosen research design to evaluate the CLTS and SanMark interventions, as implemented by WaterAid in Nigeria, is that of a randomised controlled trial. By randomly creating two groups – one that receives the intervention to be evaluated and the other that does not – in many scenarios an RCT provides a suitable comparison group to solve the counterfactual question of how the subjects of interest would have fared without the intervention being implemented.

Randomisation can happen at different levels. Researchers might decide to randomise the delivery of an intervention across individuals or households or across groups of individuals or households (typically referred to as 'clusters', which could be defined by different characteristics, such as geographical boundaries or administrative boundaries). The choice is driven by a number of considerations, but is primarily determined by the nature of the intervention under study and the questions one wants to address. In making these choices, IFS and WaterAid spent significant time carefully assessing the balance between the research needs and the impact on the implementation process. It was important to both sides that the research design allowed for the identification of causal effects, without significantly hindering the intervention implementation process. In a study of this size and complexity, decisions and compromise are necessary at an early stage in the programme design when not all variables and contextual factors are known. Flexibility to adapt the design and respond to issues was critical throughout the process. Given the different nature and implementation approaches of CLTS and the two SanMark phases, we carried out a total of three randomisations:

- *Cluster randomisation 1 (CR1)*. Groups of households, defined using administrative and geographical boundaries, that had no history of CLTS exposure were randomly assigned into either CLTS treatment or CLTS control, the latter not receiving any triggering activities during the course of the study. More details about how groups of households are defined in this case are provided in Section 4.1.

- *Cluster randomisation 2 (CR2)*. Groups of households defined using the same criteria as CR1 were randomly assigned into SanMark treatment and SanMark control. The SanMark treatment group was exposed to village-level marketing activities planned under SanMark, and this is where D2D sales agents were recruited and trained. SanMark control areas saw no SanMark marketing activities and no recruitment of D2D agents taking place during our period of study.

- *Individual randomisation (IR)*. We randomised businesses into SanMark treatment and SanMark control, with SanMark treatment businesses receiving business-level SanMark activities.

Section 4.1 provides more details on the two cluster randomisations, their rationale and execution, and the types of impacts they do, and do not, allow us to estimate. Section 4.2 does the same for the third randomisation, conducted at the business level. The main outcomes of interest we study are defined in Section 4.3. Section 4.4 gives details of our empirical strategy for estimating causal impacts.

4.1 Cluster definition and randomisations

The rationale for randomising CLTS and SanMark (phase 2) activities across groups of households or clusters (rather than individual households), defined by administrative and geographical boundaries, is driven by the nature of the interventions. Improvements in sanitation practices are believed to have community-wide benefits as fewer members of the community are exposed to potentially infectious agents. Benefits that come from one household's changes in sanitation practices might spill over into the rest of the community, generating what economists call a positive externality. At the same time, the constraints addressed by CLTS are inherently thought to be community-level constraints (social norms, coordination failure) and hence presumably better addressed by community-level interventions. SanMark marketing activities are also done at the community level, affecting the whole community, providing further rationale for taking communities as a whole rather than individual households as the unit of randomisation. Therefore, the interventions (CLTS and SanMark) were defined at the cluster level instead of at the individual or household level.

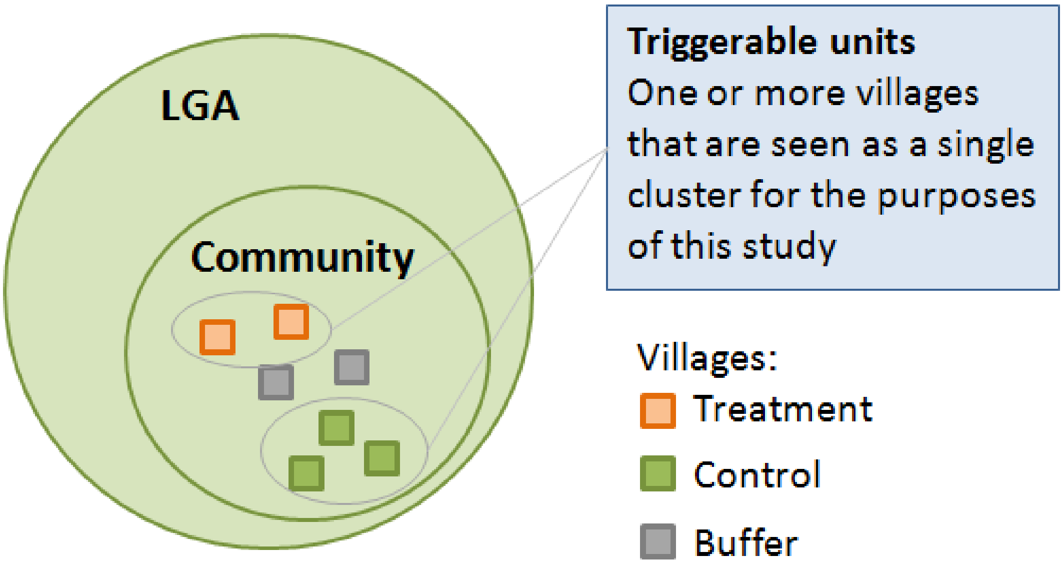
Previous evaluations of similar sanitation interventions have often used communities that are clearly defined administratively and geographically as their clusters for treatment assignment, such as rural villages. When the number of distinctively identified communities within the study areas was high, these studies also assigned treatment and control clusters in a way that ensures a minimum geographical distance between treatment and control clusters to minimise spillovers (e.g. see Pickering et al. (2015)). In the case of this research, we faced two constraints. First, detailed geographical data for our study area were unavailable, so establishing minimum distance buffers between treatment and control areas was not possible. Second, no appropriate administrative units existed: settlements or autonomous communities were too large for simultaneous triggering, while villages or quarters were too small. We chose instead to rely on WaterAid's and its partners' local knowledge of the area, as well as their previous experience in CLTS implementation, to construct geographical/administrative units in which CLTS and SanMark market-level activities could be carried out without neighbouring community members being likely to join or hear about these activities. This aligned with WaterAid's existing practice. To be conservative, buffer zones between these units were also established, to reduce the risk of any treatment spillovers.

With the help of the household census carried out for the purposes of this study, WaterAid Nigeria created 329 geographical/administrative units of similar population size (about 150 households each), at a lower level than that of settlements or autonomous communities.[16] We call these 'triggerable units' (TUs), in reference to the CLTS triggering event, and will use TU and cluster interchangeably in the remainder of this report. TUs are in fact a collection of small villages (or neighbourhoods if they belong to a small town, as in some cases in Ekiti), as shown in Figure 4.1. Each TU is formed of one to two villages on average, with a maximum of ten villages in one case.[17]

[16] See Chapter 5 for a detailed description of the household census and the data collection waves conducted as part of the study.

[17] A detailed description of TUs is available in section 2.2 of Abramovsky, Augsburg and Oteiza (2015).

Figure 4.1. Unit of randomisation: 'triggerable units'



Of these 329 TUs, 84 were located in areas with recent CLTS experience. We conducted two randomisations. The first (CR1) excluded the 84 TUs (or clusters) with recent CLTS experience, which we call pre-CLTS TUs. The remaining 245 TUs were randomly assigned to either CLTS treatment or CLTS control.[18] We stratified this randomisation by LGA in order to reduce the variance of our estimates. This first component of our research design thus allows us to estimate the causal effect of being located in a TU that is assigned to CLTS treatment, as opposed to not experiencing CLTS.

The second randomisation (CR2) randomly assigned TUs to either SanMark treatment or SanMark control. In this case, all 329 TUs in our sample were included, and assigned to either of our two treatment groups. CR2 was stratified at both the LGA and CLTS treatment group levels, ensuring that in each LGA we end up with an even distribution of TUs being exposed to only SanMark, only CLTS and a combination of both.

Importantly, these two cluster randomisations were independent from each other. CR1 was carried out between the census of households and the baseline survey (or wave 1, detailed information about survey waves and data collection more generally can be found in chapter 5). CR2 was carried out after wave 2. Further description of the survey waves used as part of this study can be found in Chapter 5.

In summary, we have six different types of clusters according to their treatment status in our sample:

1. Control (allocated to neither CLTS nor SanMark);
2. CLTS only;
3. SanMark only;
4. CLTS and SanMark;

[18] Figures A.3 and A.4 in the appendix show the approximate location of CLTS treatment and CLTS control clusters in Enugu and in Ekiti, respectively.

5. Pre-CLTS and SanMark;
6. Pre-CLTS only.

During our data collection waves, 20 households were randomly selected from each cluster for interview. The number of clusters included, and the number of households interviewed in wave 4, in each of the six treatment groups described above are summarised in Table 4.1.

Table 4.1. Number of Triggerable Units (or clusters) and households, by treatment status

	CLTS (CR1)			
SanMark (CR2)	Treatment	Control	Pre-CLTS	Total
Treatment	59	58	40	157
	(1,091)	(1,130)	(769)	(2,990)
Control	65	63	44	172
	(1,243)	(1,182)	(868)	(3,293)
Total	124	121	84	329
	(2,334)	(2,312)	(1,637)	(6,283)

Note: Table indicates the number of Triggerable Units (or clusters) assigned to each treatment arm. The number of households interviewed as part of survey wave 4 is indicated in parentheses. For the purposes of studying CLTS treatment impacts, we will compare outcomes between the first and second columns, and omit pre-CLTS households, since they are not reliable counterfactual observations. This means that the total number of TUs (households) in the CLTS analysis will be 245 (4,646).

Identification of causal impacts

The random assignment of household-level clusters to CLTS and SanMark treatment and control will, in theory, allow us to measure the causal impact of the second SanMark phase on household-level outcomes. An important condition that needs to hold is that of no spillovers from treatment to control clusters. In other words, the assignment of one cluster to CLTS, SanMark or control status should not affect the outcomes of neighbouring clusters. We discussed above the measures that were taken when designing TUs to reduce the possibility of such spillovers happening. The TU design was based on the CLTS intervention approach followed by WaterAid in the past, so we are fairly confident that we were able to keep any possible spillover between CLTS evaluation groups to a minimum.

However, we are more concerned with respect to spillovers of the SanMark phase 2 intervention, particularly related to D2D agents. While recruited and trained exclusively in SanMark areas, D2D agents were allowed to operate at their own discretion and could potentially visit households in neighbouring control clusters. This was done to mimic real market conditions and avoid imposing restrictive and unrealistic constraints. We will therefore place particular emphasis on analysing the activities of D2D agents and the number of households from control areas that were exposed to SanMark marketing events in public areas. Doing so will provide the answer to the question of whether our randomly selected SanMark control group of communities will indeed constitute a clean comparison group. In our analysis in Chapters 8 and 9, we will verify whether there is evidence of these treatment spillovers, which could weaken our identification strategy.

4.2 Individual- (business-) level randomisation

Phase 1 of the SanMark intervention is directed specifically at businesses, making them a natural choice of randomisation level. However, as discussed in Abramovsky, Augsburg and Oteiza (2015) and in line with our concerns regarding programme spillovers discussed above, businesses might operate in the same market, and this could interfere with one of the main assumptions underlying RCTs – that the treatment status of one unit does not affect the outcomes of the rest. We therefore conducted an initial business census, to help us understand the market dynamics in our study area. This exercise revealed that businesses operate in deeply connected markets and that they actively compete with each other. This means that the first-best level of randomisation would have been a local or even state-level market. Our resources allowed us to work in just two states and, within them, to only treat 50% of the businesses in our sample. We therefore decided to randomly assign treatment at the business level.

To determine the sample frame for the randomisation of businesses, after the census we conducted two survey waves on businesses that helped us to understand their entry and exit dynamics, as well as their sales profile. With this information at hand, and considering that the WET product line consists of toilet models made of cement and a plastic pan, after wave 2 a subsample of 135 businesses that regularly built and sold concrete blocks was selected as eligible for the first phase of the SanMark intervention. These were referred to as concrete block producers (CBPs) in previous reports. The newly developed SanMark intervention was then piloted in the six identified eligible CBPs located in the LGA Igbo Eze North in early 2016.

The remaining 129 CBPs constitute the basis for our phase 1 SanMark randomisation (IR) and evaluation. They were randomly assigned to either SanMark treatment or control groups, stratifying the assignment at the LGA level. This resulted in two samples of businesses, identical on average along a series of observable characteristics, which compose our experimental study sample. WaterAid staff first approached the businesses in our study group in September 2016, 9 months after the start of piloting activities in Igbo Eze North.

Non-eligible businesses were not considered part of our RCT, but were still part of our business surveys. While they did not sell cement products, they operated in a closely related market, and might react or adapt to the introduction of the WET product line in interesting and innovative ways, which we aimed to capture by surveying them.

This research design allows us to identify the causal effect of the SanMark intervention in terms of product adoption and technology spillovers. We will study the impact of the intervention on the level of adoption of the WET product line among SanMark businesses, which is a relevant result for future interventions involving the introduction of new products to small and medium businesses in developing contexts. SanMark control and non-eligible businesses, which do not have access to the WET product mould and plastic toilet pans from WaterAid, may also decide to invest in their own moulds and obtain plastic pans by other means, as a way to compete with SanMark businesses. This is a form of innovation similar to that of businesses exposed to patent-like innovations by other businesses, and is also relevant for policies such as SanMark. Our design also allows us to study these reactions by control and non-eligible businesses.

A final point worth noting is that the two SanMark-related randomisations were completely independent from each other. In other words, SanMark and control businesses may be located in either SanMark or control areas. The terminology used here is important. SanMark or control *areas* refer only to our second cluster randomisation, CR2: assigning households to SanMark's phase 2 (community-level marketing and introduction of D2D agents). Study businesses could be located anywhere in our study area, implying that SanMark treatment businesses could be located in SanMark control areas and vice versa.

Implications of the research design for the evaluation of SanMark

The individual- (business-) level randomisation described above allows us to study the impacts of assigning the SanMark intervention to a specific group of businesses, and to compare their technology adoption patterns with those of other, non-selected businesses (controls or non-eligible businesses). This, however, does not hold for other business performance indicators, given that control and non-eligible businesses are likely to be indirectly affected by the treatment of SanMark businesses since they are operating in the same market.

To be able to get an unbiased estimator of the causal impact of SanMark on businesses' performance, we need to make strong assumptions which are unlikely to hold, in light of the market characterisation revealed by the businesses census. The businesses in our study sample may sell to customers from other states, but it is likely that they also compete with control businesses for a relatively fixed pool of customers in their own district. Take the case of business revenues, for example. To assign any differences in observed revenues at endline between SanMark treatment and control groups to the causal effect of SanMark, one would need to assume that the observed performance of control businesses represents how SanMark businesses would have done, had they not been selected into SanMark. But the performance of SanMark businesses might be at the expense of control businesses, if they compete for a relatively fixed pool of customers and participation in SanMark allows them to provide a more attractive service. Control businesses, then, will not be a representative counterfactual for the performance of selected businesses had the programme not been in place. The control businesses would actually be doing worse, having lost customers to SanMark businesses. So, in this case, we would be overestimating SanMark's causal impact.

While we will study whether there is differential performance between SanMark and control businesses and will draw some careful conclusions, these should not be interpreted as the causal effect of the SanMark intervention. Our research design takes into account the geographical, budget and operational constraints of the SanMark intervention, and allows us to study causal impacts of technology adoption, but it does not provide us with causal impacts on business performance.

4.3 Outcomes of interest and heterogeneous policy impacts

Households

Sanitation interventions such as CLTS were initially motivated by the fact that subsidy-driven approaches had failed to achieve their objectives in terms of eradicating OD practices in rural communities. This was attributed, in part, to the finding that Bangladeshi households did not use the toilets they built using subsidies (Kar, 2003). Future interventions therefore acknowledged that ending OD would require a strong component of behavioural change, encouraged by community mobilisation.

This initial insight led policymakers to understand that there are at least two possible margins that sanitation interventions such as CLTS could act on. First, they could stimulate the construction of toilets by households who do not own one. Second, they could increase toilet usage (and reduce OD) among the set of households who already own, but do not always use, their private toilets and among the set of households who do not have a private toilet but have access to shared toilet facilities.[19] This motivated our choice of main outcomes at the household level with which to measure the impact of CLTS and SanMark, at both the sanitation uptake and behaviour margins:

Sanitation uptake

- Ownership of a private toilet of any kind. Households were asked whether they owned a private toilet and its type. Two of our survey waves included instructions for enumerators to ask if they could observe toilets, in order to verify their existence, state and type.
- Ownership of a functioning toilet of any kind. Toilets in rural areas may fall into disrepair for several reasons, such as the collapse or filling up of the pit. We asked households who owned toilets whether their toilets were functional, in order to build a more accurate measure of toilet access.
- Ownership of a functioning improved toilet. Improved toilets are identified following the classification used by WHO/UNICEF's Joint Monitoring Programme.[20]

Sanitation behaviour

- All members of the household use the toilet. This is measured only among those households who own functioning toilets. Children below the age of 6 are excluded from this question.
- Open defecation – main respondent. All survey waves asked the main respondent whether they performed OD (always, sometimes, never). Households are recorded as performing OD according to this outcome if the respondent indicates performing OD sometimes or always.
- Open defecation – any member. This was included in three out of four survey waves and is constructed using individual-level sanitation practices of each of the household members aged 6 or above. A household where at least one member reports performing

[19] Other relevant margins could be, for example, the amount of time and money spent on toilet and pit maintenance that households who already own toilets decide to invest.

[20] See https://washdata.org/monitoring/methods.

OD at least sometimes, close to the home or far away, is classified as performing OD under this criterion.

A significant part of the literature studying CLTS effectiveness focuses on the open-defecation-free status at village level as the main outcome. There are several reasons why we think this is not an appropriate metric for the case at hand. First of all, there is no clear consensus about how this variable is defined and measured in different contexts, particularly since ODF may involve administrative certifications with different levels of scrutiny, according to the context (Venkataramanan et al., 2018). Second, our research design was powered to detect impacts among a sample of households, not communities, and the small number of communities in our study area makes it hard for us to detect meaningful impacts at that level. Finally, our study sample is mostly composed of villages that, before the interventions, were very far from reaching ODF status. Indeed, just 65 out of the 610 villages in our sample had toilet ownership rates of 80% or above during survey wave 1 (baseline). In situations such as these, the ODF metric is too restrictive and may miss important improvements in toilet coverage rates if they fail to push villages up to 100% coverage. For these reasons, we choose to stick to household-level metrics, since they more appropriately measure improvements in toilet access and OD behaviour.

Our detailed household-level data allow us to estimate heterogeneous impacts on the outcomes of interest along household characteristics considered relevant in the literature about sanitation interventions. As stated in our research proposal, we are interested in exploring CLTS impacts among different measures of household socio-economic status, such as asset wealth and the level of education of the household head. Long-term household wealth is approximated using a relative asset wealth index as proposed by Filmer and Pritchett (2001). The authors show that combining the answers to a series of questions on household ownership of consumer durables can provide a relatively accurate proxy for a household’s long-term economic status. The questions included in this index are presented in detail in Table A.1 in the appendix. Additionally, we will estimate CLTS impacts according to the gender of the household head and the presence of children below the age of 6, since these groups of households have been singled out in the past as being particularly receptive to community-level sanitation interventions (IFS and WaterAid, 2014).

A participatory intervention such as CLTS is likely to show different results according to the types of communities it is implemented in. In the Handbook on CLTS, the authors suggest a series of community characteristics that implementing teams should look out for, arguing that they are associated with a higher likelihood of success (Kar and Chambers, 2008). Small settlements, located in remote areas and that are culturally and socially homogeneous, are hypothesised to provide more fertile ground for the collective behavioural change that CLTS is after. In a recent policy brief, Abramovsky et al. (2016c) show that CLTS triggering meetings are more likely to fail, and not be carried out at all, in areas with high population density. In a similar vein, Cameron, Olivia and Shah (2015) show that a CLTS trial in Indonesia was only effective among communities with high levels of social capital, while no improvements in terms of toilet construction and OD were observed in areas with lower levels.

In the present case, we will also investigate whether CLTS worked better in communities with different sets of characteristics. In the spirit of the recommendations in the CLTS Handbook and the evidence cited above, we will study whether CLTS had different impacts along levels of wealth, social capital and fragmentation. These three measures are taken at the level of TUs, which contain between one and two communities on average. Our measure of TU wealth will be the median household asset wealth score in each TU. Using this measure will split our sample in half (the richest half and the poorest half). Similarly, our measure of social capital will be the median household score on a social capital index built by principal component analysis.[21] Fragmentation at the TU level is measured using a religious fragmentation index that is standard in the literature on ethnolinguistic fragmentation (ELF).[22] 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.[23] In all three cases, we split the sample along the median, and compare CLTS impacts in the samples with high and low levels of each measure. For brevity, henceforth we refer to TUs with above-median asset wealth as 'rich TUs' and to those with below-median asset wealth as 'poor TUs' and we follow the same high–low distinction for the social capital score and for religious fragmentation. Assessing heterogeneous impacts of CLTS along both household and community level characteristics can point to relevant policy implications, particularly with respect to intervention targeting. Take the example of wealth: community wealth is telling us something beyond individual resource constraints and more about community characteristics that make all households in certain communities react to CLTS in different ways. Assume that individual level wealth mediates the impacts of the intervention but community wealth does not. In such a case, implementers might want to implement CLTS in all types of communities and ensure that the relevant households (poor or rich) attend the sessions. On the other hand, if intervention impacts interact with community wealth rather than individual wealth, targeting specific communities is more relevant that targeting individuals within the communities. This is however not to say that relaxing resource constraints of individual households might not be an effective complementary strategy in both rich and poor communities when aiming for complete sanitation coverage.

Businesses

Our first primary indicator at the business level will be that of technology adoption (i.e. do businesses include the WET product in their line of products offered?). We will further look at monthly sales and costs, as reported by the business and expressed in 2014 US$. Other parameters of business performance that we hypothesise might be affected by the intervention are related to innovation, such as offering new products to customers or introducing new organisational or sales-tracking methods.

4.4 Empirical strategy

Identifying the causal impact of an intervention on a set of outcomes is typically a very challenging exercise. However, given the study design described above, and the commendable cooperation and adherence to this design by WaterAid Nigeria and its

[21] The survey questions used to construct the social capital index are presented in Table A.1 in the appendix.

[22] For an early example of the use of ELF indices, see Mauro (1995).

[23] Our sample consists of two main ethnic groups, Igbos and Yorubas. These two groups account for 94% of the respondents in our survey, and their distribution follows state lines. In Enugu, 99% of our main respondents were Igbo, while in Ekiti, 86% of them declared to be Yoruba.

implementing partners, we can attribute changes in sanitation uptake by households to the CLTS and SanMark interventions.

We observe outcomes on a subset of households, randomly sampled from each study LGA, from a sample frame built by conducting a census of households. Our estimations are therefore designed to be representative at the cluster level.

CLTS household-level impacts

In our main specification, which estimates programme impacts on outcomes of interest, we compare average outcomes between CLTS and control households after the treatment, controlling for outcomes at baseline (wave 1), as follows:

$$(1) \quad y_{i,v,g,t} = \gamma CLTS_v + \theta y_{i,v,g,1} + X_i^{'}\beta + \delta_t + \mu_g + \epsilon_{i,v,g,t}$$

Each observation has a time subscript ($t = \{2,3,4\}$) according to the post-CLTS survey wave they belong to: $t = 2$ for outcomes measured at wave 2, $t = 3$ for wave 3 and $t = 4$ for wave 4. The different survey waves conducted are described in detail in Chapter 5. So, $y_{i,v,g,t}$ is the outcome variable for household i, from cluster v, located in LGA g, measured at post-treatment survey wave $t = \{2,3,4\}$. $CLTS_v$ is an indicator variable equal to 1 if cluster v was assigned to CLTS and 0 otherwise. The coefficient of interest is γ, which denotes the causal impact of the CLTS treatment.

$y_{i,v,g,1}$ is the value of the outcome variable measured before the implementation of the CLTS and SanMark interventions. The inclusion of outcomes measured before the intervention implementation as additional controls (also referred to as an ANCOVA strategy) allows us to estimate causal effects more precisely than using a standard difference-in-difference (DiD) specification (McKenzie, 2012). $X_i^{'}$ is a vector of control variables. Finally, we allow for a common time trend δ_t and for LGA fixed effects μ_g to control for time-invariant differences in the outcome across LGAs, given that our randomisation was stratified at that level.

In our preferred specification, the variables included in $X_i^{'}$ will be age, age squared, gender, employment status and level of education of the household head, household size (the only unbalanced observable characteristic between CLTS treatment and CLTS control groups, as described in Abramovsky, Augsburg and Oteiza (2016a)), an indicator variable equal to 1 if the household's main economic activity is farming and 0 otherwise, and the household's asset wealth score. All controls are measured before CLTS and SanMark implementation began. These variables were selected on the assumption that household wealth levels, female bargaining status and the presence of children might mediate in the decision to invest in, and use, a toilet.

An alternative but related approach to ours is the one used by Cameron, Olivia and Shah (2015). They estimate CLTS impacts using a DiD strategy and run it only on the subsample of households that did not have a toilet at baseline. We believe that our approach is an improvement on this for two reasons. First, our ANCOVA specification (which includes controls for baseline outcomes) has been shown to be better suited for the estimation of treatment effects in the presence of multiple survey waves (McKenzie, 2012) and is the same as the one used in the evaluation of the Bangladesh CLTS experience by Guiteras, Levinsohn and Mobarak (2015). Second, we allow CLTS to have effects on both owners and non-owners of toilets at baseline. Besides persuading non-owners to construct toilets, CLTS informs households that already own a toilet about the importance of its maintenance and usage. More than 70% of the toilets in our sample at baseline were pit latrines of different sorts. Pit latrines require either the construction of new pits or regular emptying (annual or biennial), and will sometimes collapse and become unusable. Therefore, toilet maintenance, function and usage are margins we analyse in our estimations, and we believe they might be affected by (CLTS-driven) changes in behaviour of both owners and non-owners at baseline. For this reason, we include the whole sample of households and control for baseline outcomes, and we also conduct separate impact estimations according to wave 1 toilet ownership status to tease out these two possible channels.

The parameter of interest in specification (1) is γ , which is the average impact of the intervention on our outcomes. Importantly, the impact is averaged over several points in time for which we have outcome measures. As described in Chapter 5, our study is unique in that we visited households on three separate occasions after the intervention started.

The magnitude of CLTS treatment effects may change over time, however. For example, households persuaded by CLTS to build toilets early on after the intervention was implemented could subsequently persuade even more households within their village or network to comply and build more toilets. In this example, CLTS would have a snowball effect over time, and its effects would be larger the longer the time elapsed since the intervention. The opposite scenario is also possible. For example, CLTS could operate in the short term, increasing the level of sanitation uptake but not affecting long-term trends. If this were the case, CLTS impacts would be observed immediately after triggering, but would disappear in subsequent survey waves. This would have important implications regarding the interpretation of the impacts observed under specification (1). The multiple follow-up data collection waves in this study allow us to analyse the dynamic effect of the CLTS intervention in this way.

Therefore, in a second specification, we allow impacts to vary by post-treatment survey wave:

$$(2) \quad y_{i,v,g,t} = \sum_{t=2}^{4} \gamma_t (CLTS_v \times I_t) + \theta y_{i,v,g,1} + X_i^{'}\beta + \delta_t + \mu_g + \epsilon_{i,v,g,t}$$

In specification (2), because we have three post-treatment survey waves, we now have three coefficients of interest, γ_2, γ_3 and γ_4, which represent CLTS impacts as measured at waves 2, 3 and 4, respectively. This is the only difference from specification (1). This specification will enable us to verify whether any impacts found in the first post-treatment survey round persist in the second and third.

While most studies we are aware of measure CLTS impacts in the short term, up to 2 years after the intervention, we measured household outcomes in three post-CLTS surveys, at 8 months, 24 months and 32 months after the initial CLTS meeting. This puts us in a unique position to track the evolution of CLTS impacts over time in a more detailed manner and up to almost 3 years after triggering activities took place.

In order to estimate the impact of CLTS on certain groups of households, we will use slightly modified versions of specifications (1) and (2). For example, when exploring CLTS impacts by household wealth levels, we construct a variable P_i that takes the value 1 if household i has a wealth score below the median for the whole sample and 0 otherwise. We then estimate the following regression model, for the estimation of pooled CLTS impacts:

$$(3) \quad y_{i,v,g,t} = \gamma_H CLTS_v + \gamma_D (CLTS_v \times P_i) + \mu P_i + \theta y_{i,v,g,1} + X_i^{'} \beta + \delta_t + \mu_g + \epsilon_{i,v,g,t}$$

In this case, γ_H estimates the impact of CLTS on households with above-median wealth (i.e. those with $P = 0$) and $\gamma_D$estimates the difference in impacts between the two groups of households. The estimated impact on poor households (i.e. those with $P = 1$) can be denoted γ_L and is recovered from the results of regression (3) by the following expression: $\gamma_L = \gamma_H + \gamma_D$. The key parameter in this equation though is γ_D, which tells us whether there are any significantly different impacts between the two groups being compared. Specification (2) can be modified in a similar way to estimate period-specific heterogeneous treatment effects.

The regressions presented so far will be estimated by ordinary least squares (OLS), and standard errors will be clustered at the level of randomisation, the TU, following standard practice. This report will also acknowledge the fact that we will be conducting multiple estimations. We have six main outcomes of interest, six dimensions of heterogeneity at the household level, and three at the TU level. Running all these tests simultaneously means that individual test statistics are no longer valid, and must be adjusted to correct for the increased likelihood of false rejection of the null hypothesis. We adjust *p*-values for the familywise error rate (FWR) following the method proposed by Romano and Wolf (2005).

Triggerability of villages and interpretation of coefficients

As discussed in Section 3.1 and described in Abramovsky, Augsburg and Oteiza (2016a), not all communities went through all steps of the CLTS intervention. In a separate policy brief, we documented the difficulties in achieving successful CLTS triggering in more densely populated areas, with 20,000 households or more (Abramovsky et al., 2016).

When the third step – the actual triggering activities – is not conducted, this is typically understood as CLTS not having taken place. This has important implications for the interpretation of the estimated coefficients γ, γ_2, γ_3 and γ_4. As we do not know which communities in our control group would not have reached the triggering stage, we cannot include in our sample only the triggered communities. Rather, we conduct our analysis with all communities initially selected to be part of the intervention, irrespective of whether they were eventually triggered or not.

In other words, our impact estimates γ (together with γ_2, γ_3 and γ_4) will measure what is defined as the intention-to-treat (ITT) effect – the effect of CLTS on all areas that were, at project onset, intended to be reached and not just those that were indeed reached.[24] This is a parameter of interest, given that it measures the actual effect we can expect if we were to scale up CLTS to other villages, LGAs or states. It is reasonable to assume – and in accordance with a long history of anecdotal evidence – that full compliance will never be

[24] See, for instance, Duflo, Glennerster and Kremer (2007) for a detailed technical discussion.

achieved and therefore that the ITT effect is a better measure of the expected benefit of the programme than an actual measure of CLTS impact on any given TU is.[25]

SanMark household-level impacts

SanMark was directed at both households and businesses and our empirical strategy depends on this dimension. We describe our estimation strategy for identifying the causal effects of SanMark on households in this subsection and will address our empirical strategy to assess its impacts on businesses later.

The estimation of the impact of SanMark on outcomes at the household level is carried out in a similar way as done for CLTS impacts. A few differences are worth noting, though. First, the timings of the interventions were different. SanMark was intended to be implemented right after CLTS, but problems in the intervention design delayed its roll-out for almost a year, so the first post-SanMark survey is later than the first post-CLTS survey. Second, in the estimation of SanMark impacts, we use our whole sample of households, which includes those assigned to CLTS treatment and control as well as pre-CLTS households – those excluded from the CLTS impact analysis because of their recent history of CLTS activities.[26] This increases our sample size to around 6,000 households.

Similarly to the CLTS case, we use an ANCOVA specification, as follows:

$$(4) \qquad y_{i,v,g,t} = \tau SanMark_v + \theta y_{i,v,g,1} + X_i^{'}\beta + \delta_t + \mu_g + \epsilon_{i,v,g,t}$$

Outcome variables $y_{i,v,g,t}$ are the same as those studied for the CLTS case, and the only difference from specification (1) is the SanMark treatment indicator (in place of the CLTS treatment indicator). The coefficient of interest here is τ.

As discussed in Section 4.1, we will only be able to detect the true causal impact under the condition of no spillover of D2D agents' activities. In case we detect any evidence of spillovers, our point estimates $\hat{\tau}$ will in fact be lower bounds for the causal effects of SanMark (phase 2) on households.

Interactions between SanMark and CLTS

CLTS and SanMark aim to increase the demand for toilet ownership and for improved toilet ownership among households, respectively. Our research design allows us to estimate the independent effect of each of these two interventions, and to study whether these interventions have differential impacts when combined. We study these potential combined effects by estimating SanMark treatment effects separately, within each CLTS treatment group. Recall that the first randomisation (CR1) assigned clusters of households to three groups: pre-CLTS, CLTS treatment and CLTS control. Each of these groups is composed of clusters of households that were later randomly assigned to either SanMark treatment or SanMark control. Hence, we estimate the effect of SanMark on household outcomes separately for each of the three CLTS cluster groups by estimating the following equation:

[25] In Abramovsky, Augsburg and Oteiza (2016a), we run a regression where we define treatment as successful triggering and we instrument this variable with the randomised treatment. Results are very comparable.

[26] Pre-CLTS households were only interviewed during waves 1 and 4, so these are the only waves we include in this analysis.

(5) $$y_{i,v,g,4} = \tau_1(SanMark\ \&\ CLTS) + \tau_2(SanMark\ \&\ control) + \tau_3(SanMark\ \&\ pre\text{-}CLTS) + \tau_4 CLTS_v + \tau_5 Pre\text{-}CLTS_v + \theta y_{i,v,g,1} + X_i^{'}\beta + \mu_g + \epsilon_{i,v,g,4}$$

The main coefficients of interest in this specification are τ_1, τ_2 and τ_3. We will test whether there are any statistical differences in SanMark treatment effects across CLTS samples.

The coefficients τ_4 and τ_5 are controls for average differences in outcomes among CLTS and pre-CLTS groups, while CLTS controls are the omitted category. These terms are included to account for the fact that CLTS might have, on its own, an impact on household outcomes. The only other difference between this specification and specification (1) is that we only surveyed the whole sample of households (which includes pre-CLTS clusters) in waves 1 and 4, and therefore outcomes post-treatment are measured only once, at wave 4. This is 14 months after the start of SanMark's business phase and 8 months after the start of SanMark's second phase.

The underlying hypothesis of this analysis is that SanMark's second phase, aimed at stimulating demand for a particular set of improved sanitation products, will be more effective in areas where CLTS, which aims to promote demand for private sanitation in general, was implemented. At the same time, we will be able to compare SanMark impacts between CLTS and pre-CLTS areas, to check whether the timing of the CLTS intervention is relevant when designing combined interventions. SanMark's first phase began 14 months after the last CLTS triggering meeting in our CLTS areas (late June 2015) and 25 months after the last CLTS triggering meeting in pre-CLTS areas (August 2014).

As with pure SanMark phase 2 impacts discussed above, the SanMark and CLTS interaction effects will also only be pure causal impacts under the assumption of no spillovers from treatment to control. In case of any spillovers, our point estimates $\widehat{\tau_1}$, $\widehat{\tau_2}$ and $\widehat{\tau_3}$ are in fact lower bounds for the causal effects of SanMark (phase 2) on households.

SanMark business-level impacts

As with any intervention involving business training or development, adoption of the skills or products introduced is imperfect. In other words, not all the businesses invited to participate in the SanMark programme will take part, and even fewer will eventually adopt and successfully market and sell WET products themselves. So the first step in our analysis will be to estimate the causal effect of being selected into the SanMark programme on the degree of adoption of WET products. We will estimate the share of SanMark businesses that chose to incorporate WET products into their product line, and study their WET sales performance. At the same time, we will estimate the degree of adoption of WET products by control and non-eligible businesses, which might have decided to adopt as a spillover effect of our intervention. Our detailed information on business characteristics allows us to then carry out an analysis of the characteristics that are more highly correlated with WET adoption and sales.

