The impact of an enhanced demand side sanitation and hygiene promotion on sustained behaviour change and mental well-being in Ethiopia

Maryann G Delea Emory University London School of Hygiene & Tropical Medicine

Joshua V Garn University of Nevada

Jedidiah S Snyder Emory University

Molly Linabarger Emory University

Yihenew Tesfaye Oregon State University

Matthew C Freeman Emory University

Grantee Final Report

Accepted by 3ie: February 2020





Note to readers

This final impact evaluation grantee report has been submitted in partial fulfilment of the requirements of grant TW11.1016 awarded under Sanitation Hygiene Thematic Window 11. The report is technically sound, and 3ie is making it available to the public in this final report version as it was received. No further work has been done.

All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its board of commissioners. Any errors and omissions are the sole responsibility of the authors. All affiliations of the authors listed in the title page are those that were in effect at the time the report was submitted. Please direct all comments or queries to the corresponding author, Matthew C Freeman at matthew.freeman@emory.edu.

The 3ie technical quality assurance team comprises Neeta Goel, Anmol Narain, Radhika Menon, Hugh Waddington, Marie Gaarder, an anonymous external impact evaluation design expert reviewer and an anonymous external sector expert reviewer, with overall technical supervision by Marie Gaarder.

3ie received funding for the Sanitation Hygiene Thematic Window from the United Nations Office for Project Services acting as the host agency of the Water Supply and Sanitation Collaborative Council. A complete listing of all of 3ie's donors is available on the <u>3ie website</u>.

Suggested citation: Delea, MG, Garn, JV, Synder, JS, Linabarger, M, Tesfaye, Y and Freeman, MC, 2020. *The impact of an enhanced demand side sanitation and hygiene promotion on sustained behaviour change and mental well-being in Ethiopia*, 3ie Grantee Final Report. New Delhi: International Initiative for Impact Evaluation (3ie). New Delhi: International Initiative for Impact Evaluation (3ie).

Acknowledgments

Funding for the project is provided by World Bank, International Initiative for Impact Evaluation (3ie), the Water Supply and Sanitation Collaborative Council (WSSCC), and Children's Investment Fund Foundation (CIFF). This evaluation would not have been possible without the support of our donors, the Federal Ministry of Health, particularly Biruck Kebede, Federal Democratic Republic of Ethiopia Ministry of Health's NTD Coordinator; the Amhara Regional Health Bureau; the Amhara Regional State Finance and Economic Cooperation Bureau; the South Gondar and West Gojjam Zonal Health Departments; and the Farta, Fogera, and Bahir Dar Zuria *Woreda* Health Offices. We would also like to acknowledge and thank the Health Extension Workers, Women's Development Army Leaders, *Kebele* managers and community members from our formative research and intervention communities. Special thanks to the *Andilaye* enumeration team for their hard work and dedication.

Contents

Table of figures			
Table of tables			
List of acronyms	7		
Executive summary	8		
1. Introduction	. 11		
1.1 Background and rationale	. 11		
1.2 Study aim and objectives	. 12		
1.3 Evaluation questions	. 13		
1.3 Key partners	14		
1.5 Policy context and implications for policy and practice	. 15		
1.6 Report objectives	. 15		
2. Intervention, theory of change, and research hypotheses	. 16		
2.1 Formative research	. 16		
2.2 Description of intervention	. 17		
2.3 Theory of change	. 18		
2.4 Research hypothesis	. 19		
3. Context			
3.1 Existing health programs	.22		
3.2 Study setting			
4. Timeline	.23		
5. Evaluation: Design, methods and implementation	.25		
5.1 Sample size and power analysis.	.25		
5.2 Sampling methodology	26		
5.3 Target study population	26		
5.4 Inclusion and exclusion criteria	.26		
5.5 Sampling frames and sample selection	.27		
5.6 Data collection procedures	28		
5 7 Tool development	28		
5.8 Training	29		
5.9 Field procedures	.29		
5.10 Loss to follow-up	.30		
5.11 Quality assurance	.30		
5.12 Key outcome, intermediate, and sub-group indicators	.30		
5.13 Targeting of WASH-related, NTD-preventive behaviors for intervention inclusion	.31		
5.14 Description of comparison arm			
5.15 Randomization: Allocation rule for intervention and comparison groups	.32		
5.16 Human subjects and ethics	.33		
5.17 Data analysis	.33		
6. Program or policy: Design, methods, and implementation	.34		
6.1 Key program elements and intervention activities	.34		
6.2 Process monitoring	.36		
6.3 Intervention implementation	.37		
6.4 Process data	.37		
6.5 Implementation challenges	.38		
7. Impact analysis and results of the key evaluation questions	.40		
7.1 Results	40		

7.2 Summary of baseline results	40
7.3 Summary of year one follow-up results	45
8. Discussion	52
8.1 Key findings	52
8.2 Strengths	55
8.3 Limitations	56
8.4 Conclusions	56
9. Specific findings for policy and practice	57
References	59
Appendix A. Summary documents of Andilaye formative research and intervention design (61
Appendix B. Summary of behavioral themes and domains addressed in formative research	i,
in addition to behavior change techniques	62
Appendix C. Sample size calculation	66
Appendix D. Outcome indicators and related prompts	67
Appendix E. Random allocation of study clusters	71
Appendix F. Implementation of Andilaye intervention behavioral change catalyzing and	
maintenance	74
Appendix G. Cost of implemented intervention components	78
Appendix H. Process data on the mplementation of Andilaye intervention activities to date	79
Appendix I. Dates of key activities and one year follow-up data collection	83
Appendix J. Supplemental results	84

Table of figures

Figure 1. Theory of change	21
Figure 2. Andilaye Impact Evaluation timeline	24
Figure 3. Study kebele CLTSH statu	41
Figure 4. Midline distribution of select WASH indicators	. 49

Table of tables

20
32
35
38
41
43
44
44
46
47
47
48

List of acronyms

3ie ARHB CIFF CLTSH CHC	International Initiative for Impact Evaluation Amhara Regional Health Bureau Children's Investment Fund Foundation Community-led total sanitation and hygiene Cluster Health Center
DIME	Development Impact Evaluation Group
FMoH	Federal Democratic Republic of Ethiopia Ministry of Health
GoE	Government of Ethiopia
HEP	Health Extension Programme
HEW	Health Extension Worker
HH	Household
	Hygiene Improvement Program
	Head of household
	Institutional Review Board
	Man drug administration
	Mass drug administration
	Open defection free
	Democration-nee
RANAS	The Risks, Attitudes, Norms, Abilities, Self-Regulation approach for systematic behavior change
RCT	Randomized, controlled trial
RQ	Research question
SIEF	Strategic Impact Evaluation Fund
STH	Soil-transmitted helminthiasis
тсс	The Carter Center
USU	Ultimate sampling unit
WASH	Water, sanitation, and hygiene
WHO	World Health Organization
WDAL	Women's Health Development Army Leader
WSP	Water and Sanitation Program
WSSCC	Water Supply & Sanitation Collaborative Council

Executive summary

Important gaps in knowledge related to current sanitation and hygiene programming, and effective approaches for intervention delivery exist. In particular, the effectiveness of these types of interventions in fostering progressive and sustained water, sanitation, and hygiene (WASH)-specific behavioral adoption and maintenance at the community level is unknown. As are the corresponding impacts on mental well-being. With research grants from World Bank's Strategic Impact Evaluation Fund (SIEF), the International Initiative for Impact Evaluation (3ie), the Water Supply and Sanitation Collaborative Council (WSSCC), and the Children's Investment Fund Foundation (CIFF), Emory University and its consortium partners are conducting a three-year impact evaluation, designed as an ex-ante parallel cluster-randomized controlled trail (RCT). This RCT and its preliminary formative research will serve to generate evidence to fill in these knowledge gaps.

The purpose of the three-year Andilaye Impact Evaluation project is to use behavioral theory and formative research to inform the design of a novel intervention approach (i.e., Andilave -Amharic for "togetherness") and evaluate its effectiveness on sustained behavior change and mental well-being. This WASH behavior change intervention focuses on behavioral maintenance, and was designed to be incorporated into prevailing programs, specifically community-led total sanitation and hygiene (CLTSH), and the Health Extension Programme (HEP) more generally, to complement existing efforts in Ethiopia. This project includes the development and evaluation of a theoretically-informed, evidence-based demand-side behavior change intervention and its impact on three less studied, yet critical outcomes of WASH improvements such as, (1) behaviors that prevent transmission of and exposure to neglected tropical diseases (NTDs), including reduced exposure to animal feces; (2) improvements to sanitation access on mental well-being; and (3) sustained behavior change. The study's intervention arm includes households from villages (i.e., *gotts*) within randomly selected sub-districts (i.e., kebeles) receiving the Andilaye intervention. This intervention consists of an integrated sanitation and hygiene programming approach that is enhanced through the incorporation of NTD-preventive components as well as community-oriented behavior change and maintenance strategies facilitated through the leveraging of positive motives such as nurture and status. The counterfactual arm includes households from *gotts* within randomly selected kebeles receiving the Government of Ethiopia's current standard of care WASH programming (i.e., current CLTSH programming).

The behaviors of interest included:

- **1.** Construct a long-lasting latrine that is comfortable and hygienic
- 2. All household members use a latrine every time they defecate
- 3. Immediately dispose of children's feces into the latrine
- 4. Repair your latrine whenever it is damaged
- 5. Upgrade your latrine so it becomes more long lasting, comfortable, and hygienic
- 6. Close your pit when it becomes full and reconstruct a new latrine
- **7.** All household members wash their hands with water and soap or soap substitute AFTER handling animal and human feces, even children's feces
- 8. All household members wash their hands with water and soap or soap substitute BEFORE handling food
- **9.** All household members wash their faces with water whenever they are dirty and use soap when it is available
- 10. Keep all animals separated from the house
- **11.** Keep the household compound clean by disposing of all animal feces and other waste on a DAILY basis

This is a midline (i.e., one year follow-up) report of the three-year *Andilaye Impact Evaluation*. The final evaluation will be complete in August 2019 following the collection of RCT endline

data (i.e., two year follow-up). We randomly allocated 50 *kebeles* equally to intervention or control. We selected up to two *gotts* within the center of the *kebele* for surveillance; our target was 30 households per cluster.

At our year one follow-up, we observed that 68% of the 1,496 households with completed follow-up surveys had at least one latrine, which was similar to baseline (66%). Latrine coverage was also similar, when comparing intervention and control arms at follow-up (prevalence ratio [PR]=1.01; 95% CI: 0.85, 1.22). There was no difference in the prevalence of improved latrines, when comparing intervention and control arms (PR=1.09: 95% CI: 0.77. 1.5). Similarly, there was no difference in the prevalence of households with fully constructed latrines (PR=1.14: 95% CI: 0.86, 1.50). Although there were improvements in many latrine characteristics in the intervention compared to the control arm, the conditions of latrines in the intervention arm were often still inadequate. All measures of latrine utilization were similar when comparing the intervention and control arms. This includes indicators of urination, defecation (both for respondents and other household members), disposal of child feces, and sanitation sharing. Overall, 35% of respondents' primary place of defecation during the last two days was in the open, and only 45% of respondents had defecated in any latrine during the last two days. About half of households did not leave animal feces/waste in the open. This was similar between the intervention and control arms (PR: 1.06; 95% CI: 0.80, 1.40). All other environmental sanitation measures were similar, when comparing intervention and control arms

Respondent-reported handwashing was more prevalent in the intervention arm compared to the control arm. This finding was consistently statistically significant across multiple measures of handwashing. The prevalence of stations with soap was higher in the intervention arm, although only 2.9% of households in this arm had a hand or facewashing station with soap present. The prevalence of hand and facewashing stations was similar between the two study arms (PR=1.02; 95% CI: 0.99, 1.04), and the prevalence of stations with water was also similar when comparing the two arms (PR=0.89; 95% CI: 0.66, 1.20).

The prevalence of anxiety, depression, and emotional distress among respondents was lower at midline than baseline among the overall population. However, the mean scores of these mental well-being indicators were all similar, when comparing intervention to the control arms. To date, the intervention has not reduced diarrhea in the intervention arm compared to the control arm. Diarrhea prevalence was actually higher in the intervention arm (5.5%) than the control arm (2.6%) (PR=2.10; 95% CI: 1.34, 3.30) for our primary diarrhea indicator, which measured whether index children had an episode of three or more loose stools per day over the last seven days.

Our findings do not yield any major new insights to date, likely due to low uptake of the intervention due to the short time between intervention activities and data collection. We hypothesize changes when we collect endline data in March – April 2019. The main limitation of this data collection was that it was conducted so closely following the intervention rollout. The design of the intervention dictates that behavioral modifications are spread across several potential behaviors chosen by beneficiaries, means that we were unlikely to see significant changes in any behavior.

This project is timely in that it is being carried out as Federal Democratic Republic of Ethiopia Ministry of Health (FMoH) is critically reviewing the design and execution of its HEP and constituent components, such as CLTSH. Emory University and its consortium partners have present our initial year one follow-up findings at the FMoH Second Annual NTD Symposium in Hawassa in July 2018 and we have proposed and are now co-organizing the 3rd Global WASH and NTD Roundtable to take place in Addis Ababa in September 2018. These engagements with national stakeholders have allowed us to use our initial findings on the feasibility of a novel demand-side behavior change intervention approach (i.e., *Andilaye*) to

advocate for increased attention on improving the HEW program and enhancing coordination between the WASH and NTD sectors.

1. Introduction

1.1. Background and rationale

Several key limitations endure in the design and implementation of programs aimed at improving access to and behavioral adoption of improved water, sanitation, and hygiene (WASH) in Ethiopia and beyond. For example, these programs often fail to draw upon behavioral theory to inform the selection of appropriate intervention techniques and implementation approaches (Delea et al., 2018). In addition, traditional WASH programming tends to over-simplify the promoted practices, which often represent a constellation of practices. Burden is typically placed on those tasked with implementing the WASH intervention activities at the community level, and the siloed nature of approaches within the health sector limits potentially synergistic program effects.

In Ethiopia, more specifically, the following issues prevail:

- The key approach to improving sanitation coverage and utilization community-led total sanitation and hygiene (CLTSH) - has facilitated considerable changes in coverage of basic sanitation. However, some of these gains have not been sustained, and progress up the <u>sanitation ladder</u> has not been widely achieved (Crocker et al., 2017);
- 2. WASH programs, more broadly, have focused on catalyzing behavior *change*, but not facilitating behavioral *maintenance*. Such approaches have fostered behavioral slippage, or regression back to unimproved behaviors and practices, and poor sustainability of behavioral outcomes and potentially health impacts (Garn et al., 2017);
- **3.** CLTSH largely focuses on leveraging shame to changing norms around open defecation, but these negative affective motivators may not be the most appropriate or effective drivers of change (Crocker et al., 2017);
- **4.** Behaviors and facilities promoted by existing programs are aspirational, but require considerable effort and/or capital investment amongst the programs' target audience. As such, they may not incite the small, incremental improvements that may be required to attain aspirational goals in low resource settings;
- 5. Health Extension Workers (HEWs) charged with implementing CLTSH have many responsibilities, few tools, and little capacity to continually reinforce messages. Although Cluster Health Centers (CHCs) are expected to closely support and monitor HEWs, due to a number of reasons, there is limited support extended to them;
- 6. Siloed approaches within the health sector, namely those vertical programs involved in the control and elimination of the neglected tropical diseases (NTDs), prevent the integration and harmonization of WASH behavior change initiatives (Freeman et al., 2013); and
- **7.** While the focus on diarrheal disease prevention and growth faltering have driven investments in WASH, recent evidence suggests that in sub-Saharan Africa, even substantial improvements may not be enough to impact these health outcomes (Null et al., 2018).

There is promise, however, in enhancing existing programs and programmatic approaches to facilitate improvements that may serve to address these limitations. The Government of Ethiopia's (GoE) Health Extension Programme (HEP), and its accompanying CLTSH module represent government-backed, low-cost, and locally acceptable approaches for improving sanitation and hygiene. CLTSH was originally implemented in Ethiopia through a partnership between the Amhara Regional Health Bureau (ARHB), the USAID-funded Hygiene Improvement Project (HIP), and the Water and Sanitation Program (WSP) in 2006. While an evaluation of CLTSH demonstrated a decrease in open defecation during 2008 – 2010, and an increase in unimproved latrine utilization from 19% to 46%, there was no evidence of change in coverage of improved sanitation facilities (Hernandez and Rosenbaum, 2011). The Health Extension Package being delivered via the HEP, and its accompanying CLTSH

module, are currently being scaled throughout Ethiopia. Despite the absence of key NTDpreventive WASH behavioral promotion, the Federal Democratic Republic of Ethiopia Ministry of Health (FMoH) considers CLTSH its approach for addressing WASH components of NTD programs and promoting other hygiene-related messages for control of enteric diseases.

This project focuses on the development and evaluation of a theoretically-informed, evidencebased demand-side behavior change intervention and its impact on three less studied, yet critical outcomes of WASH improvements. These include the impacts of improved WASH on: (1) behaviors that prevent transmission of and exposure to NTDs and reduced exposure to animal feces (Delahoy et al., 2018; Penakalapati et al., 2017); (2) improvements to sanitation access on mental well-being; and (3) sustained behavior change. This project is timely in that it is being carried out as FMoH is critically reviewing the design and execution of its HEP and constituent components, such as CLTSH. Our impact evaluations results will therefore be made available to government officials at the federal, regional, zonal, and district (*woreda*) levels for us to inform enhancements and modifications in programming approaches (e.g., inclusion of intervention techniques and activities that can facilitate behavioral maintenance to prevent or at least curtail behavioral slippage, leveraging positive and community-oriented motives to catalyze behavioral change at a collective level, integration of additional NTDpreventive WASH practices [e.g., personal hygiene more broadly as opposed to handwashing with soap only]).

1.2. Study aim and objectives

Important gaps in knowledge related to current sanitation and hygiene programming, and effective approaches for intervention delivery exist. In particular, the effectiveness of these types of interventions in fostering progressive and sustained WASH-specific behavioral adoption and maintenance at the community level is unknown, as are the corresponding impacts on mental well-being. With research grants from World Bank's Strategic Impact Evaluation Fund (SIEF), the International Initiative for Impact Evaluation (3ie), the Water Supply and Sanitation Collaborative Council (WSSCC), and the Children's Investment Fund Foundation (CIFF), Emory University and its consortium partners are launching a three-year impact evaluation, designed as an ex-ante parallel cluster-randomized controlled trail (RCT). This RCT and its preliminary formative research will serve to generate evidence to fill in these knowledge gaps. The purpose of the three-year Andilaye project is to use behavioral theory and formative research to inform the design of a novel intervention approach (i.e., Andilaye) and evaluate its effectiveness on sustained behavior change and mental well-being. This WASH behavior change intervention focuses on behavioral maintenance, and was designed to be incorporated into prevailing programs (specifically CLTSH, and the HEP more generally) to complement existing efforts in Ethiopia.

The study's intervention arm includes households from villages (i.e., *gotts*) within randomly selected sub-districts (i.e., *kebeles*) receiving the *Andilaye* intervention. This intervention consists of an integrated sanitation and hygiene programming approach that is enhanced through the incorporation of NTD-preventive components as well as community-oriented behavior change and maintenance strategies facilitated through the leveraging of positive motives. The counterfactual arm includes households from *gotts* within randomly selected *kebeles* receiving the GoE's current standard of care WASH programming (i.e., current CLTSH programming).

The main objectives of this work involve an examination of change along our hypothesized, yet evidence-based¹ *Andilaye Theory of Change*, including:

¹ The *Andilaye* team adapted USAID's TOPs Theory of Change approach to translate our formative research findings into our theoretically-informed and evidence-based *Andilaye Theory of Change*.

- 1. Determine the impact of *Andilaye's* enhanced intervention on NTD-preventive sanitation-and hygiene-related behavioral antecedents (e.g., internalized risk and capability perceptions, attitudes, and norms), intermediate outcomes along the causal chain;
- 2. Explore whether the promotion of the *Andilaye* intervention leads to an increase in the uptake, maintenance, and exclusive use of sanitation facilities and improved personal hygiene behaviors, behavioral outcomes along the causal chain;
- **3.** Investigate whether changes in personal hygiene, sanitation, and water behaviors are sustained; and
- **4.** Examine whether the promotion and integration of facial hygiene and other NTDpreventive practices, as components of demand-side sanitation and hygiene programming, result in behavioral benefits for NTD prevention and mental well-being, the impacts of interest at the terminal end of the causal chain.

As indicated above, our three-year study assesses and tracks changes in sanitation and hygiene-related indicators along the causal chain, including behavioral antecedents (e.g., attitudes, capabilities, collective efficacy, norms), contextual changes in household conditions that may facilitate habituation of improved behaviors (e.g., improvements in conditions of household sanitation and hygiene), and behavioral outcomes. We are also measuring and tracking changes in hypothesized behavioral mediators such as water and sanitation insecurity, and health impacts including respondent-reported diarrhea and mental well-being at multiple time points. We are examining behavioral antecedents and other behavioral mediators to establish causal determinants of behavioral change and maintenance over time, and identify mechanisms that lead to behavioral change. There is utility in employing such an approach, as it will help us determine not only *if*, but *how* and *why* the intervention was successful, and in which contexts. As part of this evaluation, we are developing and validating novel metrics and frameworks to track key outcomes along the causal pathway (e.g., behavioral antecedent indicators; innovative, objective facial and hand cleanliness indicators; sanitation insecurity, water insecurity, and collective efficacy scales refined and validated during the course of this study).

1.3. Evaluation questions

Our impact evaluation consists of two main research questions and several accompanying sub-questions. The evaluation questions reported here are identical to those presented in the pre-analysis plan submitted to 3ie in June 2017 (TW11 1016), with the exception of a slight change in RQ 1 to shift from measuring the impact of "integrating WASH-related, NTD-preventive behavior change components" to that of a holistic WASH approach. The pre-analysis plan was submitted prior to the finalization of intervention activities. The "holistic" component includes, but is not exclusively for control of NTDs. In addition, through the intervention development process, several behaviors considered critical for NTD control (e.g., shoe wearing) were not included because they were not feasible within the demand-driven context of the project. Evaluation questions addressed in this report are listed below:

RQ 1. What is the impact of *Andilaye's* enhanced, demand-side sanitation and hygiene intervention on diarrhea, mental well-being and sustained behavior change? RQ 2: How does the promotion of *Andilaye's* enhanced, demand-side sanitation and hygiene intervention effect intermediate (i.e., behavioral antecedents) and behavioral outcomes?

RQ 2a: Does promotion of the *Andilaye* intervention lead to an increase in the uptake and exclusive use of sanitation facilities and hygiene behaviors compared to the current model?

RQ 2b: Does the *Andilaye* intervention facilitate behavioral maintenance (i.e., are these changes in sanitation and hygiene sustained)?

RQ 2c. Does the *Andilaye* intervention effect sanitation security, and does water security modify the effectiveness of the intervention in changing hygiene behaviors? **RQ 2d:** What is the effect of the *Andilaye* intervention on changes to social norms and perceptions?

This is a midline (i.e., one year follow-up) report of the three-year Andilaye impact evaluation. The final evaluation will be complete in August 2019 following the collection of RCT endline data (i.e., two year follow-up). The results of our formative research and impact evaluation will support the development of a policy-relevant evidence base that indicates the most effective ways to integrate NTD prevention and control efforts within ongoing, at-scale WASH programming in Ethiopia. This work will have policy implications within Amhara and throughout Ethiopia. The study is being implemented closely with key partners (listed below), and the extended study team will work to mainstream the project outputs and findings into programmatic revisions and related government policies. The project has considerable buy-in from One WASH National Program, ARHB, FMoH, and the FMoH NTD coordinator.

1.3. Key partners

As part of *Andilaye*, our research consortium is engaging a broad cadre of partners to conduct this impact evaluation. Emory University - the team based in Atlanta, Addis Ababa, and Bahir Dar - and its consortium partners have conducted formative research and worked closely with government officials and other stakeholders to develop an evidence-based intervention approach and accompanying curricula for integrating NTD and WASH behavior change initiatives at the local (i.e., *woreda*) and community (i.e., *kebele* and *gott*) levels. Through the use of the RCT design, we are in the process of evaluating the effectiveness of this theoretically-informed and evidence-based intervention, which focuses on positive, community-oriented motivators of WASH behavior change, achievable incremental improvements, and behavioral maintenance. Throughout this impact evaluation, we are examining which aspects of the intervention are promising for NTD prevention and control, with a specific focus on soil-transmitted helminthiasis (STH), schistosomiasis, trachoma and other enteric disease control efforts in Ethiopia. Key *Andilaye* partners include the following:

Ethiopia One WASH National Program, World Bank Group, 3ie, WSSCC and Children's Investment Fund Foundation: The World Bank's SIEF, Development Impact Evaluation Group (DIME), and WSP; 3ie; WSSCC; and CIFF provide funding, subject matter expertise, technical inputs, and strategic oversight for this study. The World Bank is a key funder of the Ethiopia One WASH National Program.

Federal Ministry of Health: The FMoH NTD Coordinator and other GoE officials are engaging with the study, and have demonstrated support throughout the formative research, curriculum design and training process, and community-level implementation phases. We have conducted several workshops and meetings with the FMoH NTD-focal lead and CO-WASH lead to ensure alignment with the national strategy.

Emory University: Emory, the prime recipient of these research grants, provides research oversight and subject matter expertise related to WASH and NTDs, social and behavioral science, project costing, randomized trials, and study design. Emory is a certified 501(c)3 non-profit organization, registered in Ethiopia (Emory Ethiopia) with a permanent research team in place. Emory has several full-time staff in Ethiopia leading the project, including Dr. Abebe Gebremariam (Co-PI and Director of Emory Ethiopia), Mulusew Lijalem (Regional Manager), Kassahun Zewudie (Study Manager), Mulat Woreta and Resom Berhe (Monitoring and Evaluation Officers), Siraj Muhammed (Behavior Change and Communications Specialist), and Yihenew Tesfaye (PhD student at Oregon State University).

Oregon State University, Department of Anthropology: Anthropologists from Oregon State University who are experienced in studying the HIP and the role that HEWs and Women's

Health Development Army Leaders (WDALs) play in the deployment thereof are members of the *Andilaye* team. These investigators play a key role in the *Andilaye* process evaluation, examining the role HEWs and WDAL members play in the implementation of the *Andilaye* intervention; this team also guides the assessment of water insecurity and further refinement of existing metrics.

Amhara Regional Health Bureau: The ARHB is supporting the design and implementation of the *Andilaye* intervention. We have worked closely with both the head and deputy head of ARHB on the design and targeting of the intervention.

Zonal Health Department, Woreda Health Offices, Health Extension Workers, and Women's Health Development Army Leaders: These actors are helping to collaboratively implement the *Andilaye* intervention.

Other stakeholders such as Sightsavers, CARE, Catholic Relief Services, Fred Hollows Foundation, The Carter Center, World Vision International, Partnership for Child Survival: The Andilaye team is engaging other key WASH and NTD stakeholders working in the sector as part of the knowledge dissemination and curriculum development processes.

The principal investigators of the study are Matthew C. Freeman, PhD, MPH and Abebe Gebremariam Gobezayehu, MD. The study manager is Maryann G. Delea, MPH. The study team consists of: Frederick Goddard, MS; Mulusew Lijalem, MSc; Molly Linabarger, MPH; Gloria Sclar, MPH; Jedidiah Snyder, MPH; Hiwote Solomon, MPH; Mulat Woreta, BA, PGD; Kassahun Zewudie, MPH; ; and Yihenew Tesfaye, MS. The following are Co-Investigators: Tenagnework Antefe, BSc; Bethany Caruso, PhD, MPH; Thomas Clasen, PhD, JD; Joshua V. Garn, PhD, MS; Craig Hadley, PhD; Kenneth Maes, PhD; and Deborah McFarland, PhD, MPH

1.5. Policy context and implications for policy and practice

Research on WASH behaviors has focused predominantly on information dissemination (e.g., promoting knowledge that washing hands reduces illness); however, evidence suggests that these factual knowledge-based messages and approaches do not result in sustained changes to improved practices (Briscoe and Aboud, 2012; Webb and Sheeran, 2006; Wood and Neal, 2016). Rather, focus should be placed on changing social norms, improving demand for sanitation and hygiene infrastructure, and habit formation. This project will support existing GoE policy objectives, and serve to develop lessons learnt for potential application of integrated WASH programs elsewhere. The expected relevance of our research is farreaching for policy-making and implementing agencies, as evidence generated via this impact evaluation will provide:

- 1. Evidence of program design: testing whether a theory-informed and evidence-based WASH intervention produces a greater net impact on behavior change and behavioral maintenance and health than a traditional CLTSH intervention approach;
- 2. Evidence of program efficacy: establishing proof of concept that an enhanced, integrated WASH-NTD intervention has an effect on intermediate (i.e., behavioral antecedents) and behavioral outcomes (e.g., adoption of improved, NTD-preventive WASH practices, social norms, and sanitation-related insecurity); and
- **3.** Evidence of resource allocation decision-making: analyzing the cost-effectiveness of an enhanced WASH intervention that engages additional community stakeholders for intervention delivery relative to traditional CLTSH.

1.6. Report objectives

As part of the *Andilaye Impact Evaluation*, our team conducted a baseline assessment during March – April 2017. Our baseline assessment consisted of a household survey designed to address the *Andilaye Impact Evaluation* questions by generating data to demonstrate the

status quo (i.e., pre-intervention) of sanitation and hygiene practices (i.e., behavioral outcomes) as well as respondent-reported diarrhea and mental well-being in study communities. In addition to obtaining these baseline measures, we also captured data on other factors that may influence intervention uptake, including perceived sanitation and water insecurity and collective efficacy measures. The household survey was administered again during March – April 2018 to assess the impact of the *Andilaye* intervention and provide data for further examination of our evaluation questions one year later (i.e., follow-up 1 – reported here). A two-year follow-up will also be conducted during March – April 2019 to assess sustained behavior change resulting from the *Andilaye* intervention (i.e., follow-up 2).

This report summarizes the status and impact of the *Andilaye* project through April 2018. The objectives of the report are to:

- **1.** Describe the practicalities of the intervention, including the logistics and the intervention's underlying theory of change;
- **2.** Provide the rationale for study site selection, including discussion of relevant local trends, criteria for external validity, and sampling;
- **3.** Illustrate the planned flow of program implementation, impact evaluation, and relevant external shocks;
- 4. Describe critical details related to the study design;
- **5.** Detail how the intervention was implemented in practice and report on challenges faced during implementation;
- **6.** Provide a summary of the analysis and interpretation of the results from follow-up 1 (i.e., changes in indicators, one year post-baseline); and
- 7. Report on specific findings for policy implications.

2. Intervention, theory of change, and research hypotheses

2.1. Formative research

The *Andilaye* team executed the formative research phase of this study during September 2016 – February 2017, the end of rainy and beginning of dry seasons in Ethiopia. The formative research was grounded in several behavioral theories and frameworks, including the *Behaviour Change Wheel* (Michie et al., 2011), the *Theory of Triadic Influence* (DiClemente et al., 2002), and the *RANAS* framework (Mosler, 2012). During the formative research phase, the evaluation team conducted a series of site visits to *Woreda* Health Offices, health posts, and formative research communities to collect information relevant to the design of the *Andilaye* intervention and the refinement of various metrics that are being used throughout the course of the study (e.g., baseline and follow-up survey indicators and prompts).

Formative research included a series of qualitative and quantitative data collection methods, including focus group discussions, key informant interviews, household and community observations, and cognitive interviews. A summary of these activities and findings can be found in Appendix A, or <u>here</u>, including behavioral domains identified. The communities in which formative research activities were conducted were similar to *kebeles* eligible for the *Andilaye* trial; however, in most circumstances, the communities were not deemed eligible for inclusion in the trial given their involvement in the formative research. An exception was made for three (two intervention, one control) *kebeles* in which only two or fewer community-level formative research activities were conducted. The rationale for this decision centered around the thinking that two focus group discussions would not considerably and sustainably alter behaviors within the larger community.

In accordance with USAID's TOPs Theory of Change development process (*Designing for Behavior Change For Agriculture, Natural Resource Management, Health and Nutrition*,

2013), qualitative and quantitative data from the formative research phase were used to generate problem and accompanying solution trees. These trees formed the foundation of an intervention mapping process during which key stakeholders from FMoH, the ARHB, Zonal Health Departments, *Woreda* Health Offices, and other stakeholders from non-governmental, multi-lateral, and donor organizations provided feedback on the trees during an intervention design workshop held in Bahir Dar in April 2017 (see the meeting report in Appendix A, or <u>here</u>). Subsequent to breaking out into groups to discuss the trees, and presenting suggested modifications, the workshop participants weighed in on which of the factors presented in the solution trees were the most feasible to implement, and which were presumably the most impactful.

After the intervention design meeting, the *Andilaye* team continued working on intervention design by using the refined problem and solution trees to identify overarching behavioral antecedent and determinant categories presented in the trees as well as the behavioral factors each tapped. The team then generated a list of possible activities that tracked to intervention techniques that would appropriately address those behavioral factors, per the *Andilaye* behavioral framework.² The team then returned to formative research communities to conduct a series of behavioral trials, during which various intervention activities and approaches were tested among formative research households. These trials gave community members a voice, and an opportunity to weigh in on the initially proposed *Andilaye* intervention. Feedback from households, community change agents, and other community leaders was used to refine the *Andilaye* intervention approach and accompanying behavioral tools.

2.2. Description of intervention

The Andilaye intervention was informed by our formative research and finalized with feedback from government partners. As previously mentioned, the Andilaye intervention is a demandside sanitation and hygiene intervention that focuses on positive, community-oriented motivators of behavioral change, promotes achievable incremental improvements, and incorporates strategies that facilitate behavioral maintenance (i.e., prevention of behavioral slippage or relapse back to unimproved behaviors). The intervention addresses issues related to over-extension of HEWs, and over-saturation of messaging via the HEP's Health Extension Package through the engagement of additional community change agents as mechanisms for intervention delivery. It is based on recent work conducted by WSP, and incorporates feedback from relevant stakeholders, including FMoH, ARHB, Zonal Health Departments, *Woreda* Health Offices, WSP, and other key stakeholders, including formative research community members.

Our theory of change includes several intervention functions and approaches across multiple intervention levels. *Andilaye* intervention functions operate at four levels: (1) district, (2) community (i.e., *kebele/gott*), (3) group, and (4) household, and employ a variety of behavior change catalyzing and maintenance techniques (see Table 1 for further details regarding the aim for each intervention function):

1. District (woreda) level activities

- Sensitization and action planning workshops
- Skills-based training of the trainers for HEWs, CHC HEWs supervisors, and *Woreda* officials
- Training of community conversation (CC) facilitators
- Adaptive management workshops
- Skills-based refresher training for supervisors and facilitators
- 2. Community (kebele/gott) level activities
 - Skills-based training of WDALs

² The *Andilaye* behavioral framework is an adaptation of the *Behaviour Change Wheel* and the *RANAS* framework.

- Whole system in a room and action planning
- Community mobilization and commitment event
- Skills-based review meetings and refresher trainings for WDALs
- Household graduation and maintenance celebration ceremonies
- Cross-fertilization visits (e.g., between intervention gotts)
- 3. Group level activities
 - Community conversations
- 4. Household level activities
 - Counselling visits with caregivers
 - Focused behavioral maintenance counseling visits

2.3. Theory of change

Andilaye intervention functions were informed by theory and designed to address key drivers and barriers identified from the problem and solution trees derived from our formative research data and stakeholder-generated interpretations thereof. Through a formal intervention mapping process, we identified theoretically-informed and empirically-based intervention techniques that are specifically designed to address the behavioral antecedents and determinants represented by the problem and solution trees. Our formative work also indicated that perceptions related to self- and collective efficacy are important mediators of uptake of community-based interventions. As such, the overarching intervention motto, "Together we can be a strong, caring, healthy community", and related intervention functions work to improve individual and community-level agency (i.e., enhance self- and collective efficacy). The motto offers an aspirational message that emphasizes the need for collective action to make positive change in one's community.

We focused the *Andilaye* intervention on three behavioral domains, informed by our formative research: (1) sanitation, (2) personal hygiene, and (3) household environmental sanitation. Below, we list the constellation of specific behaviors and practices encompassed within each behavioral domain. We emphasize that while we identify 11 constituent practices of interest, they are within 3 behavioral domains. WASH interventions often target large numbers of behaviors, but often also fail to identify the constituent practices required to practice these behaviors. As such, our intervention actually focuses on *fewer* practices than many WASH interventions while also clearly specifying all necessary practices to complete that behavior.

These are the behaviors and practices that we designed the *Andilaye* intervention to target. While primary caregivers of the project's index children comprise the intervention's primary target audience, we designed the intervention in such a way so as to promote behavior change among all members of a household and community at large.

1. Sanitation

- Construct a long-lasting latrine that is comfortable and hygienic
- All household members use a latrine every time they defecate
- Immediately dispose of children's feces into the latrine
- Repair your latrine whenever it is damaged
- Upgrade your latrine so it becomes more long lasting, comfortable, and hygienic
- Close your pit when it becomes full and reconstruct a new latrine

2. Personal hygiene

- All household members wash their hands with water and soap or soap substitute AFTER handling animal and human feces, even children's feces
- All household members wash their hands with water and soap or soap substitute BEFORE handling food
- All household members wash their faces with water whenever they are dirty and use soap when it is available
- 3. Household environmental sanitation

- Keep all animals separated from the house
- Keep the household compound clean by disposing of all animal feces and other waste on a DAILY basis

A summarized theory of change for the *Andilaye* intervention is shown in Figure 1. Additional details on behavioral drivers for each domain can be found in Appendix B. Problem and solution trees that outline our detailed theory of change are in Appendix A, or <u>here</u>. The *Andilaye* intervention focuses on positive, community-oriented motivators of behavioral change and promotes achievable incremental improvements. Thus, we anticipate that the intervention's behavioral change catalyzing activities translate quickly into incremental improvements in primary outcomes.