Direct (naïve) effects of SanMark on business outcomes

As mentioned in Section 4.2, we will not be estimating the causal effect of SanMark on other business performance outcomes. This is because of the likely presence of spillovers coming from the fact that SanMark and control businesses operate in the same market. Instead, we will estimate the direct, or naïve, impact of being selected into the SanMark programme on monthly sales, monthly costs and the quantity of concrete blocks sold over the past month. As previously mentioned, the intervention might also have effects on the innovative behaviour of businesses, so we will study these as well. To estimate these differences, while controlling for other business characteristics that might be correlated with our outcomes, we use the following specification:

$$(6) \quad y_{i,g,t} = \nu SanMark_i + \tau SanMark_i \times Post_t + \theta y_{i,g,1} + X_i^{'}\beta + \delta_t + \mu_g + \epsilon_{i,g,t}$$

$SanMark_i$ is an indicator variable equal to 1 if business i was selected into the SanMark programme and 0 otherwise. $Post_i$ is an indicator equal to 1 if the period of analysis t is after the introduction of SanMark (waves 3 and 4) and 0 otherwise (wave 2). Here, τ is our coefficient of interest: the difference in outcomes between SanMark and control businesses. This first specification will pool all post-SanMark observations together (i.e. waves 3 and 4) and include controls for levels of each outcome measured at wave 1.

In a second stage, we will run a similar analysis, but allow differences to vary across survey waves. This will let us study, for example, whether SanMark businesses performed any differently from controls by March 2017, before the market-level activities had taken place, and whether these activities made any additional difference. We do this by using the following specification:

$$(7) \quad y_{i,g,t} = \sum_{t=3}^{4} \tau_t SanMark_i + \theta y_{i,g,1} + X_i^{'}\beta + \delta_t + \mu_g + \epsilon_{i,g,t}$$

Here, τ_t are our coefficients of interest and, as the subscript indicates, they will measure the difference in outcomes between SanMark and control businesses during each post-treatment survey wave.

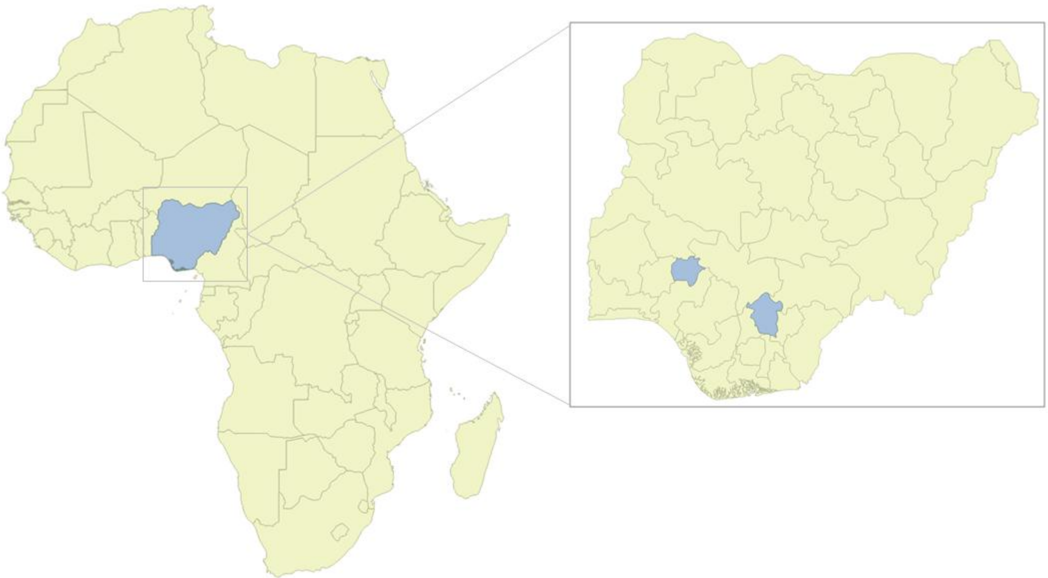
5. Data collection

5.1 Survey instruments and data collection waves

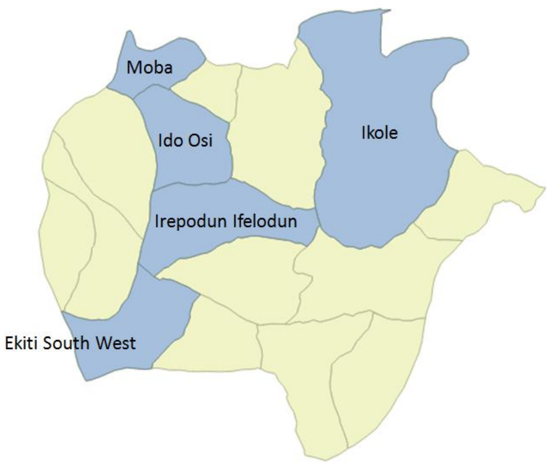
This report analyses the impact of CLTS and SanMark – as well as their interactions – in rural communities of the Nigerian states of Ekiti and Enugu. Ekiti and Enugu are two of the smallest states in Nigeria, located in the South and South West of the country, as indicated in Figure 5.1. The STS Nigeria project was also carried out in the northern state of Jigawa, but this was excluded from the research for practical and budgetary considerations, driven primarily by the security challenges faced in the North of the country.

Figure 5.1. Our study area: the Nigerian states of Ekiti and Enugu

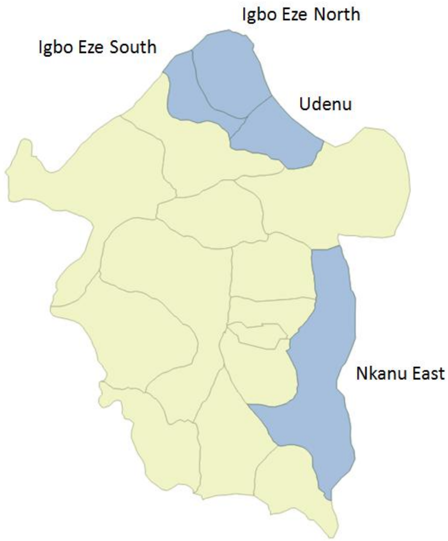
(a) Ekiti (left) and Enugu (right) highlighted in blue



(b) Ekiti LGAs in our study (blue)



(c) Enugu LGAs in our study (blue)



Source: IFS.

Ekiti and Enugu states are further divided into smaller administrative units called local government areas. The present study took place in five out of Ekiti's 16 LGAs and in four out of Enugu's 17 LGAs. These are mapped in Figures 5.1(b) and 5.1(c).

This impact evaluation required the collection of significant amounts of primary data, given the absence of any suitable secondary data sources. We work with four primary data sets to answer the research questions: (i) a survey of communities; (ii) a household panel; (iii) a panel of businesses; and (iv) a survey of SanMark sales agents. The remainder of this chapter describes the data sets and the timeline of the data collection in further detail.

Between the months of August and October 2014, we carried out a **census** of households and businesses located in our study LGAs. The need for this data collection exercise was driven by the lack of (publicly available) updated, geographically representative information for our study area. Close to 50,000 households and 150 businesses were mapped and surveyed in this first exercise, which we used to construct our sample frame.

Once the sample frame was drawn, a first set of surveys was conducted in late 2014 (**wave 1**). This data collection wave included detailed household, business and community surveys. The timing of this wave was such that it took place right before the start of the CLTS interventions and serves as a baseline for the CLTS impact analysis (therefore also referred to as our CLTS **baseline survey**). CLTS implementation took place in both our study states between January and June 2015.

In late 2015, a short survey of both households and businesses (**wave 2**) was conducted. The purpose of this wave was to measure the short-term effects of the CLTS intervention.[27] CLTS triggering was rolled out over the course of 6 months, so this survey wave measured CLTS impacts 8 months after the intervention, on average, with a minimum of 5 and a maximum of 11 months having elapsed since then.

SanMark activities – particularly phase 1 of SanMark, which focused on businesses – began in September 2016. As with CLTS, we fielded a data collection round (**wave 3**) with the purpose of analysing the progress and impacts of this SanMark phase. This third data collection wave took place in March and April 2017.[28] This wave interviewed households 24 months after the CLTS intervention, on average, with a minimum of 21 and a maximum of 28 months. It also included a third survey of businesses.

After this data collection round, SanMark phase 2 was implemented, focusing on the selection of D2D agents and marketing activities directed at consumers. During late 2017, the last data collection round (**wave 4**, also referred to as our **endline survey**) was conducted, surveying households, businesses and sales agents. The main purpose of this data collection was to capture the effects of CLTS and SanMark, as well as their interactions. This happened on average 8 months after the start of the second phase of SanMark activities, and on average 32 months after CLTS triggering, with a minimum of 28 and a maximum of 36 months.

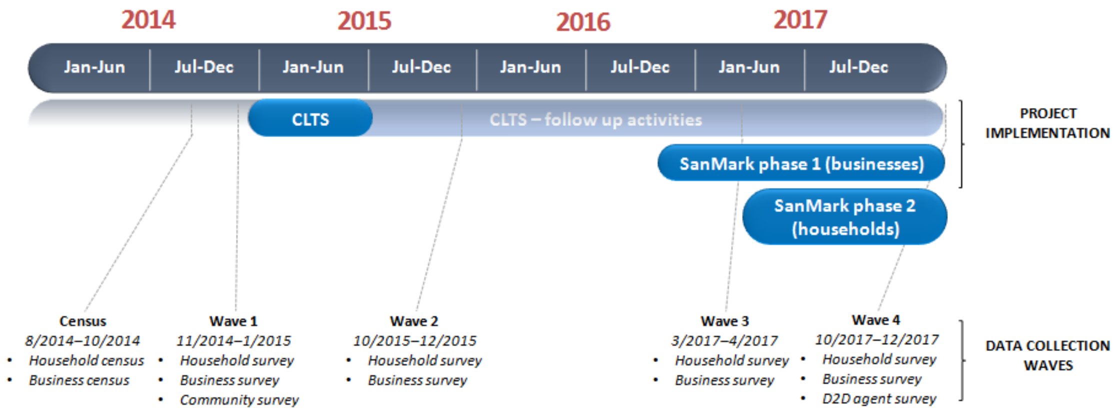
A schematic representation of the data collection timeline can be seen in Figure 5.2.

---

[27] This data collection wave was referred to in our previous reports as our first rapid assessment. For clarity purposes, here we will refer to it as wave 2.

[28] Referred to as the second rapid assessment in previous reports.

Figure 5.2. Project implementation and data collection timeline



Source: IFS.

Table 5.1 presents the sample sizes and the number of refused and failed interviews by survey instrument and wave. We now describe each survey instrument in further detail. We will discuss the levels of attrition encountered in the fourth, and final, survey wave (the endline survey) in the next chapter.

Table 5.1. Interviews conducted by survey wave

	Census	Survey wave 1	Survey wave 2	Survey wave 3	Survey wave 4
	Aug 2014 to Oct 2014	*Nov 2014 to Jan 2015*	*Oct 2015 to Dec 2015*	*Mar 2017 to Apr 2017*	*Oct 2017 to Dec 2017*
Households[a]					
Target	50,333	6,386	4,649	4,649	6,386
Completed	50,273	6,319	4,530	4,218	5,594
% of target	*99.88%*	*98.95%*	*97.44%*	*90.73%*	*87.60%*
Refused / Not available / Moved	60	67	119	431	792
% of target	*0.12%*	*1.05%*	*2.56%*	*9.27%*	*12.40%*
Businesses[b]					
Target	155	155	197	197	197
Completed	151	126	173	169	156
% of target	*97.42%*	*81.29%*	*87.82%*	*85.79%*	*79.19%*
Refused / Not available	4	29	24	28	41
% of target	*2.58%*	*18.71%*	*12.18%*	*14.21%*	*20.81%*
Communities					
Target	-	523	-	-	-
Completed	-	518	-	-	-
% of target	-	*99.04%*	-	-	-
Refused / Not available	-	5	-	-	-
% of target	-	*0.96%*	-	-	-
D2D sales agents					
Target	-	-	-	-	182
Completed	-	-	-	-	164
% of target	-	-	-	-	*90.11%*
Refused / Not available	-	-	-	-	18
% of target	-	-	-	-	*9.89%*

[a] Household surveys in waves 1 and 4 include around 1,800 observations from 'pre-CLTS' areas, which are excluded from the CLTS impact evaluation analysis but are part of the SanMark analysis (see Chapter 4). During wave 3, security concerns impeded normal data collection activities in two communities in our sample (Igbara Odo and Ogotun, located in Ekiti South West LGA), leading to 31 households not being approached at this time (they are included here as part of the 'Refused / Not available' category).
[b] The sample of businesses was boosted by 42 observations in wave 2. Interviewed businesses include SanMark pilot businesses, SanMark study businesses (treatment and controls) and non-eligible businesses. Closed businesses were also interviewed, when possible, and are included in the sample.

Source: IFS based on own primary data (census and data collection waves 1 to 4).

Household census

As mentioned above, the collection of a census was needed since there was no (publicly available) updated, geographically representative information for our study area to use as a sample frame. The census consisted of a short interview (15 minutes) which collected only basic demographic characteristics and general sanitation infrastructure access.

Household panel

To evaluate household-level impacts on the main outcomes of interest, we rely on a panel of households. Using the household census as our sample frame, we randomly selected 20 households from each cluster, to obtain a representative sample. These households were then surveyed on four occasions as described in Figure 5.2. Given the sample frame and the high re-interviewing rate, we were able to track a sample of 4,649 households with no recent experience of CLTS or SanMark, located in our study areas between waves 1 and 4.[29] This sample frame allows us to detect a minimum effect of 8 to 9 percentage points with a power of 80%.

Table 5.2 shows the different modules covered in each survey wave. The household panel survey that includes information from waves 1,2, 3 and 4 covers basic demographic characteristics, sanitation infrastructure and beliefs, and other variables of interest, spanning the course of 3 years. Wave 1 (baseline) and wave 4 (endline) surveys involved longer questionnaires, including individual-level modules with questions on education enrolment, sanitation practices, health outcomes and, for households with children, anthropometric measurements from the two youngest children below the age of 6. Wave 3 also included sections directed exclusively at women, with questions about sanitation preferences and decisions about household purchases and investments.

Household surveys in waves 1 and 4 also included an additional sample of around 1,800 households from areas in which WaterAid Nigeria had previously implemented CLTS. We call these 'pre-CLTS' households, and they are excluded from our estimation of CLTS impacts. However, because none of these areas had any experience with SanMark before this project, these households are used to supplement our analysis of SanMark impacts, resulting in a panel of around 6,400 households.

[29] We discuss attrition from wave 1 to wave 4 in more detail in Chapter 6.

Table 5.2. Modules included in the household questionnaire, by survey wave

	Survey wave 1	Survey wave 2	Survey wave 3	Survey wave 4
Household and individual characteristics				
Household characteristics	X	X	X	X
Household roster (individuals)				
Sanitation and hygiene behaviour	X		X	X
Age, gender and employment status	X			X
Literacy and education	X			X
Income, expenditure and assets	X			X
Savings, credit and debts	X			X
Shocks (health, weather, crime)	X			
Household decisions (women only)			X	
Social status				X
Contact with neighbouring communities				X
Household coordinates	X	X	X	X
Access to private sanitation facilities				
Toilet ownership (+ type)	X	X	X	X
Observation by interviewer	X	X		X
Reasons for non-ownership	X	X	X	X
Who influenced decision to build toilet		X	X	X
Assistance in toilet construction			X	X
Institutional sanctions for OD		X	X	X
Health and healthcare				
Health status (individuals)	X			X
Healthcare usage (household level)	X			X
Child health (individuals <6 years old)	X			X
Anthropometrics (individuals <6 years old)	X			X
Mediating factors				
Knowledge about sanitation and health	X			X
Costs of constructing a toilet	X			X
Sanitation beliefs, benefits and norms	X		X	X
Access to safe drinking water	X			
Participation in women's August meetings			X	
Awareness of CLTS activities		X	X	X
Awareness of WET product line			X	X

Business panel survey

Alongside the household census, we conducted a mapping of all businesses that could be eligible for the SanMark intervention. Businesses were deemed eligible if they were located within our study area and they sold concrete blocks (the main input used in the construction of toilets and pits in this area), other concrete materials (rings, collars, etc.)

or other plumbing-related products such as PVC pipes or ceramic pans. 150 such businesses were identified and briefly interviewed. Concrete block producers (CBPs) would become the main subjects for the SanMark intervention at the business level, as described in Chapter 4 and more extensively in our Supplier Baseline Report (Abramovsky, Augsburg and Oteiza, 2016b).

This sample of businesses was further approached in waves 1, 2, 3 and 4, as seen in Figure 5.2. In wave 1, detailed information on sales, costs and other business characteristics was collected, for a list of different concrete and sanitation-related products. In wave 2, 42 new businesses were identified and added to the sample. Wave 2 was also the last pre-SanMark business survey. Businesses in the sample were approached twice more. As in the case of households, we constructed a panel of businesses using the answers to survey waves 1 to 4, which covers all businesses operating in our study area during this period. Table 5.3 summarises the questionnaire modules included in each survey wave of businesses.

Table 5.3. Modules included in the business questionnaire, by survey wave

	Survey wave 1	Survey wave 2	Survey wave 3	Survey wave 4
Business characteristics	X	X	X	X
Owner characteristics (education, age,	X			
Monthly sales and prices, by product	X	X	X	X
Revenues, costs and capital expenditures	X	X	X	X
Innovative behaviour	X	X	X	X
Personality test (owners)		X		
Participation in SanMark training			X	X
WET sales and prices			X	X
Closed business module (only if out of		X	X	X

Interviews included modules on general business characteristics, such as size, number of employees and years in activity; business type, such as delivery methods available, formal registration, and provision of installation services; quantity and value of sales by product type; identification of competitors; communities most frequently serviced by the business; revenues, costs and profits; innovation and marketing activities; and access to infrastructure. During wave 2, the first survey administered to the augmented sample of businesses and the last before the start of the SanMark intervention, we also included a module on the personality of the managers of each business, using the standard 'Big Five' personality test. This test consists of 45 questions about personal preferences, and assigns a score to the following five personality traits: openness, conscientiousness, extraversion, agreeableness and neuroticism. In the past, researchers have found that individual proactive behaviour is positively correlated with openness and conscientiousness, and negatively correlated with agreeableness and neuroticism (Neal et al., 2012).

In order to understand entry and exit of businesses in the sector, waves 2 to 4 included an additional questionnaire directed only at businesses that appeared to be closed at the time of interview. In this brief questionnaire, we enquired about the main reasons for closing the business. Importantly, all businesses were approached in every wave, irrespective of their previous status. This allows for tracking of businesses that are only temporarily closed, which seems to be a common phenomenon in our study area.

Community survey

A community-level survey was conducted during wave 1, together with the household survey, and directed at community leaders. It included information on available infrastructure, including health and sanitation facilities, OD status, transportation to and from the community, and public services such as primary and secondary schools, hospitals and police stations. Access to savings instruments was also measured, as well as whether the community had any previous history of CLTS activities being carried out by other agencies in recent years. Table 5.4 lists the main modules included in this questionnaire.

Table 5.4. Modules included in the community questionnaire, by survey wave

	Survey wave 1	Survey wave 2	Survey wave 3	Survey wave 4
Road quality and travel time to urban	X			
Main economic activities of the	X			
Infrastructure, public goods and services	X			
Community activities	X			
Water access and sanitation	X			
Shocks to the community	X			
Local prices of construction materials and food	X			
Village chief characteristics (education, personality)	X			

The main purpose of this survey was to measure community characteristics that can be associated with any observed differential impact of the interventions across treated communities. Distances to other communities or markets, as well as prices for basic sanitation products, were collected for this purpose. A final module included questions directed at the community leader, with basic demographic characteristics as well as the 'Big Five' personality test.[30]

SanMark D2D agent survey

This final component of the data collection exercise was to gather information on the D2D agents recruited as part of the SanMark marketing activities. It was conducted as part of survey wave 4.

The complete list of participants in WaterAid Nigeria's training sessions was provided to us, and these people were approached in their homes for a 30-minute interview. The questionnaire included a module on individual characteristics of the agent, labour market participation, and general household economic and sanitation characteristics. A second

[30] The information collected in this community questionnaire was eventually not included in our final analysis due to data quality concerns and the fact that most community characteristics could be more reliably understood using the household questionnaire.

module included questions on the 'intensity' of the D2D activities carried out, the amount of WET sales and the commissions received.

6. Balance tests and survey attrition

In this chapter, we discuss the results of our three random assignment procedures, conducted as part of our research design, and then consider attrition from wave 1 to wave 4 of our surveys.

6.1 Balance checks

Recall that households were randomly assigned to either CLTS treatment or CLTS control clusters and, in a second stage, to either SanMark treatment or SanMark control clusters. Businesses were randomly assigned to either SanMark treatment or SanMark control at the individual level.

Households – cluster randomisations 1 and 2

In previous reports, we tested for balance across CLTS treatment and control samples, and SanMark treatment and control samples (including pre-CLTS households), over a wide range of outcomes and other observable characteristics of the household, the head of the household and the community.

As described extensively in these reports, the randomisation was successful, in that (i) CLTS and control groups were found to be observationally equivalent, on average, both at the household level and at the community level (Abramovsky, Augsburg and Oteiza, 2015)[31] and (ii) the randomisation of SanMark resulted in two balanced samples, as shown in section A.2 of Abramovsky, Augsburg and Oteiza (2016b). This allows us to conclude that any post-treatment difference between the groups cannot be attributed to pre-treatment differences (what development economists call 'selection bias'), but must be due to the treatment, as described in, for instance, Duflo, Glennerster and Kremer (2007).[32]

Rather than revisiting these results, we present here an additional check not performed before. We compare the average outcomes and other observable characteristics by treatment arm, considering the four possible combinations of CLTS and SanMark treatment status, i.e. households assigned to both CLTS and SanMark, households assigned to only one treatment, and households assigned to the control group in both interventions. Results of this exercise are presented in Table 6.1.

[31] Statistical differences appeared at the same rate as expected for any randomly designed sample. Specifically, differences were observed only in terms of household size, which is therefore included as one of our household-level controls in all our analysis.

[32] A small caveat applies here, as mentioned in Chapter 4. D2D agents were recruited from SanMark areas but allowed to work anywhere, and therefore could potentially spread to control areas as well. In this sense, we should interpret any estimates of SanMark impacts as a lower bound of its true causal effects.

Table 6.1. Balance in outcomes across treatment arms: households (HHs)

	Control both	Treatment both	CLTS only	SanMark only
	(1)	**(2)**	**(3)**	**(4)**
Toilet ownership				
HH has (or is constructing) a toilet (%)	39.808	37.129	38.111	34.862
	[2.989]	[3.149]	[3.156]	[3.419]
HH has a functioning toilet (%)	37.805	35.502	36.454	34.037
	[2.970]	[3.110]	[3.038]	[3.391]
HH has a functioning, improved toilet (%)	33.972	32.153	33.886	30.917
	[2.967]	[3.094]	[2.983]	[3.131]
Toilet usage				
All members of HH use toilet (%)	35.192	33.397	34.134	31.468
	[2.911]	[2.978]	[2.916]	[3.110]
At least one member of HH performs OD (%)	62.979	65.263	65.203	67.431
	[2.900]	[3.054]	[3.018]	[3.217]
Number of households	1,148	1,045	1,207	1,090
Number of clusters	63	59	65	58
F-test of joint significance (*F*-stat)	-	0.249	0.904	1.784

Note: Table displays group means with standard errors in square brackets. Stars indicate statistical significance: * 10%, ** 5% and *** 1%. *t*-tests for the difference in the means between each group and the 'control both' group in column 1 were carried out for each variable individually; no statistically significant differences arise. An *F*-test for joint significance for all variables predicting treatment status compared with 'control both' was also conducted, with *F*-statistics presented in the last row of the table. LGA fixed effects included in all tests and standard errors clustered at the randomisation unit (TU) level. All variables measured at wave 1.

Table 6.1 shows the means for our main outcomes of interest during wave 1.[33] We perform two types of tests to check that our samples are balanced. First, we conduct *t*-tests on the significance of the difference in means between that of each group in columns 2 to 4 with respect to that in column 1, and find no statistically significant differences. When carrying out these tests, errors were clustered at the TU level and LGA fixed effects were included, to replicate the specification we will use later on in this report to estimate programme impacts. As an additional check, in columns 2 to 4, we test whether all variables in the table can jointly predict whether a household belongs to a specific treatment arm compared with belonging to column 1. The resulting *F*-statistics are presented in the last row of the table and show no statistically significant power of the outcomes listed in predicting treatment status.

[33] The analysis was restricted to observations with no missing values for any of the variables included, which results in marginally smaller sample sizes than those presented in Table 5.1.

Table 6.2 shows the means of several variables measuring a series of characteristics of the household and the household head (all measured at wave 1) and performs the same two tests as those described for Table 6.1. Only two imbalances are observed: in the age of the household head for the CLTS only group and in the employment status for the SanMark only group, statistically significant at the 5% and 1% levels respectively. These differences could be problematic since one would assume they could both be correlated with the outcomes studied. However, the magnitude of the differences is small, and the joint test for significance of the whole set of covariates in predicting CLTS only and SanMark only treatment status is barely significant at the 10% level. Nevertheless, we include these baseline characteristics in our impact regressions to ensure that they are not driving our estimated effects. Additionally, while these comparisons are relevant for the study of the interaction between CLTS and SanMark, the evaluation of treatment impacts of the SanMark intervention on its own will be conducted separately, on a larger sample (which includes pre-CLTS households), which is indeed balanced, as shown in Abramovsky, Augsburg and Oteiza (2016b).

Overall, Tables 6.1 and 6.2 show that the samples for our combined treatment arms were mostly balanced at wave 1.

Table 6.2. Balance in other observable characteristics across treatment arms: households

	Control both	Treatment both	CLTS only	SanMark only
	(1)	**(2)**	**(3)**	**(4)**
Household head characteristics				
HH head age	56.373	55.100	53.771**	54.939
	[1.178]	[0.721]	[0.802]	[0.720]
HH head male (%)	62.021	62.488	62.469	65.780
	[1.660]	[1.937]	[1.889]	[1.775]
HH head employed (%)	63.502	65.072	64.043	70.459***
	[2.072]	[1.970]	[1.870]	[2.001]
HH head finished primary school (%)	67.944	66.603	69.843	67.523
	[1.955]	[2.416]	[2.196]	[1.921]
Household size	3.943	3.741	3.767	4.046
	[0.135]	[0.131]	[0.114]	[0.113]
Children under the age of 6	0.484	0.480	0.470	0.492
	[0.035]	[0.037]	[0.034]	[0.034]
Household characteristics				
HH primary economic activity is farming (%)	41.986	47.560	49.544	48.716
	[3.501]	[3.884]	[3.493]	[3.901]
HH has any savings (%)	24.477	21.435	24.027	20.459
	[2.120]	[2.085]	[2.222]	[1.947]
HH has any debts (%)	22.125	20.574	18.724	19.083
	[1.780]	[1.965]	[1.422]	[1.648]
Number of households	1,148	1,045	1,207	1,090
Number of clusters	63	59	65	58
F-test of joint significance (*F*-stat)	-	1.369	1.704*	1.698*

Note: Table displays group means with standard errors in square brackets. Stars indicate statistical significance: * 10%, ** 5% and *** 1%. t-tests for the difference in the means between each group and the 'control both' group in column 1 were carried out for each variable individually. An F-test for joint significance for all variables predicting treatment status compared with 'control both' was also conducted, with F-statistics presented in the last row of the table. LGA fixed effects included in all tests and standard errors clustered at the randomisation unit (TU) level. All variables measured at wave 1.

Businesses – individual-level randomisation

In our business-level analysis, we will only compare outcomes between two treatment arms. In particular, we will compare the businesses that were selected into the SanMark treatment with those assigned to the control group, which would not be approached for SanMark business-level activities during the course of the study. Balance between these two samples at wave 2, the last one before the SanMark intervention began and the moment the randomisation was conducted, has already been established in Abramovsky, Augsburg and Oteiza (2016b). Here, we reproduce that analysis, this time including all pre-treatment observations available for each business.[34] This is important because most business outcomes present a high degree of volatility, and taking only one measure introduces a significant amount of noise into our estimates. Average outcomes over the course of several survey waves will reduce this noise and allow us to arrive at a more accurate picture of the characteristics and performance of these two groups of businesses before the onset of the SanMark intervention.

[34] Since the business phase of SanMark was rolled out between waves 2 and 3, pre-treatment waves are 1 and 2.

Table 6.3. Balance in outcomes across treatment arms: businesses

	Control	SanMark
	(1)	**(2)**
Business financials		
Typical monthly sales (US$)	3,062.972	1,878.513
	[641.474]	[287.814]
Typical monthly costs (US$)	3,283.134	2,013.526
	[889.897]	[619.162]
Wages/benefits (as % of monthly costs)	*18.563*	*16.946*
	[1.176]	*[1.213]*
Raw materials (as % of monthly costs)	*40.563*	*41.176*
	[2.146]	*[2.233]*
Machinery, tools (as % of monthly costs)	*5.155*	*4.324*
	[0.771]	*[0.720]*
Electricity, fuel, etc. (as % of monthly costs)	*8.465*	*8.622*
	[0.837]	*[0.803]*
Transportation (as % of monthly costs)	*4.901*	*6.446*
	[0.566]	*[0.658]*
Maintenance, repairs (as % of monthly costs)	*5.648*	*5.851*
	[0.566]	*[0.564]*
Rental of capital (as % of monthly costs)	*0.620*	*1.135*
	[0.202]	*[0.273]*
Taxes and levies (as % of monthly costs)	*4.366*	*3.932*
	[0.373]	*[0.334]*
Sales increased in past year (%)	61.972	66.216
	[5.300]	[5.202]
Ever received a loan (%)	29.577	37.838
	[5.157]	[5.995]
Number of businesses	71	74
F-test of joint significance (*F*-stat)	-	1.060

Note: Table displays group means with standard errors in square brackets. Stars indicate statistical significance: * 10%, ** 5% and *** 1%. *t*-tests for the difference in the means between the 'control' and 'SanMark' groups were carried out for each variable individually; no statistically significant differences arise. An *F*-test for joint significance for all variables predicting treatment status was also conducted, with *F*-statistics presented in the last row of the table. LGA fixed effects included in all tests and standard errors clustered at the randomisation unit (supplier) level. All pre-treatment observations for each business were included (waves 1 and 2).

We show balance in business-level outcomes of interest in Table 6.3 and we focus on business characteristics in Table 6.4.[35] As can be seen, our SanMark and control groups of businesses are largely balanced. A single individual imbalance (significant at the 10% level) appears in the share of businesses connected to their local power grid (Table 6.4). An *F*-test for joint significance, presented at the bottom of each table, shows that all variables together have no statistically significant explanatory power to predict treatment status.

[35] As for the case of households, this analysis was restricted to observations with no missing values for any of the variables included, which results in marginally smaller sample sizes than those presented in Table 5.1.

Table 6.4. Balance in other business traits across treatment arms: businesses

	Control	SanMark
	(1)	(2)
Business characteristics		
Years in existence	6.881	7.093
	[0.758]	[0.964]
Number of full-time employees	4.376	4.773
	[0.477]	[0.542]
Some formal registration (%)	81.651	85.567
	[4.058]	[3.380]
Access to infrastructure		
Connected to power grid (%)	4.587	10.309*
	[2.365]	[2.982]
Owns electricity generator (%)	55.963	51.546
	[4.922]	[4.685]
Improved water source (%)	1.835	1.031
	[1.284]	[1.029]
Uses cell phones (%)	80.734	72.165
	[3.598]	[4.789]
Has internet connection (%)	2.752	7.216
	[1.557]	[2.928]
Other business characteristics		
Households are main customers (%)	63.303	64.948
	[4.668]	[5.094]
Owns a reliable means of transport (%)	74.312	75.258
	[4.840]	[5.387]
Customers pick up products (%)	82.569	86.598
	[3.989]	[3.260]
Number of businesses	109	97
F-test of joint significance (*F*-stat)	-	1.192

Note: Table displays group means with standard errors in square brackets. Stars indicate statistical significance: * 10%, ** 5% and *** 1%. *t*-tests for the difference in the means between the 'control' and 'SanMark' groups were carried out for each variable individually. An *F*-test for joint significance for all variables predicting treatment status was also conducted, with *F*-statistics presented in the last row of the table. LGA fixed effects included in all tests and standard errors clustered at the randomisation unit (supplier) level. All pre-treatment observations for each business were included (waves 1 and 2).

D2D agents

SanMark D2D agents were interviewed in wave 4 with the aim of understanding the effectiveness, and future potential, of the SanMark intervention for increasing WET sales and the well-being of the agents themselves. These interviews will be used as a complement to the impact evaluation.

Table 6.5 presents some average characteristics of the D2D agents interviewed. Agents are overwhelmingly male, around 40 years old and highly educated: 40% of them have completed some kind of post-secondary-school education. Most of them have children under the age of 6 living with them, and almost 60% of them have non-agricultural activities as their primary source of income. This leads us to the conclusion that D2D agents are not representative of the wider population in our study area, which we confirm by comparing their average characteristics with those from our representative sample of households, shown in column 2.

Agents are typically younger, and are more likely to have completed primary school than household heads interviewed as part of our household survey. They are also more likely to be male. The most common economic activity for households in the D2D survey is farming, but the share is lower among D2D agent households than among our wider sample of households interviewed. Importantly, D2D agents are 19pp more likely to own functioning, improved toilets than the rest of the population in our study area, and 8pp less likely to perform OD, leaving about a third of D2D agents without a toilet and 42% living in a household where at least one member performs OD.

Table 6.5. D2D agents compared with household heads in our sample

	D2D agent households	Study households	Difference (*t*-test)
	(1)	(2)	(2)–(1)
Panel A. Individual characteristics			
Age	42.8	57.8	14.957***
Male (%)	73.8	60.4	−13.4***
Education – tertiary or above (%)	40.2	9.5	−30.7***
Education – senior secondary (%)	37.2	21.3	−15.8***
Education – junior secondary (%)	4.9	5.0	0.1
Education – primary (%)	16.5	33.1	16.6***
Panel B. Household characteristics			
Size of household	5.40	4.03	−1.366***
No. of children under 6	0.64	0.50	−0.140***
Farming is primary economic activity (%)	40.2	66.9	26.7***
Owns functioning, improved toilet (%)	65.9	47.1	−18.8***
At least one member performs OD (%)	42.1	50.0	8.0**
Number of observations	164	5,579	

Note: The values displayed for *t*-tests are the differences in the means across the groups. ***, ** and * indicate significance at the 1%, 5% and 10% critical levels. Individual-level characteristics for study households represent the household head. Individual-level characteristics from the D2D sales agents show that the majority are the heads of their households (74% of cases). In 16% of the cases, the agent was the spouse of the household head and in 10% of the cases they were the biological children of the household head. Characteristics of household heads from our study sample measured at wave 4 (November 2017), so might not coincide exactly with the mean characteristics analysed in Tables 6.1 and 6.2, which use data from wave 1.

6.2 Attrition

In every panel survey, where households are repeatedly interviewed over time, researchers face the problem of losing some respondents along the way. There are two main reasons why this could generally happen. The first possibility is that the household cannot be located or the interviewers cannot reach the community itself. This could be because the household moved or for reasons that might make survey work difficult or even dangerous. For example, as mentioned in Chapter 5, during wave 3, civil unrest in Ekiti South West LGA prevented interviewers from accessing two of the communities that were under government curfew. A second possibility is that the household was located, but the respondent was not available on repeated occasions, refused to accept the interview or provided too little time to complete the questionnaire. For our purposes, interviews that could not be carried out for any reason, whether refusal or non-availability, will be considered as attrition and will not be part of our impact estimation.

In this section, we discuss attrition patterns in our two panel data sets: households and businesses. Our attention will be on attrition levels observed in the endline survey (wave 4), since attrition in previous waves has already been documented in Abramovsky, Augsburg and Oteiza (2016a, 2017a and 2017b).

Household survey

Overall, our study showed low attrition rates: 2.56% at wave 2, 9.27% at wave 3 and 12.40% at wave 4 (see Table 5.1). The fact that 3 years after the baseline survey, we successfully interviewed almost 90% of the households in our initial sample is reassuring.

Attrition can reduce the power a study has to identify causal effects of a policy, and might threaten the representativeness of the study sample if attrition rates vary significantly according to household characteristics. If attrition is asymmetric between treatment and control groups, this could be particularly problematic. A simple example illustrates this potential problem. Suppose that, after wave 2 (baseline), a large number of households with no toilets are displaced from one of our CLTS areas and cannot be located in wave 4 (endline). And suppose that all households in control areas were successfully located and interviewed in wave 4. Even if CLTS had no effect at all on toilet construction, we would find that in wave 4, (interviewed) CLTS households are, on average, more likely to own a toilet than control households. In this hypothetical case, the finding would be driven by changes in the sample representativeness due to attrition, not by the genuine impact of the programme on CLTS households.

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. A first reassuring finding is that attrition rates were not statistically significantly different across treatment arms: control (12.32%), CLTS only (14.23%), SanMark only (11.95%) and CLTS & SanMark (11.46%). We further check more formally whether treatment status (i.e. belonging to a certain treatment arm) can predict attrition, conditional on baseline characteristics. If it can, then it would suggest that attrition was, in fact, asymmetric when conditioned on observable characteristics, and we would have to account for this in our impact analysis. Table 6.6 presents the results of these regressions.

Table 6.6. Regression results for household attrition in wave 4

	Dependent variable = 1 if household is not interviewed in wave 4, = 0 otherwise			
	(1)	**(2)**	**(3)**	**(4)**
SanMark only	−0.00 (0.01)	−0.00 (0.01)	0.02* (0.01)	0.02* (0.01)
CLTS only	0.02 (0.02)	0.02 (0.02)	0.01 (0.01)	0.01 (0.01)
CLTS & SanMark	−0.01 (0.02)	−0.01 (0.01)	−0.01 (0.01)	−0.01 (0.01)
LGA fixed effects	No	Yes	Yes	Yes
Household controls	No	No	Yes	Yes
Community controls	No	No	No	Yes
Attrition rate (control group)	0.12	0.12	0.12	0.12
p-value for *F*-test on covariates			0.00	0.00
Number of observations in wave 1	6,359	6,359	6,130	6,130

Note: Attrition measured at wave 4. Dependent variable equals 1 if the household was not interviewed and equals 0 if the household was successfully interviewed. Table shows estimation results from regressions with attrition as the dependent variable and treatment status as the main coefficient of interest. Household controls include gender, age, age squared, employment status and level of education of household head, household size and farming as main economic activity. Community controls include paved inner roads, presence of primary school, presence of hospital, community population (2014) and settlement population (2014). Household- and community-level controls measured at wave 1. Standard errors, shown in parentheses, are clustered at the level of randomisation (TU). Stars indicate statistical significance: * 10%, ** 5% and *** 1%.

Column 1 shows the results of a regression of attrition on treatment indicators only, i.e. omitting any other control variables. We see that treatment arms do not predict attrition in a statistically significant way, meaning that attrition was balanced across treatment groups. This is robust to the inclusion of LGA fixed effects (column 2), household-level controls (column 3) and community-level controls (column 4). The SanMark only group seems to have slightly higher levels of attrition than the control group, but the magnitude of this is small (2pp) and only marginally statistically significant at the 10% level. Overall, the table shows that attrition in our last survey wave was balanced across treatment arms, meaning that attrition will not threaten the identification of causal impacts.[36]

The row ‘*p*-value for *F*-test on covariates’ shows the results of a test of joint significance for household-level controls in column 3 and for household- and community-level controls in column 4. The low *p*-value means that these observable characteristics have significant explanatory power in predicting attrition. This is driven mainly by the age of the

[36] Attrition was similarly balanced across treatment arms in survey waves 2 and 3, as was described in Abramovsky, Augsburg and Oteiza (2016a and 2017a).

household head (not shown), implying that households with older heads are more likely to drop from our sample. This means that our results might not be representative for all the households in our study area since a particular group – i.e. households with older heads – will be under-represented.

Business survey

Since our business survey was also administered over several waves, it too warrants analysis to ensure that there is no selective attrition. Table 6.7 presents the results of regressions of three interview outcomes of interest, in terms of attrition, on treatment status. Since we include treatment status as a regressor in these regressions, we include only businesses assigned to either SanMark treatment or SanMark control. First, we use an indicator variable equal to 1 if the business had closed between survey waves (in this case, when feasible, business owners were instead asked to respond to the closed business questionnaire). Even though closed businesses were still approached for interview, this is an interesting test because it tells us whether the composition of active versus closed businesses was different in our SanMark treatment and control groups, and gives us a first approximation to differential performance across groups. Second, we use a variable equal to 1 if the respondents refused or were not available for interview (our standard definition for attrition). Finally, some respondents may agree to be interviewed, but refuse to disclose information on sales and costs, which is crucial for our analysis of business performance. To test whether SanMark treatment status was correlated with the likelihood of cooperating with interviewers and revealing sales and costs data, we built a variable equal to 1 if either monthly sales or costs data were missing. These three indicators are measured at wave 4.

Table 6.7. Regression results for business attrition in wave 4

Dependent variable:	Closed=1, otherwise=0 in wave 4		Refused or N/A=1, otherwise=0 in wave 4		Missing sales/costs=1, otherwise=0 in wave 4	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
SanMark treatment	0.03 (0.07)	0.09 (0.07)	0.06 (0.06)	0.07 (0.07)	0.03 (0.08)	0.12 (0.10)
LGA fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Business characteristics	No	Yes	No	Yes	No	Yes
Mean dep. var. (control)	0.16	0.13	0.13	0.09	0.30	0.25
p-value for *F*-test on covariates		0.09		0.40		0.32
Number of observations	129	102	129	102	129	102

Note: Table shows estimation results from regressions with the dependent variable shown and treatment status as the main coefficient of interest. Includes observations from wave 4 only. Business characteristics, all measured before the intervention started (wave 2 or, if missing, wave 1), include business age, number of full-time employees, whether the business is registered formally or not, whether the business is connected to the power grid, whether it owns a generator, and whether it has access to water, mobile phones and internet. Standard errors, shown in parentheses, are clustered at the level of randomisation (TU). Stars indicate statistical significance: * 10%, ** 5% and *** 1%.

Although point estimates suggest slightly higher numbers of closed businesses and attrition among our SanMark group, none of the estimates is statistically different from zero. This is reassuring as one might suspect higher refusal rates particularly amongst control businesses. This does not seem to be a problem in this case, where attrition rates – if anything – are slightly higher in our SanMark group. Overall, our attrition rates are reasonable: a recent study using a panel survey of Nigerian businesses achieved a response rate of 75.6% in the first year after the intervention was conducted (McKenzie, 2017). As a final point worth noting, just 70% of the businesses in our study sample agreed to report estimates of their monthly sales and costs figures, which will reduce the power of our estimation of differential performance.

7. CLTS impacts

In this chapter, we present CLTS impacts at the household level in the short and medium run, and discuss their relationship with wider state- and country-level sanitation trends during this period.

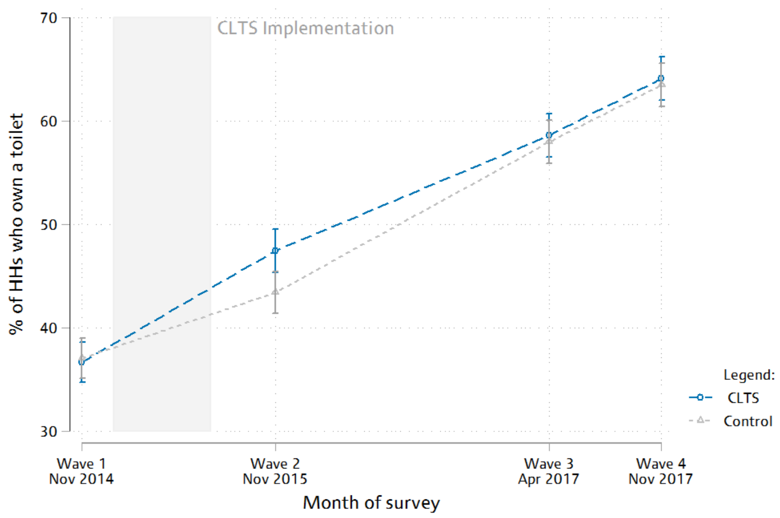
We start with a descriptive visual analysis, presenting the average values for each of our outcomes of interest, over the four survey waves that form part of this study (waves 1–4). We then concentrate on estimating regression specifications (1), (2) and (3), where we account for any differences in observable characteristics between CLTS and control households that might have been observed at baseline.

7.1 Visual evidence

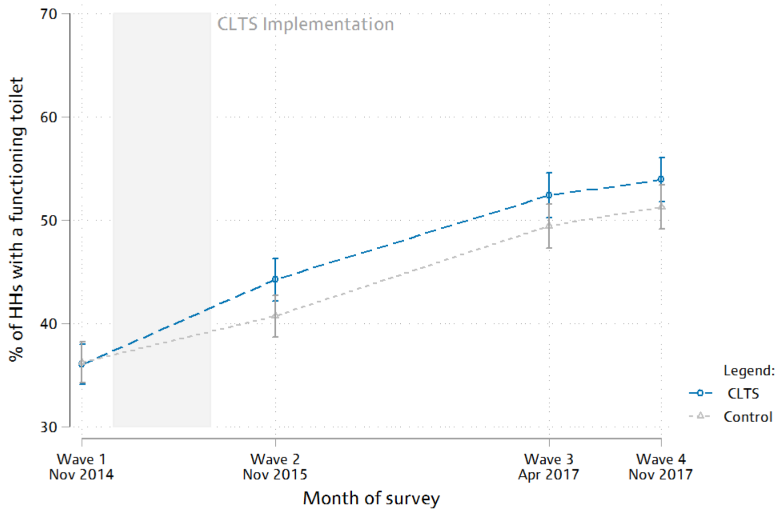
As discussed in Section 4.3, we focus on outcomes related to toilet ownership and usage. Figure 7.1 shows the evolution of outcomes related to toilet ownership (household owns a toilet of any kind; household owns a functioning toilet; household owns a functioning improved toilet) and whether the main respondent performs OD. Two further OD measures that are part of the set of six primary outcomes are considered in the regression analysis that follows but not analysed in this section.

Figure 7.1. Evolution of main outcomes by wave

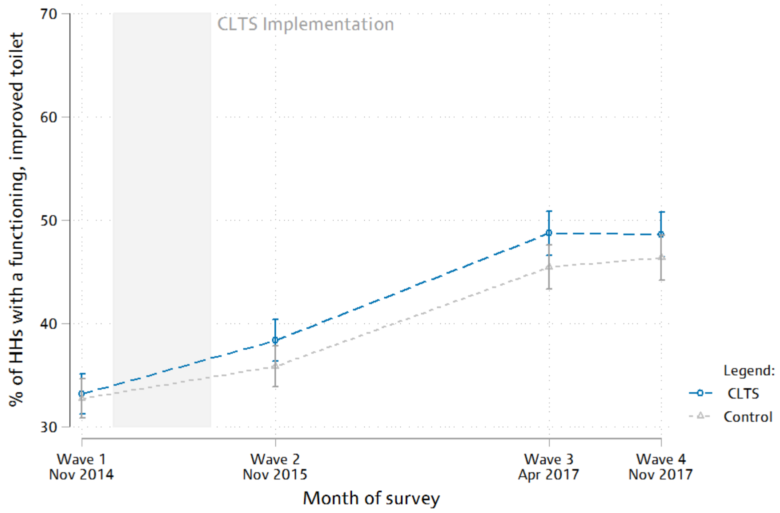
(a) Households who own a toilet (any kind)

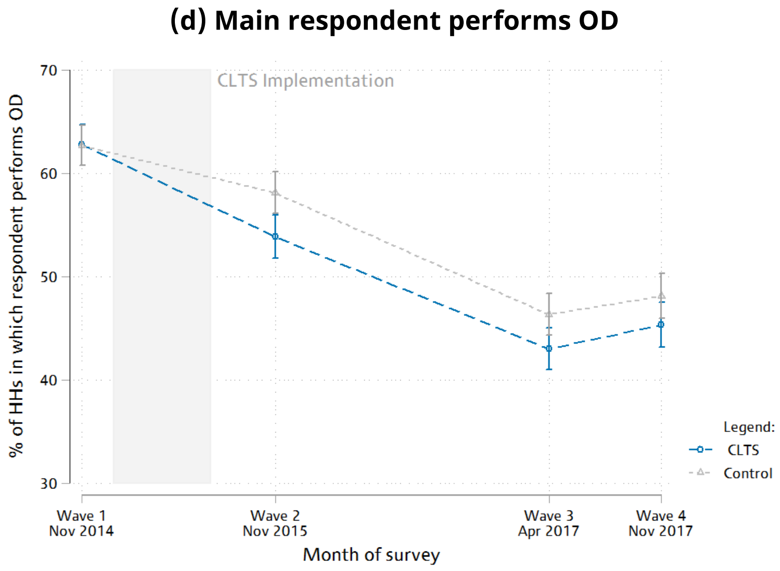


(b) Households who own a functioning toilet



(c) Households who own a functioning improved toilet





Note: The scale of the x-axis corresponds to the amount of time elapsed between each survey wave. The CLTS intervention was carried out shortly after wave 1, between January and June 2015, as indicated by the grey bar.

Source: Household survey, waves 1–4.

We see that toilet ownership (independent of what measure of toilet is considered) has increased over time, for both CLTS (blue dashed line) and control groups (grey dotted line). For the purposes of this study, we are interested in whether CLTS had any significant effect in increasing toilet ownership and/or reducing OD. This is shown in Figure 7.1 by the *additional* increase in ownership or reduction in OD observed among CLTS households, compared with control households. This difference will be the causal impact of CLTS and will be estimated more accurately in Section 7.3.

But first we should note the positive trends observed among households in control areas (grey dotted line). These areas were not assigned to CLTS and, to the best of our knowledge, were not exposed to any other sanitation promotion interventions. Nonetheless, toilet ownership has increased and OD has fallen significantly in this group, during the period of our study. For example, ownership of a functioning toilet of any kind (Figure 7.1(b)) increased from 36% to 52% in the course of just 3 years among controls. A similar finding holds when we consider the other two toilet ownership measures and also looking at our measure of OD behaviour, where we see a sharp decrease from above 60% in wave 1 to below 50% by wave 4 (Figure 7.1(d)). This improvement is impressive, particularly considering the Nigerian sanitation trends during the years previous to this study, and cannot be the result of the CLTS intervention.

In order to put these findings in context, we discuss the national and state-level sanitation trends in more detail in the next section, before proceeding with the estimation of CLTS impacts in Section 7.3.

7.2 Nigerian sanitation trends during the study period

Our data show that ownership of functioning toilets increased significantly in control areas, from an average of 36% of households in wave 1 (November 2014) to 52% in our last survey wave, 3 years later. This led us to look at recently published data on sanitation trends in both of our study states and in Nigeria in general, to better understand the wider context in which the intervention took place.

We compiled toilet coverage and OD rates at the national and state levels from all publicly available data sets at the time of writing this report.[37] We included only surveys that were representative at the state level and were carried out between 2007 and 2016, to show how sanitation levels (ownership and OD) developed over time nationally and in both Ekiti and Enugu states.[38] Figure 7.2 plots the resulting estimates of access to improved sanitation (private or shared) and rates of OD for each survey year.[39]

The first period we are interested in is the run-up to our study period, between 2007 and 2014. Figure 7.2 suggests that there is negligible improvement at both the national and state levels during these years. Improved sanitation access at the national level has decreased from 45% in 2007 to 40% in 2014, according to these sources. Indeed, Nigeria later failed to reach its Millennium Development Goal target of 63% coverage for improved sanitation in the year 2015. The OD rate has not performed much better: from a starting point of close to 30% in 2007, it decreased to 29% by 2014.

For the purposes of this study, it is also relevant to look at sanitation trends at the state level. While national sanitation trends are important to gauge the magnitude of the challenge at the national level, state-level trends will provide us with a more reasonable benchmark with which to compare our results.

State-level measures should be interpreted with care. These averages are estimated using a small number of households in each case, and therefore provide ‘noisy’ estimates of the underlying trends.[40] Keeping this in mind, Figure 7.2(a) shows opposing trends for the two states between 2007 and 2014. While access to improved toilets increased in Ekiti, from 28% to 46%, it fell from 35% to 29% in the case of Enugu. While both states reduced their overall rates of OD between 2007 and 2014, they did so to different levels: OD rates fell from 64% to 44% in Ekiti, but remained almost unchanged in Enugu, falling from 53% to 51%.

The recent publication of the 2016–17 wave of the Multiple Indicator Cluster Survey (MICS) allows us to also take a look at how these trends evolved during the period in which CLTS

[37] The report by the JMP (2017) includes a similar exercise of aggregation of data sets. However, its estimates include trends at the national level only (and rural/urban areas), but no state-level estimates. Ideally, we would look at rates in the study LGAs rather than in states. However, this disaggregation is not available and would not provide sufficient data points to draw meaningful conclusions.

[38] A total of five data sets fulfilled the conditions mentioned above: the 2007, 2011 and 2016–17 waves of the Multiple Indicator Cluster Survey, the 2013 wave of the Demographic and Health Survey (DHS) and the 2014 wave of the National Nutrition and Health Survey. The General Household Survey waves 2010, 2012 and 2015–16, as well as the 2010 and 2015 waves of the DHS and the Malaria Indicator Survey, were also used to draw the national-level trends, but these were omitted from the Enugu and Ekiti series, as they are not designed to be representative at the state level.

[39] In these surveys, OD is defined as having no access to private or shared toilets of any type. We interpret the data from these surveys as a lower-bound estimate for actual OD rates. This is slightly different from the definition we use in our household survey, in which we explicitly ask respondents whether they perform OD or not.

[40] On average, 1,113 households were interviewed from the state of Enugu and 845 from the state of Ekiti. Using data from the 2006 Nigerian census, this would represent less than 0.04% of each state’s total population.

and SanMark were carried out. In contrast to the previous disappointing performance, Ekiti and Enugu seem to have experienced some increases in access to improved sanitation between 2014 and 2016: from 46% and 29% up to 51% and 52%, respectively. This is coupled with a fall in OD rates. The OD rate in Ekiti fell from 44% to 39% and in Enugu it fell from 51% to 40%. It is in this changing context that the interventions in our study were carried out.[41]

Figure 7.3 reproduces the state-level trends observed in Figure 7.2, and also plots the evolution in improved sanitation ownership and OD rates by state from our study data. We include only non-CLTS areas in these averages, to avoid including any impacts from the CLTS intervention in the comparison.

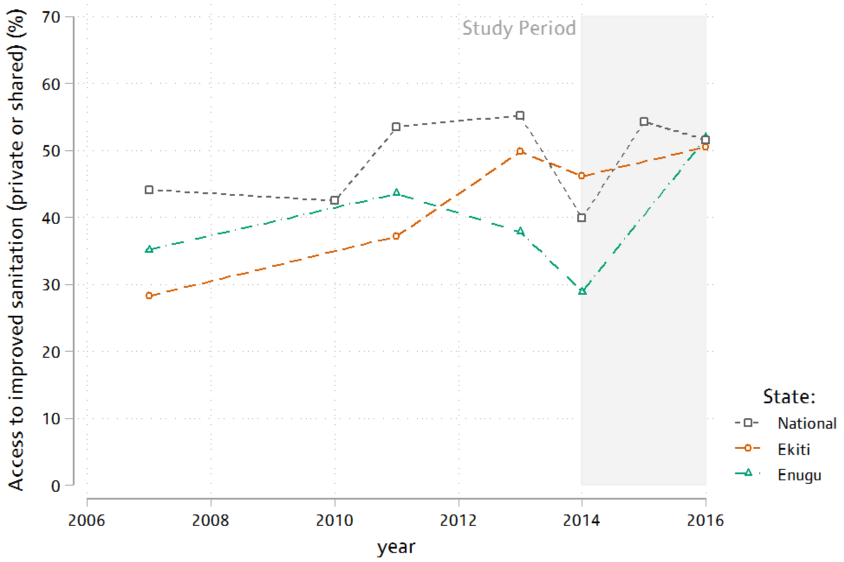
Two points are worth noting. First, because our study area includes rural and semi-rural communities only, OD rates at the start of the project (in 2014) are higher in our sample than in the survey data, which cover the whole state. This is expected as we were explicitly targeting more vulnerable and less developed areas. However, for the case of improved sanitation, the difference is not observed in Enugu, where our estimates are remarkably close to those of the 2014 National Nutrition and Health Survey conducted by the Nigerian Bureau of Statistics.

Second, and most importantly, the upward (downward) trends in improved toilet ownership (OD) from our control areas are broadly consistent with those implied by the independently collected (and state-level representative) data sets. We believe that these positive trends, particularly striking since 2014 when the STS Nigeria baseline was conducted, can therefore explain the increasing toilet ownership and decreasing OD rates observed among households in control areas that comprise part of the study communities.

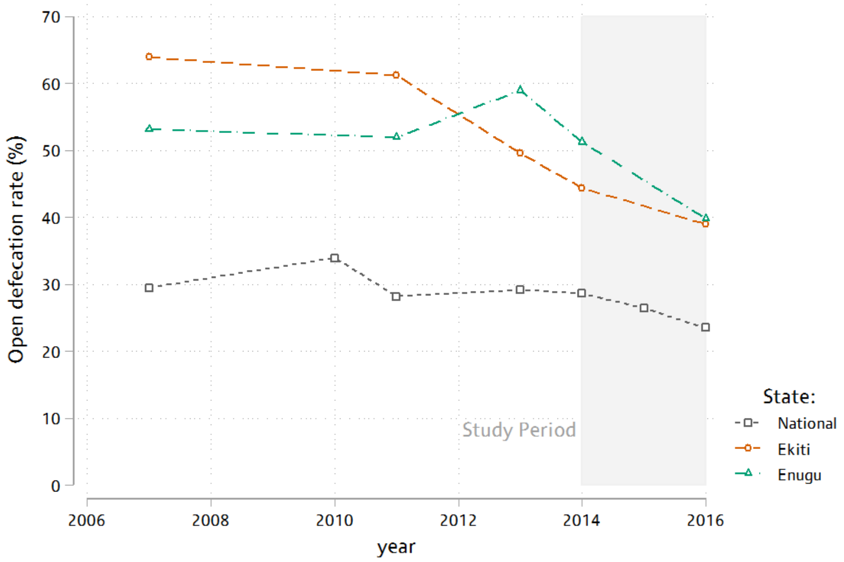
[41] In previous reports, we discussed the possibility of spillovers from the CLTS intervention being behind the trends observed in our control areas. Our intervention was designed to limit information spillovers across clusters, as described in Section 4.1, but in the absence of state-level data, this appeared as a possible explanation for the rapid pace of growth of toilet ownership in control areas. The recently released MICS data suggest instead that what we observe in control areas is in line with the general improvement of sanitation coverage and the reduction of OD at the state and national levels.

Figure 7.2. Sanitation trends nationally and in Ekiti and Enugu states

(a) Access to improved sanitation facilities



(b) OD rates

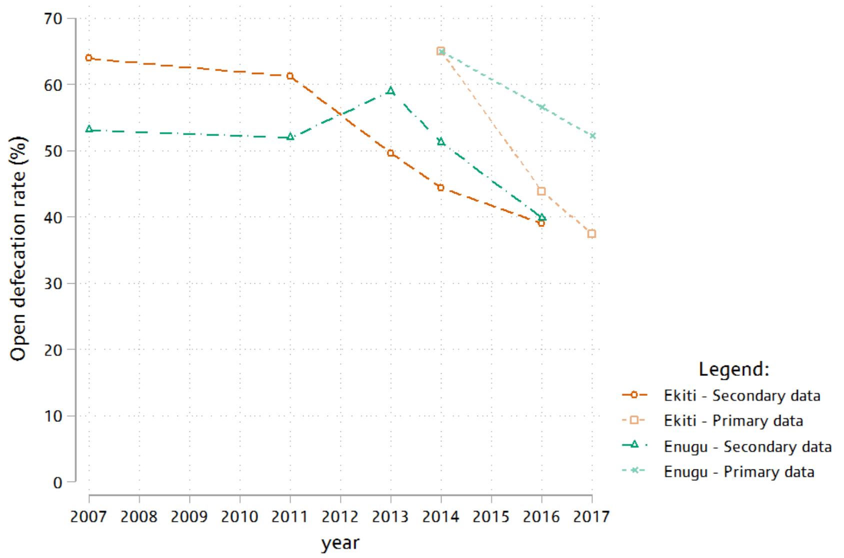


Source: Authors' calculations using: the Multiple Indicator Cluster Survey in 2007, 2011 and 2016–17; the Demographic and Health Survey (DHS) in 2010, 2013 and 2015; the Malaria Indicator Survey in 2010 and 2015; and the National Nutrition and Health Survey in 2014. National-level trend also includes data from Nigeria's General Household Survey in 2010, 2012 and 2015–16, as well as the 2010 and 2015 waves of the DHS and the Malaria Indicator Survey.

Figure 7.3. Comparison of sanitation trends by state: primary and secondary data

(a) Access to improved sanitation facilities

(b) OD rates



Source: Primary data – STS Nigeria household surveys including non-CLTS areas only, waves 1–4. Secondary data – Multiple Indicator Cluster Survey in 2007, 2011 and 2016–17; Demographic and Health Survey in 2013; and the National Nutrition and Health Survey in 2014.

7.3 CLTS impacts on toilet ownership and OD

Having established that the observed improvements in the control group are consistent with wider state-level sanitation trends, we proceed with the estimation of CLTS impacts. Table 7.1 shows the impact results for the six primary outcomes of interest concerning toilet ownership and usage. For each of these outcomes, we present the results of estimating specification (1), discussed in Section 4.4, which pools all post-treatment survey waves, in panel A. This provides us with the average impact of CLTS over the three post-treatment waves. Panel B shows the period-specific CLTS impacts estimated using specification (2).

Let us consider in detail what we can see from the table with respect to the impacts of the CLTS intervention. In the first row of results, the reported estimate of 0.02 (column 1, panel A), for example, indicates that the CLTS intervention increased the percentage of households that own a toilet of any kind by 2 percentage points (pp). The second row shows the *p*-values that are the result of individual tests of significance for each point estimate and range between 0 and 1. A *p*-value below 0.10 can be interpreted as strong evidence against the null hypothesis. The null hypothesis states that the true effect of CLTS is zero. A large *p*-value indicates weak evidence against the null hypothesis, hence suggesting that CLTS had no impact. For example, taking the first column, a *p*-value of 0.25 implies that if we reject the null hypothesis of zero CLTS effect, there is a 25% chance of us making a mistake and that the true effect is indeed zero. This is too high for statistical standards – we need the probability of making such a mistake to be 10% or smaller to have strong enough evidence to reject the null hypothesis.[42] Hence, in column 1, we cannot reject that the true effect of CLTS on toilet ownership on the whole sample is zero. The third row shows adjusted *p*-values. The adjusted *p*-value takes into account that we are testing six hypotheses simultaneously and hence, if not corrected, the chances that we find erroneously that one result of the six is statistically significant (i.e. commit a Type I error) will be higher.[43] We therefore adjust these test statistics using the procedure described in Romano and Wolf (2005).[44] So, to determine whether the point estimates are statistically significant when we test multiple hypotheses simultaneously, as in this case, we will use the adjusted *p*-values.

We observe similar point estimates of around 2–3pp for all three ownership outcomes: ownership of a toilet, ownership of a functioning toilet and ownership of an improved toilet. From baseline levels of 37%, 36% and 32%, respectively, these represent increases in toilet ownership of 5–9%.[45] It should be noted, however, that this coefficient is never statistically significant: the adjusted *p*-values in the third row all lie above the 0.10 cut-off. The fact that the observed impact is never statistically significant is unsurprising given that our study was designed to detect a minimum change in toilet ownership variables of 8–9pp (for details, see Abramovsky, Augsburg and Oteiza (2015)).

[42] In other words, the likelihood that we conclude that there is a non-zero impact of the intervention when this is not the case (what statisticians call Type I error) will be 10%. This is often referred to as statistical significance at the 10% level, and there are similar interpretations for the 5% and 1% levels (*p*-values of 0.05 and 0.01).

[43] For example, if we were testing 100 hypotheses at the 10% level, where the true effect was zero in all cases, we can still expect to find statistically significant impacts in 10 of these cases, according to the individual, unadjusted *p*-values.

[44] It is common practice to adjust standard errors for groups of variables tested, such as the group of toilet ownership and OD behaviour variables presented in this table.

[45] Baseline averages (‘control mean’) are shown in the bottom panel of Table 7.1.

Table 7.1. CLTS impacts on toilet ownership and OD: whole sample

	Toilet ownership			Toilet usage		
Outcome =1 if:	HH owns a toilet	HH owns a functioning toilet	HH owns a functioning improved toilet	All members use toilet	Main respondent performs OD	At least one member performs OD
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Pooled estimates						
Treated (γ)	0.02	0.03	0.02	0.01	−0.03	−0.03
p-value (unadjusted)	(0.25)	(0.07)	(0.13)	(0.71)	(0.04)	(0.06)
p-value (adjusted)†	(0.43)	(0.18)	(0.30)	(0.69)	(0.12)	(0.18)
Panel B. Estimates by period						
Treated × Wave 2 (γ_2)	0.04	0.04	0.02	0.01	−0.04	
p-value (unadjusted)	(0.05)	(0.07)	(0.24)	(0.65)	(0.04)	
p-value (adjusted)†	(0.14)	(0.16)	(0.41)	(0.65)	(0.14)	
Treated × Wave 3 (γ_3)	0.01	0.03	0.03	0.04	−0.03	−0.04
p-value (unadjusted)	(0.63)	(0.15)	(0.15)	(0.13)	(0.09)	(0.06)
p-value (adjusted)†	(0.64)	(0.33)	(0.33)	(0.33)	(0.25)	(0.19)
Treated × Wave 4 (γ_4)	0.01	0.03	0.02	−0.03	−0.03	−0.03
p-value (unadjusted)	(0.80)	(0.23)	(0.33)	(0.24)	(0.22)	(0.16)
p-value (adjusted)†	(0.80)	(0.54)	(0.54)	(0.54)	(0.54)	(0.45)
Control mean (wave 1)	0.37	0.36	0.32	0.93	0.63	0.63
Number of TUs	246	246	246	215	246	246
Number of households	4,166	4,166	4,166	1,420	4,232	4,259
Number of observations	12,497	12,497	12,497	4,260	12,697	8,518

† Adjusted *p*-values account for multiple hypothesis testing following the procedure described in Romano and Wolf (2005), with 1,000 clustered bootstrap samples.
Note: All specifications include a control for the value of the outcome variable at wave 1, household controls, LGA fixed effects and survey wave fixed effects. Controls include age, age squared, gender, employment status and level of education attainment of household head, household size and relative asset wealth, as well as a dummy variable indicating farming as the household's main economic activity. All controls are measured at baseline. Standard errors are clustered at the unit of randomisation (TU) level. *p*-values are given in parentheses. Stars indicate statistical significance: $*p < 0.10$, $**p < 0.05$ and $***p < 0.01$, according to the adjusted *p*-values.

Columns 4–6 in Table 7.1 present CLTS impact estimates on the set of outcomes related to sanitation behaviour: toilet usage (conditional on ownership), OD reported by the main respondent, and OD reported by any member of the household over the age of 5. Column 4 in panel A shows that CLTS had no statistically significant impact on toilet usage among households that owned toilets. Toilet usage is high in our sample among households with toilets, so this finding is not surprising. No impacts are detected either for OD by the main respondent (column 5) or any member of the household over the age of 5 (column 6).

The impacts presented in panel A of Table 7.1 are averaged over our three post-treatment surveys. For the purposes of understanding the dynamics of CLTS impacts, we can look at the evolution of impacts over time using a regression model such as specification (2) in Section 4.4. In this framework, CLTS impacts are estimated separately for each post-treatment survey wave. Panel B of Table 7.1 summarises the results of such analysis. For example, the first row, labelled 'Treated × Wave 2 (γ_2)', shows the impacts of CLTS as measured in the first post-CLTS survey, wave 2. As in the case for pooled estimates discussed above, no impacts are detected for any outcome, in any period, once we account for the fact that we are testing multiple hypotheses simultaneously.

7.4 Heterogeneous impacts

The results presented above summarise the overall CLTS impacts in our study areas. In this section, we delve into whether certain groups of people or households were more or less affected by the CLTS intervention.

Margins of CLTS impact

CLTS could reduce OD by increasing the construction of toilets that are subsequently used and/or by changing the OD behaviour of households that already own toilets. Similarly, CLTS could incentivise non-toilet-owning households to construct toilets and/or persuade toilet-owning household to improve and maintain their existing sanitation facilities. To gain an understanding of the importance of these two potential margins of CLTS impact, we split our sample according to whether households owned a functioning toilet or not in wave 1, prior to CLTS implementation. Table 7.2 presents the results for the sample of households that did not own toilets and Table 7.3 does the same for those that did.

Table 7.2. CLTS impacts on toilet ownership and OD: households with no functioning toilet in wave 1

	Toilet ownership			Toilet usage		
Outcome =1 if:	HH owns a toilet	HH owns a functioning toilet	HH owns a functioning improved toilet	All members use toilet	Main respondent performs OD	At least one member performs OD
	(1)	(2)	(3)	(4)	(5)	(6)
Pooled estimates						
Treated (γ)	0.04	0.05*	0.04	−0.00	−0.05*	−0.05
p-value (unadjusted)	(0.16)	(0.04)	(0.09)	(0.94)	(0.03)	(0.04)
p-value (adjusted)†	(0.27)	(0.10)	(0.18)	(0.93)	(0.08)	(0.11)
Control mean (wave 1)	0.01	0.00	0.00	0.00	0.97	0.97
Number of TUs	245	245	245	244	245	245
Number of households	2,666	2,666	2,666	951	2,728	1,811
Number of observations	7,998	7,998	7,998	2,853	8,185	5,432

† Adjusted *p*-values account for multiple hypothesis testing following the procedure described in Romano and Wolf (2005), with 1,000 clustered bootstrap samples.
Note: All specifications include a control for the value of the outcome variable in wave 1, household controls, LGA fixed effects and survey wave fixed effects. Controls include age, age squared, gender, employment status and level of education attainment of household head, household size and relative asset wealth, as well as a dummy variable indicating farming as the household's main economic activity. All controls are measured at baseline. Standard errors are clustered at the unit of randomisation (TU) level. *p*-values are given in parentheses. Stars indicate statistical significance: *$p < 0.10$, **$p < 0.05$ and ***$p < 0.01$, according to the adjusted *p*-values.

Table 7.2 holds a number of findings of interest. Columns 1–3 show that while no significant impact is found on the ownership of any toilet, CLTS increased the likelihood of owning a functioning toilet by 5pp. This means, for instance, that among the subset of households that did not own a functioning toilet in wave 1, those assigned to CLTS were 5pp more likely to own one than those assigned to the control group, during the post-treatment survey waves (2, 3 and 4). No detectable effect is found on ownership of an improved toilet, however.