The design of this intervention is that each household undergoes a counseling process with the WDALs to identify and plan for their own path of change. While we do think that households will quickly change some of their behaviors, this may not be detected in the statistical tests for each individual behavior at this early stage, since households may have chosen different behaviors to focus on at the outset of the process. We will learn more through our quarterly monitoring and endline findings. Sustaining improvements in primary outcomes will determine the impact of the intervention. Sustainability of these outcomes is being addressed by intervention functions that incorporate strategies that facilitate behavioral maintenance (i.e., prevention of behavioral slippage or relapse back to unimproved behaviors).

2.4. Research hypothesis

We hypothesize that:

- **RQ 1.** Improvements in sustained behavior change, reductions in the prevalence and severity of diarrheal disease, and improvements in mental well-being are more likely in communities randomized to our intervention arm receiving the *Andilaye* intervention functions compared to communities randomized to our comparison arm;
- **RQ 2a.** Improvements in the uptake, maintenance, and exclusive utilization of sanitation facilities and hygienic practices are more likely in communities randomized to our intervention arm compared to communities randomized to our comparison arm;
- **RQ 2b.** Improvements in **sustained** uptake, maintenance, and exclusive utilization of sanitation facilities and hygienic practices are more likely in communities randomized to our intervention arm compare to communities randomized to our comparison arm;
- **RQ 2c.** The *Andilaye* intervention will improve sanitation security, but water insecurity may mediate WASH-related behavior change; and
- **RQ 2d.** Improvements in social norms and perceptions and reductions in sanitationrelated psycho-social stress among women are more likely in communities randomized to our intervention arm compared to communities randomized to our comparison arm.

Table 1. Andilaye intervention behavioral change catalyzing and maintenance activities

ſ	Activity	Aim	
	Catalyzing activities		
	Sensitizing and action	To orient key stakeholders to the Andilaye intervention and engage them in intervention action	
	planning workshop	planning so as to generate buy-in and foster an enabling environment in which the intervention	
	Skills-based training of the	To provide skills-based training to HEWs/CHC HEWs supervisors/Woreda officials on	
	trainers for HEWs. CHC	household (HH)-level intervention activities, supportive supervision, and on-the-iob-training so	
e	HEWs supervisors, Woreda	HEWs can, in turn, effectively train WDALs on the implementation of HH-level activities and	
ev.	officials	provide supportive supervision.	
Ϋ́	Training of community	To provide comprehensive facilitator training to selected gott and kebele stakeholders on the	
tric	conversation facilitators	'community conversations' group-level intervention activity.	
Dis	Maintenance activities	To be a second second state to the Water second second sector list and descent stated	
-	Adaptive management	To leverage monitoring data to facilitate evidence-based, controlled, and documented	
	workshops	improve intervention outcomes and resource management by learning from monitored program	
		outcomes.	
-	Skills-based refresher	To reinforce previously acquired knowledge and skills and address trainer/facilitator turnover.	
	training for supervisors and	Prior experience indicates that such trainings serve to sustain actor motivation and further	
	facilitators	strengthen capacity.	
	Catalyzing activities		
	Skills-based training of	To provide skills-based training to WDALs on HH-level intervention activities, as detailed in the	
	Women's Development	training of the trainers for HEWs, CHC HEWs supervisors, Woreda officials.	
	Army Leaders		
	whole system in the room	to engage key community stakenoiders, orient them to the Andilaye Intervention, and facilitate	
	and action planning	buy-in and foster an enabling environment (i.e. social opportunity) in which the Andilave	
		intervention can be supported and effectively implemented for a "strong, caring, healthy	
<u>e</u>		community."	
Le	Community mobilization	To shift social norms (including community by-laws and sanctions), and improve action	
ty-	and commitment event	knowledge, barrier identification and planning, and attitudes regarding targeted WASH	
uni		behaviors through a form of contextually appropriate and interactive edutainment.	
Ē	Skills based review	To reinforce providually acquired knowledge and skills, address M/DAL turnover, and review	
ion	meetings and refresher	successes and address challenges faced in implementing counseling visits with caregivers	
0	trainings for Women's	Prior experience indicates that such trainings serve to sustain actor motivation and further	
	Development Army Leaders	strengthen capacity.	
	Household graduation and	To hold a celebration to reward households/communities and to motivate one another to	
	maintenance celebration	sustain well-earned gains. Celebrations foster motivation and also help reinforce improved	
-	events	behaviors and promote healthy competition among communities.	
	Cross-tertilization visits	I o provide an opportunity to share experiences across different intervention communities – to	
		preliminary behavior change and health outcomes	
	Catalyzing activity		
	Community conversations	To change factual beliefs and attitudes, enhance action knowledge, improve percentions of	
_	community conversations	capability, identify and make plans to overcome barriers, and shift social norms regarding	
eve		targeted behaviors through community group dialogue. To carry out demonstrations that	
Ļ		address key factors associated with both breaking unimproved practices and adopting	
dnc		improved ones.	
อี	Maintenance activity	To generate community level dialogue regarding successful issues accessized with resistances	
	Community conversations	of improved practices and barriers thereof through a follow-up round of community group	
		dialog. To carry-out demonstrations related to behavioral maintenance issues.	
	Catalyzing activity		
I-Level	Counseling visits with	To provide personalized counseling to caregivers to equip them with the knowledge skills and	
	caregivers	motivation necessary to develop improved WASH practices. To foster action capacity, self-	
		efficacy, and barrier planning so caregivers maintain the improved WASH practices.	
old	Maintenance activity		
Househo	Focused behavioral	To provide continuous follow-up to households such that the house graduates from counseling	
	maintenance counseling	related to initial adoption of improved behaviors to counseling related to behavioral	
	V13113	namenance skills so the caregiver can maintain his/her improved WASH practices, especially as	
_			



Figure 1. Theory of chang

3. Context

3.1. Existing health programs

The Andilaye Impact Evaluation is being carried out in Amhara National Regional State, a region of Ethiopia in which WASH conditions are poor, behavioral slippage has been documented, and several NTDs are hyperendemic. As with the rest of Ethiopia, where CLTSH is being scaled nationally, study communities have either been triggered with CLTSH or are scheduled to be triggered in the near future. Despite the absence of key NTD-preventive WASH behavioral promotion, FMoH considers CLTSH its approach for addressing WASH components of NTD programs and promoting other hygiene-related messages for control of enteric diseases.

As this study is operating in an area where CLTSH is being rolled out nationally, we will not interfere with established CLTSH roll-out and implementation protocols. While we cannot be sure that our comparison communities will not receive further CLTSH interventions during the course of the trial, any such further implementation of the current CLTSH interventions would only bias effect estimates toward the null. Communities allocated to the comparison arm may receive the *Andilaye* intervention at a later time point. Interestingly, our baseline results indicated that 78% (39 of 50) of *kebele* clusters randomly selected for inclusion in the *Andilaye Impact Evaluation* have been triggered with CLTSH, and certified as open defecation free (ODF) according to *Woreda* Health Office records. Another 14% (7 of 50) of study clusters have been triggered with CLTSH but have not yet been certified ODF, and the remaining 8% (4 of 50) have not yet been triggered with CLTSH (see Appendix E). These statistics, along with our findings at baseline (only 66% of households reporting at least one latrine) provide strong evidence that behavioral slippage is, indeed, an issue that needs to be addressed in Amhara and perhaps elsewhere in Ethiopia.

Our study was designed, and is being executed at a time when GoE and FMoH are critically evaluating the nationally scaled HEP. As a result, our intervention design considered demandside sanitation and hygiene intervention approaches that could be considered as refinements within the HEP if they demonstrated impact. However, we were cognizant of the fact that the Health Extension Package utilized by the HEP has become saturated, and that HEWs are constantly having more work added to their plates via the HEP. As a result, our intervention is exploring the engagement and potential of alternative community change agents for intervention delivery at the community level.

This evaluation does not include a mass drug administration (MDA) component, and has not sought to influence the timing of such activities in study communities. While MDA campaigns are active in Amhara, specifically the administration of Zithromax® for Trachoma management, no parasitologic health outcomes are included in our evaluation (see section entitled *Key outcome, intermediate, and sub-group indicators* for further details). Thus, the presence of MDA in the study area will not alter the research questions assessing the impact of the *Andilaye* intervention on targeted health impacts, behavior change, or sustainability of improved sanitation and hygiene practices.

3.2. Study setting

Three districts (i.e., *woredas*) in South Gondar and West Gojjam Zones were targeted for this study, as they provided a range of topographical conditions present in the Amhara in specific, and Ethiopia in general. Emory Ethiopia's previous work in these zones and districts allowed us to leverage important political relationships for the successful launch of the study. Along with in-country partners, we felt targeting these topographically diverse areas would help improve the external validity of our results within and beyond Ethiopia. Farta is a mountainous area that is more rural than Fogera, a low-land, marshy area in close proximity to Lake Tana.

A large damn is located in Farta. Bahir Dar Zuria is a relatively flat area on the rural outskirts of Bahir Dar town.

Rural and peri-urban *kebeles* (i.e., sub-district study sites) within Bahir Dar Zuria, Fogera, and Farta *woredas* that are accessible throughout the course of the year were eligible for inclusion in the RCT. This decision was made in partnership with relevant *Woreda* Health Offices and One WASH National Program representatives, as officials from these areas were meant to facilitate and supervise the implementation of the *Andilaye* intervention. Officials from these *Woreda* Health Offices helped study staff identify *kebeles* meeting this eligibility criterion. Once the *kebele* sampling frame was identified, a random number generator was employed to randomize *kebeles* to intervention and comparison arms. See *Section 5* for further details.

Given our study includes quarterly monitoring that will result in data collection on targeted behavioral antecedents and improved WASH practices throughout the course of the year, our study also presents an opportunity to examine seasonal trends. We collect a variety of directly observed, objective measures as well as self- and respondent-reported information on a variety of individuals within our study households. Our study *kebeles* represent not just rural, but also peri-urban sub-districts. Such information can be extrapolated on a larger scale to estimate and predict similar behaviors and practices amongst various population segments throughout Amhara. One limitation of our study, however, is that it does not capture data on life, intervention implementation, and uptake thereof in urban contexts. Therefore, we are uncertain how externally valid our results are for urban contexts.

4. Timeline

The Andilaye Impact Evaluation consists of three major phases: (1) intervention design, (2) intervention implementation and process evaluation, and (3) and impact evaluation (see Figure 2 for details). As mentioned above, the Andilaye intervention was informed by our formative research (September 2016 – February 2017), and was finalized with feedback from government partners (April 2017). Once feedback was received, behavioral trials were completed (May – June 2017) to test out aspects of the Andilaye intervention. Findings from the behavioral trials were used to finalize the Andilaye intervention design and materials. Implementation of intervention activities began in September 2017 and will continue, with a focus on behavior change catalyzing activities, through and beyond follow-up 1 (March – April 2018). Intervention activities will transition to behavior change maintenance activities, as dictated by household and community progress, and continue until follow-up 2 (March – April 2019).

Ethiopia declared a State of Emergency in October 2016 that continued through August 2017. A subsequent State of Emergency was declared in February 2018, and was recently lifted. While these conditions have created heightened security and barriers in communication among *Andilaye* team members, the states of emergency had limited bearing on intervention implementation. Aside from the protracted drought that caused delays in the launching of field activities in 2016, to date, no major unanticipated or unexpected events have negatively influenced intervention implementation or impact evaluation.

Figure 2. Andilaye Impact Evaluation timeline



Impact evaluation

5. Evaluation: Design, methods and implementation

5.1. Sample size and power analysis

We are interested in understanding the impact the *Andilaye* intervention has on sustained WASH behavior change, diarrhea, and mental well-being. As such, we conducted a sample size determination for mental well-being, as this impact was deemed the most restrictive in terms of required sample size.

Various studies from Ethiopia and East Africa suggest that approximately 20-35% of rural women experience elevated symptoms of common mental disorders such as anxiety and depression. Drawing on two studies that have used the <u>Hopkins Symptom Checklist</u> in East Africa, we estimate that average scores on this tool are around 1.5 (SD 0.5) on a scale of 1-4 (Hadley et al., 2008b; Hadley and Patil, 2006). Using unpublished data from a large on-going study of young people in Ethiopia (Hadley et al., 2008a), we estimate that the intra-cluster correlation (ICC) for a measure of mental health at approximately 0.05, although we suspect this is low given that the sample are young people and the measure used captured more severe mental health symptoms. 33% of women in South Oromo had psychosocial distress, and it was positively correlated with water insecurity, indicating that women who experienced more water insecurity also reported more symptoms of common mental disorders (Stevenson and Yohannes, 2014). There are no data on the impact of changes to sanitation access on mental well-being, so we will use a similar difference to estimate our impact.

Appendix C presents the results of our sample size calculations and sensitivity of the effect size in changes in parameters. We assumed a relative reduction of 30% for poor well-being as measured by the WHO-5, an α of 0.05 and at least 80% power. Our sample size analysis indicated that we should recruit and enroll a total of 25 households from each of 50 study clusters (25 clusters per study arm). Our final sample size accommodates for 20% of households being lost to follow-up and/or household level loss to follow-up. Poor compliance to intervention allocation and inconsistent adherence to the intervention is also accommodated within this loss to follow-up, as we calculated the expected difference based on average change on the intervention (since this is an effectiveness study). Our sample therefore included 30 households in each study *kebele*.

Post hoc power analysis

Following baseline, we conducted an ex-post power calculation for our main outcome, mental well-being, using the World Health Organization (WHO)-5. Given the prevalence of poor well-being (33.2%) and a calculated ICC of 0.026, we are well powered (>99%) to detect the expected difference of 37%. Our detectible difference is 8% points (24% reduction) to a value of 30.5%.

For our secondary health outcomes, we conducted an ex-post power calculation, and determined that we are able to detect the following:

- 1. Anxiety with baseline of 29.7% (ICC: 0.007) 7% point reduction;
- **2.** Depression with baseline of 20.8% (ICC: 0.01) 6% point reduction; and
- **3.** Emotional distress with baseline of 17.3% (ICC: 0.005) 5% point reduction.

Given the low prevalence of diarrhea and the recent large-scale studies that showed mixed effects of the impact of WASH on diarrhea (Luby et al., 2018; Null et al., 2018), we consider this a secondary outcome, and not one we are powering to detect. However, given a baseline reported diarrheal prevalence of 9.1% for the index child during the past 7 days (ICC: 0.077), we are powered to see a 6% point reduction to 3.1%.

5.2. Sampling methodology

The Andilaye team employed a structured sampling strategy to randomly select eligible kebele clusters and study households. The primary sampling unit (PSU) for this study was the *kebele*; specifically, any rural or peri-urban kebele that is accessible throughout the course of the year. The ultimate sampling unit (USU) for this study is the household; specifically, any household residing in a targeted, sentinel gott within a randomly selected study kebele. While we randomly selected eligible study clusters (i.e., *kebeles*), we purposively selected *gott(s)* from which we randomly selected study households. We utilized a 'fried egg' ("Cluster Randomised Trials," 2017) approach to purposively select one or two gotts that are either situated in/near the center of the kebele (if there are centric *qotts*) or are not adjacent to any other study kebele (in the event there are no centric *gotts*). The number of targeted *gotts* depended only on the number of eligible households that consented to participate in the study. The purposive selection of data collection sites within study clusters via the 'fried egg' approach is justified, as it minimizes spill-over of intervention effects and other externalities associated with the research between intervention and control clusters, especially those adjacent to each other. In accordance with our sample size calculation, we randomly selected approximately 30 households total per kebele cluster. Not all households met eligibility criteria for inclusion in the study sample, and some households refused to participate in the study.

5.3. Target study population

Our target study population included all households residing in randomly selected rural and peri-urban *kebeles* that are accessible throughout the course of the year in Farta, Fogera, and Bahir Dar Zuria *woredas*. These *woredas* are located within South Gondar and West Gojjam Zones in Amhara National Regional State.

5.4. Inclusion and exclusion criteria

Kebele-level criteria: Rural and peri-urban *kebeles* in Farta, Fogera, and Bahir Dar Zuria *woredas* that are accessible throughout the course of the year were targeted for selection into the evaluation. Given intervention implementation is being supervised by local government officials (e.g., *woreda* health officers, CHC HEWs supervisors), it was necessary for the *kebeles* to be accessible throughout the course of the year, to demonstrate proof of concept regarding the effectiveness of the *Andilaye* intervention.

While sanitation coverage and utilization were originally incorporated as inclusion criteria, the veracity of those data were questionable in many *kebeles* in which initial visits were made (i.e., only one latrine observed in a community in which sanitation coverage was reportedly over 80%, and community reports of this being the case for as long as people could recall). Due to uncertainty with regard to the sanitation coverage and utilization data, and the fact that it became apparent during formative research that behavioral slippage was rampant even in *kebeles* previously declared as ODF, we decided to drop those criteria from inclusion requirements. The *Andilaye* team did discuss this change in selection criteria with relevant donors, who agreed the study would demonstrate added value if it could include otherwise eligible *kebeles*, regardless of their sanitation coverage and use or previous CLTSH triggering status, to explore issues related to behavioral maintenance and prevention of behavioral slippage.

Household-level criteria: Inclusion criteria for the *Andilaye Impact Evaluation* included any household randomly selected from the *gott* census book that resides in the target *gott*(s) that:

1. Had at least one child aged 1-9 years at baseline (1-10 years during follow-up 1, one year post-baseline) and consented to allowing study staff to observe the children, specifically their faces and hands;

2. Provided consent to participate in our study, with at least one adult household member consenting to serve as the primary survey respondent.³

We excluded from enrollment in our study any household that:

- 1. Refused to provide consent to participate in our survey;
- 2. Was repeatedly vacant or does not have an appropriate member of the household (capable adult) home to serve as the household's respondent after three attempts to engage the household; and
- **3.** Did not have a household member aged 1-9 years (at baseline, 1-10 years at this data collection moment one year post-baseline) living in the household.

At baseline, after consulting with the field supervisor, the enumerator replaced excluded households with the next randomly selected household on the eligible household register. Field supervisors and study supervisors from Emory Ethiopia supervised field activities and ensured enumerators were only surveying households within the eligible household sampling frame in order to guarantee the sample was random and standardized.

Actual recruitment of households selected for the *Andilaye Impact Evaluation* took place within the home compounds. The enumerator made contact with adult members of the household; she explained the purpose of the visit, the purpose of the study, and asked the respondent if s/he would be willing to consent to participate in the study. Enumerators assessed household level eligibility by asking potential survey respondents a series of questions that lead to a determination of eligibility. Potential survey respondents were informed that they had the right to choose not to participate in the study, the right to refuse to answer any question, and the right to stop the survey for any reason at any point in time.