Column 4 of Table 7.2 shows that CLTS did not affect usage among the households that constructed new toilets after wave 1. In other words, the number of households that report using their toilets, as a share of the number of households with toilets, is not significantly different between the CLTS and control groups. OD by the main respondent fell by a statistically significant 5pp, while no impact is found for the third measure, which looks at OD over the whole household. Note that the two significant impacts presented in the table – namely, the increase in ownership of a functioning toilet and the reduction in OD by the main respondent – are of the same magnitude.

Table 7.3. CLTS impacts on toilet ownership and OD: households with functioning toilets in wave 1

	Toilet ownership			Toilet usage		
Outcome =1 if:	HH owns a toilet	HH owns a functioning toilet	HH owns a functioning improved toilet	All members use toilet	Main respondent performs OD	At least one member performs OD
	(1)	(2)	(3)	(4)	(5)	(6)
Pooled estimates						
Treated (γ)	0.00	0.01	0.02	0.01	−0.02	−0.01
p-value (unadjusted)	(0.69)	(0.41)	(0.30)	(0.72)	(0.29)	(0.53)
p-value (adjusted)†	(0.91)	(0.80)	(0.72)	(0.91)	(0.72)	(0.88)
Control mean (wave 1)	1.00	1.00	0.90	0.93	0.01	0.02
Number of TUs	215	215	215	215	215	215
Number of households	1,500	1,500	1,500	1,416	1,543	1,025
Number of observations	4,499	4,499	4,499	4,249	4,629	3,076

† Adjusted *p*-values account for multiple hypothesis testing following the procedure described in Romano and Wolf (2005), with 1,000 clustered bootstrap samples.
Note: All specifications include a control for the value of the outcome variable in wave 1, household controls, LGA fixed effects and survey wave fixed effects. Controls include age, age squared, gender, employment status and level of education attainment of household head, household size and relative asset wealth, as well as a dummy variable indicating farming as the household's main economic activity. All controls are measured at baseline. Standard errors are clustered at the unit of randomisation (TU) level. *p*-values are given in parentheses. Stars indicate statistical significance: $*p < 0.10$, $**p < 0.05$ and $***p < 0.01$, according to the adjusted *p*-values.

Table 7.3 conducts the same analysis on the subsample of households that already owned a functioning toilet when surveyed in wave 1. It shows that there were no detectable effects of the intervention on this group: coefficient estimates are very small and insignificant. In other words, CLTS did not affect toilet ownership, maintenance, usage or OD among toilet owners.

Taken together, Tables 7.2 and 7.3 show that, in the context of this study, the impacts CLTS had on reducing OD were closely tied to increases in ownership of functioning toilets. This is supported by the fact that by the time of our last data collection wave, only 5% of households who owned a functioning toilet declared that they have at least one household member performing OD at least sometimes. On the contrary, 96% of the households with no functioning toilet declared that they practise OD. This means that 95% of households who perform OD do not own toilets. That is, there is a high correlation between toilet ownership and toilet usage in Nigeria. The correlation between toilet ownership and everybody in the household using the toilet is 0.95. In other words, CLTS affected only households with no functioning toilet before the intervention, and the magnitude of its impact on toilet ownership was almost identical to the magnitude of its impact on OD.

CLTS impacts by household and household head characteristics

Abramovsky, Augsburg and Oteiza (2016a) documented how, by wave 2, CLTS appeared to have a stronger impact on a particularly vulnerable set of households. Among them were households whose heads were either women or not highly educated (incomplete primary school or illiterate). Programme impacts were also exclusively concentrated among poorer households (those with below-median wealth levels at baseline), households with children under the age of 5, households with seniors (over 65 years old) and households with no debts or no savings (i.e. no access to channels of formal or informal financial services). We interpreted this as evidence that CLTS had been most effective in mobilising toilet construction and usage among vulnerable households. Importantly, these results did not account for multiple hypothesis testing.

In this subsection, we build on this previous work in two ways. First, we correct standard errors and *p*-values for the fact that we are running multiple hypotheses and that, as a result, the individual unadjusted significance tests are not valid. Second, we include results from wave 4 and test whether any of these impacts in selected populations are still present then. For simplicity, we focus our discussion on a single outcome: ownership of a functioning toilet. As discussed in Abramovsky, Augsburg and Oteiza (2016a), and in Section 7.3, CLTS had impacts of similar magnitudes on all main outcomes of interest.

Table 7.4. CLTS impact on toilet ownership by household characteristics

Outcome = 1 if HH owns a functioning toilet	**Restricted to households with the following characteristics at wave 1:**			
	HH head not completed primary education (1)	HH head is female (2)	Children <5 years old in HH (3)	Seniors >65 years old in HH (4)
Panel A. Pooled estimates				
Treated (γ)	0.06*	0.04*	0.05*	0.06**
p-value (unadjusted)	(0.03)	(0.09)	(0.02)	(0.01)
p-value (adjusted)†	(0.06)	(0.09)	(0.06)	(0.02)
Panel B. Estimates by period				
Treated × Wave 2 (γ_2)	0.08**	0.06*	0.05*	0.07*
p-value (unadjusted)	(0.01)	(0.02)	(0.06)	(0.02)
p-value (adjusted)†	(0.02)	(0.06)	(0.06)	(0.06)
Treated × Wave 3 (γ_3)	0.06	0.04	0.04	0.07*
p-value (unadjusted)	(0.03)	(0.11)	(0.23)	(0.02)
p-value (adjusted)†	(0.17)	(0.21)	(0.23)	(0.08)
Treated × Wave 4 (γ_4)	0.03	0.00	0.00	0.05
p-value (unadjusted)	(0.30)	(0.92)	(0.03)	(0.07)
p-value (adjusted)†	(0.52)	(0.92)	(0.12)	(0.20)
Control mean (wave 1)	0.29	0.33	0.37	0.38
Number of TUs	239	237	238	236
Number of households	1,329	1,511	1,241	1,352
Number of observations	3,986	4,534	3,723	4,057

† Adjusted *p*-values account for multiple hypothesis testing following the procedure described in Romano and Wolf (2005), with 1,000 clustered bootstrap samples.
Note: All specifications include household controls, LGA fixed effects and survey wave fixed effects. Controls include ownership of a functioning toilet in wave 1, age, age squared and employment status of the household head, household size and a dummy variable indicating farming as the household's main economic activity. All controls are measured at baseline. Standard errors are clustered at the unit of randomisation (TU) level. *p*-values are given in parentheses. Stars indicate statistical significance: $*p < 0.10$, $**p < 0.05$ and $***p < 0.01$, according to the adjusted *p*-values.

Table 7.4 shows the results of estimating either pooled (panel A) or period-specific (panel B) treatment effects of CLTS on the probability of owning a functioning toilet. Each column shows the results from restricting the sample to a specific group of interest. For example, column 1 shows the impact of CLTS on toilet ownership for households whose head had not completed primary education at baseline. We see positive and significant programme impacts among this group, where CLTS households had 6pp higher ownership rates than their counterparts in the control group, when looking at the pooled estimate (compared with 3pp, not statistically significant, for the whole sample). Looking at panel B, we see that impacts by period were concentrated in the short term: CLTS households are 8pp more likely than control households to own a functioning toilet by wave 2, and no statistically significant impacts are detected as of wave 3. These findings mirror those discussed in Abramovsky, Augsburg and Oteiza (2016a), with the sole difference being the reduced level of significance of the coefficients, brought about by using adjusted *p*-values.

This pattern of findings is similar for other subsamples considered. In column 2 of Table 7.4, we see pooled impacts of 4pp, significant at the 10% level, for the households with female heads. Stronger impacts are shown for wave 2, but these cease to be significant by wave 3. Column 3 shows statistically significant impacts from CLTS of 5pp among households with young children, for both pooled estimates and wave 2. These also fade by wave 3. In column 4, we observe that CLTS households with seniors show statistically significantly higher levels of functioning toilet coverage than households with seniors in control areas during waves 2 and 3, but once again these become statistically indistinguishable from zero by wave 4.

These findings support the key message in our previous reports that the CLTS intervention seems to have mobilised more vulnerable groups to construct and use toilets in the short term. However, as of wave 3, these types of households in control areas also constructed toilets, and eventually caught up with those in CLTS areas, leaving no significant intervention impacts almost 3 years after intervention implementation.

Shifting our attention to economic variables at the household level, Table 7.5 performs a similar exercise, now splitting households according to whether they report having any savings or any debts and according to their level of wealth. CLTS impacts of similar magnitude to those observed for the whole sample can be observed for households with no savings. These again fade out after wave 2. A similar pattern, with a slightly higher impact estimate of 4pp, can be seen in column 2 among households with no debts. Stronger programme impacts, almost twice the magnitude of those for the whole sample, are observed among the subsample of asset-poor households, as shown in column 3. As mentioned in Section 4.3, wealth is measured by a composite index constructed using the answers to a series of questions on ownership of certain durable assets, at wave 1.[46] The poorest half of our sample in wave 1 shows CLTS impacts of 7pp in wave 2, but these become statistically insignificant by wave 3.

[46] The set of survey questions used to create the relative asset wealth index used throughout this report can be found in Table A.1 in the appendix.

Table 7.5. CLTS impact on toilet ownership by economic situation of the household

Outcome =1 if HH owns a functioning toilet	**Restricted to households with the following characteristics at wave 1:**		
	HH has no savings (1)	HH has no debts (2)	HH has < median wealth (3)
Panel A. Pooled estimates			
Treated (γ)	0.03*	0.04*	0.06**
p-value (unadjusted)	(0.08)	(0.05)	(0.01)
p-value (adjusted)†	(0.08)	(0.06)	(0.04)
Panel B. Estimates by period			
Treated × Wave 2 (γ_2)	0.04*	0.04*	0.07**
p-value (unadjusted)	(0.06)	(0.04)	(0.03)
p-value (adjusted)†	(0.06)	(0.05)	(0.05)
Treated × Wave 3 (γ_3)	0.03	0.04	0.07
p-value (unadjusted)	(0.14)	(0.11)	(0.05)
p-value (adjusted)†	(0.15)	(0.15)	(0.11)
Treated × Wave 4 (γ_4)	0.02	0.03	0.04
p-value (unadjusted)	(0.31)	(0.19)	(0.13)
p-value (adjusted)†	(0.31)	(0.29)	(0.29)
Control mean (wave 1)	0.32	0.37	0.21
Number of TUs	246	246	245
Number of households	3,182	3,281	2,060
Number of observations	9,547	9,844	6,181

† Adjusted *p*-values account for multiple hypothesis testing following the procedure described in Romano and Wolf (2005), with 1,000 clustered bootstrap samples.
Note: All specifications include household controls, LGA fixed effects and survey wave fixed effects. Controls include ownership of a functioning toilet in wave 1, age, age squared and employment status of the household head, household size and a dummy variable indicating farming as the household's main economic activity. All controls are measured at baseline. Standard errors are clustered at the unit of randomisation (TU) level. *p*-values are given in parentheses. Stars indicate statistical significance: *$p < 0.10$, **$p < 0.05$ and ***$p < 0.01$, according to the adjusted *p*-values.

This set of findings suggests that CLTS is most effective among a specific group of households that could be thought of as the most vulnerable. These households are vulnerable in several dimensions: their household heads have no education at all (primary school not completed), they have children and senior members (making them more susceptible to disease) and they are relatively less wealthy than the rest of the households in the sample. Consequently, they are also less likely to have access to financial instruments (no savings, no debts). Nonetheless, almost 3 years after the intervention, none of the effects are observed, and CLTS and control households exhibit statistically indistinguishable rates of toilet ownership. No long-term effects from CLTS are observed along these dimensions, suggesting that we should interpret these results carefully. While it seems that this particularly vulnerable group reacted to CLTS by increasing toilet construction and usage, control households caught up over time, implying that no detectable difference is observed two years after the triggering activities happened.

CLTS impacts by community characteristics

Given that a participatory intervention such as CLTS relies on community mobilization, it is of interest to explore whether community characteristics are important drivers of intervention success. In this subsection, we study whether this was the case using the three dimensions described in Section 4.3 and postulated as important intervention mediators by the creators of CLTS: community-level wealth, social capital and religious fragmentation. It is hypothesised that CLTS might be more effective in poor, socially cohesive and homogeneous communities, which is what these variables proxy for, respectively. Take the first of these three indicators: community level wealth. Community wealth tells something about community characteristics (that is proxied combining data from all households sampled in a community); whereas household level wealth tells something about the household and its ability to pay or their behaviour, regardless of the community the household lives in. It is possible that individual wealth does not mediate the impact of the intervention but community level wealth does. Such a case would support the hypothesis that relatively poor and relatively rich households can both benefit from CLTS, if they live in a poor community, but neither type of household benefits if they live in a relatively rich community. This is not to say that relatively poor households in relatively rich communities might not benefit from interventions, including those that might alleviate any resource or other constraints these households face in improving their sanitation behaviour.

Table 7.6 presents the results of this analysis. We remind the reader that all these measures are taken in wave 1.

Table 7.6. CLTS impact on toilet ownership by characteristics of the TU

Outcome: =1 if HH owns a functioning toilet	**Asset wealth (1)**	**Social capital score (2)**	**Religious fragmentation (3)**
Pooled estimates			
(a) Treated x High (γ)	–0.02	0.05*	0.02
p-value (unadjusted)	(0.22)	(0.04)	(0.46)
p-value (adjusted)†	(0.36)	(0.10)	(0.46)
(b) Treated x Low (γ)	0.10***	0.01	0.04
p-value (unadjusted)	(0.00)	(0.68)	(0.08)
p-value (adjusted)†	(0.00)	(0.68)	(0.15)
(c) Difference (b – a)	0.12***	–0.04	0.03
p-value (unadjusted)	(0.00)	(0.24)	(0.45)
p-value (adjusted)†	(0.00)	(0.41)	(0.45)
Control mean (wave 1), high	0.45	0.32	0.36
Control mean (wave 1), low	0.24	0.40	0.36
Number of TUs	246	246	246
Number of households	4,166	4,166	4,166
Number of observations	12,497	12,497	12,497

† Adjusted *p*-values account for multiple hypothesis testing following the procedure described in Romano and Wolf (2005), with 1,000 clustered bootstrap samples.
Note: All specifications include household controls, LGA fixed effects and survey wave fixed effects. Controls include ownership of a functioning toilet in wave 1, age, age squared, gender, employment status and level of education attainment of household head, household size and relative asset wealth, as well as a dummy variable indicating farming as the household's main economic activity. All controls are measured at baseline. Standard errors are clustered at the unit of randomisation (TU) level. *p*-values are given in parentheses. Stars indicate statistical significance: *$p < 0.10$, **$p < 0.05$ and ***$p < 0.01$, according to the adjusted *p*-values.

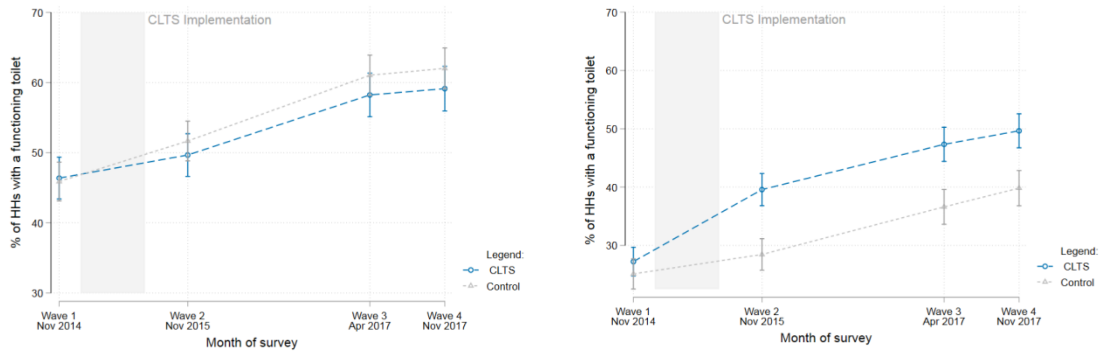
We focus here on a single outcome: ownership of a functioning toilet. The three columns of the table study heterogeneous CLTS impacts along the three dimensions described above: asset wealth, social capital and fragmentation. Row (a) shows the estimates of CLTS impact on the sample with high scores on each measure (s.a. above median asset wealth level), while row (b) shows the same for the sample with low values. Row (c) shows the difference between the coefficients for the two samples. From column 1, we see that CLTS had no detectable impact on toilet ownership among households living in rich TUs (i.e. TUs with above-median wealth levels). The point estimate for this group is –0.02 and not statistically significant. Among households in poor TUs, however, CLTS appears to have had an impact of 10pp, significant at the 1% level even after adjusting for multiple hypothesis testing. The difference between the two point-estimates is also large, 12pp, and significant. This stark difference in CLTS effectiveness is even more striking when we consider the baseline toilet coverage rates in these two samples. The first row of the table's bottom panel shows that 45% of the households in rich TUs owned toilets at baseline, meaning there was still a significant margin for growth in this group. Toilet ownership was 24% in poorer TUs.

Columns 2 and 3 of Table 7.6 conduct a similar analysis using social capital and religious fragmentation, both measured at the TU level, again considering above and below median values for each. Column 2 suggests that CLTS had an impact of 5pp among households living in communities with high levels of social capital, and no impact in areas with low levels of this index, consistent with the findings presented in Cameron, Olivia and Shah (2015). Nonetheless, in our case, the evidence is not extremely robust: the difference between the coefficients for these two groups is small, just 4pp, and not statistically significant. In other words, our study does not allow us to reject the hypothesis that CLTS impacts were identical in both groups. Finally, in column 3, we see that communities with low levels of religious fragmentation (more homogeneous) experienced CLTS impacts that are statistically indistinguishable from those of their counterparts in more fragmented communities, and neither of these groups seems to be associated with significant CLTS impacts.

The results presented in column 1 of Table 7.6 motivate a closer look at the evolution of toilet ownership along TU wealth lines. Figure 7.4 plots the unconditional trends in the ownership of functioning toilets by CLTS treatment group in rich TUs (panel (a)) and poor TUs (panel (b)). Panel (a) shows that toilet ownership in the two experimental groups in this subsample did not differ substantially over time. In fact, households in the CLTS group appear, on average, less likely to own a functioning toilet than those in the control group, although this difference is not statistically significant, as shown by the vertical lines representing 95% confidence intervals. Panel b presents a very different picture. Households in the CLTS group that live in poor communities experienced a large increase in toilet ownership between waves 1 and 2 compared with control households, and this gap persisted throughout the length of our study.

Figure 7.4. Ownership of functioning toilets by community wealth level

(a) Communities with above-median wealth | **(b) Communities with below-median wealth**



Note: The vertical lines represent the 95% confidence intervals. The scale of the x-axis corresponds to the amount of time elapsed between each survey wave. The CLTS intervention was carried out shortly after wave 1, between January and June 2015, as indicated by the grey bar.

Source: Household survey, waves 1–4.

Having established that communities with low levels of asset wealth were associated with higher CLTS impacts on ownership of functioning toilets, we might ask whether this pattern of starkly different CLTS effectiveness across TU wealth levels is observed for the rest of the outcomes as well. Table 7.7 presents the results of CLTS impact estimations by TU wealth level, for each of our six outcomes of interest. It includes *p*-values that are adjusted for the fact that we are testing six hypotheses simultaneously.

Starting in columns 1–3, we see that CLTS was ineffective in shifting toilet ownership rates among rich TUs. On the other hand, CLTS had strongly significant impacts of 8pp, 10pp and 7pp on the probability of owning a toilet, a functioning toilet and an improved toilet respectively among poor TUs. Row (c) shows that these coefficients are significantly different from the ones estimated for the rich TU sample, confirming that CLTS had differential impacts in these two types of communities.

Table 7.7. CLTS impact on toilet ownership and OD, by TU wealth group

	Toilet ownership			Toilet usage		
Outcome =1 if :	HH owns a toilet	HH owns a functioning toilet	HH owns a functioning improved toilet	All members use toilet	Main respondent performs OD	At least one member performs OD
	(1)	(2)	(3)	(4)	(5)	(6)
Pooled estimates						
a) Treated x Rich TU (γ)	−0.03	−0.02	−0.01	−0.03	0.01	0.02
p-value (unadjusted)	(0.10)	(0.21)	(0.58)	(0.16)	(0.51)	(0.41)
p-value (adjusted)†	(0.28)	(0.49)	(0.81)	(0.30)	(0.81)	(0.75)
b) Treated x Poor TU (γ)	0.08***	0.10***	0.07**	0.08***	−0.09***	−0.10***
p-value (unadjusted)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
p-value (adjusted)†	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
c) Difference (b – a)	0.11***	0.12***	0.08**	0.11***	−0.10***	−0.11***
p-value (unadjusted)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)
p-value (adjusted)†	(0.00)	(0.00)	(0.03)	(0.00)	(0.01)	(0.00)
Control mean (wave 1), rich	0.47	0.46	0.44	0.43	0.52	0.52
Control mean (wave 1), poor	0.25	0.25	0.20	0.22	0.74	0.74
Number of TUs	246	246	246	246	246	246
Number of households	4,166	4,166	4,166	4,166	4,232	4,259
Number of observations	12,497	12,497	12,497	12,497	12,697	8,518

† Adjusted *p*-values account for multiple hypothesis testing following the procedure described in Romano and Wolf (2005), with 1,000 clustered bootstrap samples.
Note: All specifications include household controls, LGA fixed effects and survey wave fixed effects. Controls include the value of the outcome variable in wave 1, age, age squared, gender, employment status and level of education attainment of household head, household size and relative asset wealth, as well as a dummy variable indicating farming as the household's main economic activity. All controls are measured at baseline. Standard errors are clustered at the unit of randomisation (TU) level. *p*-values are given in parentheses. Stars indicate statistical significance: *$p < 0.10$, **$p < 0.05$ and ***$p < 0.01$, according to the adjusted *p*-values.

Columns 4–6 in Table 7.7 show the results for the three outcomes of toilet usage. Column 4 indicates that toilets in poor communities are significantly more likely to be used by all household members at endline: we estimate a coefficient of 0.08, implying that at endline 8pp more households report that all members use their toilet. This finding is significant at the 1% level even after adjusting our *p*-values for multiple hypothesis testing. Columns 5 and 6 mirror the findings discussed for the toilet ownership outcomes: no impacts among rich TUs, but strong and significant reductions of 9–10pp in OD among households living in poor TUs. These impacts, and the difference between them and the impacts estimated for households in rich TUs, are significant at the 1% level, even after adjusting for multiple hypothesis testing.

So CLTS had strong impacts in poor TUs on all of the main outcomes of interest. One question remains to be explored, though. Tables 7.6 and 7.7 presented the results for pooled CLTS impacts, which are averaged over all post-treatment surveys. It is also of interest to study how these impacts evolved over time. Were they sustained over the course of the 3 years after the CLTS interventions, or were they short-lived? Figure 7.4 provides a first pictorial assessment, which is formalised in results presented in Table 7.8. , which shows period-specific treatment effects. It performs the same analysis as panel B of Table 7.1, but restricting the sample to those households living in poor TUs only, since this is the population where CLTS had the strongest impact.[47]

We focus first on columns 1–3 of Table 7.8, which show the results for toilet ownership outcomes. Interestingly, the pattern of CLTS impacts over time varies by outcome. For example, ownership of a toilet of any kind and in any condition increased strongly in the first survey wave (column 1), less than a year after CLTS triggering meetings. At that point, CLTS households were 12pp more likely to own a toilet than households in control areas. But by wave 3, this difference became statistically insignificant. On the other hand, ownership of a functioning toilet is 10–11pp higher in all survey waves in CLTS areas, and this is significant at least at the 5% level. To verify that the point estimates are indeed different between columns 1 and 2, we conduct a Wald test of equality of coefficients and find that while the coefficients for waves 2 and 3 may be identical, those for wave 4 are statistically different at the 10% level. Finally, improved toilet ownership seems to have evolved in the inverse way: small and barely significant differences between CLTS and control areas in waves 2 and 3, and significant impacts only by wave 4. The only coefficients for which we can reject equality between columns 2 and 3 are those corresponding to wave 2.

This first set of findings from Table 7.8 paints a potential picture of the dynamics of CLTS impacts in poor communities. In the short term, right after the triggering meetings, ownership of toilets of any kind and in any condition increased. The impacts on ownership of improved toilets (column 3) were at this point (wave 2) smaller and less statistically significant that the impacts on any toilet or functioning toilet (columns 1 and 2). By the end of the study period, almost 3 years after the intervention started, two different patterns emerge. On the one hand, we find no statistically meaningful difference in ownership of any type of toilets (this is shown in column 1, which includes whether a toilet is functioning or not; or whether the toilet is improved or not) between CLTS and control areas. At the same time, functioning toilets were 10pp more prevalent in CLTS areas during both waves 3 and 4 and by wave 4, households in CLTS areas were 8pp more likely to own functioning improved toilets than control areas. These results suggest that CLTS induced the construction of more sustainable toilets in the first place, and/or induced households to maintain their (newly) constructed toilets better. .

[47] Table A.2 in the appendix shows the results for rich TUs.

Table 7.8. CLTS impacts on toilet ownership and OD, by period: poor TUs

	Toilet ownership			Toilet usage		
Outcome =1 if:	HH owns a toilet (1)	HH owns a functioning toilet (2)	HH owns a functioning improved toilet (3)	All members use toilet (4)	Main respondent performs OD (5)	At least one member performs OD (6)
Estimates by period						
Treated × Wave 2 (γ_2)	0.12***	0.11***	0.06*	0.12***	−0.09**	
p-value (unadjusted)	(0.00)	(0.00)	(0.04)	(0.00)	(0.01)	
p-value (adjusted)†	(0.00)	(0.00)	(0.07)	(0.00)	(0.02)	
Treated × Wave 3 (γ_3)	0.07	0.10**	0.07	0.10**	−0.10***	−0.10**
p-value (unadjusted)	(0.06)	(0.01)	(0.06)	(0.00)	(0.00)	(0.00)
p-value (adjusted)†	(0.15)	(0.02)	(0.15)	(0.01)	(0.01)	(0.01)
Treated × Wave 4 (γ_4)	0.07	0.10**	0.08**	0.04	−0.09**	−0.10**
p-value (unadjusted)	(0.07)	(0.01)	(0.01)	(0.02)	(0.01)	(0.00)
p-value (adjusted)†	(0.14)	(0.03)	(0.03)	(0. 153)	(0.03)	(0.01)
Control mean (wave 1)	0.25	0.24	0.19	0.34	0.75	0.75
Number of TUs	123	123	123	123	123	123
Number of households	2,045	2,045	2,045	2,045	2,060	2,073
Number of observations	6,136	6,136	6,136	6,546	6,182	4,174

† Adjusted *p*-values account for multiple hypothesis testing following the procedure described in Romano and Wolf (2005), with 1,000 clustered bootstrap samples.
Note: All specifications include household controls, LGA fixed effects and survey wave fixed effects. Controls include the value of the outcome variable in wave 1, age, age squared, gender, employment status and level of education attainment of household head, household size and relative asset wealth, as well as a dummy variable indicating farming as the household's main economic activity. All controls are measured at baseline. Standard errors are clustered at the unit of randomisation (TU) level. *p*-values are given in parentheses. Stars indicate statistical significance: *$p < 0.10$, **$p < 0.05$ and ***$p < 0.01$, according to the adjusted *p*-values.

Columns 4–6 of Table 7.8 show period-specific estimates of the impact on the three toilet usage outcomes. Column 4 shows that usage, increased by 12pp in wave 2, an impact that sustained into wave 3 but was not detectable anymore by the time wave 4 was fielded. Given that usage was extremely high before the intervention (91% when conditioning on ownership), it is likely that further improvements in this dimension were just too difficult to achieve from wave 2 onwards. Columns 5 and 6 show a stable pattern of sustained CLTS impacts on OD practice. Both the main respondent and the whole household are 9–10pp less likely to report performing OD in CLTS households in all survey waves. This result adds to the important result mentioned above: the early (wave 2) impacts achieved for ownership of functioning toilets and OD reduction are sustained and do not show any evidence of faltering.

Community wealth and pre-intervention toilet coverage

We have seen so far that CLTS treatment effects are concentrated among households with no toilets at baseline and among households living in poor TUs. As it happens, the TU wealth is correlated with TU average toilet ownership at baseline. As seen in Figure 7.5, before the intervention occurred, toilet coverage increases with TU-level wealth. Among the poorest 25% of TUs, the average toilet coverage rate was below 20%, while it was above 50% for the wealthiest 25%. A question then naturally arises: what characteristic of the TUs is more correlated with high CLTS impacts – wealth or baseline toilet coverage?

Figure 7.5. Toilet ownership at the TU level by TU wealth quartile: wave 1

Note: Bars plot the average rate of toilet ownership at the TU level, by TU wealth quartile. Quartile 1 includes the poorest 25% of TUs, while quartile 4 includes the richest 25%. Both wealth and toilet ownership are measured in wave 1, before CLTS was implemented.

Source: Household survey, wave 1.

We provide evidence to answer this question in Table 7.9. As before, we estimate CLTS impacts for rich (row a) and poor TUs (row b) separately. The coefficient of most interest here is presented in row c: the difference in CLTS impacts between these two groups. We run this analysis over the whole sample (column 1) and separately according to the level of toilet coverage in the TU before the CLTS intervention began. The median level of toilet coverage back then was 30%, so we split the sample into two groups – TUs with 30% or lower toilet coverage and TUs with more than 30% toilet coverage. If baseline toilet coverage were the true driving force behind the CLTS impacts discussed above, then we should expect to see large and significant point estimates for both rich and poor TUs with low toilet coverage (column 2), and small or non-existent effects in TUs with higher toilet coverage (column 3).

Table 7.9. CLTS impacts by TU-level wealth and baseline toilet ownership

Outcome: =1 if HH owns a functioning toilet	**Whole sample (1)**	**TUs with low toilet ownership (2)**	**TUs with high toilet ownership (3)**
Pooled estimates			
a) Treated x Rich TU (γ)	−0.02	−0.03	−0.01
p-value (unadjusted)	(0.20)	(0.34)	(0.59)
p-value (adjusted)†	(0.50)	(0.56)	(0.59)
b) Treated x Poor TU (γ)	0.10***	0.10**	0.07**
p-value (unadjusted)	(0.00)	(0.01)	(0.03)
p-value (adjusted)†	(0.00)	(0.02)	(0.03)
c) Difference (b – a)	0.12***	0.13**	0.08**
p-value (unadjusted)	(0.00)	(0.01)	(0.02)
p-value (adjusted)†	(0.00)	(0.02)	(0.02)
Number of TUs	122	122	124
Number of households	2,137	2,137	2,028
Number of observations	6,412	6,412	6,085

† Adjusted *p*-values account for multiple hypothesis testing following the procedure described in Romano and Wolf (2005), with 1,000 clustered bootstrap samples.
Note: All specifications include household controls, LGA fixed effects and survey wave fixed effects. Controls include ownership of a functioning toilet in wave 1, age, age squared, gender, employment status and level of education attainment of household head, household size and relative asset wealth, as well as a dummy variable indicating farming as the household's main economic activity. All controls are measured at baseline. Standard errors are clustered at the unit of randomisation (TU) level. *p*-values are given in parentheses. Stars indicate statistical significance: $*p < 0.10$, $**p < 0.05$ and $***p < 0.01$, according to the adjusted *p*-values.

Our estimates suggest the opposite. Row a shows that among households in rich TUs, CLTS had no detectable impacts, regardless of the level of baseline toilet ownership. In row b, we see that households from poor TUs experienced statistically significant increases in toilet ownership, again regardless of their TU's baseline toilet coverage. The point estimate is slightly lower for TUs with high toilet ownership, at 7pp instead of 10pp, but this is nevertheless large and significant. Finally, in row c, we see that the difference in impacts between households from poor TUs and those from rich TUs is large and significant among both low (column 2) and high (column 3) baseline coverage groups. In essence, Table 7.9 suggests that, while baseline toilet coverage may play a role in mediating CLTS impacts (as evidenced by the smaller coefficients for poor TU households with high toilet ownership), the role of TU wealth is important. TU wealth is correlated with CLTS effectiveness above and beyond what can be explained by baseline differences in toilet ownership, and seems to play a key role in mediating CLTS impacts.

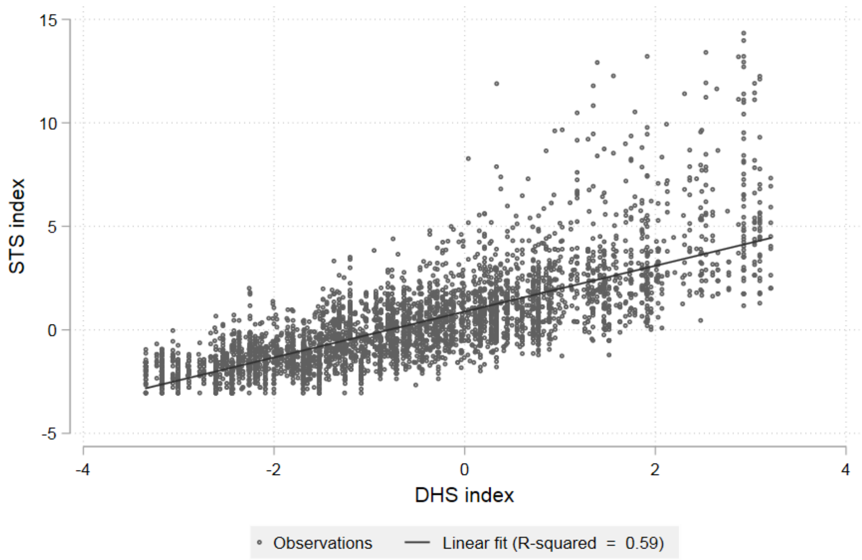
How do these poor households compare with households in Nigeria as a whole?

We have established that households in the poorest half of the TUs in our sample experienced strong and sustained CLTS impacts, while those in the wealthier communities showed no sign of improvement. This is a policy-relevant finding, since it can inform future decisions about the suitability of CLTS roll-out in Nigeria and elsewhere. However, the measure of asset ownership used in our study is a relative ranking within study households. To further understand the characteristics of the households living in these poor communities, and how they compare with households in Nigeria as a whole, we conduct an analysis using our 2014/15 baseline survey of households and the 2013 Demographic and Health Survey (DHS) from Nigeria. The latter survey includes a module on asset ownership, general quality of the dwelling, and access to basic services such as drinking water and sanitation. Importantly, the DHS is representative at the national level, so it provides a suitable benchmark against which to compare the households in our survey.

To start with, we focus only on our survey data and construct a new asset wealth index (the 'DHS index'), using the sub-set of asset information collect in our survey, which is also asked about in the DHS survey.[48] Comparing this new index to the previously used data (based on more information) reveals a high correlation between the two wealth measures (ρ=0.77, significant at the 1% level). Figure 7.6 plots the scores for both indices for the households in the STS sample (the only ones for which both indices are available). The figure not only shows the positive correlation between the asset wealth indices, but also that the newly created DHS index explains up to 60% of the variation in the previously used STS index (R^2=0.59). The match is not perfect, but a reasonable fit considering that only some of the questions used to construct each index overlap. In other words, households with a high asset wealth score for our study STS index are likely to have a high asset wealth score using the DHS index constructed to compare with the DHS sample, and vice versa.

[48] The questions used for this second wealth index are shown in column 2 of Table A.1 in the appendix.