5.5. Sampling frames and sample selection

All *kebeles* that are rural or peri-urban, and are accessible throughout the course of the year, per *Woreda* Health Office definition, situated in the three targeted *woredas* were eligible for inclusion in our study. The enumerated list of all *kebeles* the three respective *Woreda* Health Offices maintain served as the first level sampling frame. From this sampling frame, we employed a random number generator and a stratified (at the *woreda* level) selection approach to identify 50 eligible *kebele* level clusters from across the three *woredas* for inclusion in our study. Given each of the three *woredas* vary with regard to their hydrogeological conditions and the size and number of *kebeles*, we deemed a stratified selection approach appropriate, and used it to select study clusters. Of the 50 clusters, 22 were selected from Farta, 12 from Fogera, and 16 from Bahir Dar Zuria. Proportionally, these selected *kebeles* represented 51 (22/43), 38 (12/32), and 50 (16/32) percent of all *kebeles* in Farta, Fogera, and Bahir Dar Zuria, respectively. An even number of clusters were selected from each *woreda* to ensure an equivalent sample size between the intervention and control clusters selected from each *woreda*. Fogera had a slightly lower proportion of selected *kebeles* due to accessibility concerns given the frequency of floods in the low-land, marshy areas close to Lake Tana.

Once an appropriate *gott* was selected from each *kebele*, the team worked with the HEW to obtain a list of all of the households within the *gott*, specifically those with a child aged 1-9 years, as per study inclusion criteria. In order to operationalize this in a standardized manner, we used The Carter Center's (TCC) household census books, which are kept at the Health Post and enumerate all households and all household members (by age) residing within the household. At each Health Post, we obtained all TCC census books pertaining to the relevant *gott(s)*. When there was more than one version of the TCC census book (i.e., books from censuses conducted during different years), the book with the latest census data was used as the *gott* sampling frame. After all relevant books were gathered, the total number of households in the *gott* was determined (by counting up the number of households from each

³ The adult individual had to be capable of understanding and providing informed consent.

of the *gott*'s TCC census books), and a random number generator was used to generate a list of 60 households per *gott* that reportedly had a child between the ages of 1-9 years. The list exceeded the total number of households that would be required for enrollment in each *gott* in order to allow for replacement in the field if the household did not, in reality, meet inclusion criteria (i.e., have at least one child aged 1-9 years), have an eligible respondent available after three attempts, or have an eligible respondent who consented to participate in the study.

During data collection, each field supervisor and enumerator were provided a list of households, and instructed to visit each household, starting with the first household on the list, to invite them for enrollment in the study, ask an eligible adult representative to provide informed consent, and administer the household survey. If households were absent, no eligible adult respondent was available or refused to consent, or upon further conversation with the household it became apparent that the household was not eligible (e.g., the death of the only child between 1-9 years, the child was actually older than 9 years), the enumerator electronically recorded the information and notified the field supervisor. If the household was absent or no eligible adult was available, the enumerator visited the household three times prior to replacing that household with the next household on the list - after contacting the field supervisor to confirm the replacement. If the household was otherwise ineligible (e.g., refused consent, no children within the targeted age range), the enumerator replaced that household on the list – after contacting the field with the next household on the list – to confirm the replacement.

5.6. Data collection procedures

Formative research commenced in September 2016, after the receipt of the appropriate ethical approvals from relevant University Institutional Review Boards (IRBs) and the ARHB (see section entitled *Human subjects and ethics* for further details). Subsequent to the formative research phase, the *Andilaye* team launched the official RCT with baseline data collection during March – April 2017.

An important part of the baseline data collection included the characterization of baseline conditions. This is important for monitoring long-term intervention fidelity. During baseline data collection, the *Andilaye* team obtained data on latrine construction and use and hygiene behaviors via spot checks. The team will continue collecting these data throughout the study period, for two years following treatment allocation. The first round of follow-up data collection (follow-up 1), presented here, took place in each location one year after the baseline data collection (March – April 2018). A second round of follow-up data collection (follow-up 2) will take place one year later in March – April 2019 to assess behavior change sustainability, with two rounds of monitoring throughout (July and November 2018).

5.7. Tool development

When developing the baseline and follow-up survey instrument, the *Andilaye* team pulled from its reserve of existing WASH and NTD survey instruments, and leveraged formative research data to contextually adapt the survey prompts and answer choices. To the greatest extent possible, the team included validated metrics for assessment. Prior to enumerator training, the *Andilaye* survey instrument was translated into Amharic, and back-translated by two independent Amharic speakers. The team discussed any discrepancies noted between the intended English prompts and the Amharic translations (identified via the back-translations). Survey prompts were revised accordingly. In order to ensure face validity, the vast majority of the survey instrument was tested via cognitive interviews. Through the use of this qualitative method, which included a 'think-aloud' technique, the *Andilaye* team obtained feedback from formative research households about the meaning, comprehensiveness, and appropriateness of survey prompts and their related answer choices. Once the Amharic version of the tool was complete, four enumerators were trained on the tool, and brought on to conduct a week-long field pilot in targeted formative research communities. At the end of each day of piloting, the

team discussed issues related to respondent comprehension of survey prompts and answer choices, survey logic and skip patterns, and suggested revisions. At the end of the piloting period, key data were checked and analyzed, and related modifications were made to the tool prior to finalizing for supervisor and enumerator training.

5.8. Training

Below, we outline details related to the training approach we employed for our RCT. We used the same training curriculum, approach, and lead trainer for both baseline and follow-up 1. To the greatest extent possible, baseline enumerators were re-hired and re-trained to collect follow-up 1 data.

Supervisor training: Prior to engaging with the full team of enumerators, senior technical research staff oriented field supervisors to the *Andilaye* project, their roles and responsibilities as supervisors, field and debriefing check-lists, and the supervisor validation survey (i.e., a sub-set of the survey instrument). The supervisors were also involved in the enumerator training, so they received additional information on the survey tool during that time so as to ensure the entire enumeration team had the same understanding of the baseline survey prompts.

Enumerator training: Subsequent to enumerator recruitment, senior research staff conducted a training with enumerators that included topics related to research ethics, rights and protection of research participants, the informed consent process, data collection tools and procedures, and the use of electronic mobile data collection applications. The training was conducted in both English and Amharic (i.e., some more technical topics were first presented in English, but to ensure thorough comprehension among the enumeration team, were also summarized in Amharic immediately after the English explanation was presented). In order to ensure enumerators had a thorough understanding of the survey instrument, senior staff facilitated a group training and discussion with all enumerators. After the purpose and meaning of each prompt and set of answer choices had been discussed, enumerators broke out into teams of two to practice administering the survey. Upon the conclusion of the officebased practice session, the team came back together to discuss any questions, concerns, or suggested revisions. Only after that point did the entire enumeration team move to the field for piloting. Survey piloting occurred iteratively, with an initial visit to the field, followed by revisions to the survey, and a subsequent day of piloting. Prior to the initial of official survey data, all enumerator and supervisor survey concerns had been resolved.

Trainings on data collection tools and field procedures included discussion among all field supervisors and enumerators to ensure a standardized approach to data collection via household surveys and structured observation. Pilot testing of the instrument ensured enumerators were familiar and comfortable with the survey prior to official data collection, while also allowing senior research staff to check that the instrument was an appropriate tool for collecting desired data.

5.9. Field procedures

During the household surveys, enumerators sought out adult respondents within households enrolled in the *Andilaye* trial, with preference going first to the primary female caretaker of the index child (i.e., the youngest child between the ages of 1-9 years residing in the household at baseline; the youngest child aged 1-10 years at this data collection moment, one year postbaseline), as she would tend to know the most about the latrine use, defecation, and personal hygiene practices of most members of her household. If she was not available, enumerators sought out other household members in the following order: eldest available female caretaker, eldest available female household member, eldest available male caretaker, or eldest available male household member. All household members present during survey administration were asked to self-report on their own sanitation and hygiene habits, and hand cleanliness and facial and hand cleanliness were assessed for the primary survey respondent and all children under the age of ten years, respectively.

5.10. Loss to follow-up

Households enrolled in the study that were lost to follow-up one year post-baseline (follow-up 1) were defined as those in which:

- 1. Consent to participate in our follow-up survey was refused;
- 2. The entire household moved out of the study kebele; or
- **3.** No child aged 1-10 years (at follow-up 1) resides in the household any longer (e.g., death, divorce, moved).

Households with no eligible respondent available after three attempts during follow-up 1 data collection (i.e., household/eligible respondents repeatedly vacant after three attempts) will continue to be enrolled in the study, and will be visited for subsequent quarterly monitoring and follow-up 2 data collection. However, follow-up data from these households are not represented within this report.

5.11. Quality assurance

Field data were captured electronically on password-protected mobile phones (stored securely when not in use) to improve accuracy of data entry and enable immediate review of results. Logic, range, and consistency checks were incorporated into the electronic data collection file to further improve data quality and minimize data entry errors. To ensure data quality, the supervisory team, comprised of faculty and staff from Emory University and Emory Ethiopia, coordinated and supervised data collection along with field supervisors. Field supervisors independently assessed all objective measures (simultaneous to enumerator assessment) at 10% of households in each study *kebele*. These data will be analyzed to determine inter-rater reliability of related metrics.

5.12. Key outcome, intermediate, and sub-group indicators

The follow-up survey instrument consisted of several modules that aimed to collect information from respondents regarding:

- 1. Their basic demographics;
- 2. Reported defecation, urination, and latrine use practices for the respondent, head of household, and individuals aged 4-17 years, including use for child feces disposal for children aged 0-3;
- **3.** Recent illness/symptom reports (e.g., diarrhea, dysentery and other negative control indicators) for all children aged 0-10 years, and assessments of facial and hand cleanliness of each child aged 1-10 years who is present at the household during the time of the survey;
- 4. Animal husbandry and hygiene practices;
- 5. Water utilization practices and water insecurity;
- **6.** Personal hygiene practices (e.g., facewashing, handwashing, bathing, fomite washing);
- 7. Sanitation insecurity;
- 8. Mental health (e.g., anxiety, depression, mental well-being); and
- 9. Waste management (e.g., solid and liquid waste); and
- **10.** Latrine construction and repair; structure and maintenance; and functionality.

Andilaye enumerators also collected data via direct observation of objective measures of the following:

- 1. Hand and facial cleanliness;
- 2. Sanitation and hygiene facilities (i.e., latrines, washing stations); and
- **3.** Waste management.

Structured facilities observations included visual inspections of all household latrines and hand/facewashing stations with a prescribed checklist that was used to assess type of latrine, functionality, evidence of use, maintenance, and cleanliness. Structured observations also included an examination of the household compound for evidence of open defecation and/or inappropriate disposal of child feces.

Enumerators administered all survey modules among all households enrolled in the study, such that all households in the sample provided self- or respondent-reported data, as well as data on other objective measures. All households in the sample had a latrine spot check if they were willing to consent to the enumerator observing the household latrine.

As indicated above, enumerators completed facial and hand cleanliness observations via visual inspections of the faces and hands of all children aged 1-9 years present at and residing within the household during survey administration. Enumerators also conducted hand cleanliness assessments on the primary survey respondent (as indicated above – this person was ideally a primary caregiver). The team assessed facial cleanliness systematically, through an evaluation of the absence of flies, dirt, and/or debris on the face; and eye, nose, or other discharge emitted from facial orifices. One child aged 1-9 years identified at baseline served as the household's index child, and provided additional proxy measures for practices targeted by the intervention. This child is being followed throughout the duration of the study, and provides additional measures at every major data collection moment (e.g., baseline and 1 and 2 year follow-ups).

Appendix D shows the survey prompts used to generate respondent-reported data and guide direct observations that were made in households. These indicators relate to several of our key behavioral outcomes and indicators thereof. The *Andilaye* team has also captured data on behavioral antecedents specifically related to our *Andilaye* intervention and the behaviors the intervention intends to target for change and maintenance. Data on these behavioral outcome and antecedent indicators collected during baseline and 1 and 2 year follow-ups provide longitudinal tracking which increases study power, reveals seasonal and temporal variations in intermediate and behavioral outcomes, provides more time to assess key health impacts, and offers a longer perspective for examining externalities, cost-effectiveness, and the sustainability of sanitation and hygiene behavior change.

5.13. Targeting of WASH-related, NTD-preventive behaviors for intervention inclusion

Our Andilaye intervention is designed to address and improve specific WASH-related, NTDpreventive behaviors, as indicated by our formative research and findings from our baseline analyses. In order for interventions to be effective, it is necessary to focus on a select number of behaviors and related practices to address. Attempting to change a multitude of behaviors all at once can prove ineffective. As such, the *Andilaye* team used a systematic approach to leverage findings from our problem and solution tree analyses (i.e., data synthesized from our focus group discussions, key informant interviews, community and household observations), feedback from government and other stakeholders, and our baseline data to identify a parsimonious, yet influential set of behaviors the intervention is designed to improve. The overarching behavioral themes the intervention will address include:

- **1.** Exclusive utilization (by all family members) of sanitation facilities for defecation and urination;
- 2. Improved personal hygiene practices (i.e., handwashing with soap/substitute after handling animal and human feces, including after defecation and handling child feces, and before handling food; facewashing whenever faces are dirty); and
- **3.** Other household and environmental sanitation (i.e., animal husbandry/hygiene, feces, and waste disposal).

To draw attention to results regarding assessments of these behaviors and related practices, data for these key indicators are solely presented in the results tables (see section entitled *Summary of year one follow-up results* for further details). Supplemental indicators along with key indicators (**in blue bolded text**) are reported in Appendix J.

5.14. Description of comparison arm

The comparison group will receive current standard of care sanitation and hygiene programming (i.e., interventions related to FMoH's current CLTSH model). Any other intervention in comparison communities will be limited, and we will be working with government partners to ensure this is the case. Table 2 outlines differences between the *Andilaye* intervention and the current CLTSH model (see Table E.2 for more details).

Table 2. Sampling,	per treatment arm
--------------------	-------------------

Intervention arm	Comparison arm
Andilaye intervention	Standard of care sanitation & hygiene (i.e., current CLTSH)
1-2 sentinel <i>gott</i> (s) from each of 25 randomly selected <i>kebele</i> s, ~30 randomly selected households enrolled in each <i>kebele</i>	1-2 sentinel <i>gott</i> (s) from each of 25 randomly selected <i>kebele</i> s, ~30 randomly selected households enrolled in each <i>kebele</i>
A demand-side sanitation and hygiene intervention that incorporates NTD-preventive behaviors, focuses on positive, community-oriented motivators of behavioral change, promotes achievable incremental improvements, and incorporates strategies that facilitate behavioral maintenance (i.e., prevention of behavioral slippage or relapse back to unimproved behaviors).	The <i>Andilaye</i> team will not attempt to modify the government's roll-out of its CLTSH model. Therefore, comparison communities may receive interventions related to FMoH's CLTSH model. We will work with our government partners to minimize other WASH interventions in our comparison communities, to the greatest extent possible.

5.15. Randomization: Allocation rule for intervention and comparison groups

Following baseline data collection, we used a stratified random design to assign study *kebeles* to either the intervention or comparison arm. Within each stratum (*woreda*), we used a random number generator to generate a random number between zero and one for each *kebele* (cluster), and then ordered each *kebele* by the randomly generated number in ascending order. We then partitioned the communities within each *woreda* into two equal sizes, assigning the first half of *kebeles* to the intervention arm and the second half to the control arm. We used replacement randomization (Lachin, 1988) to secure balance across three key variables (latrine coverage, washing station with soap coverage, and head of household education). Cluster randomized trials, particularly trials with a small number of clusters, often have individual-level imbalances between arms. Therefore, we established *a priori* that the intervention and control mean values for these three variables should be within two standard deviations of the overall mean for these variables. The randomization process described above was repeated (twice) until these variables were balanced according to that *a prior* criterion. Appendix E enumerates *kebele* treatment allocation (Table E.1), and provides a summary of the study flow (Figure E.1).

While allocation occurs at the *kebele* level, we only collect data in one to two sentinel *gotts* per *kebele*, purposively selected to minimize spillover. The number of sentinel *gotts* per *kebele* in which data are collected depends solely on the number of eligible and consenting households (i.e., if less than 30 eligible and consenting households are present in one *gott*, data collection took place in a second sentinel *gott* as well in order to obtain the minimal required sample size per *kebele* cluster).

After enrollment and baseline survey administration, we randomly allocated *kebeles* to either the intervention arm to receive the intervention, implemented in collaboration with the *Woreda* Health Offices, HEWs, WDALs, and other community change agents, or to the counterfactual (comparison) arm to receive "standard of care" (i.e., current sanitation and hygiene programming, including potential roll-out of the current CLTSH intervention in *woredas* and *kebeles* in Amhara per the existing FMoH implementation protocol).

5.16. Human subjects and ethics

Ethical approval for the *Andilaye Impact Evaluation* was provided by Emory University (IRB00076141), the London School of Hygiene & Tropical Medicine (9595), and locally by the ARHB (HRTT0135909). In addition, we registered the trial on clinicaltrials.gov (NCT03075436). We provided study participants with full details regarding the study as well as their rights as a participant in the study prior to inquiring about consent to participate. This process took place in Amharic. The *Andilaye* team took appropriate steps to ensure confidentiality for all study participants.

5.17. Data analysis

All data were cleaned and analyzed using Stata 14 (StataCorp, College Station, TX).

5.17.1. Descriptive analyses design

Our study design is a parallel cluster-RCT. Population percentages or means were calculated and shown for key indicators at follow-up. We present data on the overall prevalence of indicators, and compare results between intervention and comparison arms to assess balance between randomization arms.

Binary variables (those with "yes/no" response options) were presented as the percentage of households, communities, or water sources that meet the criterion indicated. Continuous variables were presented as population means.

5.17.2. Validity of randomization: Balance of variables between arms

Our baseline analyses built the infrastructure for follow-up analyses. We did a bivariate assessment of balance between intervention arms at baseline. These results serve as the assessment of balance in our intervention studies. Here, we used log-linear binomial models to estimate prevalence ratios comparing the outcomes between the intervention and control arms. All models accounted for the stratified design, through the inclusion of the *woreda* indicator variable (Kahan and Morris, 2012), as well as clustering within *kebeles*, through the use of generalized estimating equations with robust standard errors.

The primary analysis of the overall *Andilaye* trial reflect an "intention-to-treat" analysis, which compares the intervention arm to the control arm, without regard to intervention fidelity or adherence. The models we used for these follow-up analyses are similar to those we will use for the overall trial analyses. We will preferentially use log-linear binomial models to estimate prevalence ratios, comparing the outcomes between the intervention and control arms. Log-linear binomial models often have difficulty converging, and so we may instead use modified Poisson regression, if we encounter problems with convergence (Zou, 2004). We will preferentially use mixed effects models to account for clustering and correlation of the repeated observations over time. Mixed models also often have issues with convergence, and so we may instead use generalized estimating equations with robust standard errors, if we encounter this problem. For continuous outcomes, we will instead use linear models, and for count outcomes, we will instead use either negative binomial or Poisson regression. All models will include the intervention variable, and account for the stratified design, through the inclusion of the *woreda* indicator variable (Kahan and Morris, 2012), as well as clustering within *kebeles*, as described above.

There is often interest in determining the absolute difference in outcomes between the intervention and control arms, as opposed to a relative ratio (e.g., a prevalence ratio). For key outcomes, we will also present a difference measure. We will use the same models as described above (e.g., log-linear binomial models), but use post-estimation commands to estimate the average marginal effects.

Given no imbalances were detected between arms at baseline for any of the primary variable of interest, we did not need to perform supplementary analyses, as outlined in our pre-analysis plan, to control for the baseline levels of these imbalanced variables in more fully adjusted models.

For many of our outcomes, there is interest in determining the impact of the intervention across various sub-groups, such as sex, follow-up round (once multiple rounds of collection are completed), exposure to previous triggering, and modifiers such as water and sanitation insecurity. For select key outcomes, we will use interaction terms and/or stratification, and we will present the impact of the intervention at each level of the sub-group variable (e.g., separately for boys and girls).

Sub-optimal adherence, due perhaps to low uptake of interventions by participants or to poor implementation of the interventions, can lead to trial results that do not reflect the true efficacy of WASH. The intention-to treat analyses show a valid and unbiased causal effect for the effectiveness of a specific WASH program/intervention, but do not show the efficacy of WASH under ideal circumstances. If fidelity and adherence are heterogeneous, we will supplement our intention-to-treat analyses with an instrumental variable analysis, a per-protocol analysis, or an as-treated analysis, at the end of the RCT. Given we need to accommodate for the time required for intervention roll-out, and the collection of process evaluation and monitoring data to assess adherence, these analyses will not be possible until the end of the RCT (after followup 2). When we are able to perform these analyses, each will attempt to assess the impact of optimal adherence to the Andilave intervention on our various outcomes of interest. We have performed such analyses in other studies (Garn et al., 2017, 2016). We will preferentially use instrumental variable analyses if the instrumental variable study assumptions appear to be met, but will default to the as-treated and/or per-protocol analyses if the assumptions are not met. Both as-treated and per-protocol approaches have limitations with regard to potential confounding by unknown or unmeasured factors.

6. Program or policy: Design, methods, and implementation

6.1. Key program elements and intervention activities

The Andilaye intervention was developed by Emory University, in partnership with the FMoH, the ARHB, and relevant Zonal Health Departments and *Woreda* Health Offices. All behavior change illustrations developed for *Andilaye* were produced by an artist based in Ethiopia, and were informed by our formative research. We also employed a cognitive validation process through which we obtained feedback from potential participants and implementers (e.g., WDALs, HEWs) regarding their comprehension and ease of use of the materials and to ensure the images and tools were locally appropriate and acceptable. Emory University oversaw the development of all tools and material used for the intervention. The description and function of key tools and materials developed for the *Andilaye* intervention are provided in Table 3. The methods of delivery of intervention content and further details on the implementation of behavioral change catalyzing and maintenance activities are shown in Appendix F.

The facilitation and overall supervision of the intervention at each level was led by a qualified Ethiopian-based *Andilaye* team (Emory Ethiopia). These members of the *Andilaye* team were heavily involved in the formative research, design, and trialing of the *Andilaye* intervention

and, thus, possessed high capacity to facilitate the implementation of activities when local government officials failed to take on this role.

While the primary target participants of the *Andilaye* intervention are community members, the intervention also builds capacity amongst *kebele* and *woreda* level stakeholders (e.g., WDALs, HEWs, CHC HEWs supervisors, *woreda* officials). Participation, adherence, and compliance of these stakeholders to the *Andilaye* intervention were designed to be facilitated through a chain of supervision – starting with the *Andilaye* team. These *kebele* and *woreda* level stakeholders were incentivized to partake in *Andilaye* trainings and workshops by receiving a regional standard per diem to compensate them for travel and accommodation to attend. Outside of the provision of an *Andilaye Household Goal Card*, no incentives were provided to primary caregivers or household members for partaking in the counseling caregiver visits.

The purpose of the *Andilaye* intervention, to design and test an enhanced, demand-side sanitation and hygiene intervention to complement the existing efforts – specifically CLTSH, the HEP more generally – was communicated to *kebele* and *woreda* level stakeholders.

Tools and material ^a	Description/function	Key activities
Overview of	PowerPoint presentation that provides an overview of: (1)	 Sensitizing and action planning workshop
Andilaye	the project and its purpose, (2) target behaviors/domains	
presentation	and intervention activities, and (3) intervention tools.	
Training guides	Detailed protocols and agendas used to guide activities and maximize quality and fidelity of <i>Andilaye</i> intervention trainings and workshops.	 Sensitizing and action planning workshop Skills-based training of the trainers for HEWs, CHC HEWs supervisors, <i>Woreda</i> officials Training of community conversation facilitators Whole system in the room and action planning
Training handout	A four-page training handout that provides a summary of: (1) <i>Andilaye</i> behavior change messaging, (2) purpose and desired outcomes of <i>Andilaye</i> counseling visits with caregivers, (3) roles and responsibilities of WDALs, HEWs, CHC HEWs supervisors, <i>Woreda</i> officials and the <i>Andilaye</i> team, (4) key steps for performing <i>Andilaye</i> counseling visits with caregivers, and (5) <i>Andilaye</i> team contact information.	 Skills-based training of the trainers for HEWs, CHC HEWs supervisors, <i>Woreda</i> officials
Supportive supervision checklists	Two-page tool that provide HEWs, CHC HEWs supervisors, and <i>Woreda</i> officials with a supervision checklist for on the job training, which guides them to provide WDALs with constructive feedback, first delivering positive feedback regarding what she is doing well and reinforcing any good counselling strategies, then to provide constructive feedback by identifying areas for improvement and specific strategies for how to improve.	 Skills-based training of the trainers for HEWs, CHC HEWs supervisors, <i>Woreda</i> officials
Action planning tools	Action planning templates completed during workshops and trainings to identify the timing of subsequent <i>Woreda</i> and community-level intervention activities, and delegate roles and responsibilities amongst the various <i>Woreda</i> and <i>Kebele</i> stakeholders in implementing the <i>Andilaye</i> intervention.	 Sensitizing and action planning workshop Skills-based training of the trainers for HEWs, CHC HEWs supervisors, <i>Woreda</i> officials Training of community conversation facilitators Whole system in the room and action planning
Playbill	One-page handout distributed to the <i>Andilaye</i> team, selected <i>Gott/Kebele</i> stakeholders from the 'whole system in the room and action planning' activity, and performance groups to plan and guide the community mobilization and commitment event and assure that all aspects of the event are completed and done in the correct order.	Community mobilization and commitment event
WASH role play scripts	Three role play skits, aligning with each <i>Andilaye</i> intervention behavioral domain, that acknowledge real-life behavioral barriers to the target WASH behaviors while utilizing motives and social messaging to emphasize the benefits.	Community mobilization and commitment event

 Table 3. Description and function of key tools and materials developed for Andilaye

Andilaye commitment banner	3x6 foot banner that provides a reminder of community commitment to adopting improved WASH practices and hangs in a central location in the community following the community mobilization and commitment event.	 Community mobilization and commitment event
Andilaye CC Facilitator Flipbook	40-page illustrative flipbook used by selected <i>Gottl/Kebele</i> stakeholders to guide community conversations. The flipbook contains information regarding the ideal way to perform the 11 targeted WASH practices of the <i>Andilaye</i> intervention, and aims to change attitudes toward improved practices by discussing the benefits of practicing the target WASH behaviors and dispel misconceptions through demonstrations.	Community conversations
Andilaye 'Gobez!' Flipbook	40-page illustrative flipbook used by WDALs to guide monthly counseling visits with caregivers at the household. The flipbook contains information regarding the ideal way to perform the 11 targeted WASH practices of the <i>Andilaye</i> intervention, identifies benefits related to the improved practices, identifies barriers, and provides counseling for caregivers on how to plan for, cope with, and overcome barriers.	Counseling visits with caregivers
Andilaye Household Goal Card	1x3 foot goal card used by household caregivers and WDALs to set and monitor household level goals for improved WASH behaviors. The goal card hangs in the caregiver's household and goals are reviewed by the WDAL and household caregiver each month during the counseling visits with caregivers.	 Counseling visits with caregivers
Andilaye Household Monitoring Matrix Card	Letter-sized card used by WDALs to indicate which Andilaye intervention behavioral domain(s) were covered during each counseling visit with caregivers. The monitoring matrix hangs in the caregiver's household next to the goal card.	Counseling visits with caregivers

^a Key tools and materials developed for Andilaye can be found <u>here</u>

6.2. Process monitoring

For the purposes of our study, we are monitoring capacity and performance indicators amongst local and district level officials (e.g., CHC HEWs supervisors, *woreda* officials), community-based change agents (e.g., HEWs, WDALs), and organizational staff (e.g., Emory Ethiopia), with a particular focus on their ability to support the *Andilaye* intervention. We are collecting process evaluation data over time, therefore allowing us to conduct trend analyses to determine whether capacities change amongst these various cohorts over time. Note, as intervention activities will continue until follow-up 2 (March – April 2019), process data collection will continue as well. We classified key process evaluation components into three domains, as explained below and in further detailed in Appendix H.

1. Intervention implementation fidelity

Objective: To document the extent to which the intervention was delivered, as planned. This attribute will be used to assess the quality and integrity of the intervention as originally conceived by the study team.

2. Participation and dose response

Objective: To assess the extent of engagement of participants with the intervention. This attribute will also help us pinpoint any variations among our different study communities. This enables us to better understand the extent to which intervention participation and dose response are associated with the delivery of the intervention (dose delivered).

3. Context

Objective: To understand which environmental factors (social, political, and economic) may influence intervention implementation. This attribute will help us to understand
potentially important contextual factors that may influence the delivery or outcome of our intervention.

We are using a mixed methodological approach to capture process data. We are collecting quantitative data via semi-structured interviews and activity observations, post-assessments of trainings, systematic audits of key records, and structured household-level surveys. We are collecting qualitative data via semi-structured interviews and informal discussions with key informants (e.g., government stakeholders, HEWs, WDALs, and community members at large) and participant observations during relevant trainings and intervention activities. Caregivers, WDALs, HEWs, CHC HEWs supervisors, and CHC heads are key informants engaged to investigate community level uptake and context of the intervention. *Woreda* administration office heads, *kebele* administration heads, hygiene and sanitation officers and heads from *woreda* level education offices, agricultural offices, and health offices are being engaged to investigate government stakeholder motivation. The primary focus of interviews with key informants is to capture the extent to which the participants actively engage with, are receptive to, and conducted behaviors/activities promoted by the intervention and to gather participation feedback on the implementation of the intervention.

6.3. Intervention implementation

As discussed above, the implementation of *Andilaye* intervention activities began in September 2017 and will continue, with a focus on behavior change catalyzing activities, through follow-up 1 (March – April 2018). The focus of intervention activities will transition to behavior change maintenance activities in accordance with household and community progress. As such, the intervention will continue supporting both catalyzing (when and where the need dictates) and maintenance focused activities through the end of the study (i.e., follow-up 2, March – April 2019). As of the writing of this follow-up report, initial rounds of all behavior change catalyzing activities have been implemented, apart from the training of community conversation facilitators and the subsequent group-level community conversations (see process data in Appendix H).

To date, there have been no major changes to the design and implementation of the *Andilaye* intervention compared to what was outlined in our intervention manual and accompanying field protocols. Following the intervention design phase of the project, our team developed strict protocols and accompanying tools and materials (see Table 3) that were adhered to for each intervention activity. These protocols were prescriptive in how the *Andilaye* team and community-based change agents facilitated intervention activities. On-the-spot-innovations were deterred from during the implementation of behavior change catalyzing activities as protocols, tools, and material for these activities incorporated comprehensive feedback from key stakeholders and study participants (e.g., WDALs, HEWs, community members) during intervention design. Future skills-based review meetings and refresher trainings with government stakeholders, HEWs, and WDALs will allow stakeholders to review successes and discuss how to address challenges faced in implementation.

6.4. Process data

Extensive action planning, at all levels of intervention (e.g., *woreda, kebele*, group, household), was completed during the roll out of intervention activities. While coordinating with stakeholders to conduct district and community-level intervention activities caused slight delays in the roll out of the intervention (e.g., trainings and workshops), the outcome of this planning resulted in high participation of target audiences across all intervention activities. Furthermore, the capacity of implementers and their ability to follow tools and materials developed for the *Andilaye* intervention are reflected in the high levels of dose response amongst activity participants (e.g., action planning tool completed and high post assessment scores). See an overview of process data for completed activities in Table 4 and further details in Appendix H.

Activity	Intervent	ion implem fidelity	entation	Participat	tion and	Context			
	Date(s)	All study sites received activity	All study sites received activity as planned	All study sites had target population participate in activity as planned	All study sites had activity outcome as planned	Comments			
Sensitizing and action planning workshop	September – October 2017	\checkmark	~		\checkmark	 One study <i>kebele</i> had an administrator that was not in attendance Some regional representatives and <i>woreda</i> administrators did not attend the afternoon session of the workshop 			
Skills-based training of the trainers for HEWs, CHC HEWs supervisors, Woreda officials	December – January 2018	✓	✓	~	✓	• Participants scored well (average 65-72%) when tested on key <i>Andilaye</i> intervention components during post-training assessment			
Skills-based training of Women's Development Army Leaders	January – February 2018	✓	✓	~	✓	 In some communities, Emory Ethiopia members acted as trainers and trained WDALs together with HEWs 96% of WDALs in targeted communities were trained Trained WDALs scored well (average 80-95%) when tested on key <i>Andilaye</i> intervention components during post training assessment 			
Whole system in the room and action planning	January – March 2018	\checkmark	~	~	\checkmark	 Expectation of per diem associated with activity attendance was a source of disappointment for some participants. However, facilitators were able to lead community stakeholders to completing an action planning tool 			
Community mobilization and commitment event	March – April 2018	~	~			 The per diem issue associated with the 'whole system in the room and action planning' activity might have impacted the involvement of some event coordinating committee members identified during action planning. This may have impacted mobilization of the community members Setting community bylaws and, most importantly, determining sanctions (deciding regulations) for people who violated bylaws was challenging in some communities 			
Counseling visits with caregivers	January 2018 – ongoing	✓	~			 In some communities, WDALs performed few counselling visits with caregivers (frequency) and there are cases where WDALs performed counselling visits with caregivers only to few households (not to all households in their development network) Very limited, or in some communities, complete absence of supportive supervision and on the job training given to WDALs by HEWs Likely due to the above points, the proportion of study intervention households with an <i>Andilaye Household Goal Card</i> observed during follow-up 1 (54%) was lower than expected 			

Table 4. Overview of process data for completed Andilaye intervention activities

6.5. Implementation challenges

As indicated in Table 4, weak links in the implementation of activities were observed. The details, impact, and efforts to mitigate these weak links are outlined below:

1. Turnover of HEWs

While at least one HEW per intervention *kebele* was trained during the skills-based training of the trainers that targeted HEWs, CHC HEWs supervisors, and *woreda*

officials, HEW turnover is a reality. This turnover impacts the capacity of the WDALs to receive effective supportive supervision from HEWs when they conduct their *Andilaye* household counselling visits. To mitigate this, the *Andilaye* team identified new or untrained HEWs in intervention *kebeles*, and conducted a "mop up training" to ensure HEWs are able to support the *Andilaye* intervention, as planned. The mop up training took place six months after the initial 'skills-based training of the trainers for HEWs, CHC HEWs supervisors, *woreda* officials.' Turnover of HEWs, CHC HEWs supervisors and facilitators' which is a district-level maintenance activity scheduled for 8-10 months after the initial training.

2. Local government officials as facilitators of implementation activities

There has been limited integration of *Andilaye* activities within the larger HEP, and supportive supervision to HEWs and WDALs by CHC HEWs Supervisors and *Woreda* Health Office officials has not met targets set by those actors. When the team communicates to district level stakeholders, enthusiasm and motivation to participate in the project is relatively low. HEWs have multiple tasks given by the government and partners, which limits their ability to partake in their roles and responsibilities in *Andilaye* activities as planned. Overlapping of different tasks is likely due to challenges in completing activities based on an endorsed plan, as HEWs are often engaged by unplanned activities given by different authorities at different levels. As a result, the level of supportive supervision and on the job training given to WDALs by HEWs has been limited in some *kebeles*. To overcome these challenges, the *Andilaye* team has increased its role in facilitating and monitoring the implementation of activities, as needed, when local government officials failed to take on this role.

3. Coordinating committee for community mobilization and commitment events

Only 9 of 25 intervention *kebeles* had the entire coordinating committee in attendance of the day of their *Andilaye* community mobilization and commitment event. These events began shortly after Ethiopia declared a State of Emergency in February 2018. As a result, many government officials who were part of the coordinating committees were absent, as they were required to attend other government meetings. The absence of key community members during these events may have impacted the outcome of this intervention activity. However, other key *Andilaye* stakeholders (e.g., HEWs, WDALs, the *Andilaye* team) were present to facilitate these events. These stakeholders helped maintain fidelity, as indicated by the high dose response of study *kebeles* that: (1) determined unimproved WASH practices were no longer acceptable by the community, (2) determined improved WASH behaviors needed to be adopted by the entire community, and (3) set community by-laws and regulations for monitoring the by-laws by the end of the community event.

4. Andilaye Household Goal Card coverage

HEWs from all intervention *kebeles* indicated that WDALs have conducted at least 1-2 *Andilaye* household counseling visits with caregivers 3-4 weeks prior to follow-up 1 data collection. However, the proportion of study intervention households with an *Andilaye Household Goal Card* observed during follow-up 1 (419/780, [54%]) was lower than expected. This could have resulted from three major factors: (1) the goal card was destroyed or moved within the household, (2) goal cards were never distributed by WDALs during initial visits, or (3) WDALs did not have enough goal cards for all households in their catchment area.

The goal setting and commitment process executed by caregivers during the *Andilaye* household counseling visits allows households to set their own goals for the coming month (until the WDAL's next visit). The *Andilaye Household Goal Card* is used to facilitate the goal setting and commitment process. The absence of the goal card would

impact the goal setting behavior change technique targeted in our theory of change. To mitigate negative consequence of this implementation gap, the *Andilaye* team has worked with HEWs to provide them with additional goal cards and a list of households that were not observed to have goal cards during follow-up 1. The impact of these mitigation efforts will be measured during quarterly monitoring visits and follow-up 2.

7. Impact analysis and results of the key evaluation questions

7.1 Results

We highlight key output and outcome findings from the year one follow-up assessment. We provide prevalence ratios (Abele et al., 2011) and associated confidence intervals.

7.2 Summary of baseline results

7.2.1. Baseline demographic information

At baseline, we collected data from a total of 1,589 households (Table 5). Of 1,691 initiated household surveys, 1,589 (94%) met all study inclusion criteria (i.e., inclusion in the analytical sample). This resulted in data on 1,944 children aged 1-9 years. Of those households excluded from the baseline analytical sample, 81 did not have a member between 1-9 years of age, 17 had no eligible respondent available, three initiated but did not fully complete the survey, and one did not provide consent to participate. Given the primary female caregiver of the index child was targeted for baseline survey administration, a large majority (91%) of the respondents were female, by design. Of these 1,589 respondents, 85% were the mother of the index child, only 13% had completed at least some secondary education, and the average age respondent age was 34 years. All respondents reported being of Amharic ethnicity, and 95% reported being Orthodox Christian (data not shown). Approximately one-third (29%) of respondents were the head of the household, while 67% of respondents were the spouse or wife of the head of household (data not shown). Only 17% of the heads of household had completed at least some secondary education. The average age of heads of household was 41 years. The demographic variables appeared balanced across treatment arms, with no meaningful differences in the prevalence of key demographic variables between arms.