Figure 7.6. Asset wealth indices are highly correlated with each other



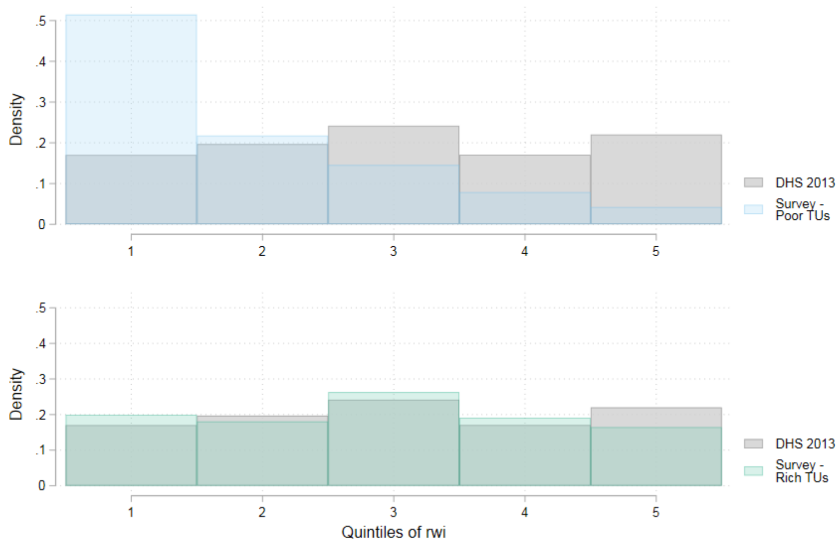
Note: Scatter plot showing the asset wealth scores obtained for each household in our study sample during wave 1. The y-axis shows the value for the asset wealth index used in our study, built using asset ownership questions from our household survey. The x-axis shows the score obtained in the asset wealth index constructed using only those questions that were included in both our household survey and the DHS, for the same sample of households. Each dot represents one household.

Source: Household survey, wave 1; Nigeria Demographic and Health Survey, 2013.

We then turn to the DHS survey data and split the households in the DHS sample into wealth quintiles, based on a wealth indicator using the same questions as for the DHS index discussed above. , We calculate the average DHS wealth index level for each quintile using this countrywide distribution. Then we assign the households in our survey to the countrywide corresponding quintiles using their DHS index score. The top panel of Figure 7.7 shows the distribution of the households in our sample living in poor TUs along the country-wide distribution of wealth (quintiles), and the bottom panel does the same for those households in our survey living in rich TUs. By construction, the DHS sample should be distributed uniformly over each quintile, but in reality the distribution is slightly uneven due to discrete values of the index.

Two important points are worth noting from Figure 7.7. First, the households in our sample who live in poor TUs are heavily concentrated in the bottom three quintiles of the Nigerian asset wealth index distribution, as shown by the blue histogram in the top panel. Half of them belong to the poorest 20% of Nigerian households. In this sense, the households living in poor TUs in our sample are generally poor compared with the national distribution. On the other hand, the households in our survey who live in the wealthiest 50% of TUs are much more evenly distributed over the five wealth quintiles, as seen in the green histogram of the bottom panel of Figure 7.7. In terms of asset wealth at least, households in rich TUs seem to be a fairly representative sample of Nigerian households and not particularly rich once the national distribution of wealth is considered.

Figure 7.7. Distribution of the households in our sample

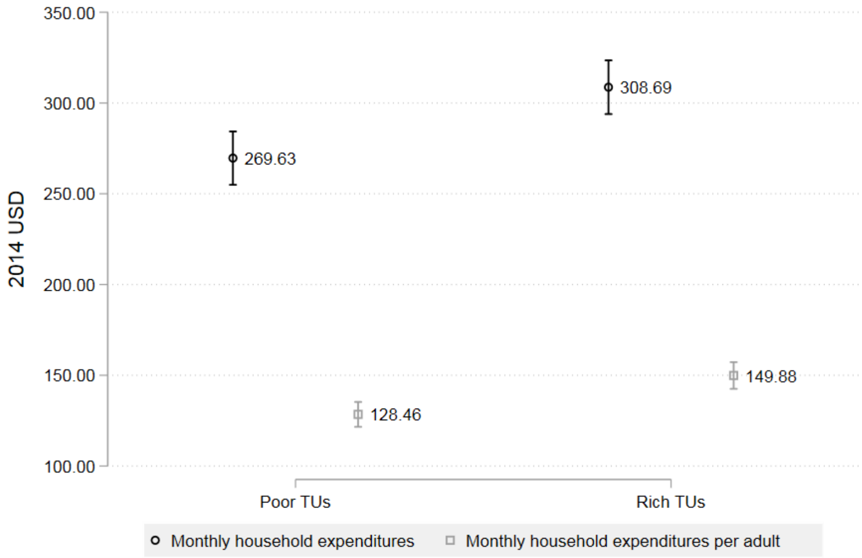


Note: Distribution of the households in our survey along wealth quintiles for the whole of Nigeria, estimated using the 2013 DHS (rwi = relative wealth index). The top panel plots households from our sample living in poor TUs (blue) and all the households from the DHS sample (grey). The bottom panel plots households from our sample living in rich TUs (green) and all the households from the DHS sample (grey). 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. These questions are included in Table A.1 of the appendix.

Source: Household survey, wave 1; Nigeria Demographic and Health Survey, 2013.

In the next section, we will discuss in more detail the distinction between household and community level wealth. To motivate that discussion, we first turn here to household level monthly expenditures reported by the households in our sample as part of our household survey in wave 1. Reported expenditures in household surveys are often more reliable than reported income in countries such as Nigeria for measuring total resources and purchasing power at the household level, given the multiple incentives respondents might have for under-reporting the latter and the problems with valuing own-production and agricultural incomes. In addition, consumption expenditures may give a better sense of long-term resources than income, if there is consumption smoothing. Nonetheless, our survey has a significant number of households that did not report expenditures. Of the 2,282 households interviewed in wave 1 who live in poor TUs, 569 (25%) did not report expenses. The share of missing values for households from rich TUs is very similar: 525 out of 2,440 (22%). A *t*-test indicates we cannot reject the hypothesis that the share of missing values is the same in the two samples, which is reassuring.

Figure 7.8. Average monthly household expenditures by TU wealth group



Note: Average household expenditures and expenditures per adult, expressed in US$ of December 2014. Standard errors used to construct the 95% confidence intervals (capped bars) account for clustering at the TU level.

Source: Household survey, wave 1.

Figure 7.8 plots the mean and 95% confidence intervals of two household expenditure variables for each TU wealth group (with poor TUs on the left and rich ones on the right). The first variable reports the average of total monthly household expenditures and the second reports the average of monthly household expenditures divided by the number of adults living in the household. Both are expressed in US$ of December 2014, the mid-point of survey wave 1 interviews. On average, households living in rich TUs spent around US$40 more per month or around US$20 more per month per adult than households in poor TUs. It is interesting that, given the significant, but not huge, difference in average household expenses between the two groups, CLTS had such different impacts in poor and rich TUs. This suggests that the CLTS's effectiveness is likely to be associated with community-level characteristics rather than household characteristics, and that it is unlikely to be linked to differential purchasing power at the household level. In other words, it suggests that effective CLTS targeting could be improved by focusing on community-level characteristics, particularly median wealth levels, rather than on household-level wealth. We analyse this matter in more detail in the next subsection.

Community- or household-level wealth?

As we saw above, CLTS impacts vary significantly by community-level wealth group (Table 7.7). At the same time, when looking at household-level wealth, by wave 3 no significant differences remain between CLTS and control households, even among those with below-median wealth, who initially showed significant CLTS impacts (column 3 of Table 7.5). One might point out, though, that poor TUs are likely to consist mostly of poor households, and vice versa. Indeed, 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. But although the measures of wealth at the TU and household levels are correlated with each other, it is important, for policy purposes, to understand which measure is more correlated with high CLTS impacts. In other words, should CLTS implementing agencies target poor communities or poor households?

To provide some evidence in this respect, we conducted a regression analysis of CLTS impacts using three possible categorisations of asset wealth: (i) using TU level wealth to split the sample into high and low wealth levels (referred to in the results table as 'TU'); (ii) using HH level wealth to split the sample (referred to as 'HH'); (iii) using HH level wealth and split the sample within each TU (referred to as 'HR'). This latter categorisation implies that a HH is poor or rich with respect to other households in the TU and not with respect to other household within the whole sample. The results of this analysis are presented in Table 7.10. We again concentrate on a single outcome – ownership of a functioning toilet – and pool all post-CLTS survey waves to estimate average CLTS impacts during our study period. The first two columns of the table reproduce the results already discussed above. In column 1, we see that CLTS had no impacts among rich TUs and impacts of 10pp among poor TUs, and that the difference in impacts between the two samples is large and strongly significant. Similarly, in column 2, we see that CLTS had no impacts among rich households (those with above-median asset wealth) and impacts of 6pp among poor households. Importantly, the difference in estimated coefficients for rich and poor households in column 2 is less than half the size of that estimated in column 1 and is not statistically significant. In short, CLTS impact estimates are more correlated with low TU wealth levels than with low household wealth levels. We discuss the interpretation of these results and why this could be the case in section 7.8.

Table 7.10. CLTS impacts on toilet ownership, by TU- and household-level wealth

Outcome =1 if HH owns a functioning toilet	**Whole sample**		**Poor TUs**		**Rich TUs only**	
Wealth measure used:	TU	HH	HH	HR	HH	HR
	(1)	(2)	(3)	(4)	(5)	(6)
Pooled estimates						
a) Treated x High (γ)	−0.02	0.01	0.07	0.09**	−0.01	−0.01
p-value (unadjusted)	(0.22)	(0.63)	(0.05)	(0.01)	(0.51)	(0.77)
p-value (adjusted)†	(0.53)	(0.87)	(0.17)	(0.04)	(0.82)	(0.87)
b) Treated x Low (γ)	0.10***	0.06*	0.12***	0.12***	−0.04	−0.04
p-value (unadjusted)	(0.00)	(0.02)	(0.00)	(0.00)	(0.18)	(0.11)
p-value (adjusted)†	(0.01)	(0.05)	(0.01)	(0.00)	(0.18)	(0.16)
c) Difference (b – a)	0.12***	0.05	0.05	0.03	−0.02	−0.03
p-value (unadjusted)	(0.00)	(0.06)	(0.16)	(0.28)	(0.41)	(0.26)
p-value (adjusted)†	(0.01)	(0.26)	(0.46)	(0.56)	(0.56)	(0.56)
Number of TUs	246	246	123	123	123	123
Number of households	4,166	4,166	2,045	2,045	2,120	2,120
Number of observations	12,497	12,497	6,136	6,136	6,361	6,361

† Adjusted *p*-values account for multiple hypothesis testing following the procedure described in Romano and Wolf (2005), with 1,000 clustered bootstrap samples.
Note: Heterogeneous CLTS impacts using three measures of asset wealth. TU – TU wealth is the median level of household asset wealth within each TU. TUs are then sorted by wealth and split into two identically sized groups, high (above-median TU wealth) and low (below median TU wealth). HH – households are sorted by asset wealth and split into two identically sized groups, high (above-median household wealth) and low (below median asset wealth). HR – within each TU, households are sorted by asset wealth. They are then split into two identically sized groups, high (higher asset wealth than the median for their TU) and low (lower asset wealth than the median for their TU). All specifications include household controls, LGA fixed effects and survey wave fixed effects. Controls include age, age squared, gender, employment status and level of education attainment of household head, household size and relative asset wealth, as well as a dummy variable indicating farming as the household's main economic activity. All controls are measured at baseline. Standard errors are clustered at the unit of randomisation (TU) level. *p*-values are given in parentheses. Stars indicate statistical significance: *$p < 0.10$, **$p < 0.05$ and ***$p < 0.01$, according to the adjusted *p*-values.

Columns 3–6 conduct additional comparisons, running the analysis separately in poor (columns 3 and 4) and rich TUs (columns 5 and 6). This is to understand whether it is community characteristics or households characteristics that mediate CLTS impacts, or both. Column 3 shows that, in poor TUs, CLTS had large impacts among both rich (7pp) and poor (12pp) households. The point estimate on the difference between these two coefficients, 5pp, is not statistically significant, meaning that we cannot reject the hypothesis that impacts were identical among both types of households. Likewise, column 5 shows that, in rich TUs, CLTS had no impact among both rich and poor households. These two columns lend support to the idea that CLTS targeting should consider TU- or community-level wealth levels, we discuss the interpretation of these results further in section 7.8 below.

Finally, columns 4 and 6 of Table 7.10 test whether relative poverty plays a role in mediating CLTS impacts. A household's absolute poverty level might be important in determining investments if their lack of resources is a binding constraint. A relative poverty measure, on the other hand, can be used to test whether a household's *position in the wealth distribution* is what is mediating CLTS impacts. We rank households from richest to poorest within each TU, and assign those in the top half of the wealth distribution of their own TU to the 'high' wealth group and those in the bottom to the 'low' wealth group. In contrast to previous indices used, this measure ensures that *within* each community, half of the respondents are classified as 'high wealth' and half are classified as 'low wealth'. The results using this measure are almost identical to those estimated using our previous household wealth measure: CLTS had strong and statistically identical impacts among both rich and poor households in poor TUs (column 4) and no impacts in rich TUs (column 6). This finding is in line with the nature of the CLTS intervention, which aims to transform social norms and behaviours regarding sanitation at the community level.

7.5 Improved sanitation

A reasonable concern raised by Abramovsky, Augsburg and Oteiza (2016a), among others, relates to the quality of the additional toilets constructed as a result of CLTS. The Handbook on CLTS does not instruct facilitators to promote the construction of any toilet models in particular (Kar and Chambers, 2008). However, the question of toilet quality is important. The nature of CLTS impacts would be different if the construction of unimproved, unsafe toilets was promoted. While toilet use is considered safer than the practice of OD, unimproved toilets do not provide such an effective barrier between people and faeces, and thus these are unlikely to provide the same health benefits as safe toilets (or, in the worst case, any health benefits compared with OD). If the toilets built because of CLTS had a higher proportion of unimproved units than those constructed in control areas, then CLTS areas would, on average, experience reduced health benefits from CLTS. Additionally, low-quality toilets are more likely to collapse and go out of use, implying that the gains achieved through the CLTS intervention could be short-lived.

Table 7.11. CLTS impact on toilet quality

Outcome: % of households in the TU who own an improved toilet (conditional on owning a toilet)	**Whole sample** (1)	**Poor TUs** (2)	**Rich TUs** (3)
Panel A. Pooled estimates			
Treated (γ)	−0.01	−0.03	0.01
p-value (unadjusted)	(0.56)	(0.24)	(0.33)
p-value (adjusted)†	(0.58)	(0.58)	(0.58)
Panel B. Estimates by period			
Treated × Wave 2 (γ_2)	−0.06*	−0.13**	−0.00
p-value (unadjusted)	(0.03)	(0.01)	(0.98)
p-value (adjusted)†	(0.06)	(0.02)	(0.98)
Treated × Wave 3 (γ_3)	0.03	0.02	0.03
p-value (unadjusted)	(0.25)	(0.92)	(0.04)
p-value (adjusted)†	(0.43)	(0.92)	(0.12)
Treated × Wave 4 (γ_4)	0.01	0.03	0.00
p-value (unadjusted)	(0.46)	(0.52)	(0.81)
p-value (adjusted)†	(0.84)	(0.84)	(0.84)
Control mean (wave 1)	0.90	0.81	0.95
Number of TUs	235	112	123
Number of observations	680	314	366

† Adjusted *p*-values account for multiple hypothesis testing following the procedure described in Romano and Wolf (2005), with 1,000 clustered bootstrap samples.
Note: Regressions at the TU level, using the share of toilets that are improved in each TU, conditional on owning a toilet, as the outcome. In each wave, TUs in which no households had toilets were excluded. All specifications include LGA and survey wave fixed effects and control for the value of the outcome variable in wave 1. *p*-values are given in parentheses. Stars indicate statistical significance: *$p < 0.10$, **$p < 0.05$ and ***$p < 0.01$, according to the adjusted *p*-values.

Table 7.11 addresses these concerns. It shows the results of a cluster-level analysis, where the variable of interest is the share of existing toilets that are classified as improved. If CLTS had promoted the construction of simple, unimproved toilets, then we would expect to see negative coefficients for CLTS treatment effects on this outcome. We run this analysis for the whole sample (column 1) and for poor and rich TUs separately (columns 2 and 3). Column 1 shows that there is a negative and significant coefficient for the first survey wave after CLTS implementation took place. The same pattern of findings, albeit with slightly higher point estimates for wave 2, can be found among the subsample of poor TUs. This suggests that 6 months after the intervention, CLTS indeed resulted in the construction of simpler, unimproved toilets in programme communities. Nonetheless, by waves 3 and 4, no observable differences between CLTS and control TUs remained in terms of the proportion of toilets that are improved toilets. In fact, by wave 4, treatment and control groups had almost identical proportions of improved toilets to those exhibited in wave 1: 90% for control TUs (same as in wave 1) and 92% for CLTS TUs (up from 91% in wave 1).

These findings imply that there is no evidence of CLTS distorting the ratio of improved to unimproved toilets in a significant manner, and equally no evidence that CLTS is promoting improved toilets to a greater extent.

7.6 Constraints on toilet adoption

The household survey conducted during wave 4 included a question that may shed additional light on the constraints faced by households when they decide whether to invest in a toilet or not. Households who did not own a functioning toilet were asked what the main reason for this was. Respondents were offered a series of alternative responses, and were also free to input their own. The answers to this question are presented in Figure 7.9, grouped by theme.

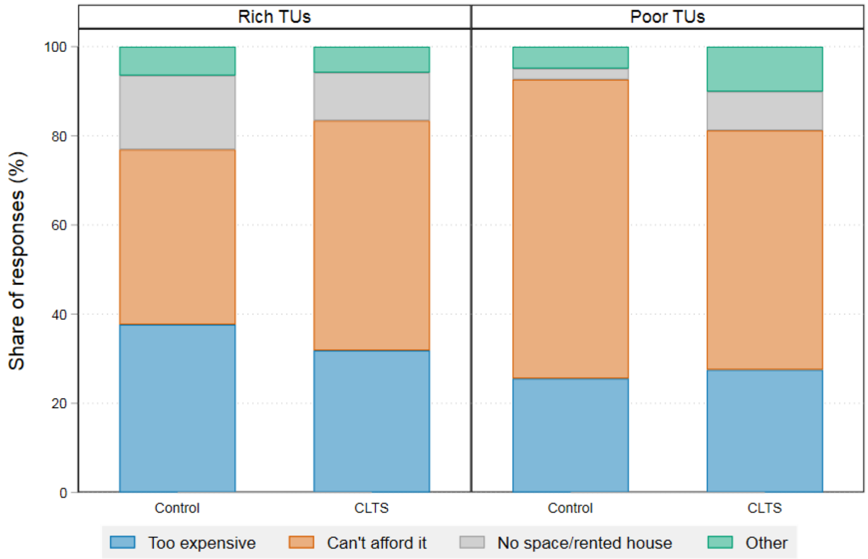
From Figure 7.9, we see that financial constraints on toilet adoption, represented by the ‘Too expensive’ and ‘Can’t afford it’ categories, were the most important reason mentioned by the households in our sample. Even the wealthiest households in our sample (two leftmost columns) refer to financial constraints when explaining the reasons why they do not own toilets. Among the households who did not own a toilet and belonged to rich TUs, 80% said that they could not afford one or that they were too expensive. However, during wave 1, the simplest pit latrine model (with no slab or superstructure) had an expected cost of construction of US$180. This amounted to less than 5% of the median yearly expenses reported by the households in the top wealth quintile (US$3,600). These estimated construction costs include the hiring of labour, which some households may be able to avoid by digging the pit themselves. The facts that simple pit latrines are not out of reach, at least of the wealthiest households in our sample, and that the strongest CLTS impacts were detected among households living in, on average, poorer communities, suggest that we need to take a more nuanced approach to financial constraints.

For this purpose, let us compare the answers provided by *control* households, from rich and poor TUs. Regression analysis tells us that control households in rich TUs were 18pp more likely to declare that toilets were ‘too expensive’ than control households in poor TUs. On the other hand, they were 21pp less likely to say that they could not afford one.[49] When households report that toilets are too expensive, they do not refer only to the absolute cost of constructing one, but to the relationship between its costs and its expected benefits. Imagine being offered a US$10 bottle of water and a US$10 car. A household can decide that, although the two items cost the same, the latter is cheap given its benefits and the former is too expensive. The same thing happens with toilets. When the benefits from their usage are underestimated, they appear as onerous investments with little or no return.

[49] These results are significant at the 5% level and robust to adjustments for multiple hypothesis testing.

An alternative explanation to the differences observed between control households in the two TU wealth groups is that they might have different toilets in mind, which is just a hypothesis that we cannot test since we do not have the relevant data. Households in wealthy TUs might be thinking about the costs and benefits of a ceramic pour flush toilet, while households in poor TUs might have simpler, more affordable models in mind. Unfortunately, we cannot distinguish between these two alternative explanations in this case. Nonetheless, what is clear is that households report facing significant financial constraints to investing in the toilets they aim for.

Figure 7.9. What is the main reason why you decided not to construct a / repair your toilet?



Note: The question was directed only at households who did not own functioning toilets.

Source: Household survey, wave 4.

Figure 7.9 also shows noticeable differences in the pattern of responses to the question by CLTS treatment status in poor TUs, which may shed light on the mechanisms by which CLTS persuaded households to construct and use toilets. One clear candidate mechanism is that of increasing the expected benefits of toilet ownership. By stressing the disadvantages of OD and the importance and advantages of toilet ownership, CLTS may have led households to recalculate the benefits of toilet ownership relative to OD. It may have tipped the balance of a household's cost–benefit calculation and pushed them towards investment, while leaving the costs of toilet construction unaffected. If this were the case, we would expect a lower share of households reporting to find toilets 'too expensive' in the CLTS group than in the control group. However, this is not what happened. Regression analysis shows that, when we look at poor communities, CLTS households with no toilets are just as likely as controls to report that toilets are 'too expensive'. However, they are 10pp less likely to declare that they *cannot afford* toilets and

5pp more likely to say that they do not have enough space or live in rented accommodation.[50]

This evidence may seem counterintuitive. Fewer households reporting that they cannot afford a toilet in CLTS areas suggests that CLTS relaxed some financial constraint among this group of households and led them to invest in a toilet. But CLTS provides no subsidies or credit, and therefore does not help households overcome financial constraints in any meaningful way. In principle, CLTS also does not provide information about the costs of specific toilet models, although facilitators are free to discuss this during triggering meetings. While the evidence is only tentative, these results suggest that CLTS directed households in poor communities towards cheaper, more affordable toilet models, or that it corrected their perceptions of the costs of construction downwards, thus showing some of the households that in fact they could afford to invest in a toilet. Meanwhile, this was not achieved by CLTS in rich TUs.

7.7 CLTS impacts on child health

Having established that CLTS increased toilet ownership and reduced OD in poor TUs, we ask whether CLTS had any impact on child health in these areas. We restrict this analysis to children under the age of 5 living in households in poor TUs only, since this is where we detected any CLTS impacts in terms of toilet ownership and reductions in OD. Following standard practice, anthropometric measurements were taken for the two youngest children under 5 in each household during wave 4. Using standard measurements provided by the World Health Organisation (WHO), these were then converted into *z*-scores that reflect the distance that each child is from the global average, expressed in terms of their standard deviation, according to their age and gender. We study three outcomes commonly used in the health literature: length-for-age, weight-for-age and weight-for-length.

[50] These findings are significant at the 5% and 10% levels, respectively, after adjusting errors for multiple hypothesis testing. By wave 4, when this question was asked, toilet ownership in rich TUs was statistically indistinguishable between control (62%) and CLTS areas (59%). The difference in poor TUs was significantly larger and statistically significant: ownership in CLTS areas reached 50%, while in control areas it lagged at 38%.

Table 7.12. CLTS impacts on child health: poor TUs

Outcome: *z*-score for each in wave 4	**Anthropometric measurements: children under the age of 5**		
	Length-for-age	Weight-for-age	Weight-for-length
	(1)	(2)	(3)
Treated (γ)	–0.00	0.10	–0.03
p-value (unadjusted)	(0.57)	(0.20)	(0.98)
p-value (adjusted)†	(0.83)	(0.47)	(0.98)
Number of TUs	102	105	102
Number of children	572	614	566

† Adjusted *p*-values account for multiple hypothesis testing following the procedure described in Romano and Wolf (2005), with 1,000 clustered bootstrap samples.
Note: All specifications include household controls, LGA fixed effects and fixed effects for child age (in months). Controls include age, age squared, gender, employment status and level of education attainment of household head, household size and relative asset wealth, as well as a dummy variable indicating farming as the household's main economic activity. All controls are measured at baseline. Standard errors are clustered at the unit of randomisation (TU) level. *p*-values are given in parentheses. Stars indicate statistical significance: *p < 0.10, **p < 0.05 and ***p < 0.01, according to the adjusted *p*-values.

Table 7.12 shows that CLTS appears to have had no impact on any of the three outcome measures. The increase in toilet ownership and the reductions in OD experienced by CLTS households in poor TUs did not translate into significant average increases in child health.

Significant gaps remain in the water sanitation and hygiene (WASH) literature in our understanding of faecal contamination pathways and the evidence on the impact of CLTS on health outcomes is mixed. In three recent studies (Indonesia see Borja-Vega, (2014), based on Cameron, et al., (2013), Mali see Pickering et al., (2015) and Mozambique see Godfrey, et al., (2014)), CLTS had positive effects on child health, reducing self-reported diarrhoea incidence, reducing the proportion of stunted and underweight children, or reducing the incidence of water-borne diseases in general. Interestingly, these positive outcomes appear even in contexts where the impact on toilet construction was modest, at best, as is the case of Indonesia. Other studies in different developing countries that looked at health outcomes reported no significant improvements in these health outcomes, even after significant increases in toilet ownership were achieved. For example, a set of recent randomised trials of WASH interventions conducted in Bangladesh and Kenya failed to detect large impacts on child health (Tofail, et al., 2018; Stewart, et al., 2018). At the same time, using data from Northern India, Geruso & Spears (2018) showed that a 10 percentage point reduction in the fraction of neighbours performing OD is associated with a reduction in infant mortality of 8% of the population mean. A similar finding is shown by Augsburg & Lesmes (2018). Other recent studies show that not only is sanitation coverage the variable that should be used as explanatory variable for health outcomes, but that health benefits will not materialise or will not be large enough to be measurable until this toilet coverage is brought above a certain threshold. This is consistent with the idea of strong positive externalities of toilet coverage or negative externalities of open defecation (see, for instance, Cronin et al (2017), and Gertler et al (2015) among others), and highlights the importance of considering the level of toilet coverage and OD behaviour in the community when looking at the impact of WASH interventions on health outcomes. Considering that the coverage in terms of households owning a functioning latrine was below 50% at the time of the endline survey (45% in the whole sample and 46% in the CLTS group of villages), this may help explain the different and sometimes inconsistent findings from the literature and the fact that, in this study, we do not find an impact on child health.

7.8 Discussion of CLTS results

How can we interpret the results presented above for the Nigerian case in the light of the existing evidence from other contexts? Table 7.1 suggests that CLTS had no statistically significant impacts on toilet construction, ownership or usage on average over the whole sample. We then split the sample according to toilet ownership status in wave 1, before CLTS was implemented. Table 7.2 shows that CLTS increased the likelihood of owning a functioning toilet and reduced OD (reported by the main respondent) among households with no toilets, by a statistically significant 5pp. No impacts on OD were detected among households who already owned toilets in wave 1.

These results are reasonable in light of the ownership and usage patterns observed in our study sample during wave 1. Back then, 36% of households reported owning a private, functioning toilet, meaning that 64% of the households in our sample had no private toilet. When asked whether they performed OD, 63% of households in our sample responded that at least one adult member did. Looking at the households that did not own functioning toilets at baseline, OD rates were 97%, while only 2% of the households that owned toilets reported that they performed OD. This suggests that in the Nigerian context, the gap between ownership and usage was extremely small, and the margin for improvement in this dimension was narrow. In other words, while there was an important margin of improvement in terms of toilet construction (64% of the households in our sample, 97% of which declared that they performed OD), there was a small margin of possible improvement in terms of OD among households that already owned a toilet (just 2% of them performed OD). Consequently, it should not be surprising to observe that CLTS reduced households’ OD by prompting them to build a new toilet (the first margin) and did not reduce OD in households that already had a toilet prior to the intervention (the second margin), which is in line with the previous evidence discussed above. To the best of our knowledge, CLTS has only been shown to affect OD rates among households that own a toilet in the case of Mali, where baseline OD rates among this subset of households was as high as 37% (Pickering et al., 2015). In light of this evidence, the construction of new toilets seems to be the main constraint on OD reduction in the Nigerian context, as opposed to the usage of existing toilets.

Where was CLTS more effective? Table 7.7 shows that CLTS had strong and sustained impacts only among the poorest communities in our sample. This group experienced increases in toilet ownership, ownership of a functioning toilet and ownership of improved toilets of 8pp, 10pp and 7pp respectively, all significant at least at the 5% level. These improvements in ownership were accompanied by equivalent reductions in OD as reported by the main respondent, and by any adult member of the household, by 9pp and 10pp respectively, both significant at the 1% level. At the same time, households living in the richest communities in our sample experienced no significant change as a result of CLTS. Wealth at the community level appears to be a key factor mediating CLTS effectiveness. Other community-level characteristics, such as social capital, religious fragmentation or initial conditions in community-level toilet coverage, are not correlated with significantly different CLTS impacts (as seen in Tables 7.6 and 7.9). Moreover, community-level wealth appears to be more important than household-level wealth. Table

7.10 shows that CLTS impacts were statistically identical for rich and poor households in both poor communities and rich communities. In each of these cases, the magnitude of CLTS impacts is driven by wealth at the community level instead.

Our results highlight the importance of community characteristics – in particular, community-level wealth – in mediating CLTS impacts. The Handbook on CLTS (Kar and Chambers, 2008) acknowledges the importance of community characteristics when deciding where to target the intervention, but emphasises traits such as homogeneity, low toilet coverage and high levels of social capital that, in the case of this study, do not appear to play a significant role. The evidence provided by this intervention is intended to motivate CLTS researchers to inquire further into the possible mechanisms of CLTS effectiveness. CLTS seems to have convinced households in poor TUs that they could in fact afford toilets and carry out the investment necessary to install them and use them. How CLTS managed this, and why this did not happen in rich TUs, remain important questions to be explored in future studies. Practitioners should take note and further refine CLTS targeting to achieve more cost-effective delivery of sanitation policy.

8. SanMark results at the household level

In this chapter, we present some evidence on the degree to which the SanMark intervention reached households, and whether they indeed adopted WET products. As discussed in Section 3.2, this relates to both SanMark phases, as households could have decided to invest in WET products after interacting with SanMark businesses (phase 1) or after D2D sales activities and community-level marketing events (phase 2). We then estimate the impacts of the SanMark intervention, both on its own and in areas where CLTS had also taken place.

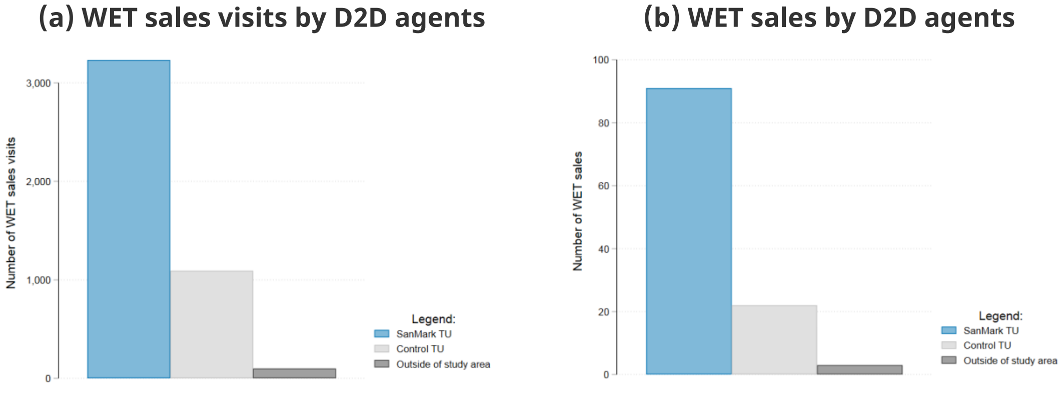
8.1 Awareness and adoption of WET products

The first margin of interest is whether awareness of the WET product line was increased as a result of the SanMark intervention. Abramovsky, Augsburg and Oteiza (2017b) showed that in March 2017, 6 months after introducing the new product to local businesses, only around 7% of the households in our sample knew about the products. After that, marketing activities and D2D sales agents were deployed. Our wave 4 data reveal that in the 8-month period between waves 3 and 4, during which businesses had additional time to market their products and the SanMark community-level marketing activities / D2D sales agents were introduced, household-level awareness of WET products increased from an average of 7% to 20%. Regression analysis (not shown) reveals that this increase of 13pp is statistically significant at the 1% level.

A question of interest is whether this increase is driven by the additional time businesses had to market the new product they offer, or whether it is induced by the market-level activities and deployment of D2D sales agents. D2D sales agents were randomly allocated to communities (TUs or clusters) defined as SanMark treatment TUs. In SanMark control TUs, no market-level promotional activities were conducted and no D2D agents were recruited and trained. A comparison of awareness levels between SanMark treatment and control communities could tell us whether they were affected by the work of D2D agents or not. However, as mentioned in Section 4.1, D2D agents were allowed to market WET products freely, in any areas they saw fit. We assumed they would mostly target households in their own (SanMark treatment) communities, but there were no constraints on their work: they could just as well visit households in nearby SanMark control TUs. And indeed, we do find that three quarter of all sales are conducted in their own (treatment) communities, leaving though some sales in control, which weakens our identification strategy, suggesting any estimated impacts to be lower bounds. So before analysing their effectiveness at raising awareness of WET products, we must check whether D2D agents concentrated their efforts in SanMark areas. If they did not, then we will not be able to identify their impact, if any.

The D2D agent survey asked respondents to identify the three main communities they performed household visits and WET sales in. This could include, of course, the agent's own community, which by definition would have been a SanMark treatment TU. When these communities fell within our study area, we then matched these TUs with their corresponding SanMark treatment status. Thus we can study the total number of household visits and WET sales performed by D2D agents, according to the SanMark treatment status of the TUs where the visits/sales took place. As we can see from Figure 8.1(a), WET sales visits to households located in SanMark treatment TUs outnumbered visits in control TUs by more than three to one. This relationship is even more skewed for WET sales, as seen in Figure 8.1(b): four out of every five WET sales were conducted in SanMark treatment areas.[51]

Figure 8.1. WET household visits and sales by D2D agents, by community's SanMark treatment status



Source: D2D agent survey, wave 4.

This first result allows us to proceed with the second step of our analysis, i.e. to study whether the work of D2D agents was responsible for the growing awareness of WET products among households. Given that SanMark control and treatment areas share similar levels of exposure to SanMark businesses as a result of random assignment, uneven levels of awareness of WET products between these two groups could only be the result of D2D agent visits. Figure 8.2(a) shows the levels of WET awareness by treatment arm. Comparison of the grey and blue bars, representing awareness rates at waves 3 and 4 respectively, shows the observed increase from 7% to 20% over time. The figure further reveals an almost identical, and statistically indistinguishable, increase across SanMark treatment and control communities. Activities conducted by D2D agents, which took place mostly in SanMark treatment TUs, appear to have had no significant effect on WET awareness.