7.2.2. Previous triggering at baseline

Of the 50 *kebele* clusters randomly selected for inclusion in the *Andilaye Impact Evaluation*, 78% (n=39) had previously been triggered with CLTSH, and certified as ODF according to *Woreda* Health Office records. Another 14% (7 of 50 clusters) had been triggered with CLTSH, but not yet certified ODF; the remaining 8% (4 of 50 clusters) had not yet been triggered with CLTSH (Figure 3, Appendix E, Table E.1). Results from our equivalence analyses indicate balance in the number of previously triggered and ODF certified *kebeles*, between treatment arms, with 20 triggered and ODF communities in the intervention arm (80%, 20 of 25 study clusters) and 19 in the control arm (76%, 19 of 25 study clusters).

Table 5. Baseline levels of respondent-report	ed key demographic variables
---	------------------------------

	Overall			rvention	Co	ontrol
Demographic information	N	%	Ν	%	Ν	%
Respondent sex was female (%)	1589	90.7	793	91.3	796	90.1
Respondent was mother of index child (%)	1589	84.6	793	85.4	796	83.9
Respondent is of Amharic ethnicity (%)	1589	100	793	100	796	100
Primary caregiver/mother has at least secondary education (%)	1589	12.8	793	11.9	796	13.8
The primary caregiver/mother is married (%)	1587	89.3	792	91.2	795	87.4
Head of household has at least some secondary education (%)	1579	16.7	791	15.2	788	18.2
Head of household has fetched water in past 7 days (%)	1589	27.3	793	28.6	796	26.0
	Ν	mean (SE)	Ν	mean (SE)	Ν	mean (SE)
Respondents age (mean)	1589	33.5 (0.38)	793	33.7 (0.52)	796	33.3 (0.55)
Head of household age (mean)	1589	41.3 (0.46)	793	41.6 (0.54)	796	41.1 (0.73)
Number of household members (mean)	1589	5.3 (0.08)	793	5.3 (0.10)	796	5.3 (0.12)

Figure 3. Study kebele CLTSH status



7.2.3. Baseline WASH information

Table 6 shows the prevalence of our primary WASH indicators at baseline, both overall and by intervention assignment; additional secondary indicators are shown in Appendix J, Table J.1. At baseline, 66% of households had at least one latrine (Table 6). Among those households with a sanitation facility, 40% met criteria of an improved facility based on the JMP definition (World Health Organization and UNICEF, 2013). Only 31% of households had a fully constructed latrine, and only 13% of households had added or improved anything on their latrine since the original construction. Thirteen percent of latrines had a drop hole cover, and two-thirds of those had the cover situated over the hole. There were a number of commonly observed inadequacies or problems with these latrines, including 80% of latrine with flies present, 52% had feces on the floor or slab or some other place in the latrine (besides the pit), and 20% had a pit that was full or close to full (Appendix J, Table J.1). Only 29% of the latrines had anal cleansing materials in or near the latrine, and 4% had water available inside or near the latrine for flushing or self-cleansing.

Our assessment of latrine utilization at baseline showed that 38% of respondents' primary place of defecation during the last two days was in the open; only 46% of respondents had defecated in any latrine during the last two days (Table 6). There was observed evidence of open defecation in or near 57% of the household compounds. However, there was also evidence of use among many of the primary household latrines, for example, 92% had a well-worn path to the latrine, 81% had fresh feces in the pit or pan, and 81% had an odor from stool or urine (Appendix J, Table J.1). Latrines were often shared with other households, with 1.5 households using each latrine. Less than half (40%) of households safely disposed of child feces (Table 6). Animal feces were present in the compound in 85% of the households (Appendix J, Table J.1), and these feces were left out in the open 56% of the time (Table 6).

All of the sanitation variables (e.g., coverage, latrine characteristics, latrine utilization, latrine sharing, environmental sanitation) appeared to be balanced between the intervention and control communities at baseline (Table 6, Appendix J, Table J.1).

Among children aged 1-9 years, observations of facial cleanliness showed ocular discharge among 40% of children, wet nasal discharge among 47%, dry nasal discharge among 65%, and dust, dirt or other debris on 70% of children (Table 6). When restricting to just the 1,385 index children who were present, results for observed facial cleanliness were very similar to those among all children aged 1-9 years in the study households (Appendix J, Table J.1). This suggests that our index children serve as acceptable sentinels of behavioral outcomes for children of similar ages within the larger household. The mean number of times that a fly landed on index children's faces over a one minute observation period was 4 times (Table 6).

Hand or face wash stations were reported in 79% of the households (Table 6). Of the 1,589 respondents, most reported washing their own hands yesterday (97%), but generally without soapy water (36%). Respondents washed their hand with soapy water after defecation 37% of the time, and they washed before food preparation 40% of the time. It was reported that index children's hands were washed yesterday 98% of the time, but only 39% of the time with soap (Appendix J, Table J.1). Only 26% of children washed their hands with soapy water after the last time they defecated. The index children's finger nails were observed to be clean approximately 91% of the time, their finger pads were clean 86% of the time, and their palms were clean 86% of the time. Nearly all of these hygiene variables were balanced comparing the intervention and control arms. However, there was a slightly higher (but not significant) prevalence of handwashing with soap in the control arms.

Table 6. Baseline levels of the WASH indicators of primary interest

	(Overall	Int	ervention		Control	
Latrine coverage	Ν	%	Ν	%	Ν	%	
Households with at least one latrine (%)	1589	65.5	793	64.1	796	66.8	
Households with improved latrine (%) ^a	1553	39.8	775	39.9	778	39.7	
Households with fully constructed latrine (%)	1583	30.7	792	29.6	791	31.9	
Sanitation facility operation and maintenance	Ν	%	Ν	%	Ν	%	
HH has added or improved anything on this latrine since its original construction (%)	1028	12.9	504	11.9	524	13.9	
Latrine characteristics	Ν	%	Ν	%	Ν	%	
Presence of drop hole cover in the latrine (%)	1033	13.4	505	12.5	528	14.2	
Among those with a drop hole, a cover was situated over drop hole (%)	138	66.7	63	65.1	75	68.0	
Defecation	Ν	%	Ν	%	Ν	%	
Respondent's primary place of defecation was OD during last 2 days (%)	1589	37.5	793	39.5	796	35.6	
Respondent defecated in any latrine during last 2 days (%)	1589	45.6	793	46.0	796	45.1	
Safely disposed of child feces (%)	961	40.2	463	38.9	498	41.4	
Latrine sharing	Ν	mean (SE)	Ν	mean (SE)	Ν	mean (SE)	
Given HH has a latrine, number of people who used this latrine from ANOTHER HH during last 7 days, not including your HH members	1037	0.94 (0.12)	506	1.08 (0.18)	530	0.79 (0.52, 1.06)	
Animal husbandry / other HH sanitation	Ν	%	Ν	%	Ν	%	
Animal feces/waste not left out in open in compound (%)	1589	44.1	793	42.0	796	46.2	
Facial cleanliness among children ages 1-9	Ν	%	Ν	%	Ν	%	
Ocular discharge is present (%)	1944	40.3	932	42.2	1012	38.6	
Wet nasal discharge is present (%)	1944	47.3	932	47.6	1012	46.9	
Dry nasal discharge is present (%)	1944	65.4	932	64.7	1012	66.1	
Dirt/dust/other debris is present (%)	1944	69.9	932	68.7	1012	71.0	
Fly observations	Ν	mean (SE)	Ν	mean (SE)	Ν	mean (SE)	
Number of times a fly land on the index child's face during a 1 minute observation	1382	4.2 (0.23)	669	4.1 (0.34)	713	4.3 (0.32)	
Number on the color scale that most closely matches the darkest part of the eye swab	1385	6.1 (0.09)	668	6.1 (0.09)	717	6.1 (0.15)	
Washing station coverage	Ν	%	Ν	%	Ν	%	
HH hand or facewashing station(s) (%)	1589	78.9	793	77.1	796	78.8	
Handwashing practices	Ν	%	Ν	%	Ν	%	
The last time the respondent washed he/she used soap/ash/soapy water (%)	1588	36.4	793	35.1	795	37.7	
The last time the respondent defecated, he/she cleaned hands with water and soap, substitute (%)	1585	37.2	791	36.3	794	38.0	
The last time the respondent prepared food, he/she cleaned hands with water and soap, substitute before beginning food preparations (%)	1586	39.7	791	41.0	795	38.5	

^a Improved based on JMP definition; (World Health Organization and UNICEF, 2013) see Figure 4A for all latrine type categories

Table 7. Baseline levels of respondent-reported mental health outcomes

		Over	all		Inter	vention		Control		
Anxiety and depression	N	%	Mean (SD)	Ν	%	Mean (SD)	Ν	%	Mean (SD)	
Anxiety ^a	1584	29.7	1.56 (.62)	790	29.6	1.56 (.60)	794	29.7	1.56 (.63)	
Depression ^a	1588	20.8	1.46 (.52)	793	21.3	1.45 (.51)	795	20.4	1.46 (.52)	
Emotional distress ^a	1583	17.3	1.38 (.48)	790	17.2	1.38 (.47)	793	17.4	1.38 (.48)	
Anxiety and depression										
Poor well-being ^b	1586	33.2	16.0 (7.0)	792	31.1	15.6 (7.1)	794	35.3	16.3 (6.8)	

^a We asked respondents to indicate how much the symptoms bothered them in the previous week with four potential response options (not at all (1) to extremely (4)). The first ten symptoms assess anxiety (i.e. 'suddenly scared for no reason', 'nervousness or shakiness inside'), the next 13 assess depression (i.e. 'feeling low in energy', 'feeling hopeless about the future'), and the 23 collectively assess non-specific emotional distress. For each outcome, the score is the sum of the responses divided by the number of items. Each of these scores was dichotomized, with scores greater than 1.75 indicating a positive status for any of the three outcomes. ^b We asked respondents about well-being, and responses ranged from '(0) At no time' to (5) All of the time'. Scores were summed, and range from 0- 25; the higher the score, the better the well-being. Each of these scores was dichotomized with scores below 13 indicating poor well-being.

Table 8. Baseline levels of respondent-reported diarrheal outcomes

	Ov	Overall		Intervention		ntrol
Diarrhea	Ν	%	N	%	Ν	%
Among all HH members 0-9 years						
In the LAST 2 days, HH member had three or more loose stools per day (%)	2789	6.1	1368	5.3	1421	6.8
During the last 7 days, including today, HH member had three or more loose stools per day (%)	2792	8.3	1369	7.5	1423	9.0
During the last 7 days, including today, HH member had blood in the stool (%)	2755	2.1	1356	1.5	1399	2.7
Among index children						
In the LAST 2 days, child had three or more loose stools per day (%)	1577	6.3	782	5.6	795	6.9
During the last 7 days, including today, child had three or more loose stools per day (%)	1575	9.1	778	8.1	797	10.0
During the last 7 days, including today, child had blood in the stool (%)	1558	2.1	773	1.2	785	3.1

During the week (i.e., seven days) preceding the survey, respondents reported washing the index child's clothes 87% of the time, their own clothes 79% of the time, towels 65% of the time, bedsheets 30% of the time, a baby carrier 28% of the time, and children's toys 2% of the time. In 76% of households, the index child was bathed at least once during the week preceding the survey. Just over half (57%) of the respondents reported bathing their own bodies at least once during the week preceding the survey, and 57% reported using surface water for bathing during that time. The prevalences of respondent reported washing and bathing practices were similar comparing the intervention and control arms at baseline.

7.2.4. Baseline health information

The baseline prevalence of anxiety and depression among respondents was 29.7% and 20.8%, respectively (Table 7). Emotional distress was indicated among 17.3% of respondents. Poor well-being, per WHO-5 scale, was indicated among 33.2% of respondents. These distributions of scores were generally balanced, when comparing the intervention and control arms.

At baseline, the prevalence of self-reported diarrhea during the two days preceding the survey was 6%, when considering all children under the age of 10 years (Table 8). The prevalence of diarrhea during the week (i.e., seven days) preceding the survey was 9% among index children. Diarrhea prevalence was generally similar comparing the intervention to the control, except there was a slightly lower prevalence of index children who had blood in their stool during the week preceding the survey compared to control households.

7.3. Summary of year one follow-up results

At midline (i.e., follow-up 1), we collected a total of 1,589 household surveys; nineteen households did not have a member between 1-10 years of age, 53 had no eligible respondent available, 3 did not consent to take the survey, and 11 had moved away from the *kebele*. Six surveys were initiated but not finished to completion, and four other surveys were completed but using the pilot version of the survey. The midline results reflect complete data from the remaining 1,496 households (94%).

7.3.1. Impacts on sanitation coverage

At midline, we observed that 68% of households had at least one latrine (Table 9), which was similar to baseline (66%; Table 6). Latrine coverage was also similar, when comparing intervention and control arms at midline (PR=1.01; 95% CI: 0.85, 1.22). Only seven households had more than one latrine; we focus all our analyses on the primary household latrine. Among those households with a sanitation facility, 30% met criteria of an improved facility based on the JMP definition (World Health Organization and UNICEF, 2013), which was 10 percentage points lower than at baseline. There was no difference in the prevalence of improved latrines, when comparing intervention and control arms (PR=1.09; 95% CI: 0.77, 1.5). Similarly, there was no difference in the prevalence of households with fully constructed latrines (PR=1.14; 95% CI: 0.86, 1.50). As for secondary outcomes, there were no differences in the prevalence of latrines with smooth, cleanable surfaces, when comparing intervention to control, with only 13% of households overall having had a latrine platform with a smooth and clean surface.

The distribution of different sanitation technologies, both overall and by intervention group, is shown in Figure 4A. This graphic indicates that there were very few latrines whose construction was within the upper, more improved rungs of the sanitation ladder. The distribution of sanitation technologies is similar between intervention arms.

Table 9. Midline levels of the WASH indicators of p	orimary	interest
---	---------	----------

	Overall Intervention		(Control				
Latrine coverage observations	Ν	%	Ν	%	Ν	%	PR (95% CI) ^a	PD (95% CI) ^b
Households with at least one latrine (%)	1496	67.7	751	67.8	745	67.5	1.01 (0.85, 1.22)	0.0098 (-0.11, 0.13)
Households with improved latrine (%) ^c	1485	29.8	743	31.0	742	28.4	1.09 (0.77, 1.55)	0.0254 (-0.08, 0.13)
Households with fully constructed latrine (%)	1490	31.9	749	33.8	741	30.0	1.14 (0.86, 1.50)	0.0405 (-0.05, 0.13)
Sanitation facility operation and maintenance	Ν	%	Ν	%	Ν	%	PR (95% CI) ^a	PD (95% CI) ^b
HH has reported adding or improving anything on this latrine since its original construction (%)	1006	24.3	506	23.3	500	25.2	0.91 (0.72, 1.15)	-0.0223 (-0.08, 0.04)
Observed latrine characteristics	Ν	%	Ν	%	Ν	%	PR (95% CI) ^a	PD (95% CI) ^b
Presence of drop hole cover in the latrine (%)	1011	19.5	508	23.8	503	15.1	1.56 (1.07, 2.28)	0.0851 (0.00, 0.16)
Among those with a drop hole, a cover was situated over drop hole (%)	197	70.6	121	68.6	76	75.0	0.90 (0.76, 1.08)	-0.0727 (-0.20, 0.50)
Respondent reported defecation	Ν	%	Ν	%	Ν	%	PR (95% CI) ^a	PD (95% CI) ^b
Respondent's primary place of defecation was OD during last 2 days (%)	1496	34.6	751	34.5	745	34.6	0.98 (0.69, 1.41)	-0.0065 (-0.13, 0.12)
Respondent defecated in any latrine during last 2 days (%)	1491	44.9	749	47.3	742	42.5	1.12 (0.89, 1.41)	0.0521 (-0.05, 0.16)
Safely disposed of child feces (%)	873	46.4	438	43.8	435	45.0	0.97 (0.73, 1.27)	-0.0160 (-0.14, 0.11)
Respondent reported latrine sharing	Ν	mean (SE)	Ν	mean (SE)	Ν	mean (SE)	-	difference (95% CI) ^d
Given HH has a latrine, number of people who used this latrine from ANOTHER HH during last 7 days, not including your HH members	1001	1.10 (0.14)	505	0.94 (0.17)	496	1.27 (0.22)	-	-0.41 (-0.88, 0.06)
Animal husbandry/other HH sanitation	Ν	%	Ν	%	Ν	%	PR (95% CI) ^a	PD (95% CI) ^b
Animal feces/waste not observed out in open in compound (%)	1496	51.1	751	52.9	745	49.3	1.06 (0.80, 1.40)	0.03 (-0.12, 0.17)
Observed facial cleanliness among children ages 1-9	Ν	%	Ν	%	Ν	%	PR (95% CI) ^a	PD (95% CI) ^b
Ocular discharge is present (%)	1441	37.7	682	37.4	759	37.9	0.97 (0.81, 1.16)	-0.0122 (-0.08, 0.06)
Wet nasal discharge is present (%)	1441	45.8	682	45.0	759	46.5	0.96 (0.82, 1.12)	-0.0183 (-0.09, 0.05)
Dry nasal discharge is present (%)	1441	55.3	682	55.1	759	55.5	0.99 (0.89, 1.10)	-0.0077 (-0.07, 0.05)
Dirt/dust/other debris is present (%)	1441	53.0	682	52.2	759	53.6	0.97 (0.83, 1.13)	-0.0151 (-0.10, 0.06)
Fly observations	Ν	mean (SE)	Ν	mean (SE)	Ν	mean (SE)	-	difference (95% CI) ^d
Number of times flies landed on index child's face during minute observation	921	3.1 (0.19)	444	3.3 (0.32)	477	2.9 (0.22)	-	0.41 (-0.27, 1.08)
Washing station coverage	Ν	%	Ν	%	Ν	%	PR (95% CI)ª	PD (95% CI) ^b
Reported HH hand or facewashing station(s) (%)	1494	95.0	750	97.3	744	92.6	1.02 (0.99, 1.04)	0.0154 (-0.01, 0.04)
Respondent reported handwashing practices	Ν	%	N	%	N	%	PR (95% CI) ^a	PD (95% CI) ^b
The last time the respondent washed he/she used soap/ash/soapy water (%)	1492	44.0	750	48.1	742	39.8	1.22 (1.00, 1.47)	0.0865 (0.00, 0.17)
The last time the respondent defecated, he/she cleaned hands with water and soap, substitute (%)	1486	43.3	747	49.1	739	37.4	1.30 (1.03, 1.66)	0.1153 (0.01, 0.22)
The last time the respondent prepared food, he/she cleaned hands with water and soap, substitute before beginning food preparations (%)	1407	45.3	712	50.8	695	39.6	1.29 (1.02, 1.62)	0.1139 (0.01, 0.22)
Respondent reported water insecurity	Ν	%	Ν	%	Ν	%	PR (95% CI) ^a	PD (95% CI) ^b
Water insecure for any of the four reported insecurity measures (%)	1420	36.4	716	35.2	704	37.6	0.89 (0.64, 1.25)	0.0424 (-0.17, 0.08)

^a We used log-linear binomial models to estimate prevalence ratios comparing the outcomes between the intervention and control arms. Models accounted the stratified design by including woreda indicator variables, (Kahan and Morris, 2012) and accounted for clustering within kebeles by using generalized estimating equations with robust standard errors. ^b We used similar linear regression models to estimate difference comparing the outcomes between the intervention and control arms. ^c Improved based on JMP definition; (World Health Organization and UNICEF, 2013) see Figure 4A for all latrine type categories. ^d We used similar linear regression models to estimate difference comparing the outcomes between the intervention and control arms.