To test whether SanMark business-level activities were behind this increase, we look at whether the increase in awareness was higher in clusters where SanMark businesses were operating than in clusters where no SanMark businesses were located, as shown in Figure 8.2(b). Here we study whether the presence of SanMark points of sale was the factor driving the increase in awareness we observe between waves 3 and 4, as opposed to the SanMark D2D visits and marketing events studied in Figure 8.1(a). In order to compare two similar sets of clusters, here we restrict the sample to clusters with either SanMark or control businesses, and we drop the clusters that have none. We find that although

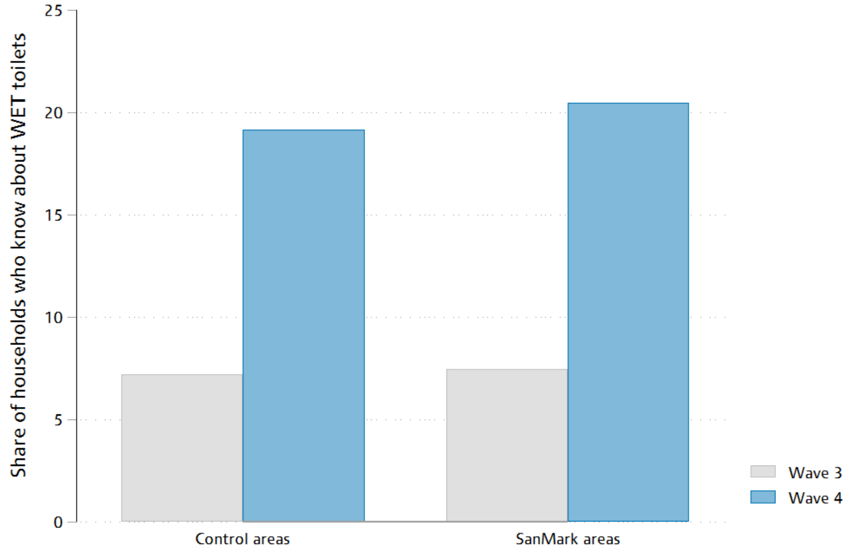
[51] Relating this analysis back to intervention effectiveness by community wealth levels, we find that the number of visits is more or less evenly distributed amongst rich and poor communities. However, actual sales are twice as high in rich than they are in poor communities. The ratio of sales in treatment and control are not different by where the D2D agents are based.

average WET awareness levels in wave 4 seem to be slightly higher in areas with SanMark businesses, this difference is not statistically significant.

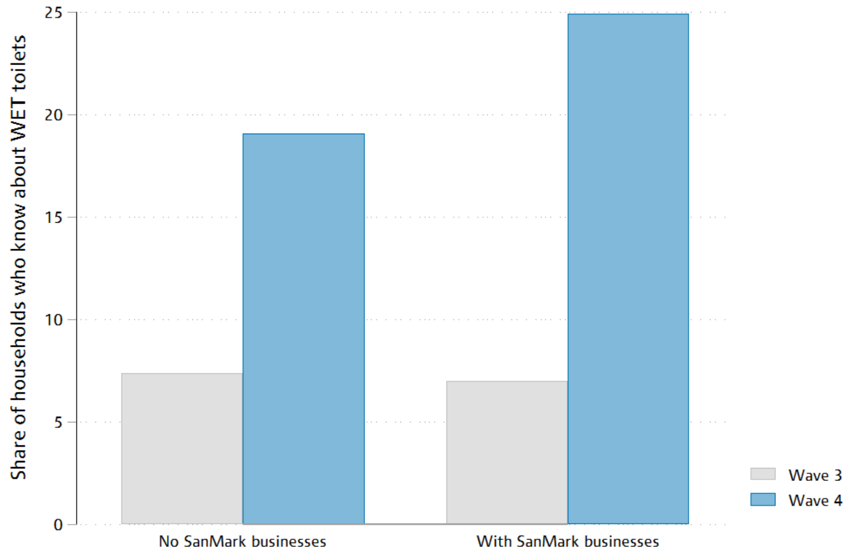
The two exercises in Figure 8.2 lead us to conclude that there is no robust evidence indicating that the increase in WET awareness observed between waves 3 and 4 was mainly driven either by clusters being assigned to SanMark's second phase of marketing activities or by the presence of SanMark businesses. Regression analysis confirms these results, and also suggests that there is no significant difference relative to areas assigned to both interventions. Rather, it seems that information travelled fast across clusters, possibly partly due to D2D agents visiting at times also control areas. This makes it impossible for us to disentangle the relative importance of these two channels, but it is clear that awareness of SanMark products is higher in wave 4 than in wave 3 in all study areas.

Figure 8.2. Awareness of WET products by survey wave

(a) WET awareness by SanMark treatment status (cluster)



(b) WET awareness by presence of SanMark businesses



Note: The top panel shows the number of households who responded that they knew about or had seen WETs, as a percentage of the number of households interviewed in each survey wave. Control and SanMark areas refer to the cluster randomisation of households to either SanMark or control groups for SanMark's phase 2 activities. The bottom panel shows similar estimates, but compares households located in clusters with SanMark businesses and households in clusters where only control businesses were located. Clusters in which neither SanMark nor control businesses were located were removed from this panel to obtain two comparable samples.

Source: Household survey, waves 3 and 4.

Next, we look at whether this higher awareness translated into adoption of WET products. We do not find high levels of adoption of WET products by the households in our sample. Out of the 5,594 households who agreed to be interviewed during our last survey wave, in late 2017, only 13 owned a WET of any kind, while three of them owned one in the previous wave.[52]

Six of these 13 households provided further information on their investment.[53] Given the small sample, it is difficult to make broader inferences, but we report their responses here nevertheless. Four of the six households invested in the replacement of a non-functioning toilet, one replaced a functioning toilet and one had no toilet at all previous to the WET installation. Four of the six households said they paid for the pit digging and installation, and two said they also paid for the construction of the superstructure. No households reported that they installed a WET as a simple conversion from existing VIP latrines or other existing models on the market. One of the six households reported to be satisfied and the remaining five reported to be very satisfied with the toilet. Finally, four of these households reported that they made a single payment for the whole sum, while only one had paid in instalments.

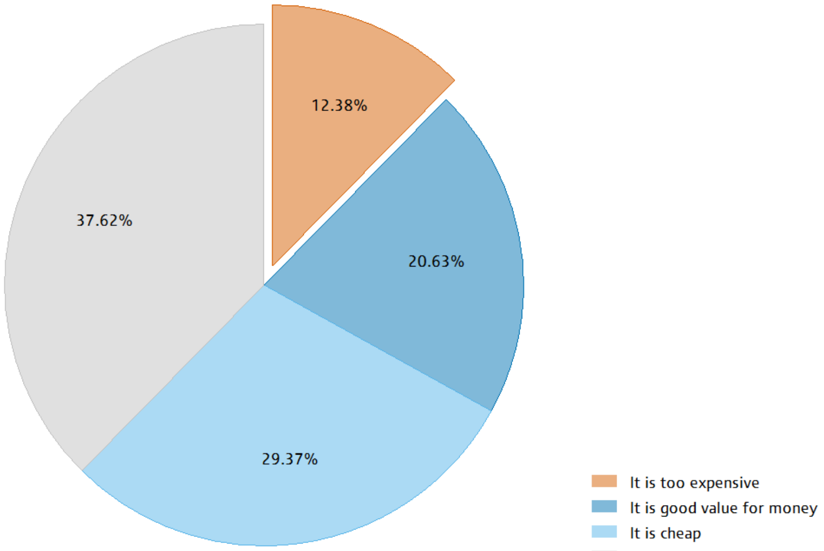
Possible constraints on WET adoption

A more in-depth look into the possible reasons for the low rate of adoption at the time of wave 4 is presented in Figures 8.3 and 8.4, using questions related to WET affordability and thoughts about future purchases. We look at the responses given by households who were aware of WET products but did not own them. Because these are a small subset of the households in our sample, we indicate the total number of valid responses in each case in the figure notes.

[52] In this last survey, as opposed to the one carried out in wave 3, enumerators asked to see the toilets, and they confirmed that these were indeed WET products. Toilet observations also allowed us to review our findings from wave 3 discussed in Abramovsky, Augsburg and Oteiza (2017a) and to correct the last figure from two to three units.

[53] The remaining seven households did not provide any additional information regarding the WET purchase and installation process, possibly because the main respondent to the interview was not involved in them.

Figure 8.3. Opinion regarding the affordability of WET product line: non-owners of WET products



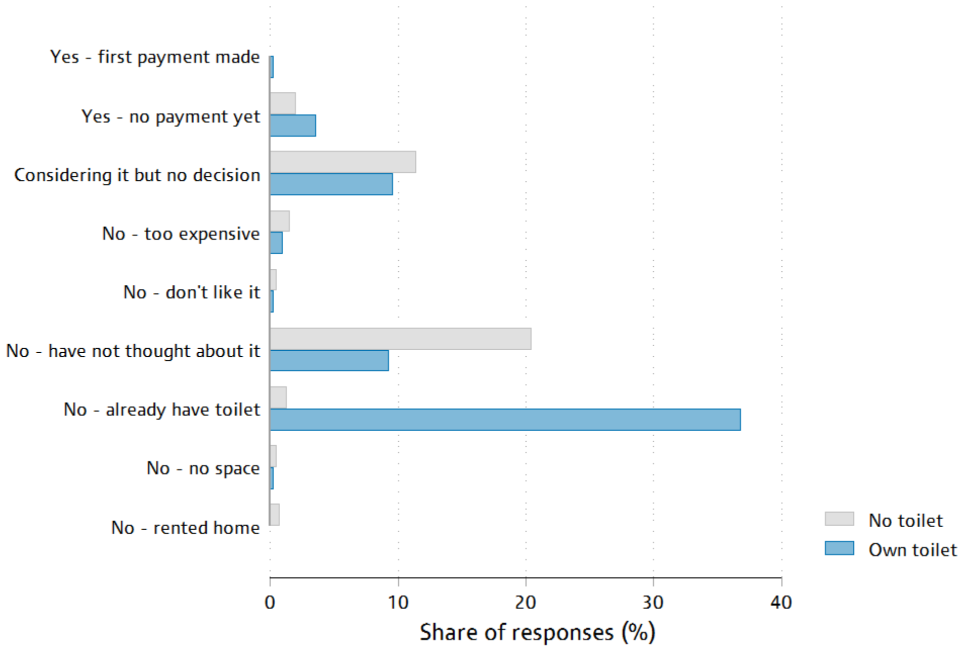
Note: Restricted to households who said they were aware of the WET product line and who did not already own a WET (412 valid responses).

Source: Household survey, wave 4.

Figure 8.3 shows the answers to a question directed to elicit exclusively the respondent's thoughts on the affordability of the WET product line. Just 12% out of the 412 households who responded to this question said they found the products too expensive, compared with 50% who thought they were either cheap or good value for money. In addition, we find that the majority of households consider WET to be superior to other toilet options: out of the 420 households who responded to this survey module, 63% believe WET products are safer than other models on the market, 68% believe WET products use less water and 73% believe they look good (not shown). Given this positive impression of WET products and that only 12% find them too expensive, it is rather puzzling that not more households have invested in the new toilet model so far.

Figure 8.4 shows the possible answers to the question 'Are you considering buying and installing a WET product in the future?'. Answers are presented separately for households who own functioning toilets and those who do not. We can see that few households responded affirmatively to this question, and it appears that especially households who already own toilets see little reason to replace their existing facilities. At the same time, amongst the households who did not already own toilets, the two most common responses were that they had not considered it and that they had considered it but had made no investment decision yet.

Figure 8.4. Household plans for investment in WET products: non-owners of WET products



Note: Restricted to households who said they were aware of the WET product line and who did not already own a WET (385 valid responses).

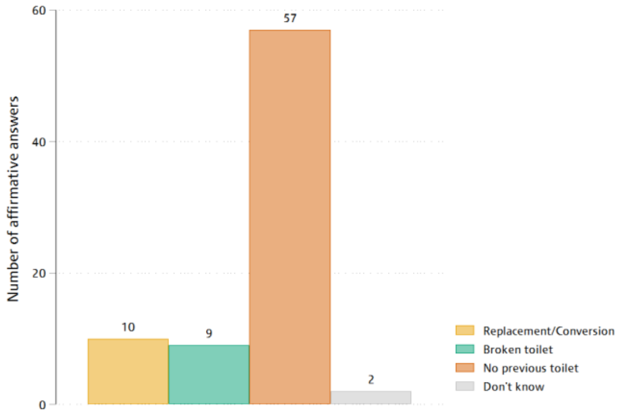
Source: Household survey, wave 4.

While the impression among households appears to be that WET products are affordable and attractive overall, investment in them has been, so far, underwhelming, and few households report having plans to invest in them in the future. How do we reconcile these two findings? Two important points should be taken into consideration. First, overall awareness of WET products is still just around 20%, and is correlated with household wealth. The rate of awareness was 16% among the poorest 20% of our sample and 27% among the wealthiest 20%. A second related point is that the responses described above reflect the views of a selected group of households and may not be representative of the views from the whole sample. For example, the 420 households who agreed to respond to the questions presented in Figures 8.3 and 8.4 are wealthier than the rest of the sample: 32% of these respondents belong to the wealthiest 20% of our sample, while only 13% of them belong to the poorest 20%. Their views, while informative on their own, should not be used to make inferences about the impressions held by the rest of the households in our sample.

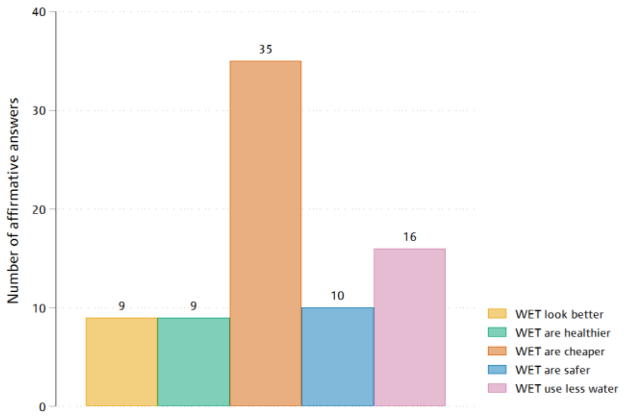
These views from household survey respondents can be contrasted with those of D2D sales agents, also interviewed as part of survey wave 4. We analyse their responses to three questions about the main purpose of the WET sales they successfully conducted, the reasons why households chose to invest in WET products and the reasons why some households did not. Figure 8.5 shows the answers to these questions.

Figure 8.5. D2D sales agents' impressions of drivers and constraints behind WET adoption

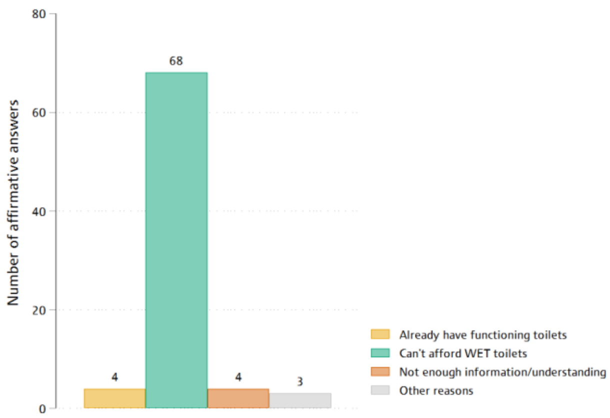
(a) What is the main purpose of WET investments in your experience?



(b) What are the main reasons households invest in WET models?



(c) What are the main reasons households do not invest in WET models?



Note: Multiple answers were possible (78 valid responses).

Source: D2D agent survey, wave 4.

From Figure 8.5(a), we see that D2D sales agents are primarily involved in the sale of WET products to households who previously did not own private sanitation facilities of any kind. They report that the main reason why households invest in a WET is the construction of a new toilet (orange bar). This is encouraging, as it shows that SanMark has the potential to tackle the lack of access to sanitation and is not only acting as an affordable way to improve existing facilities. At the same time, we know from Abramovsky, Augsburg and Oteiza (2016a) that households who do not own toilets are generally poorer than households who already own toilets, so they are also the most likely to be financially constrained. This is confirmed by evidence shown in Figure 8.5(b), which suggests that affordability is the main reason behind the adoption of WET products as opposed to other toilet models in the market. Figure 8.5(c) is also consistent with this hypothesis, as it shows that financial constraints are indeed the main limiting factor behind investment in WET products, according to D2D sales agents.

The slow uptake of WET facilities by households described in this section indicates that the SanMark intervention still faces an important challenge. While it is recognised by potential users as a more affordable product line than other models on the market, financial constraints remain the key factor behind low levels of WET adoption, together with an increasing, but still far from complete, awareness at the household level.

8.2 Impacts of SanMark at the household level

In this section, we present the overall impacts of the SanMark intervention on sanitation outcomes at the household level. As of March 2017, SanMark community-level marketing activities began in clusters (communities) assigned to SanMark phase 2 activities. These activities included marketing events in public areas, as well as the deployment of SanMark D2D sales agents who canvassed households in their clusters, promoting the WET product line and working on commission. To evaluate this SanMark component, we exploit the random assignment of communities to receiving the phase 2 SanMark activities or not. In the next section, we will expand the analysis, looking at the interaction between CLTS and SanMark, defined by the intersection of SanMark treatment and control areas with CLTS treatment areas, CLTS control areas, and areas where CLTS had occurred before our study period, known here as pre-CLTS areas.

We focus on the five primary outcomes for which CLTS was shown to have had significant impacts in poor communities: toilet ownership, ownership of a functioning toilet, ownership of a functioning and improved toilet, whether the main respondent declares that they perform OD, and whether they declare that at least one adult member of the household performs OD. Note that here we look at the impact of SanMark on toilet uptake of any kind and not specifically the promoted WET product.

First, we compare outcomes between SanMark treatment and control areas, using waves 1 and 4 only, as this allows us to work with the larger sample of 6,386 households, i.e. including study areas where CLTS activities had taken place before the start of the STS Nigeria project and hence this evaluation. Of these, 5,594 agreed to be interviewed in wave 4, so this is the set of households for which we will be estimating impacts. Table 8.1 presents the results. There are no statistically significant differences in any of the outcomes between households assigned to SanMark treatment and control areas.

Table 8.1. SanMark impact on toilet ownership and OD

Outcome =1 if:	HH owns a toilet (1)	HH owns a functioning toilet (2)	HH owns a functioning improved toilet (3)	Main respondent performs OD (4)	Any household member performs OD (5)
Treated (τ)	−0.02	−0.02	−0.02	0.02	0.01
p-value (unadjusted)	(0.25)	(0.36)	(0.23)	(0.30)	(0.47)
p-value (adjusted)†	(0.44)	(0.46)	(0.44)	(0.44)	(0.49)
Control mean (wave 1)	0.37	0.36	0.33	0.62	0.65
Number of TUs	327	327	327	327	327
Number of households	5,368	5,368	5,368	5,373	5,373

† Adjusted *p*-values account for multiple hypothesis testing following the procedure described in Romano and Wolf (2005), with 1,000 clustered bootstrap samples.
Note: Results presented were estimated using specification (4) described in Section 4.4, including LGA fixed effects. Controls include age, age squared, gender, employment status and level of education attainment of household head, household size and relative asset wealth, as well as a dummy variable indicating farming as the household's main economic activity. All controls are measured at wave 1 and outcomes are measured at wave 4. Standard errors are clustered at the unit of randomisation (TU) level. *p*-values are given in parentheses. Stars indicate statistical significance: *$p < 0.10$, **$p < 0.05$ and ***$p < 0.01$, according to the adjusted *p*-values.

There are multiple possible reasons why SanMark does not appear to have affected toilet ownership and OD rates at the household level, as we have discussed in the previous section. In particular, one of these is the low uptake of the WET product line by households. A second possible factor that could be biasing our impact estimates downwards is the action of D2D sales agents. While they were recruited and trained from SanMark clusters, their work was not limited to these areas. If they decided to operate in SanMark control clusters, then we might be underestimating their effectiveness, as both SanMark treatment and control communities could be experiencing increases in ownership rates due to D2D sales agents.

8.3 Interaction between CLTS and SanMark

The SanMark intervention might have had better results where demand for improved sanitation was already higher. This demand is affected by a multitude of factors such as the level of urbanisation and the average wealth of each location. In addition to this, the CLTS intervention aimed to increase demand for private sanitation facilities. Because the SanMark intervention aimed directly at households was randomly assigned across communities, using the same definition of clusters as the one used for CLTS, we can estimate the impacts of each intervention on its own, as well as their combined effect.

In order to study the possible combined impact of CLTS and SanMark, we use regression specification (5) described in Section 4.4. In this approach, we exploit the independent and random assignment of households to both CLTS and SanMark. Households in our sample belong to one of three CLTS groups: CLTS, control and pre-CLTS. The last group is composed of households located in areas where WaterAid conducted CLTS activities before the start of this study. Within each of these three groups, households were also randomly assigned to either SanMark treatment or control, as described in Chapter 4. Thus, in order to understand whether SanMark had differential impacts by CLTS treatment status, we estimate SanMark impacts by CLTS treatment group. We present the results in Table 8.2.

The table shows the treatment effects of SanMark in each of the three CLTS household groups: those that received CLTS as part of this study (first row), those that received CLTS before the start of the study (fourth row) and those that did not receive any CLTS (seventh row). We find no statistically significant effects of SanMark, or of SanMark combined with CLTS, on any of the outcomes studied.

The row labelled '*p*-value of τ_{CLTS}=$\tau_{pre\text{-}CLTS}$=$\tau_{no\ CLTS}$=0' shows the result of a joint test for statistical significance, which tests the likelihood that the three estimates for each column are different from zero. The null hypothesis of this test is never rejected to standard levels of confidence.

Table 8.2. Impact of interactions between CLTS and SanMark on toilet ownership and OD

Outcome =1 if:	HH owns a toilet (1)	HH owns a functioning toilet (2)	HH owns a functioning improved toilet (3)	Main respondent performs OD (4)	Any household member performs OD (5)
SanMark and CLTS (τ_{CLTS})	−0.04	−0.02	−0.03	0.01	0.01
p-value (unadjusted)	(0.24)	(0.63)	(0.38)	(0.72)	(0.73)
p-value (adjusted)†	(0.42)	(0.75)	(0.58)	(0.76)	(0.76)
SanMark and pre-CLTS ($\tau_{pre\text{-}CLTS}$)	0.04	0.02	0.03	−0.02	−0.03
p-value (unadjusted)	(0.36)	(0.59)	(0.44)	(0.52)	(0.41)
p-value (adjusted)†	(0.57)	(0.61)	(0.61)	(0.61)	(0.59)
SanMark only ($\tau_{no\ CLTS}$)	−0.05	−0.04	−0.05	0.06	0.05
p-value (unadjusted)	(0.12)	(0.15)	(0.11)	(0.05)	(0.12)
p-value (adjusted)†	(0.20)	(0.20)	(0.20)	(0.10)	(0.20)
p-value of $\tau_{CLTS}=\tau_{pre\text{-}CLTS}=\tau_{no\ CLTS}=0$	0.20	0.46	0.27	0.22	0.35
Control mean (wave 1)	0.39	0.38	0.34	0.61	0.61
Number of TUs	329	329	329	329	329
Number of households	5,391	5,391	5,391	5,244	5,396

† Adjusted *p*-values account for multiple hypothesis testing following the procedure described in Romano and Wolf (2005), with 1,000 clustered bootstrap samples.
Note: All specifications include household-level controls and both LGA and survey wave fixed effects. Controls include age, age squared, gender, employment status and level of education attainment of household head, household size and relative asset wealth, as well as a dummy variable indicating farming as the household's main economic activity. All controls are measured at baseline. Standard errors are clustered at the unit of randomisation (TU) level. *p*-values are given in parentheses. Stars indicate statistical significance: *$p < 0.10$, **$p < 0.05$ and ***$p < 0.01$, according to the adjusted *p*-values.

Hence, at the time of the endline data collection, we do not observe that the creation of CLTS demand would have sped up the uptake of the SanMark product marketed. The inverse is also true: we do not observe that households previously exposed to CLTS achieve higher levels of ownership of improved sanitation when also selected for the SanMark intervention. Additionally, given that CLTS impacts were detected only among the poorest half of the communities in our sample, we conducted a final test of interacted impacts between CLTS and SanMark among this subset of households (not shown). Once again, we find no evidence of higher toilet ownership in areas exposed to both interventions compared with those exposed only to CLTS. We cannot discard the possibility that the absence of impacts when CLTS and SanMark are combined is due to the particular timing of the interventions in this case, or due to the fact that the SanMark intervention was not developed completely by the time the endline survey was collected, and the fact that it takes sometimes years for a new product like the WET product to penetrate the relevant market.

9. SanMark business-level results

As described in Section 3.2, the first phase of the SanMark intervention was aimed at introducing the WET product line to businesses. This phase was rolled out after piloting in the Igbo Eze North LGA, in September 2016, and it ran until December 2017. In this chapter, we present evidence on the degree of adoption of the WET products by businesses as part of their product mix (Section 9.1); we provide some basic characteristics of the work done by D2D sales agents (Section 9.2); and we compare the performance of SanMark and control businesses and study whether there is any evidence of benefits accrued from the intervention (Section 9.3).

Before proceeding with our results, it is important to clarify the data sources used. Unless otherwise stated, results come from our business surveys conducted in waves 1–4 and the D2D sales agent survey conducted in wave 4. These interviews were conducted by an independent party, is not related in any way to the intervention, and no incentive was provided to participants for their responses. Interviewers stated clearly in every survey wave that they were not related to WaterAid. Therefore, we believe that these data, although prone to measurement error, were collected in a way that minimised incentives to over-report or under-report sales performance on the part of interviewed businesses.

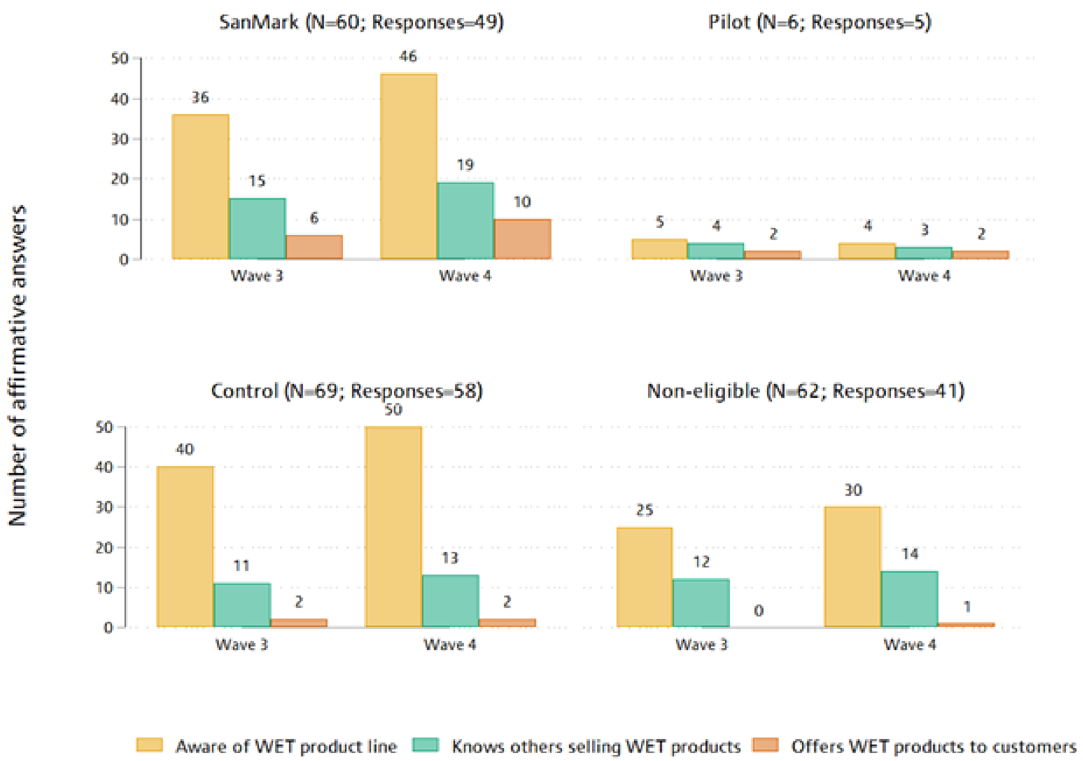
We also rely on internal sales tracking data collected by WaterAid Nigeria, as part of its monitoring and evaluation efforts within the SanMark project. These data have the advantage of being recorded in multiple intervals between our two post-SanMark survey waves, and therefore we can take a close look at the evolution of sales over time. These two sources of data have some differences. In particular, total WET sales according to our business survey, of around 400 WET units, are significantly lower than total WET sales according to WaterAid sales tracking, which reached 600 WET units by January 2018. These differences may be a result of standard measurement error, imperfect recall of sales in our two business surveys, or participating businesses possibly feeling that there is an advantage to overstate their sales to WaterAid. Additionally, as mentioned in Section 6.2, in wave 4, 13% of the businesses in our sample either refused to take part in our business survey or were not available for interview, so our sales figures are likely to underestimate total sales. Finally, WaterAid's tracking of sales focuses on businesses selected to participate in SanMark only, and therefore excludes control and non-eligible businesses. For these reasons, we do not make any attempt in this report to reconcile these two sources of information. We are reassured, however, by the fact that they do not present qualitatively different pictures of the extent of WET sales and WET adoption among businesses, as we see below.

9.1 WET product line: awareness, sales and technology adoption

Figure 9.1 shows the number of suppliers who report that they know about WET products, that they know of other businesses selling WET products and that they offer WET products to their customers, by treatment group and survey wave.

As discussed in Chapter 5, we initially surveyed all those businesses that sold inputs related to the sanitation sector, including cement, concrete blocks, PVC tubes and ceramic squat pans. During wave 2, we defined the subsample of eligible businesses for the SanMark intervention as those that were selling concrete blocks, as these were the main component of the WET models. These eligible suppliers were then randomly assigned to either SanMark treatment or control groups. While these are the businesses we focus on in the evaluation study, we also show findings related to awareness and adoption for those businesses that sold sanitation-related products but were not eligible for the intervention, as well as for those businesses located in Igbo Eze North, where SanMark was first piloted, and hence also excluded from the randomisation and related impact analysis. The results for each of these four groups are presented separately in Figure 9.1.

Figure 9.1. Business awareness and adoption of WET products, by SanMark treatment group



Note: Affirmative answers to three questions related to WET product awareness measured at waves 3 and 4. Pilot businesses are those classified as eligible for the SanMark intervention (i.e. producing concrete blocks of any type during wave 2) and located in the LGA of Igbo Eze North, where SanMark was piloted. Non-eligible businesses are those that were not eligible for the intervention during wave 2 because they did not sell concrete blocks. The numbers of completed interviews in each group were: SanMark, 49; pilot, 5; control, 58; non-eligible, 41.

Source: Business survey, waves 3 and 4.

The leftmost (yellow) bar shows the number of suppliers who declare that they know about WET sanitation products. As can be seen in the upper-left panel, at the time of wave 3, 36 businesses allocated to the SanMark treatment group reported that they were aware of the WET product. By the time of wave 4, 10 additional suppliers responded affirmatively, meaning that at the time of our last survey, 46 of the 60 businesses in the SanMark group had at least heard of the WET products. In light of the fact that only 49 of the 60 businesses were interviewed successfully in this fourth wave, this represents 94% of the respondents in this group.[54] Awareness grew significantly among control (lower-left panel) and non-eligible businesses (lower-right panel), for which we see 50 and 30 affirmative responses respectively, equal to 86% and 73% of total successful interviews in each group. In the pilot business group (upper-right panel), we have one fewer observation than in wave 3, attributable to one declined interview, but all cooperative businesses responded affirmatively otherwise.

Monthly WET unit sales

Has this (increased) awareness translated into (increased) adoption of the WET product line? As can be seen from the rightmost (red) bars in Figure 9.1, a positive trend is also observed for the case of sales. By wave 4, 10 SanMark businesses reported offering WET products, an increase of 4 additional businesses. The middle (green) bars in the figure further report whether businesses know of other businesses selling WET products. We only see a small increase in this dimension.

The fact that control and non-eligible businesses also report that they offer WET products is a particularly interesting and encouraging finding, as it suggests speedy technology spillovers within the market. At the same time, it raises questions because, as described earlier in this report, WET products are made using a metal mould that is available only to SanMark businesses at no cost. Hence, in order to carry out their sales, these other businesses would have to either make the mould themselves, rent one out from competing SanMark businesses or design WET versions that do not require a mould.[55] Unfortunately, our survey does not have information to enable us to say which of these strategies they adopted.

Overall, a total of 15 businesses in our sample declare to be selling WET products at the time of wave 4, 10% of the 153 who responded to our survey. The percentage is higher when focusing on SanMark treatment businesses, reaching 20% (10 out of 49 businesses), compared with 3% in the control group (2 out of 58). More than a year after the roll-out of the SanMark intervention, it seems that adoption is still lagging well behind awareness. However, this finding needs to be interpreted with caution. First, it is only the SanMark group from which significant levels of adoption should be expected: they are the only ones with free access to the metal mould and which can buy Sato pans from WaterAid. Even within the SanMark business group, we ought not to expect 100% adoption necessarily. The introduction of a new product into a business’s product line involves fixed costs, an investment some may find is not profitable. The efficient number of WET producers in each market will depend on the size of the market, the number of competitors, and other characteristics of both businesses and potential customers. As a result, we cannot say that the optimal outcome would have been to observe all businesses in our study offering and selling WET products.

Another margin of interest is unit sales performance. Concentrating our attention on the businesses that reported offering WET products, we now turn our attention to the number of WET units that they declare to sell in a typical month. As we see in Figure 9.2, reported monthly sales increased significantly among SanMark suppliers. Combining the reports of

[54] The characteristics of businesses that agreed to be interviewed (82% of the businesses in our sample) may differ, on average, from those of the whole sample. Nonetheless, it is important to note that the probability of agreeing to an interview was not different across treatment groups, as noted in Chapter 6.

[55] The cost of developing a new mould was estimated at US$400 by WaterAid.

all SanMark businesses, we find that their monthly turnover is 135 units, an almost sevenfold increase compared with the amounts reported at wave 3, when the reported turnover was 20 units per month. Sales increased among pilot businesses as well, but at a slower pace, which is expected given that this group includes only 6 businesses. Control and non-eligible businesses reportedly sell a joint total of 19 WET units per month.

What do these figures tell us about the average performance of WET unit sales of businesses that offer WET units? The 135 WET units reportedly sold in the SanMark group are split among ten businesses, which are then averaging 13.5 sales each. The two pilot businesses sell, on average, 7 units each, while the two control businesses reportedly sell 6 units each every month. The sole non-eligible business has a similar performance, selling 7 units in a typical month.

This suggests two important points. First, although we observe WET sales in all of the four study groups, not surprisingly SanMark businesses are performing better in terms of WET unit sales, on average. This could be due to the mould and technology transfer and the business training and support components of the intervention, which control and non-eligible businesses did not have access to. A second point worth noting is related to the profitability of WET products.[56] Anecdotal evidence from the field suggests that profit margins are not large for these toilets, so the sale of 6–14 units a month might not be an attractive option for suppliers who still do not offer WET products, particularly given the investment necessary in terms of time and resources devoted to setting up the new product line. Businesses would have to speculate and expect an increase in monthly sales over time for the investment to become viable.

While accurate profitability data regarding WET products are unavailable, we can use sales tracking data to take a more in-depth look at this question. WaterAid has tracked WET sales by participating businesses in a regular manner since the start of the intervention. There are 13 SanMark or pilot businesses which have reported to WaterAid that they have made WET sales at least once during this period, which means that they incurred the cost of producing the new WET toilet (which involves learning a new skill) and managed to make successful sales as a result.[57] However, five of them had stopped recording WET sales by the last sales tracking period, which ran from November 2017 to January 2018. This suggests that, in addition to barriers to entry, the sustained adoption of WET products faces other challenges, one of which could be profitability.

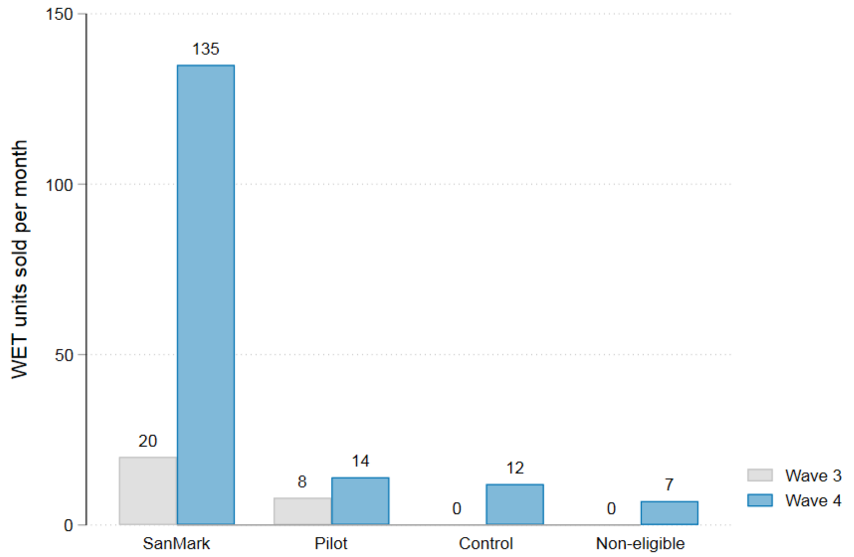
Looking in more detail at what types of WET products were reportedly sold by businesses, Figure 9.2(b) shows that the major share of these transactions is explained by sales of the WET offset model. The most affordable model – the WET direct pit version – is the second most popular product, while the onerous WET dual set, which is in fact two toilets with a shared pit, was the product most rarely sold.