Table 10. Midline respondent-reported sanitation insecurity

	Overall		In	tervention		Control	difference (95% CI) ^a
Sanitation insecurity ^b	N	Mean (SD)	Ν	Mean (SD)	Ν	Mean (SD)	
Factor 1-Potential harms	1494	0.54 (0.027)	750	0.54 (0.042)	744	0.54 (0.035)	-0.0038 (-0.10, 0.09)
Factor 2-Social expectations resultant repercussions	1491	0.27 (0.017)	750	0.27 (0.028)	741	0.27 (0.021)	-0.0079 (-0.07, 0.05)
Factor 3-Physical exertion or strain	1490	0.43 (0.035)	748	0.43 (0.053)	742	0.44 (0.045)	-0.0108 (-0.15, 0.13)
Factor 4-Night concerns	1496	0.34 (0.025)	751	0.32 (0.038)	745	0.35 (0.033)	-0.0306 (-0.12, 0.06)
Factor 5-Social support	1495	0.16 (0.016)	750	0.14 (0.022)	745	0.17 (0.021)	-0.0299 (-0.09, 0.03)
Factor 6-Physical agility	1492	0.15 (0.011)	747	0.14 (0.015)	745	0.15 (0.016)	-0.0097 (-0.05, 0.03)
Factor 7-Defecation place	1494	0.31 (0.024)	750	0.30 (0.034)	744	0.32 (0.032)	-0.0252 (-0.11, 0.06)

^a We used linear regression models to estimate the change in the outcomes comparing the intervention and control arms. Models accounted the stratified design by including woreda indicator variables, (Kahan and Morris, 2012) and accounted for clustering within kebeles by using generalized estimating equations with robust standard errors. ^b We asked respondents to indicate how often they felt some form of sanitation insecurity (never, sometimes, often, always). These items were then summed with all other items in that factor and divided by the numbers of items to create a score. The factors were predesignated, and based on a validation that was done in another study (Caruso et al., 2017). A higher score represents higher sanitation insecurity.

Table 11. Midline respondent-reported mental health outcomes

	Overall				Inter	vention		Cor	ntrol	difference (95% CI)
Anxiety and depression	Ν	%	Mean (SD)	Ν	%	Mean (SD)	Ν	%	Mean (SD)	
Anxiety ^b	1467	22.2	1.45 (.57)	729	21.2	1.43 (.55)	734	23.3	1.47 (.59)	-0.37 (-0.12, 0.045)
Depression ^b	1435	15.3	1.35 (.44)	714	14.3	1.34 (.42)	717	16.3	1.37 (.45)	-0.030 (-0.096, 0.036)
Emotional distress ^b	1429	13.8	1.28 (.42)	710	13.5	1.26 (.41)	715	14.1	1.29 (.43)	-0.028 (-0.093, 0.037)
WHO-5 well-being										
Poor well-being ^c	1488	22.9	17.7 (6.7)	749	20.8	17.9 (6.6)	739	25.0	17.5 (6.7)	0.47 (-0.62, 1.56)

^a We used linear regression models to estimate the change in the outcomes comparing the intervention and control arms. Models accounted the stratified design by including woreda indicator variables, (Kahan and Morris, 2012) and accounted for clustering within kebeles by using generalized estimating equations with robust standard errors. ^b We asked respondents to indicate how much the symptoms bothered them in the previous week with four potential response options (not at all (1) to extremely (4)). The first ten symptoms assess anxiety (i.e. 'suddenly scared for no reason', 'nervousness or shakiness inside'), the next 13 assess depression (i.e. 'feeling low in energy', 'feeling hopeless about the future'), and the 23 collectively assess non-specific emotional distress. For each outcome, the score is the sum of the responses divided by the number of items. Each of these scores was dichotomized, with scores greater than 1.75 indicating a positive status for any of the three outcomes. ^c We asked respondents about well-being, and responses ranged from '(0) At no time' to (5) All of the time'. Scores were summed, and range from 0- 25; the higher the score, the better the well-being. Each of these scores was dichotomized with scores below 13 indicating poor well-being.

Table 12. Midline respondent-reported diarrheal outcomes

	Overall		Intervention		Control		PR (95% CI) ^a	PD (95% CI) ^b
Diarrhea	Ν	%	Ν	%	Ν	%		
Among all HH members 0-10 years								
In the LAST 2 days, HH member had three or more loose stools per day (%)	2884	4.9	1432	5.7	1452	4.1	1.38 (0.97, 1.96)	0.0158 (0.00, 0.03)
During the last 7 days, including today, HH member had three or more loose stools per day (%)	2891	6.4	1436	7.3	1455	5.4	1.35 (0.94, 1.92)	0.0189 (0.00, 0.04)
During the last 7 days, including today, HH member had blood in the stool (%)	2878	1.7	1426	1.9	1452	1.6	1.19 (0.60, 2.37)	0.0030 (-0.01, 0.02)
Among index children								
In the LAST 2 days, child had three or more loose stools per day (%)	1454	4.1	728	5.5	726	2.6	2.09 (1.18, 3.70)	0.0301 (0.00, 0.06)
During the last 7 days, including today, child had three or more loose stools per day (%)	1461	5.5	732	7.5	729	3.6	2.10 (1.34, 3.30)	0.0413 (0.01, 0.07)
During the last 7 days, including today, child had blood in the stool (%)	1456	1.4	727	1.8	729	1.0	1.82 (0.66, 4.99)	0.0082 (-0.01, 0.02)

^a We used log-linear binomial models to estimate prevalence ratios comparing the outcomes between the intervention and control arms. Models accounted the stratified design by including woreda indicator variables,(Kahan and Morris, 2012) and accounted for clustering within kebeles by using generalized estimating equations with robust standard errors. ^b We used similar linear regression models to estimate difference comparing the outcomes between the intervention and control arms.



A. Latrine type distribution







□ With water only

□ Wiping eyes □ With saliva

preceding the survey

B. Disposal of child feces for children under 4 years of age (n=873)



□ The faeces were put/rinsed into the latrine

□ The faeces were buried

□ The faeces were thrown into the garbage

The faeces were put/rinsed into a surface water source, drain, ditch or the bush/yard

D. Method used to clean the index child's face during the day preceding the survey



With water and soap

7.3.2. Impacts on latrine characteristics

We present the assessment of latrine characteristics only among the households that both had a latrine, and allowed the enumerator to observe their latrine (N=1.011; Table 9). Among our key latrine characteristic indicators, we observed that the prevalence of latrines with a drop hole cover was higher in the intervention arm than the control arm (PR=1.56; 95% CI: 1.07, 2.28). However, when a drop hole cover was present, there was no significant difference in the actual coverage of the drop hole (i.e., appropriate utilization of the drop hole) between intervention and control arms (PR=0.90; 95% CI: 0.76, 1.08). In other words, while the presence of a drop hole cover was higher in intervention communities, it was not necessarily being used to cover the drop hole. Among secondary latrine characteristic indicators, latrines in the intervention arm were in better condition than those in the control arm. Latrines from intervention communities had: a lower prevalence of flies in them (PR=0.87; 95% CI: 0.76, 1.00), less feces on the floor/walls (PR=0.08; 95% CI: 0.68, 0.93), and less odor from stool or urine (PR=0.88; 95% CI: 0.79, 0.98). Intervention households were more likely to have water available for handwashing near or inside the latrine (PR=2.57; 95% CI: 1.03, 6.41), cleansing agents for handwashing near or inside the latrines (PR=5.75; 95% CI: 1.77, 18.70), and water available inside or near the latrines for flushing or self-cleansing (PR=2.67; 95% CI: 1.12, 6.34). Although there were improvements in many latrine characteristics in the intervention compared to the control arm, the conditions of latrines in the intervention arm were often still inadequate. For example, 72% of intervention latrines still had flies present, only 24% had a drop hole cover, and 51% had feces on the floor or slab or some other place in the latrine (besides the pit). All of the sanitation operation and maintenance indicators were similar, when comparing intervention and control arms.

7.3.3. Impacts on washing station coverage

Presence of hand and facewashing stations were reported and observe in 95% of households, although presence of water, and presence of soap were observed in only 20% and 1.9% of stations, respectively (Table 9). The prevalence of hand and facewashing stations was similar between the two study arms (PR=1.02; 95% CI: 0.99, 1.04), and the prevalence of stations with water was also similar when comparing the two arms (PR=0.89; 95% CI: 0.66, 1.20). The prevalence of stations with soap was higher in the intervention arm, although only 2.9% of households in this arm had a hand or facewashing station with soap present.

7.3.4. Behavioral outcome indicators

We assessed a number of different behavioral outcomes, including latrine utilization, facilities operation and maintenance, animal husbandry, personal hygiene, and safe water practices. We discuss each below in *sections* 7.3.5 - 7.3.7.

7.3.5. Impacts on latrine utilization (defecation and urination practices)

All measures of latrine utilization were similar when comparing the intervention and control arms. This includes indicators of urination, defecation (both for respondents and other household members), disposal of child feces (Figure 4B), and sanitation sharing. Overall, 35% of respondents' primary place of defecation during the last two days was in the open, and only 45% of respondents had defecated in any latrine during the last two days (Table 9).

7.3.6. Impacts on other environmental sanitation

Across both arms, the majority of both respondents and heads of household had animal herding responsibilities (~91% for each), and animal feces were present in the compound in 83% of the households (Appendix J, Table J.7). Absence of animal feces from the household compound (i.e., no animal feces left out in the open) was our primary environmental sanitation outcome. About half of households did not leave animal feces/waste in the open (Table 9). This was similar between the intervention and control arms (PR: 1.06; 95% CI: 0.80, 1.40). All other environmental sanitation measures were quite similar, when comparing intervention and control arms.

7.3.7. Impacts on personal hygiene

The primary personal hygiene outcomes of interest were respondent-reported handwashing, observed facial cleanliness among children, and the number of flies observed on children's faces during a one-minute observation period. Respondent-reported handwashing was more prevalent in the intervention arm compared to the control arm. This finding was consistently statistically significant across multiple measures of handwashing (Table 9). When reporting on their last handwashing experience, respondents reported washing their hands with soap/ash/soapy water 44% of the time (Figure 4C); they reported washing their hands after defecation 43% of the time; and they reported washing their hands before food preparation 45% of the time. However, handwashing practices for other members of the household (beside the respondent) were similar comparing the intervention and control arms.

When reporting on handwashing practices carried out on/by the index child during the day prior to survey administration, 98% of respondents reported that the index children's hands had been washed. However, soap was reportedly used only 45% of the time (Figure 4D). The prevalence of these reported practices were similar in both study arms. A higher prevalence of children had reportedly washed their hands after defecation in the intervention arm, compared to the control arm (PR: 1.41; 95% CI: 1.09, 1.83), although only 37% of children reportedly had their hands washed overall.

There were 1,441 children aged 1-9 years whose facial cleanliness was observed. Overall, observations of facial cleanliness showed ocular discharge among 38% of children, wet nasal discharge among 46%, dry nasal discharge among 55%, and dust, dirt or other debris on 53% of children. There were no differences between the intervention and control arm for any of these facial cleanliness measures. When restricting these analyses to just the 921 index children who were present for midline observation, facial cleanliness results were similar to those provided by all children aged 1-9 years. Overall, 94% of the index children reportedly had their faces cleaned yesterday, which was similar for both arms. The overall mean number of times that a fly landed on index children's faces over a one minute observation period was three, and was similar in both the intervention and control groups (difference=0.41; 95% CI: - 0.27, 1.08).

To date, the intervention has had little impact on the washing of fomites and on bathing practices. This finding is not surprising given these practices are not the focus of the *Andilaye* intervention. The only statistically significant finding was that respondents used surface water for bathing less in the intervention arm than in the control arm (PR=0.54; 95% CI: 0.31, 0.96) (see Appendix J, Table J.6).

7.3.8. Impacts on mental well-being and water and sanitation insecurity

The mean scores of both water insecurity (Table 9) and sanitation insecurity (Table 10) variables were all similar, when comparing intervention to the control arms, although each of the estimates trended in the direction of better security in the intervention group. The prevalence of anxiety, depression, and emotional distress among respondents was lower at midline than baseline among the overall population (Table 7, Table 11). However, the mean scores of these mental well-being indicators were all similar, when comparing intervention to the control arms, although each of the estimates trended in the preventive direction (Table 11). Scores indicating lower prevalence in anxiety, depression and emotional distress in the overall population from baseline suggests an influence of non-controllable factors related to year on year changes, such as drought, food security, etc. Box and whisker plots indicate similar distributions between the two arms (Appendix J, Figure J.1). Well-being scores and distributions were also similar, when comparing intervention and control arms (Table 11; Appendix J, Figure J.1).

7.3.9. Impacts on respondent-reported diarrhea

To date, the intervention has not reduced diarrhea in the intervention arm compared to the control arm. Diarrhea prevalence was actually higher in the intervention arm than the control arm (PR=2.10; 95% CI: 1.34, 3.30) for our primary diarrhea indicator, which measured whether index children had an episode of three or more loose stools per day over the last seven days (Table 12). Diarrhea prevalence was also higher among index children when assessing episodes over the last two days (PR=2.09; 95% CI: 1.18, 3.70). Among index children, the prevalence of diarrhea with blood in the stool was similar, when comparing intervention and control arms. Diarrhea prevalence using various diarrhea measures was similar comparing the intervention to control among all HH members age 0-10 years, including for diarrhea in the last two days, diarrhea in the last seven days, and blood in the stool over the last seven days.

7.3.10. Heterogeneity across sub-groups

For each of our primary outcomes of interest, we assessed if there was interaction across various sub-groups, including exposure to previous triggering and sex. We also assessed if water insecurity modified the effectiveness on hygiene behaviors. For all of these analyses, we included interaction terms, and if effect modification was present (i.e., the interaction term had a p-value of <0.05), we stratified by the sub-group variable. We did not detect interaction of the intervention by previous triggering for any of the primary outcome variables of interest. We assessed interaction by the child's sex, and did not detect interaction of the intervention by sex for any of those primary outcome variables related to children. We also did not detect interaction between the intervention and water insecurity on any of the primary the handwashing or facewashing variables,

8. Discussion

8.1. Key findings

Our three-year impact evaluation leverages behavioral theory and formative research to inform the design of a novel demand-side sanitation and hygiene intervention (i.e., *Andilaye*), and evaluate its impact on sustained behavior change and mental well-being in Amhara, Ethiopia. This study addresses several remaining knowledge gaps related to:

- 1. The effectiveness of theoretically-informed and evidence-based demand-side WASH intervention approaches;
- 2. The types of strategies that may be successful for improving the integration of disease control programming, health, and development efforts; and
- **3.** The sustainability of behavioral outcomes when community-level behavioral control (i.e., collective efficacy) and individual- and household-level behavioral maintenance techniques are incorporated and emphasized within the design of community-based interventions.

In Ethiopia, recent scaled intervention approaches have focused on leveraging negative motives to drive open defecation cessation initiatives. Personal hygiene initiatives incorporated into community-level WASH programming have largely focused on handwashing with soap or soap alternative after defecation. Little to no emphasis has been placed on other personal hygiene practices (e.g., facewashing, bathing) despite Amhara's high prevalence of NTDs that are likely propagated due to poor personal hygiene practices (e.g., trachoma, STH, schistosomiasis). In addition, we add validated mental well-being measure to this WASH impact evaluation to determine if changes to sanitation access and security generated by this intervention leads to improved mental well-being. This measure has not been used in the WASH sector previously to assess intervention impact.

Our formative research results provide evidence that communities negatively perceived of prior demand-side sanitation and hygiene intervention activities, particularly those that

focused on shame and disgust. Leveraging feedback received from community members and key stakeholders, we designed a theoretically-informed and evidence-based demand-side sanitation and hygiene intervention that focuses on positive, community-oriented motivators of behavioral change. The *Andilaye* intervention promotes achievable incremental improvements, incorporates strategies that facilitate behavioral maintenance, and addresses the over-extension of HEWs and the over-saturation of HEP's Health Extension Package messaging. The overarching intervention approach was designed to be incorporated into prevailing programs (e.g., HEP) to demonstrate potential for scale-up.

As indicated in our summary of intervention activities, to date, nearly all of *Andilaye*'s behavior change catalyzing activities have been implemented. Early feedback from WDALs and HEWs suggest that *Andilaye* tools and skills-based trainings have been well received by the community-level change agents. These actors have indicated that our *Andilaye* intervention materials and trainings help organize and facilitate household visits. We question whether social desirability bias influenced this feedback, as we have observed gaps in effective utilization of study materials during our process evaluation activities. These shortcomings can be improved upon, however, through the provision of supportive supervision to WDALs by HEWs and CHC HEWs supervisors.

We have already observed early indications that additional resource considerations need to be addressed when determining whether this intervention approach can and should go to scale. For example, despite several orientation meetings, action planning workshops, provision of supportive supervision and on-the-iob-training tools. Woreda Health Office officials, CHC HEWs supervisors, and HEWs themselves have not been making supportive supervisory visits to WDALs in accordance with their action plans. The system is overburdened, with the same individuals being tasked with supporting numerous and diverse programming initiatives, some of which continue to operate in a relatively siloed manner. Even a limited number of supportive supervisory visits seem to be prohibitive. It is also worth noting that the intervention approaches promoted by the Andilaye intervention are different than the standard behavior change communication activities that focus on dissemination of information and messages. Rather than focusing on dissemination of information and messages focused on health considerations, our Andilaye intervention was specifically designed to incorporate intervention techniques that address other behavioral factors such as action knowledge, personal/household barrier identification and planning, behavioral control perceptions, attitudes, and social norms amongst others. Evidence suggests that it is important to move away from information-based interventions to address an array of behavioral factors and determinants that operate at various levels of influence (Briscoe and Aboud, 2012; Hulland et al., 2015; Marteau et al., 2012; Rabbi and Dey, 2013; Wood and Neal, 2016). However, further capacity building of federal, regional, and local level government officials as well as community-level change agents may be necessary for the successful implementation of approaches that move beyond dissemination of information and messages, particularly those solely focused on health.

The data presented in this report were collected approximately one to two months after key community-level intervention activities were implemented at the community level (Appendix I). While these midline results do signal some promising trends, it is too early to make definitive statements about the impact of the *Andilaye* intervention on targeted practices along the intervention's three behavioral themes – sanitation, personal hygiene, and household environmental sanitation. That said, we do present a summary of these midline findings below. Upon the conclusion of our full trial in 2019, we will execute our full pre-analysis plan, at which time we will have complete datasets that will allow for definitive assessments of intervention fidelity, adherence, and impact.