[56] While businesses are free to set their own prices, WaterAid gave a suggested price structure during business training sessions.

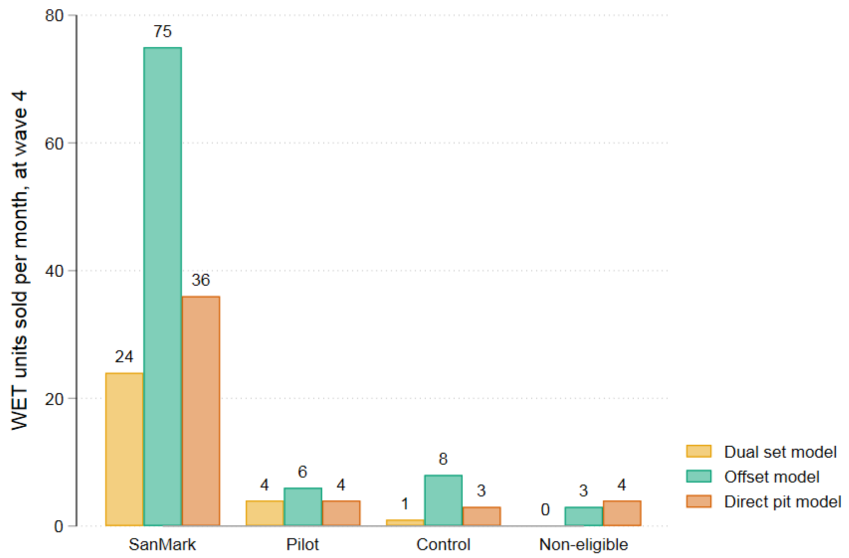
[57] As we discuss in the following subsection, only 12 SanMark or pilot businesses reported having made at least one WET sale in the past, according to our business survey. The missing business did not agree to be interviewed in our survey.

Figure 9.2. Monthly unit sales of WET products, by SanMark treatment group

(a) Total monthly WET sales, by SanMark experimental group and survey wave



(b) Total monthly WET sales, by model at wave 4



Note: We use the reported monthly sales of the three WET products proposed in the SanMark training sessions: dual set, offset and direct pit toilets. Total monthly sales in Figure 9.2(a) include all three WET products. Figure 9.2(b) shows details of sales reported at wave 4 by toilet model. A non-eligible business declared that it makes monthly WET sales, but did not tell us how many it had sold in total, so it was omitted from both panels.

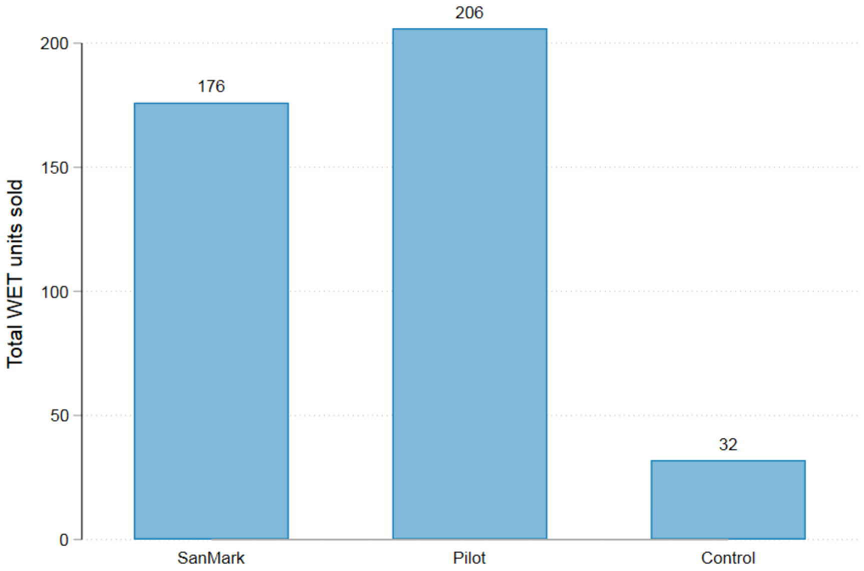
Source: Business survey, waves 3 and 4.

Total WET unit sales since the start of the intervention

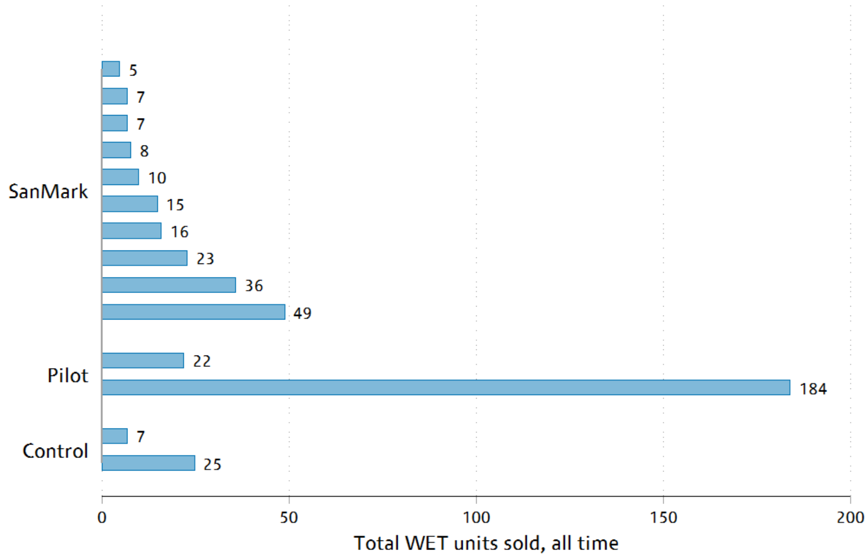
So far, we have discussed the number of WET sales that businesses reported making, on average, over the course of a month. The last survey wave also asked businesses about the total number of WET products sold since the start of the intervention. Figure 9.3(a) shows the aggregate results by SanMark treatment group and Figure 9.3(b) presents the results by individual businesses. Although only two out of six pilot businesses currently sell WET products, their cumulative sales (206 units) are higher than those of the ten SanMark businesses as a whole (176). This is partly because of the exceptional performance of one particular business, which alone accounts for 184 sales. Nonetheless, we should keep in mind that pilot businesses had been working with WaterAid under the SanMark framework for 9 months before the intervention was rolled out to the rest of the study area. Thus, they have received more support and guidance than any other businesses in our sample. Total sales in the SanMark group appear to be more evenly distributed than among pilot businesses, with total sales ranging from 5 to 49 units.

Figure 9.3. Total unit sales of WET products

(a) Total WET sales, by SanMark treatment group



(b) Total WET sales, by business



Note: We use reported total sales of the three WET products since the start of the intervention.

Source: Business survey waves 3 and 4.

Aggregate sales figures allow us to make inferences about the degree of success that the intervention may have had in our study area, and they contribute to our understanding of the lack of statistically detectable impacts at the household level, as discussed in Chapter 8. In total, businesses declare that they have sold 414 WET units since the start of the intervention. This is likely to be a lower-bound estimate, given that not all businesses agreed to participate in our last survey wave and because, as we will see in Section 9.2, some businesses that reported no sales appear, in fact, to have made WET sales via D2D sales agents. Nonetheless, the census conducted during 2014 showed that at least 50,000 households lived in our study area, meaning that less than one WET unit was sold for every 100 inhabitants.

While there are reasons to expect stronger sales performances in the future, as the project matures and businesses further adapt their operations to more seamlessly incorporate the WET product line, total WET sales at their current stage seem unlikely to make a significant dent in improving the sanitation gap of our study area. This is confirmed in our household-level analysis presented in Section 8.2.

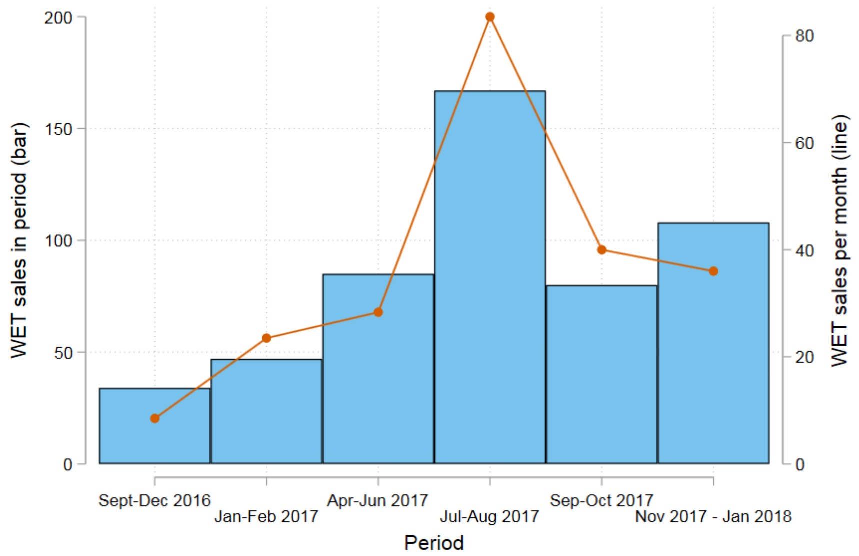
Evolution of WET sales over the study period

A final dimension of the analysis of SanMark's effectiveness is to study the evolution of WET sales over the period of study. Survey data include information about typical monthly WET sales and total WET sales since the start of the project, but these data do not allow us to look at how sales evolved between our two business survey waves. In order to do this, we use WaterAid's internal sales tracking data; WaterAid has collected total WET sales from SanMark businesses at relatively regular time intervals since September 2016.

WaterAid's sales tracking data include reported sales by both SanMark and pilot businesses over the course of six consecutive periods: September–December 2016, January–February 2017, April–June 2017, July–August 2017, September–October 2017 and November 2017 to January 2018. Sales tracking data are not available for the month of March 2017. Two concerns arise when using these data. The first is that, because the measurement periods differ in length, comparisons over time may be deceiving. We overcome this problem by presenting both total and average WET monthly sales in each period. The second concern is related to the internal programme nature of the data. This means that the data exclude sales conducted by control and non-eligible businesses, and the data might be measured with error if businesses believe they have incentives to over-report or under-report their sales performance to WaterAid. These concerns notwithstanding, we believe that studying the evolution over time using this data set is still an informative exercise.

Figure 9.4 plots the total and average monthly WET sales in each period, from WaterAid's internal sales tracking data set. WET sales show a slow improvement until April 2017, after which sales increase and peak during the July–August 2017 period. This increase follows the introduction of SanMark D2D sales agents and market-level activities. The fall in the pace of WET sales as of September 2017 should be not necessarily be a reason for concern, since it could be due to seasonality in households investments in toilets, however there is no longer term data to explore this further. Furthermore, having established above that five out of the 13 active businesses had stopped selling WET products by January 2018, it is important to assess further the sustainability of the programme, and whether sales are expected to increase in the future. Having said this, it is important to note that according to WaterAid practitioners working on SanMark, some of the initial "active businesses" were part-time businesses who dropped out due to their primary employment. Those who have stayed on are those businesses whose primary activity is concrete block producing, and running the business is the primary employment for the owners. This could mean that those businesses that stayed on will be the most motivated among those recruited and hence will continue to be engaged in the selling of WET toilets. At this stage, these considerations are speculative.

Figure 9.4. WET sales evolution from internal WaterAid sales tracking data



Note: Total WET sales (any model) reported by SanMark businesses to WaterAid as part of its internal sales tracking system. Blue bars (left axis) show total sales for each period of analysis. Because periods have different durations, the red line (right axis) presents average monthly sales within each period, for comparison purposes. Trends in total and monthly sales are almost identical if we exclude sales conducted by pilot businesses.

Source: WaterAid Nigeria Monitoring and Evaluation.

Drivers of technology adoption

Our survey of businesses revealed that, by December 2017, one out of every six SanMark businesses had made at least one WET sale in the past. Two of the six eligible businesses in the pilot LGA of Igbo Eze North, two of the businesses in the control group and one non-eligible business also reported that they have made sales. Using our detailed data on business characteristics, we can now look at whether any traits are particularly correlated with the adoption of WET technologies (which equals 1 if a business has made at least one successful sale of a WET model).

This analysis is conducted separately for two groups of businesses: SanMark and pilot businesses on one side, and control and non-eligible businesses on the other. This is because the determinants of adoption might be different for businesses invited to participate in the programme and for those actively excluded from participation. We look at the characteristics of the business owner (age, level of education and personality traits) as well as business-level markers such as business age, size, formality and access to infrastructure.

Table 9.1 presents the results of this analysis. Column 1 shows the correlates for WET adoption among businesses selected to participate in the SanMark programme. We find no correlation among owner characteristics or business characteristics. In column 2, we perform the same exercise for non-selected businesses, finding evidence to suggest that extraversion, one of the 'Big Five' personality traits assessed by our questionnaire, is positively correlated with WET adoption.[58] Nevertheless, the magnitude of this correlation is extremely small at 0.01 as can be seen in the table.

In addition to the individual tests for each coefficient, we conducted joint tests of statistical significance for different sets of characteristics that might have significant predictive power as a group. We tested whether the owner's characteristics (i.e. those included in Panel A of Table 9.1) are jointly significant, finding that they are not: the *p*-value for this set of regressors is 0.72 and 0.40 for selected (SanMark and pilot) and not selected (control and non-eligible) businesses, respectively. The same is true when considering the 'Big Five' personality traits of the owners on their own (i.e. excluding owner age and education), with *p*-values of 0.63 and 0.28. Business characteristics are also jointly statistically insignificant in predicting WET adoption, with *p*-values of 0.79 and 0.81. Finally, we find that we cannot reject the hypothesis that the whole set of characteristics included in Table 9.1 has no significant correlation with WET adoption, with *p*-values of 0.57 and 0.67 for each set of businesses. All in all, the table shows little evidence that WET adoption is correlated with any of the observable characteristics considered.

[58] The 'Big Five' personality traits are: (i) Openness to experience; (ii) Conscientiousness; (iii) Extraversion; (iv) Agreeableness; and (v) Neuroticism.

Table 9.1. Correlates of technology adoption

Outcome: sale of at least one WET unit	**SanMark and pilot businesses (1)**	**Control and non-eligible businesses (2)**
Panel A. Owner characteristics		
Age of owner	0.08 (0.10)	0.01 (0.02)
Highest education level completed: secondary school	−0.13 (0.27)	0.04 (0.07)
Highest education level completed: tertiary or above	0.06 (0.29)	0.06 (0.07)
Big Five: openness	0.00 (0.02)	−0.01 (0.00)
Big Five: conscientiousness	0.02 (0.03)	0.00 (0.01)
Big Five: extraversion	−0.03 (0.02)	0.01* (0.01)
Big Five: agreeableness	0.01 (0.03)	−0.01 (0.01)
Big Five: neuroticism	0.03 (0.02)	−0.00 (0.01)
p-value *F*-test (owner characteristics)	*0.72*	*0.40*
p-value *F*-test (owner personality only)	*0.63*	*0.28*
Panel B. Business characteristics		
Years in existence	0.00 (0.01)	0.00 (0.00)
Number of full-time employees	0.01 (0.02)	0.00 (0.00)
No means of delivering products (pick-up only)	−0.39 (0.28)	0.04 (0.05)
Business has sales book	−0.15 (0.20)	−0.01 (0.05)
Some formal registration	−0.04 (0.32)	0.04 (0.05)
Connected to power grid	−0.14 (0.21)	−0.05 (0.06)
p-value *F*-test (business characteristics)	*0.79*	*0.81*
Number of businesses	46	86
p-value *F*-test (owner and business characteristics)	*0.57*	*0.67*

Note: All variables are measured at wave 2, before the roll-out of SanMark. The dependent variable is a dummy equal to 1 if the business has reported having sold at least one WET product in any survey wave. The regressions include LGA fixed effects. Standard errors in parenthesis. Stars indicate statistical significance: *$p < 0.10$, **$p < 0.05$ and ***$p < 0.01$.

9.2 The role of D2D sales agents

The second phase of the SanMark intervention, rolled out in March 2017, included the recruitment, training and deployment of WET D2D sales agents. They worked on commission with any SanMark business of their choice and they visited households to inform them about the risks of unimproved sanitation and to promote the advantages of the WET product line using aspirational drivers. They were recruited from SanMark communities (i.e. areas where SanMark marketing events were also carried out), but they were allowed to visit households in any community they wanted. We interviewed D2D sales agents during our last survey wave, to measure the success of the recruitment and training stages and to gauge their importance as vehicles of WET awareness and sales promotion.

Table 9.2 presents a series of descriptive statistics for the D2D sales agents interviewed as part of wave 4. Almost all of them report that they have attended WaterAid training sessions, but only half (78 respondents) actually conducted any visits during the last month. We classify these respondents as active agents for the purposes of the following four questions, which are only directed at them. Active agents report that D2D sales are rarely their primary economic activity; however, they do this activity for an average of 2.5 days a week and for 3 hours per day on average. Repeat visits appear to be an important part of their job, possibly due to the follow-up and persistence required to persuade a household to proceed with the investment.

Table 9.2. Sales activities of D2D sales agents

Variable	Mean (1)	SD (2)	Minimum (3)	Maximum (4)
Panel A. D2D sales training and activity				
Attended D2D sales training sessions (%)	97.56	–	–	–
Active = at least one D2D visit in past month (%)	47.56	–	–	–
If active: is this their primary economic activity? (%)	5.13	–	–	–
If active: days a week they perform D2D visits	2.51	1.09	1	6
If active: hours spent in D2D visits, per day	3.24	1.44	1	8
If active: visited households more than once (%)	94.87	–	–	–
Panel B. WET sales conducted by agent				
Participated in the sale of a WET product (%)	29.88	–	–	–
Number of WET products sold through agent's activities	3.80	4.42	1	25
Sales done through market-level events	1.71	3.43	0	13
Number of suppliers that agent worked with	1.65	1.30	1	7
Panel C. Commissions from WET sales				
Received a commission for at least one sale (%)	81.63	–	–	–
Direct pit model – commission per unit sold (in US$)	2.80	3.65	0	13.06
as % of final price	*5.55*	*4.65*	*0*	*10*
Offset model – commission per unit sold (in US$)	3.36	4.04	0	20.00
as % of final price	*6.56*	*3.92*	*0*	*10*
Dual set model – commission per unit sold (in US$)	3.32	3.80	0	10.56
as % of final price	*4.40*	*4.71*	*0*	*10*
Average monthly full-time wage equivalent (in US$)	7.98	6.31	1.59	27.66
Number of observations	164			

Note: Mean values from the D2D agent survey conducted as part of wave 4. US$ amounts are expressed in dollar equivalents of reported amounts received in naira (₦), converted at the prevailing exchange rate during November 2017 of 360₦ = US$1. Monthly full-time wage equivalent calculated using reported hours and days worked, and assuming an average commission per WET unit sold of US$3.30. For reference, the Nigerian minimum wage now stands at US$60.

Out of the whole sample of agents, 30% (49 agents) have participated in a successful sale, meaning one in which the household completed the transaction. These successful sales agents report that they have participated in the sale of 4 units, on average, and a total of almost 200 WET sales over the whole study period.[59] This is equal to around half of all WET sales recorded in our business survey, or one out of every three sales recorded by

[59] The distribution of WET sales by D2D agent is highly skewed. Thus the average sales figure might not be representative of the true performance of the average agent. The median and modal number of sales by D2D agents was in fact 2 units. Mean sales rise to 4 units due to the influence of two highly performing agents who reportedly sold 15 and 25 units.

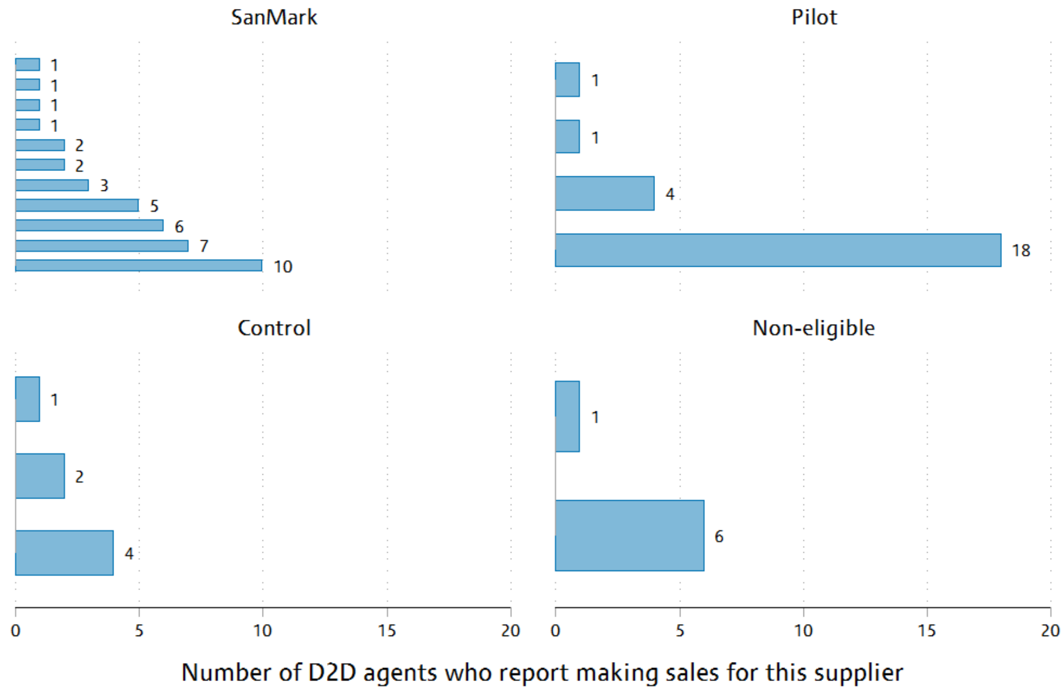
WaterAid's internal tracking data. This highlights the potential importance of D2D agents' work in supporting WET distribution. D2D sales agents further report that around half of the sales they are involved in resulted from community-level marketing activities organised as part of the SanMark intervention, and not individual household visits. Finally, agents report that they work with an average of two WET businesses – and up to seven different businesses in the case of one agent – highlighting the freedom agents have in terms of whom they work with.

As intended, the sales are mostly resulting in commissions being paid to the agents, with over 80% of them reporting that they are paid after a successful transaction. This is reassuring, as it is the means of providing incentives for the sustainability of their work. Average commissions range from 4% to 7% of the final price of the product.

The last row of panel C in Table 9.2 shows an estimated equivalent of the monthly wage being received by these agents. Using the hours and days worked by each agent, we construct a variable of total time worked, assuming that their marketing activities began on 14 April 2017 and lasted until the day of the interview. Based on information about the commissions paid to the agent, which includes the value of the commissions received, we further assume that the average commission paid for each WET unit sold is US$3.30. Using these values, we estimate the agent's hourly return (i.e. the level of commission an agent received for each hour worked) and we multiply it by the standard number of hours in a working month, in order to calculate the full-time monthly equivalent wage. We find that the mean monthly wage equivalent for the agents in our sample is US$8, less than a fifth of the current legal minimum wage of US$60 (₦18,000). Although we know that the job a D2D sales agent carries out does not generally constitute a household's main economic activity, and therefore WET commissions are only supplemental income, the low rate of pay per hour worked raises questions about the sustainability of the scheme. This may be one of the reasons behind the fact that 52% of agents reported that they had not conducted any marketing activities in the month before the interview.

Agents were also asked to identify the three main businesses for which they had made sales. Using this information, we can get an understanding of which businesses make use of D2D sales agents and the number used. Figure 9.5 synthesises these findings, splitting businesses named by D2D sales agents according to their SanMark treatment status. It is interesting to see that many businesses work with more than one D2D agent. For example, the maximum is 10 agents for a SanMark business, 18 for a pilot business, 4 for a control business and 6 for a non-eligible business. We note that while SanMark and pilot businesses were at the front line in adopting D2D sales agents (accounting for 39 and 24 agents, respectively), control and non-eligible businesses have also resorted to them for their sales.

Figure 9.5. Number of agents who reported selling WET products from each business



Note: D2D sales agents were asked to name the three businesses for which they sold most WET products. Each bar above represents a different business. The length of the bar is equal to the number of agents that mentioned that business as one of these three. Businesses are grouped according to their SanMark treatment assignment. In the top-right panel, we show suppliers located in Igbo Eze North, where SanMark was piloted.

Source: D2D agent survey, wave 4.

Not all businesses that reported selling WET products have worked with D2D sales agents. This is not problematic in itself; the collaboration was not compulsory for any of the parties involved. However, we also find that agents report that they have made sales for businesses that show no record of WET sales in the business survey. This misreporting may be intentional, or it may originate from the agents (who may have mistaken the name of the business they worked for) or from businesses (which either declined to be interviewed or did not remember making WET sales that they indeed made).

Table 9.3 uses information from both survey instruments, the business survey and the D2D sales agent survey, and tabulates the number of suppliers with successful WET sales by survey instrument, assuming that the reporting by each party was accurate. Combining both sources can improve the measurement of the total number of businesses that have positive sales of WET products. Using these figures, we find that out of the 60 businesses invited to participate in the SanMark intervention (excluding pilots), a total of 14 (23%) of them are reported to have conducted at least one sale of a WET toilet, according to the information from either the business (10 businesses with sales) or D2D agent surveys (12 businesses with sales). Two of the four SanMark businesses captured by the D2D agent survey and not by the business survey were not captured in our business survey because they did not agree to be interviewed in survey wave 4. In the same way, four pilot businesses, four control businesses and two non-eligible businesses appear to have made

WET sales, when combining data from both surveys.[60] This moderate increase in uptake rates when using the combined sources of information still leaves WET adoption rates at relatively low levels, however, and it does not alter the impression that there are still significant hurdles to the adoption of the WET product line by the businesses in our study area.

D2D sales agents were also asked to declare the number of sales made via each of their three main business partners. Using these answers, we calculated the total number of sales that D2D sales agents reported to have made through each business. A total of 213 WET sales were reported by agents in this way, equivalent to around half of the total sales reported by businesses themselves, as seen in the previous section.[61] Figure 9.6 presents businesses' WET sales by SanMark treatment status, and by individual business.

[60] Most of these businesses selling WET products that did not show up in our business survey can be explained by businesses not agreeing to be interviewed in wave 4. The two control businesses and the two non-eligible businesses that reported to sell WET products via D2D sales agents and which are missing in our business survey indeed failed to respond it. This is also the case for one of the two pilot businesses.

[61] This could be a lower-bound estimate, as five of the agents reported to have worked with more than three businesses, and these sales would have not been recorded in this set of questions. However, when asked about the total number of successful WET sales they were involved in for any business, the same set of agents report that they have sold a total of 186 units; so, some level of misreporting is indeed present.

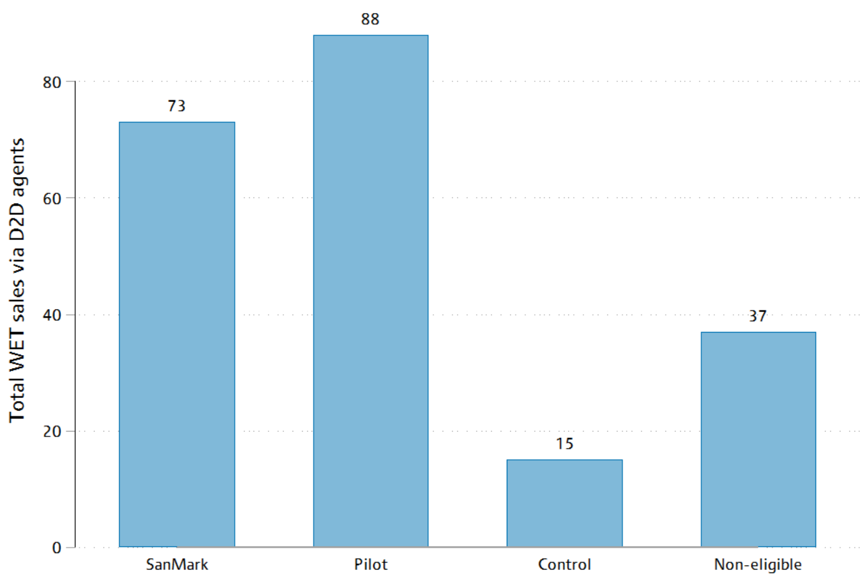
Table 9.3. Businesses with recorded WET sales by data source

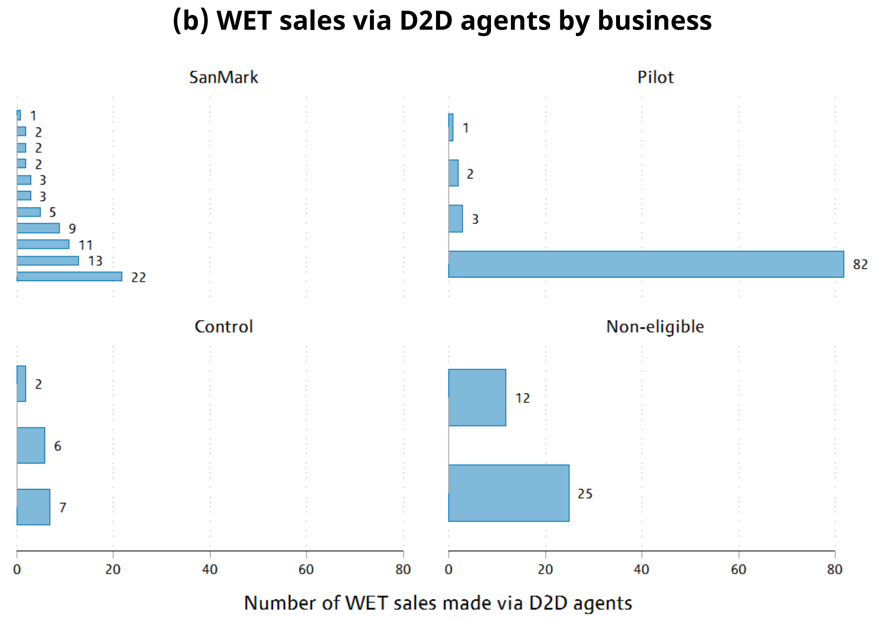
Reported WET sales in survey of businesses	Reported WET sales in survey of D2D sales agents	
	Yes	No
SanMark businesses		
Yes	8	2
No	4	46
Pilot businesses		
Yes	2	0
No	2	2
Control businesses		
Yes	1	1
No	2	65
Non-eligible businesses		
Yes	0	2
No	0	59

Note: Each cell shows the number of businesses reported to have made WET sales according to our two available data sources: the survey of businesses (rows) and the survey of D2D agents (columns). Businesses are presented by SanMark treatment group. For every category, businesses that reported no WET sales or refused to be interviewed are classified as 'No'.

Figure 9.6. Total unit sales of WET products via D2D agents, by SanMark group

(a) WET sales via D2D agents by treatment group





Note: D2D sales agents were asked to name the three businesses for which they sold most WET products, and the number of units sold from each. In (a), we plot the total number of sales made via D2D agents by business treatment group. In (b), we present disaggregated sales figures by individual businesses in each group. Businesses are grouped according to their SanMark treatment assignment.

Source: D2D agent survey, wave 4.

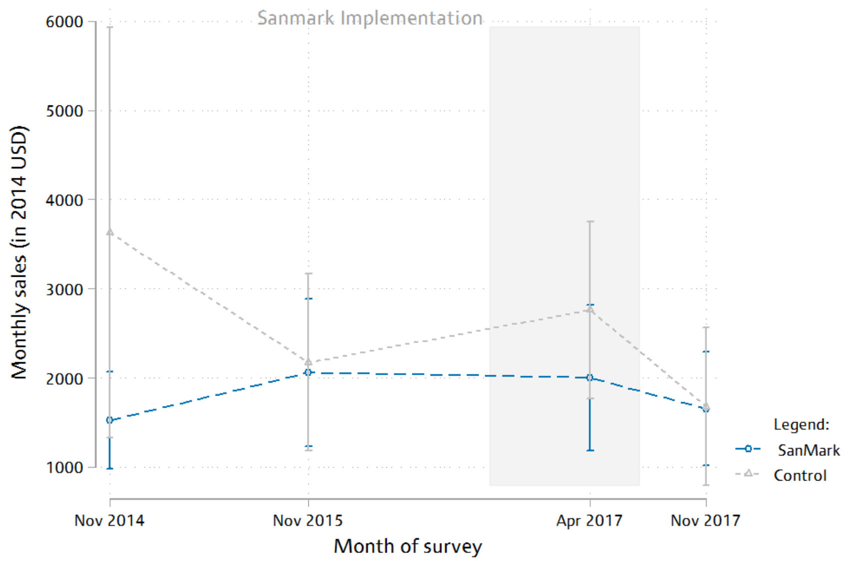
9.3 Comparing SanMark and control businesses

Figure 9.7 plots the unconditional means of monthly revenues and costs, for both SanMark and control businesses, over the course of four survey waves. We can observe no statistically significant differences in these four indicators for business performance, between businesses invited to participate in SanMark, and control businesses, before, during and after the intervention. The large confidence intervals reveal a high degree of variation across businesses, particularly in self-reported monthly costs and monthly concrete block sales.

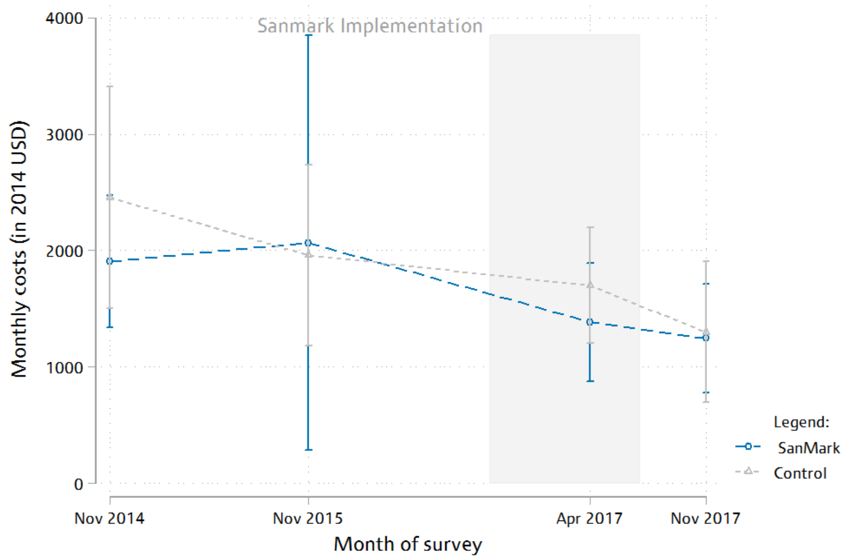
To investigate this further, we next carry out a regression analysis using specifications (6) and (7) in Section 4.4. We are interested in whether being selected into the SanMark intervention was associated with any improvements in business performance measured by revenues and costs. We use pooled estimates over the whole study period in a first instance, as well as the period-specific estimates, as they could potentially shed light on the relative effectiveness of the different phases of the intervention. Recall that the supplier phase of SanMark was rolled out in September 2016, market-level activities and D2D sales agents only began operating in April 2017, after our third supplier interview had been conducted.

Figure 9.7. Evolution of main business outcomes by wave

(a) Monthly revenues, in 2014 US$



(b) Monthly costs, in 2014 US$



Note: Lines plot average values for each variable, by survey wave and SanMark treatment group, in 2014 US$. Revenues and costs refer to the entire business, not limited to WET sales. The scale of the x-axis corresponds to the amount of time elapsed between each survey wave. The SanMark intervention, highlighted in grey, was conducted between September 2016 and June 2017. Capped bars indicate 95% confidence intervals.

Source: Business survey waves 1-4.

It is important to note that, because the small businesses in our sample operate in connected markets, and potentially compete with each other, we are just comparing the outcomes between SanMark and control businesses. This is different from estimating programme impacts, because improvements in the performance of one business might (or might not) result in a reduced performance among competitors. A positive difference in the performance of SanMark businesses relative to control businesses could be therefore driven by SanMark businesses capturing the market share from control businesses. So, we cannot distinguish between positive causal impacts on SanMark businesses and negative externalities on control businesses. Generally, the presence of negative externalities depends on whether businesses compete over a single, relatively isolated market, or whether they can tap into other markets. However, if there were no differences between SanMark and control businesses, then this would also be an informative result; it would mean that there are no significant negative externalities of the intervention on control businesses. And indeed, we cannot rule out that the programme had identical impacts (positive or negative) on both SanMark treatment and control groups.

Table 9.4. Firm performance by SanMark treatment group

Outcome =:	**Monthly sales (US$), in logs**	**Monthly costs (US$), in logs**	**Quantity sold (concrete blocks), in logs**	**Closure (=1 if not active)**
	(1)	**(2)**	**(3)**	**(4)**
Panel A. Pooled estimates				
Treated (τ)	0.06	0.03	−0.04	0.05
p-value (unadjusted)	(0.78)	(0.84)	(0.82)	(0.32)
p-value (adjusted)†	(0.98)	(0.98)	(0.98)	(0.74)
Panel B. Estimates by period				
Treated × Wave 3 (τ_3)	-0.17	-0.26	-0.09	0.03
p-value (unadjusted)	(0.53)	(0.24)	(0.71)	(0.67)
p-value (adjusted)†	(0.89)	(0.59)	(0.89)	(0.89)
Treated × Wave 4(τ_4)	0.30	0.36	0.01	0.07
p-value (unadjusted)	(0.27)	(0.12)	(0.98)	(0.30)
p-value (adjusted)†	(0.60)	(0.34)	(0.98)	(0.60)
Number of observations	146	159	164	204
From wave 3	*77*	*83*	*83*	*102*
From wave 4	*69*	*76*	*81*	*102*

Note: p-values in parenthesis. † Adjusted p-values account for multiple hypothesis testing following the procedure described in Romano & Wolf (2005), with 1,000 clustered bootstrap samples. Stars indicate statistical significance: *p < 0.10, **p < 0.05 and ***p < 0.01, according to the adjusted p-values.

The first row of Table 9.4 presents our pooled estimates and shows no statistically significant differences between SanMark and control businesses. Disaggregated estimates, in Panel B, show similar, non-significant differences between the two groups. At wave 4, SanMark businesses seem to experience both higher sales and costs than their control counterparts, although the difference is not statistically significant at conventional levels.

A large share of the businesses interviewed agreed to answer most questions, but were reluctant to provide (or did not have knowledge of) sales and costs figures. For this reason, sample size is reduced when studying these outcomes, leading to noisy, imprecise estimates. Of the 129 businesses in our study assigned to either SanMark or control groups, the highest number of valid responses obtained was 83, as seen in the last two rows of the table. This represents a response rate of 64% and is balanced across treatment groups. The response rate falls to a minimum of 53% for the case of sales data in wave 4. The problem of non-response leads to lack of statistical precision. However, other modules in our supplier questionnaire had higher response rates and may allow us to take a second look at the evolution of the two groups of businesses.

The final column of Table 9.4 presents results from an outcome for which more valid responses are available – whether a business has closed. Note that this is different from not being available for interview, as in this case the enumerators could not confirm whether the business was still in operation or not. We have 102 valid responses in each wave for this module, out of a total of 129 study businesses, which represents a response of 79%. No significant differences appear between SanMark and control groups, meaning that the SanMark intervention did not affect the rates of exit from the sector differently across treatment groups.

An area of interest when introducing a new technology into a market is whether this sparks innovation in the sector. Innovation may include offering a wider variety of products, adopting a different business model, or embarking on new marketing activities. In the case of SanMark, the mould needed to cast the offset and dual set WET products was accessible, at no cost, for SanMark businesses only. It might be interesting, then, to see whether this resulted in higher levels of innovation in other fronts among this group, and whether businesses in the control group reacted by increasing innovation on other fronts, as a response to the new product available to their competitors.

Table 9.5. Firm behaviour by SanMark treatment group

Outcome:	**Introduced new products (1=Yes)** (1)	**Product variety (1=Min, 13=Max)** (2)	**Innovation score (1=Low, 13=High)** (3)	**Sells in instalments (1=Yes)** (4)
Panel A. Pooled estimates				
Treated (τ)	0.05	0.05	0.02	0.05
p-value (unadjusted)	(0.25)	(0.88)	(0.97)	(0.50)
p-value (adjusted)†	(0.63)	(0.99)	(0.99)	(0.85)
Panel B. Estimates by period				
Treated × Wave 3 (τ_3)	-0.01	0.30	-0.01	0.04
p-value (unadjusted)	(0.92)	(0.53)	(0.99)	(0.69)
p-value (adjusted)†	(0.99)	(0.93)	(0.99)	(0.96)
Treated × Wave 4(τ_4)	0.11	-0.19	0.05	0.06
p-value (unadjusted)	(0.07)	(0.68)	(0.94)	(0.54)
p-value (adjusted)†	(0.21)	(0.90)	(0.95)	(0.87)
Number of observations	204	204	204	204
From wave 3	*102*	*102*	*102*	*102*
From wave 4	*102*	*102*	*102*	*102*

Note: p-values in parenthesis. † Adjusted p-values account for multiple hypothesis testing following the procedure described in Romano & Wolf (2005), with 1,000 clustered bootstrap samples. Stars indicate statistical significance: *p < 0.10, **p < 0.05 and ***p < 0.01, according to the adjusted p-values.

Error! Reference source not found. explores whether SanMark generated a statistically significant difference in innovation outcomes between treatment and control, using several outcomes related to business innovation. The results show no difference between treatment and control businesses along a number of innovation indicators at no point in time measurements were taken. The first column presents the results of using a simple binary variable equal to 1 if the respondent declares that they have introduced a new product to their product line over the past 12 months. Column 2 conducts a similar ordinary least squares regression on the number of products offered by the business. We construct this measure by counting the number of different products that each business declared to sell on a regular basis, on each survey wave, from a minimum of 1 to a maximum of 13. In column 3, we construct a similar indicator but using 13 questions on different innovations, both at the product level and at the business administration level. These include questions about the addition of new products, changes in the organisation of sales and relationships with customers, and changes in the business model. The results in column 4 are from the use of a simple binary variable = 1 if the business offers customers the option of paying for their products in instalments, and we interpret this as a measure of innovation in sales strategies. None of these measures seems to show any statistically significant differences, on average, between SanMark and control businesses.

10. Limitations of our study

Our research design allowed us to estimate the causal impact of the two interventions on household toilet ownership and usage. In survey waves 1 and 4, the self-reporting of toilet ownership was validated by observations from enumerators. However, the practice of OD was not observed or measured objectively, which is a limitation of our data collection exercise. However, the similar evolutions of toilet ownership, which was verified, and OD, which was not, reassure us that the error in our OD measure is small.

Our findings regarding CLTS will nonetheless be limited to the specific version of CLTS implemented in our study area. As discussed in Section 3.1, there was significant variation in the choice of activities conducted by CLTS facilitators, and follow-up activities were directed at monitoring improvement, not at reinforcing the CLTS message. Finally, while the design of the household clusters for our study was intended to minimise information spillovers, the lack of accurate GPS data for the study area before the project made it impossible to construct precise distance buffers between clusters. This should be considered when comparing the results presented in this study with those of studies of similar CLTS interventions conducted in different contexts.

Regarding SanMark, we should note that this is still a young intervention, and that the last data collection wave occurred only 8 months after the roll-out of SanMark's marketing activities component. Additionally, the delays in the SanMark roll-out meant that it could not be in place straight after the CLTS triggering meetings, possibly reducing the chance of observing any positive impacts from combining both interventions.

11. Conclusions

This report discusses the design and results of the impact evaluation of two of WaterAid Nigeria's main interventions within the STS Nigeria project, CLTS and SanMark, implemented in the states of Ekiti and Enugu. CLTS was rolled out in our study areas during the first half of 2015. It was followed by the introduction of SanMark's first phase in September 2016 and SanMark's second phase 7 months later. Both interventions aim at increasing the level of improved toilet ownership and its sustained usage, with the final goal of eliminating community-wide OD.

Our analysis reveals no impacts of CLTS on average. However, when looking at the poorest half of the studied communities (defined based on an asset wealth index), we find that CLTS had strong and sustained positive impacts on toilet ownership and negative impacts on OD. In the last survey wave, conducted almost 3 years after the start of CLTS activities, households in poor treated TUs were 10pp more likely to own a functioning toilet (of any kind), 7pp more likely to own an improved toilet, and 9–10pp less likely to report that the main respondent or any member of the household performs OD, than households in non-treated communities. No effects are detected among richer communities, which results in the lack of impacts over the whole sample.

To the best of our knowledge, community-level wealth as a mediating factor for CLTS effectiveness has not been established in the CLTS literature thus far. In the context of this study, we find that community-level wealth is a stronger predictor of CLTS impacts than community characteristics that are deemed beneficial in the CLTS Handbook (Kar and Chambers, 2008), such as initial conditions of toilet coverage, social capital and religious fragmentation. Additionally, policymakers should note that CLTS effectiveness is more related to low community-level wealth than to low household-level wealth. This finding is in line with the Handbook's observation that community characteristics play a crucial part in determining the effectiveness of CLTS, a community-level intervention in spirit.

Interestingly, the impacts of CLTS on toilet ownership are mirrored inversely by the impacts of CLTS on OD. In other words, no reductions in OD are detected among households who already had toilets when the intervention began. This finding is primarily driven by high usage rates rather than by CLTS being inefficient in getting toilet owners but non-users to use their latrine. By the time of our last data collection wave, only 5% of households who owned a functioning toilet declared that they have at least one household member performing OD at least sometimes. On the contrary, 96% of the households with no functioning toilet declared that they practise OD. This means that 95% of households who perform OD do not own toilets.

SanMark's phase 1, business-level, activities were rolled out 15 months after the last CLTS triggering meeting, and SanMark's phase 2 activities started 7 months after that. By the time of our last survey, 13 months after the roll-out of the SanMark intervention at the business level and 7–9 months after the deployment of D2D sales agents, we found no significant impacts of the SanMark intervention on toilet ownership at the household level. This is also true for areas exposed to both CLTS and SanMark activities, which, by the end of 2017, exhibit toilet ownership and OD rates statistically indistinguishable from those observed in similar areas that were not exposed to either intervention. When restricting the analysis to poor TUs only, where CLTS was found to have had large impacts, we also find that SanMark did not contribute in a statistically significant way to increase toilet ownership or ownership of improved toilets. We believe that the difficulties in successfully introducing the SanMark intervention and the relatively short implementation period are likely to be important drivers of the lack of impact at the time of the endline survey.

Our descriptive analysis shows that at the time of writing this report, while uptake of the WET amongst households was low, WET products were increasingly being adopted by businesses, with one out of every six businesses selected to participate in the programme now declaring that they make WET sales as part of their regular monthly sales. Awareness of the WET product line is high among all businesses in our sample, and indeed a small number of non-selected (control and non-eligible) businesses appear to have started selling WET products as well. D2D sales agents played a significant role in facilitating sales: they were involved in one out of two successful sales, and they have generally been compensated for their work by the businesses through commission. At the same time, while WET products among households in our study area remain rare (less than 1% of households in our sample owned WET products by December 2017), we also find that awareness of the product is increasing over time. These observations leave room for optimism that uptake of the WET products might accelerate in the months following our study period. Studies from developed countries support our impression that one year might still be too short a horizon to appropriately gauge the success of the SanMark intervention in diffusing the WET product in this context. In their seminal study on product life cycles in the US, Golder and Tellis (1997) estimate a hazard model using a data set of new consumer durables and find that sales take off, on average, 6 years after their introduction to the market.

At the same time, the new evidence generated in this study on the patterns of toilet adoption suggests that households continue to face constraints to adopting sanitation technology. WET products are more affordable than other improved toilet models on the market and are more water efficient. The majority of households who were aware of WET products indeed perceived them as more affordable, water efficient and aesthetically pleasing. However, financial constraints seem to be one of the main reasons behind the low adoption of WET products. Indeed, D2D sales agents report that the main reason why households they visited do not purchase WET products is their lack of funds. This might, of course, be influenced by the sample of households these agents chose to visit, but it is nonetheless an important data point, which is confirmed by our household survey data: 85% of households without toilets declared that the main reason why they have not constructed a toilet (of any kind) is that it is too expensive or they cannot afford it. Other demand-side constraints might also be important, as suggested by the fact that 39% of the households with above-median asset wealth still lacked a functioning toilet by December 2017.

Given that close to half the households in our sample still perform OD, what does this evidence tell us about the constraints for reducing OD in Nigeria? Our results point towards a nuanced understanding of the effectiveness of the interventions studied and of the sanitation challenge ahead. Eliminating OD in rural Nigeria is intimately tied to increasing the rates of toilet ownership. We have also established the importance of the enabling environment in which the interventions are conducted. CLTS and other community-driven policies may not be appropriate for all contexts: half of the households in our sample experienced no detectable improvements as a result of CLTS. Better targeting of sanitation policies such as CLTS should take into account their limitations, as well as their advantages.

A majority of the households without a toilet in both rich and poor TUs report that toilets are too expensive or that they cannot afford to build one. This is true even after the new WET model was introduced by the SanMark intervention. At the same time, the share of households saying they cannot afford toilets *fell* as a result of CLTS in poor communities, suggesting that CLTS has corrected their cost-benefit ratio of owning a toilet downwards or has nudged households towards more affordable toilet models. We cannot say whether additional reinforcement of the CLTS message in poor communities would lead to further growth in toilet ownership rates, but it is certainly unlikely to be effective in rich communities, where CLTS had no impact. Other policy alternatives should be considered for these other areas, if serious improvements are sought. Policymakers may also want to explore interventions that alleviate financial constraints (for example, targeted or non-targeted subsidies, perhaps tied to toilet construction, or subsidised credit) in order to foster toilet ownership and eliminate OD in rural households. This type of intervention may complement or substitute for CLTS or SanMark interventions.

References

Abramovsky, L., Augsburg, B., & Oteiza, F. (2015). *Sustainable Total Sanitation - Nigeria: Baseline Report.* Institute for Fiscal Studies , Centre for the Evaluation of Development Policies, London.

Abramovsky, L., Augsburg, B., & Oteiza, F. (2017b). *Sustainable Total Sanitation – Nigeria: First SanMark Impact Evaluation Report.* Institute for Fiscal Studies, Centre for the Evaluation of Development Policies, London.

Abramovsky, L., Augsburg, B., & Oteiza, F. (2016a). *Sustainable Total Sanitation - Nigeria: Rapid Assessment Report.* Institute for Fiscal Studies, Centre for the Evaluation of Development Policies, London.

Abramovsky, L., Augsburg, B., & Oteiza, F. (2017a). *Sustainable Total Sanitation – Nigeria: Second CLTS Impact Evaluation Report.* Institute for Fiscal Studies, Centre for the Evaluation of Development Policies, London.

Abramovsky, L., Augsburg, B., & Oteiza, F. (2016b). *Sustainable Total Sanitation - Nigeria: Supplier Baseline Report.* Institute for Fiscal Studies, Centre for the Evaluation of Development Policies, London.

Abramovsky, L., Augsburg, B., Flynn, E., & Oteiza, F. (2016c). *Improving CLTS targeting: evidence from Nigeria.* London: Institute for Fiscal Studies.

Borja-Vega, C. (2014). "The effects of the Total Sanitation and Sanitation Marketing programme on gender and ethnic groups in Indonesia". Waterlines. Vol. 33, No. 1, Assessing impact in WASH (January 2014), pp. 55-70

Cairncross, S. (2004). The case for marketing sanitation. *Water and Sanitation Program-Africa* .

Cameron, L. A., Olivia, S., & Shah, M. (December de 2015). Initial Conditions Matter: Social Capital and Participatory Development. *IZA Discussion Paper no. 9563* .

Cameron, L., Shah, M., & Olivia, S. (2013). Impact Evaluation of a Large-Scale Rural Sanitation Project in Indonesia. *World Bank Policy Research Working Paper* .

Clasen, T., Boisson, S., Routray, P., Torondel, B., Bell, M., Cumming, O., et al. (2014). Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: a cluster-randomised trial. *Lancet Global Health , 2*, 645-653.

Cronin, A. A., Gnilo, M. E., Odagiri, M., & Wijesekera, S. (2017). Equity implications for sanitation from recent health and nutrition evidence. *International journal for equity in health*.

Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: a toolkit. En T. P. Schultz, & S. J. A., *Handbook of Development Economics* (Vol. 4, págs. 3895-3962). Amsterdam: North-Holland.

Evans, W., Pattanayak, S., Young, S., Buszin, J., Rai, S., & Bihm, J. W. (2014). Social marketing of water and sanitation products: A systematic review of peer-reviewed literature. *Social Science and Medicine* , 18-25.

Filmer, D., & Pritchett, L. (2001). Estimating wealth effects without expenditure data-or tears: an application to educational enrollments in states of India. *Demography* , 115-132.

Gertler, P., Shah, M., Alzua, M. L., Cameron, L., Martinez, S., & Patil, S. (2016). How Does Health Promotion Work? Evidence from the dirty business of eliminating open defecation. *CCPR Working Paper Series* .

Geruso, M. & D. Spears. (2018). "Neighborhood Sanitation and Infant Mortality," *American Economic Journal: Applied Economics*, vol 10(2), pages 125-162.

Godfrey, S., Van der Velden, M., Muianga, A., Vigh, M., Gunning, J. W., & Elbers, C. (2014). Impact study of the One Million Initiative rural water and sanitation programme in Mozambique. *Waterlines*.

Golder, P. N., & Tellis, G. J. (1997). Will it ever fly? Modelling the takeoff of really new consumer durables. *Marketing Science* .

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

Hammer, J., & Spears, D. (2013). Village sanitation and children's human capital : evidence from a randomized experiment by the Maharashtra government. *WB Impact Evaluation series , WP 6580*.

IFS, & WaterAid. (2014). *Grant proposal: formal research component of the project "Sustainable Total Sanitation in Nigeria - implementation, learning, research, and influence on practice and policy".* London.

JMP. (22 de 11 de 2017). *Data - Nigeria*. (WHO; UNICEF). Available from Joint Monitoring Programme: https://washdata.org/data#!/nga

Kar, K. (2003). Subsidy or self-respect? Participatory total community sanitation in Bangladesh. *IDS Working Paper , 184*.

Kar, K., & Chambers, R. (2008). *Handbook on Community-Led Total Sanitation.* Brighton: Institute of Development Studies.

Mara, D., Lane, J., Scott, B., & Trouba, D. (2010). Sanitation and health. *PLOS Medicine* , 1-7.

Mauro, P. (1995). Corruption and Growth. *The Quarterly Journal of Economics* , 681-712.

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

McKenzie, D. (2012). Beyond Baseline and Follow-up: The Case for more T in Experiments. *Journal of Development Economics* , 210-221.

McKenzie, D. (2017). Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition. *American Economic Review , 107* (8), 2278-2307.

Mondal, D., Minak, J., Alam, M., Liu, Y., Dai, J., Korpe, P., et al. (2011). Contribution of enteric infection, altered intestinal barrier function, and maternal malnutrition to infant malnutrition in Bangladesh. *Clinical Infectious Diseases , 54* (2), 185-192.

Neal, A., Yeo, G., Koy, A., & Xiao, T. (2012). Predicting the form and direction of work role performance from the Big 5 model of personality traits . *Journal of Oganizational Behaviour* , 175-192.

Patil, S., Arnold, B., Salvatore, A. L., Briceno, B., Ganguly, S., Colford, J., et al. (2014). The Effect of India’s Total Sanitation Campaign on Defecation Behaviours and Child Health in Rural Madhya Pradesh: A Cluster Randomized Controlled Trial. *PLOS Medicine* .

Petri Jr, W. A., Miller, M., Binder, H. J., Levine, M. M., Dillingham, R., Guerrant, et al. (2008). Enteric infections, diarrhea, and their impact on function and development. *The Journal of Clinical Investigation , 118* (4), 1277.

Pickering, A. J., Djebbari, H., Lopez, C., Coulibaly, M., & Alzua, M. L. (2015). Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali: a cluster-randomised controlled trial. *Lancet Global Health , 3*, 701-711.

Prüss, A., Kay, D., Fewtrell, L., & Bartram, J. (2002). Estimating the Burden of Disease from Water, Sanitation, and Hygiene at a Global Level. *Environmental Health Perspectives* , 537-542.

Robinson, A. (2009). *Sustainability and equity aspects of total sanitation programmes: A study of recent WaterAid-supported programmes in Nigeria.* London: WaterAid.

Romano, J. P., & Wolf, M. (2005). Stepwise Multiple Testing as a Formalized Data Snooping. *Econometrica* , 1237-1282.

Spears, D. (2012). Effects of Rural Sanitation on Infant Mortality and Human Capital: Evidence from India's Total Sanitation Campaign.

Stewart, C., Kariger, P., Fernald, L., Pickering, A., Arnold, C., Arnold, B., et al. (2018). Effects of water quality, sanitation, handwashing, and nutritional interventions on child development in rural Kenya (WASH Benefits Kenya): a cluster-randomised controlled trial. *The Lancet Child & Adolescent Health*, 269-280.

Tofail, F., Fernald, L., Das, K., Rahman, M., Ahmed, T., Jannat, K., et al. (2018). Effect of water quality, sanitation, hand washing, and nutritional interventions on child development in rural Bangladesh (WASH Benefits Bangladesh): a cluster-randomised controlled trial. *The Lancet Child & Adolescent Health*, 255-268.

Venkataramanan, V., Crocker, J., Karon, A., & Bartram, J. (2018). Community-Led Total Sanitation: A Mixed-Methods Systematic Review of Evidence and Its Quality. *Environmental Health Perspectives* .

WHO/UNICEF. (20 de 3 de 2008). *Poor sanitation threatens public health*. Available at the WHO Media centre:
http://www.who.int/mediacentre/news/releases/2008/pr08/en/

WSP. (March de 2012). *Nigeria loses NGN455 billion annualy due to poor sanitation*. Available from Economic Impacts of Poor Sanitation in Africa:
https://www.wsp.org/sites/wsp.org/files/publications/WSP-ESI-Nigeria-brochure.pdf

Appendix

Figure A.1. The three WET products

(a) Direct pit model

(b) Offset model

(c) Dual set model (offset + direct pit)

Source: WaterAid Nigeria.

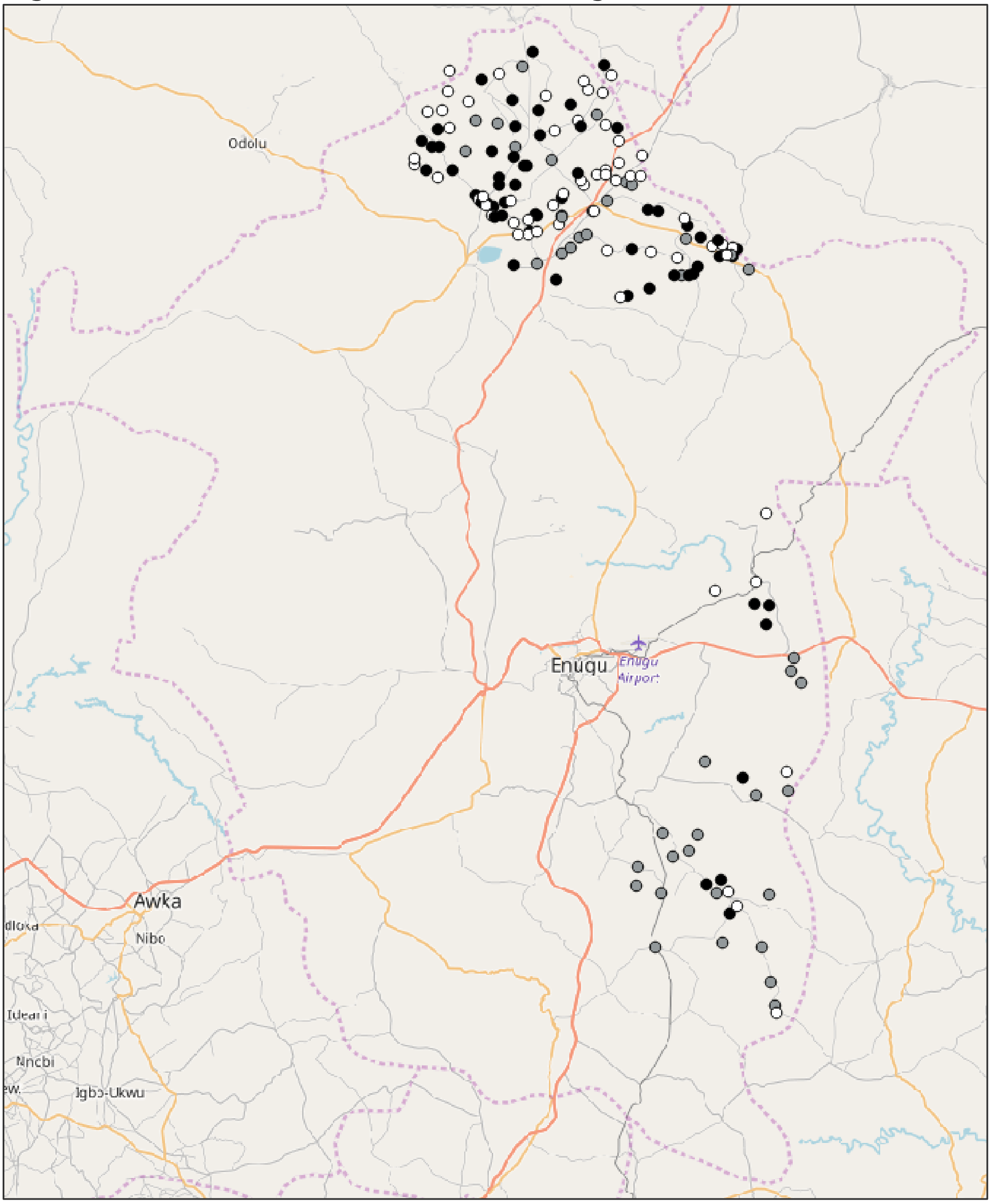
Figure A.2. Components of the WET product line

(a) Offset component of the WET metal mould

(b) A SATO toilet pan

Source: WaterAid Nigeria.

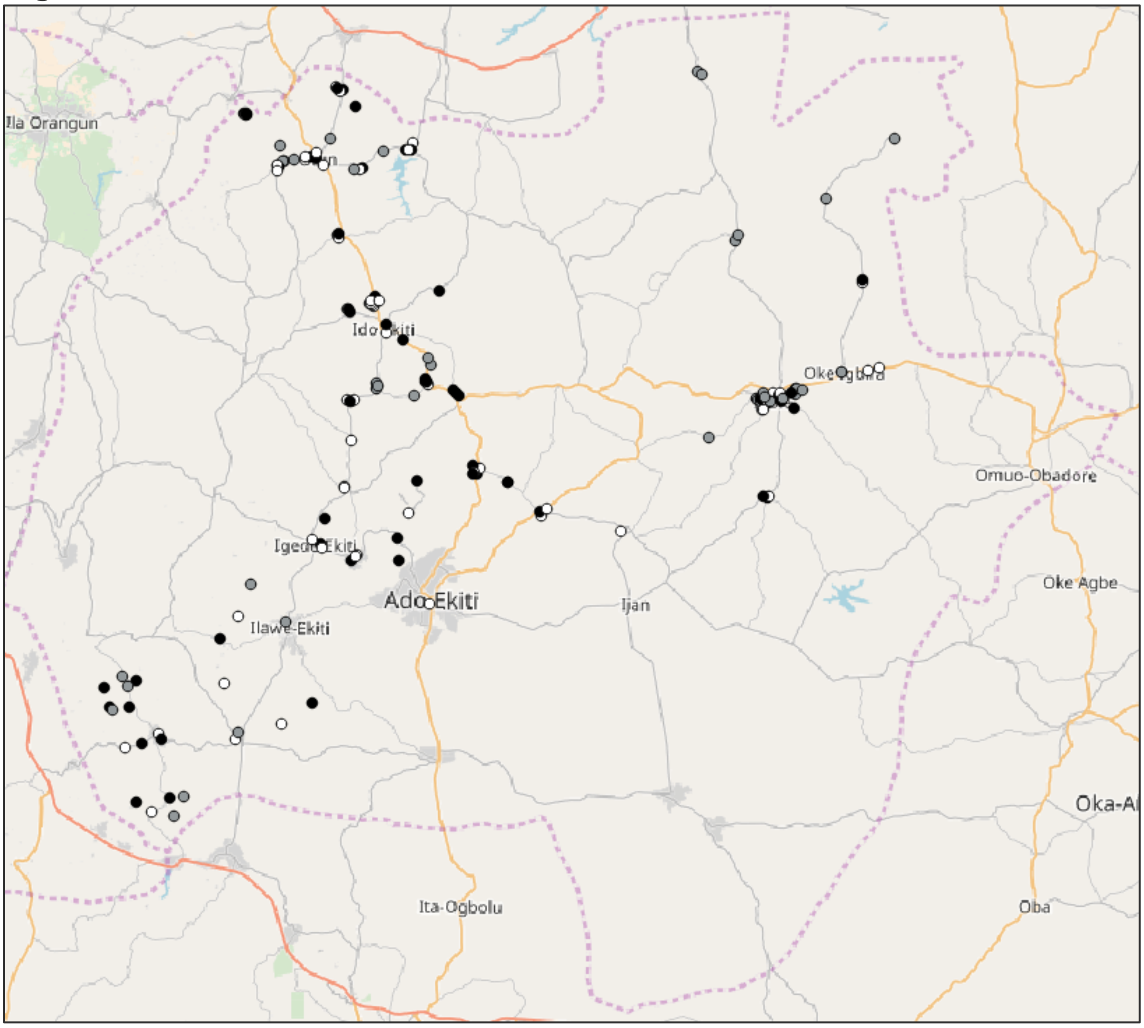
Figure A.3. Location of CLTS and control TUs in Enugu state



Note: CLTS treatment (black), pre-CLTS (grey) and CLTS control (white) TUs in the state of Enugu. Pre-CLTS TUs were omitted in the estimation of CLTS impacts, but were part of the sample used to estimate SanMark impacts. Locations indicate the centroid of a polygon formed by all of the households included in each TU, according to a census conducted before wave 1, during mid 2014. TUs that contained 10 households or fewer were omitted from the map for confidentiality purposes.

Source: IFS using household census data.

Figure A.4. Location of CLTS and control TUs in Ekiti state



Note: CLTS treatment (black), pre-CLTS (grey) and CLTS control (white) TUs in the state of Ekiti. Pre-CLTS TUs were omitted in the estimation of CLTS impacts, but were part of the sample used to estimate SanMark impacts. Locations indicate the centroid of a polygon formed by all of the households included in each TU, according to a census conducted before wave 1, during mid 2014. TUs that contained 10 households or fewer were omitted from the map for confidentiality purposes.

Source: IFS using household census data.

Table A.1. Variables included in the asset wealth and social capital indices

Household index: / Survey questions	Asset wealth (our study) (1)	Asset wealth (DHS comparison) (2)	Social capital (our study) (3)
Ownership of the following durable assets			
Motorcycle/scooter/tricycle	X	X	
Furniture: chairs	X		
Furniture: tables	X		
Furniture: beds	X		
Refrigerator	X	X	
Washing machine	X		
Micro-wave	X		
Gas cooker	X		
Plasma(Flat Screen) TV	X		
Other TV	X	X	
Satellite dish (monthly subscription)	X		
Other satellite dish (DSTV, etc)	X		
Radio/CD/DVD Player	X	X	
Smart phones	X		
Other Telephone / phones	X		
Computer	X		
Air conditioner	X		
Power generator	X		
Sewing machine	X		
Electric iron	X		
Pressure cooker	X		
Electric fans	X		
Bicycle		X	
Car/truck		X	
Household characteristics			
Someone in the household owns the dwelling		X	
Access to improved water source		X	
Access to improved sanitation		X	
Agriculture is the main economic activity		X	
Social capital†			
How many times in the past 12 months have you …			
… donated blood?			X
…worked on a community project?			X

...attended any public meeting in which there was discussion of town or school affairs			X
...attended a political meeting or rally?			X
...attended any club or organizational meeting (not including meetings for work)?			X
...had friends over to your home?			X
...been in the home of a friend of a different race or had them in your home?			X
...been in the home of someone of a different neighbourhood or had them in your home?			X
...been in the home of someone you consider to be a community leader or had one in your home?			X
...volunteered?			X
...served as an official or served on a committee of any local club or community association?			X
Not including weddings and funerals, how often do you attend religious services?			X
Number of households included (wave 1)			
From the CLTS study sample (N=4,722)	*4,622*	*4,487*	*4,227*
From the whole sample (N=6,388)	*6,240*	*6,082*	*5,655*

Note: † 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.

Table A.2. CLTS impacts on toilet ownership and OD, by period: rich TUs

	Toilet ownership			Toilet usage		
Outcome =1 if:	HH owns a toilet	HH owns a functioning toilet	HH owns a functioning improved toilet	All members use toilet	Main respondent performs OD	At least one member performs OD
	(1)	(2)	(3)	(4)	(5)	(6)
Estimates by period						
Treated × Wave 2 (γ_2)	−0.02	−0.02	−0.00	−0.02	−0.01	
p-value (unadjusted)	(0.53)	(0.47)	(0.91)	(0.58)	(0.68)	
p-value (adjusted)†	(0.91)	(0.87)	(0.91)	(0.91)	(0.91)	
Treated × Wave 3 (γ_3)	−0.03	−0.02	0.01	−0.03	0.02	0.00
p-value (unadjusted)	(0.16)	(0.50)	(0.71)	(0.28)	(0.46)	(0.85)
p-value (adjusted)†	(0.42)	(0.73)	(0.89)	(0.63)	(0.71)	(0.89)
Treated × Wave 4 (γ_4)	−0.05	−0.03	−0.03	−0.05	0.03	0.02
p-value (unadjusted)	(0.04)	(0.18)	(0.19)	(0.20)	(0.24)	(0.43)
p-value (adjusted)†	(0.12)	(0.42)	(0.43)	(0.43)	(0.43)	(0.43)
Control mean (wave 1)	0.47	0.45	0.43	0.94	0.53	0.53
Number of TUs	123	123	123	120	123	123
Number of households	2,120	2,120	2,120	949	2,171	2,172
Number of observations	6,361	6,361	6,361	2,847	6,515	4,344

† Adjusted *p*-values account for multiple hypothesis testing following the procedure described in Romano and Wolf (2005), with 1,000 clustered bootstrap samples.
Note: All specifications include a control for the value of the outcome variable in wave 1, household controls, LGA fixed effects and survey wave fixed effects. Controls include age, age squared, gender, employment status and level of education attainment of household head, household size and relative asset wealth, as well as a dummy variable indicating farming as the household's main economic activity. All controls are measured at baseline. Standard errors are clustered at the unit of randomisation (TU) level. *p*-values are given in parentheses. Stars indicate statistical significance: *p < 0.10, **p < 0.05 and ***p < 0.01, according to the adjusted *p*-values.