To a large extent, observed and reported sanitation coverage and utilization remain similar between intervention and control communities at midline. These findings are not surprising

given the short timeframe between after the implementation of key community and household intervention activities and data collection. Access to any household latrine is still deficient, with only 68% of all study households having access. This indicates that household latrine coverage is similar to baseline levels. Amongst households with any latrine, only 30% met criteria of an improved facility, per the JMP definition – this was 10% higher than baseline. There was also no significant difference in the proportion of household latrines with a smooth and cleanable platform surface between baseline and midline. There were, however, significant improvements in the prevalence of latrine drop hole covers in household latrines in intervention communities. Unfortunately, these drop hole covers were often not being used appropriately to cover the actual drop hole. Therefore, there was no significant difference in the prevalence of latrine drop holes actually being covered in intervention versus control communities. This is not entirely surprising, as the presence of a drop hole cover indicates the intention to keep the drop hole sealed, while the actual presence of the cover over the drop hole indicates the actual adoption of the improved behavior. This phenomenon is promising in that it is suggestive of the stages of change (Trans-theoretical Model) taking hold (Prochaska and DiClemente, 1982). Household latrines in intervention communities were also in better condition, overall, than those in control communities, as indicated by the lower prevalence of latrines with flies, feces on the floor/walls, and odor in them. Intervention latrines also had a higher prevalence of water and soap present. That said, the conditions of the household latrine in intervention communities are still inadequate. All measures of latrine utilization were similar in intervention and control communities. Only 45% of midline respondents had defecated in any latrine during the last two days.

Personal hygiene facilities coverage and practices remain similar between intervention and control communities. While presence of washings station was observed to be high (95%), the presence of the materials need to perform washing at these stations was starkly low, with only 20% of washing facilities equipped with water, and less than 2% equipped with soap or soap alternative. Presence of soap at washing stations was higher in intervention communities, but remains inadequate at just under 3%. Respondent-reported handwashing outcomes was more prevalent in intervention communities compared to control, and was consistently statistically significant across multiple handwashing metrics (e.g., washing with soap/soap alternative, washing before food preparation, washing after defecation). A higher prevalence of children had reportedly washed their hands after defecation in the intervention arm, compared to the control arm; only 37% of children reportedly had washed their hands overall. There were no differences between the intervention and control arm for any facial cleanliness measures. Facial cleanliness indicators also demonstrated a need for improvement, with ocular discharge present amongst 38% of 1,441 children aged 1-9 years, wet nasal discharge present amongst 46%, and dry nasal discharge amongst 55%. It is worth noting that observations from our index children were not significantly different than other children from the household within the same age range. This indicates that the index child can serve as a sentinel for observed personal hygiene outcomes for all children within study households.

Household environmental sanitation indicators were also similar between intervention and control communities, and signals the need for further improvement. Animal feces were present in the compound in 83% of the households. However, about half of households managed the feces in some way (i.e., did not leave it out in the open).

Overall, the prevalence of anxiety, depression, and emotional distress was lower at midline than baseline amongst all survey respondents, regardless of treatment allocation. The mental well-being scores were similar between intervention and control arms, although each of the estimates trended toward the preventive direction. Mental well-being is an impact, at the farthest end of our theory of change. Changes in other factors at the output and outcome stages along our hypothesized causal chain are presumably required in order to trigger meaningful changes in mental well-being. Therefore, we expected that it would be too early for respondents to manifest indications of impact on these mental well-being indicators at a one year follow-up.

8.2. Strengths

Our study had several strengths, including the inclusion of theory-informed formative research and behavioral trials as part of our intervention design process. Our rigorous study design, with attention to external validity, and triangulation of data via objective indicators are also strengths.

We employed behavioral and change theories and systematically mapped our formative research, intervention development, and behavioral trials against three theories: Michie's COM-B model (Michie et al., 2011), the Theory of Triadic Influence (DiClemente et al., 2002), and the Trans-theoretical Model (Prochaska and DiClemente, 1982). Findings from formative work were leveraged to design our overarching intervention approach, intervention components, and tools and materials. For example, given feedback from community-level stakeholders, we avoided leveraging negative motives such as shame and disgust, and instead carried out a motive analysis to identify and leverage positive motives such as nurture and status. With a few notable exceptions (Dreibelbis et al., 2013), few WASH interventions are designed and evaluated with specific behavioral theory in mind. The incorporation of theory yielded a deliberate set of intervention techniques and behavioral factors such as action planning at woreda, community, and health post (e.g., HEW) levels. We have also paid close attention to the use of appropriate actors within the health system and at the community level. While the Andilaye team facilitated this process, utilizing standardized action planning templates, the action planning itself was led by key actors who were responsible for carrying out the plans. It was assumed that having the actors create their own action plans would improve buy-in and increase the likelihood of the execution thereof. We also received anecdotal evidence from the World Bank that the whole system in the room and action planning activities had been well received and executed in their previous programming in Amhara.

Our study utilized a randomized study design, in which intervention and control communities were allocated to treatment arms by chance. While RCTs tend to emphasize internal validity, we made considerable effort to enhance external validity of our study to better influence replicability and scale of the intervention and to influence policy. Our study was spread over three *woredas* in two zones, yielding a heterogeneous mix of contexts, including low-lying areas bordering Lake Tana, higher more arid areas, and sites closer to the regional capital of Bahir Dar. This heterogeneity improves the external validity of the study. We used a "fried egg" approach to select central *gotts* within intervention and control *kebeles* to minimize spillover and we have not revealed any evidence of spillover. We targeted both rural and peri-urban communities, and are collecting behavioral outcome data on a variety of household members (e.g., primary female caregiver of index child, head of household, all children aged 0-17 years).

We have collected myriad types of data. We conducted continuous data collection as part of an on-going process evaluation and quarterly monitoring visits (conducted at half of the households in each study community per quarter to minimize reactivity), which will allow us to monitor seasonal trends, and therefore improve the precision of our inferences and external validity across time. We are conducting a process evaluation alongside our impact evaluation to help contextualize and interpret results. Ongoing collection of qualitative and quantitative data was conducted at various levels, to obtain different perspectives on intervention implementation, fidelity, adherence, and behavioral adoption. Data collection instruments have triangulated self-report and objective data along the causal pathway of our theory of change, including data on behavioral antecedents and determinants, reported intentions and commitment, as well as directly observed and respondent-reported behavioral outcomes linked to the intervention. Our mental well-being scale is widely used and validated across several contexts.

8.3. Limitations

The study has several notable limitations. First, the year one follow-up data were collected only one to two months after key intervention activities were carried out at the community level (e.g., community mobilization and commitment events, skills-based training of WDALs). This was a result of significant delays in local ethical approval to start the project, but also a longer than anticipated intervention design process during which we emphasized the solicitation and incorporation of feedback from key stakeholders at regional, zonal, *woreda*, and community levels. In our original conceptualization, we were to study a government-led intervention. However, this plan became problematic as it became evident that rolling out a government sponsored program in a select number of areas was both politically and logistically infeasible. Our alternative approach, to develop an intervention that complemented the existing government approach will yield important policy and programmatic findings, but did result in a delayed intervention timeline.

A second limitation is that key actors have been less involved than planned (e.g., *Woreda* Health Office officials, CHC HEWs supervisors, HEWs have not conducted supportive supervision, as planned), and they have not fully utilized the intervention tools provided to them to facilitate supportive supervision and on-the-job training. We will continue to work with stakeholders at the national and regional level to disseminate key learnings as we enter the final year of the study. Our team proposed the initial idea to hold the 3rd WASH-NTD Roundtable in Addis Ababa in 2018, sponsored by the FMoH. Dr. Freeman is currently on the coordinating committee to organize the meeting in September and we expect that our *Andilaye* findings will be presented at this international forum, giving us an important policy influencing platform.

The results should be interpreted cautiously as there were \sim 30 outcomes of interest, and therefore a high possibility of some false positives (e.g., \sim 1.5 false positives would be expected). We did not do any multiple testing correction, as these tests reduce false positives at the expense of inflating the rate of false negatives (i.e., they reduce our ability to detect important effects). Several of our behavioral outcomes were self-reported, and these types of outcomes may be prone to reporting biases. Finally, we used parametric methods (e.g., linear regression) to test some of our continuous scores, and the assumptions for these methods may not have always be met.

8.4. Conclusions

The *Andilaye* intervention is theoretically-grounded and empirically-informed. Despite some gaps, fidelity of the rollout remains relatively high. Attrition of the study has been low, and we have successfully conducted midline data collection. Given the timeline, it is too early to assess the impact of the intervention on the sustainability of improvements in key WASH practices; however, some midline findings provide early indications of the potential promise of the intervention. Endline data collection in early 2019 will provide important, actionable findings.

The GoE is currently examining their HEP, with an eye toward further integration and improvement. We are in contact with the teams writing this guidance, and will continue to play an important role in influencing the policy. We believe that the final results of this study will contribute important evidence to the sector in terms of: holistic approaches to addressing sustained WASH behavior change, proof of principle for theory-informed intervention development, findings of the impact of sanitation improvements on mental well-being, and provision of tools that complement current CLTS(H) models that have revealed limitations with regard to sustainability of improved WASH behaviors.

9. Specific findings for policy and practice

Below we outline several key findings for policy and practice that we have uncovered to date. Our team will prepare more formal policy and practice recommendations with key government stakeholders when our endline data have been collected and analyzed.

- Our one year follow-up evaluation was conducted only a few months after our initial intervention rollout. Still, we noted some gains in key behaviors. We will look to the two year follow-up in March – April 2019 to fully assess the effectiveness of our approach. We saw some suggested trend towards the intervention impacting mental well-being that we will need to explore further at endline.
- CLTSH has resulted in dramatic increases in latrine coverage and reductions in open defecation. However, in many communities these gains have not been sustained. There is a need for approaches that complement CLTSH focused on behavioral maintenance. Our two year follow-up data collection will attempt to assess not only behavior change, but behavioral maintenance.
- We have developed tools that address behavioral antecedents to our targeted behaviors. We expect that further analysis will yield usable tools that could be used for future WASH evaluations.
- The use of behavioral theory and a structured approach to intervention development that incorporated stakeholder feedback yielded important guidance in selecting behaviors to target, identifying leading indicators of behavior change, and informed our approach.
- There is a need to coordinate WASH activities and incorporate behaviors that can mitigate NTDs endemic to Ethiopia. We have shown that it is feasible to develop a holistic WASH intervention.
- Some key behaviors, such as shoe wearing, though important, were not included because our formative research revealed that changing these behaviors require a supply-side intervention. In other contexts, or with the inclusions of supply-side approaches, it may be possible to target these NTD-related behaviors.
- The Andilaye intervention focused in part on capacitating strengthening of the WDALs. However, of the 25 intervention kebeles, the WDALs structure is fully functional in less than half. Woreda officials and health center supervisors state that though this structure should be in place throughout the region, the WDAL structure is not well established in some areas. This is not unique for Andilaye and is an issue for all types of community health delivery modalities. Observed differences between the intervention and control arms may come due to the availability of functional WDALs. Data from the midline and feedbacks from WDALs and HEWs, suggest that there is no support from health center supervisors and Woreda officials for the WDALs. There is a need to further strengthen the WDAL system and explain what duties and responsibilities they should have as a volunteer.
- Improving community change-agents (e.g., WDALs) motivation and participation: The project might have improved motivations of community change-agents through enhancing their capacity by providing counselling training and counseling tools, and small per diems during training days. Qualitative process evaluation preliminary findings showed better enthusiasm, participation and engagement of community change-agents (e.g., WDALs) in intervention kebeles.
- Improving motivation and participation of community change agents (e.g., WDALs) through providing trainings and incentives/motivation during trainings can help lift some burdens away from the already overburdened HEWs.
- There may be a need for more integration of activities during provision of supportive supervision to HEWs by CHC HEWs supervisors and *woreda* health officials. CHC HEWs supervisors and *woreda* health officials seem to focus on specific activates (e.g., vaccination and maternal and child health care) when doing supportive supervision to HEWs.

 It is important to consider stakeholders' expectations while extending invitations to participate in some intervention activities. For example, we noticed that lack of per diem for 'whole system in the room and action planning' participants (especially communitylevel leaders or administration workers and non-health government stakeholders) was a big concern for HEW who invited these participants. This may have implication on stakeholders' participation on other activities.

References

- Abele, A., Vjaters, E., Irmejs, A., Trofimovičs, G., Miklaševičs, E., Gardovskis, J., 2011. Epidemiologic, clinical, and molecular characteristics of hereditary prostate cancer in Latvia. Medicina (Kaunas) 47, 579–585.
- Briscoe, C., Aboud, F., 2012. Behaviour change communication targeting four health behaviours in developing countries: A review of change techniques. Social Science & Medicine, Part Special Issue: Challenges to changing health behaviours in developing countries 75, 612–621. https://doi.org/10.1016/j.socscimed.2012.03.016
- Cluster Randomised Trials [WWW Document], 2017. CRC Press. URL https://www.crcpress.com/Cluster-Randomised-Trials-Second-Edition/Hayes-Moulton/p/book/9781498728225 (accessed 8.28.18).
- Crocker, J., Saywell, D., Bartram, J., 2017. Sustainability of community-led total sanitation outcomes: Evidence from Ethiopia and Ghana. Int J Hyg Environ Health 220, 551–557. https://doi.org/10.1016/j.ijheh.2017.02.011
- Delahoy, M.J., Wodnik, B., McAliley, L., Penakalapati, G., Swarthout, J., Freeman, M.C., Levy, K., 2018. Pathogens transmitted in animal feces in low- and middle-income countries. Int J Hyg Environ Health 221, 661–676. https://doi.org/10.1016/j.ijheh.2018.03.005
- Delea, M.G., Solomon, H., Solomon, A.W., Freeman, M.C., 2018. Interventions to maximize facial cleanliness and achieve environmental improvement for trachoma elimination: A review of the grey literature. PLoS Negl Trop Dis 12, e0006178. https://doi.org/10.1371/journal.pntd.0006178
- Designing for Behavior Change For Agriculture, Natural Resource Management, Health and Nutrition, 2013., Operational Performance Support (TOPS) Program. Food Security and Nutrition Network Social and Behavioral Change Task Force, Washington, DC.
- DiClemente, R.J., Crosby, R.A., Kegler, M.C., 2002. Emerging Theories in Health Promotion Practice and Research: Strategies for Improving Public Health. John Wiley & Sons.
- Freeman, M.C., Ogden, S., Jacobson, J., Abbott, D., Addiss, D.G., Amnie, A.G., Beckwith, C., Cairncross, S., Callejas, R., Jr, J.M.C., Emerson, P.M., Fenwick, A., Fishman, R., Gallo, K., Grimes, J., Karapetyan, G., Keene, B., Lammie, P.J., MacArthur, C., Lochery, P., Petach, H., Platt, J., Prabasi, S., Rosenboom, J.W., Roy, S., Saywell, D., Schechtman, L., Tantri, A., Velleman, Y., Utzinger, J., 2013. Integration of Water, Sanitation, and Hygiene for the Prevention and Control of Neglected Tropical Diseases: A Rationale for Inter-Sectoral Collaboration. PLOS Neglected Tropical Diseases 7, e2439. https://doi.org/10.1371/journal.pntd.0002439
- Garn, J.V., Brumback, B.A., Drews-Botsch, C.D., Lash, T.L., Kramer, M.R., Freeman, M.C., 2016. Estimating the Effect of School Water, Sanitation, and Hygiene Improvements on Pupil Health Outcomes. Epidemiology 27, 752–760. https://doi.org/10.1097/EDE.000000000000222
- Garn, J.V., Trinies, V., Toubkiss, J., Freeman, M.C., 2017. The Role of Adherence on the Impact of a School-Based Water, Sanitation, and Hygiene Intervention in Mali. Am. J. Trop. Med. Hyg. 96, 984–993. https://doi.org/10.4269/ajtmh.16-0558
- Hadley, C, Lindstrom, D., Tessema, F., Belachew, T., 2008. Gender bias in the food insecurity experience of Ethiopian adolescents. Soc Sci Med 66, 427–438. https://doi.org/10.1016/j.socscimed.2007.08.025
- Hadley, C., Patil, C.L., 2006. Food insecurity in rural Tanzania is associated with maternal anxiety and depression. Am. J. Hum. Biol. 18, 359–368. https://doi.org/10.1002/ajhb.20505
- Hadley, C., Tegegn, A., Tessema, F., Cowan, J.A., Asefa, M., Galea, S., 2008. Food insecurity, stressful life events and symptoms of anxiety and depression in east Africa: evidence from the Gilgel Gibe growth and development study. J Epidemiol Community Health 62, 980–986. https://doi.org/10.1136/jech.2007.068460
- Hernandez, O., Rosenbaum, J., 2011. Results from working at scale for better sanitation and hygiene in Amhara, Ethiopia: Baseline and endline comparisons of institutional, household, and school surveys. WSP and USAID.
- Hulland, K., Martin, N., Dreibelbis, R., DeBruicker Valliant, J., Winch, P., 2015. What factors affect sustained adoption of safe water, hygiene and sanitation technologies? A systematic review of literature Hulland K, Martin N, Dreibelbis R, DeBruicker Valliant J, Winch P (2015) What factors affect sustained adoption of safe water, hygiene and sanitation technologies? A systematic review of literature. London: EPPI-Centre, Social Science Research Unit, UCL

Institute of Education, University College London. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.

- Kahan, B.C., Morris, T.P., 2012. Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis. BMJ 345, e5840. https://doi.org/10.1136/bmj.e5840
- Lachin, J.M., 1988. Statistical properties of randomization in clinical trials. Control Clin Trials 9, 289– 311.
- Luby, S.P., Rahman, M., Arnold, B.F., Unicomb, L., Ashraf, S., Winch, P.J., Stewart, C.P., Begum, F., Hussain, F., Benjamin-Chung, J., Leontsini, E., Naser, A.M., Parvez, S.M., Hubbard, A.E., Lin, A., Nizame, F.A., Jannat, K., Ercumen, A., Ram, P.K., Das, K.K., Abedin, J., Clasen, T.F., Dewey, K.G., Fernald, L.C., Null, C., Ahmed, T., Colford, J.M., 2018. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Bangladesh: a cluster randomised controlled trial. The Lancet Global Health 6, e302–e315. https://doi.org/10.1016/S2214-109X(17)30490-4
- Marteau, T.M., Hollands, G.J., Fletcher, P.C., 2012. Changing human behavior to prevent disease: the importance of targeting automatic processes. Science 337, 1492–1495. https://doi.org/10.1126/science.1226918
- Michie, S., van Stralen, M.M., West, R., 2011. The behaviour change wheel: A new method for characterising and designing behaviour change interventions. Implement Sci 6, 42. https://doi.org/10.1186/1748-5908-6-42
- Mosler, H.-J., 2012. A systematic approach to behavior change interventions for the water and sanitation sector in developing countries: a conceptual model, a review, and a guideline. Int J Environ Health Res 22, 431–449. https://doi.org/10.1080/09603123.2011.650156
- Null, C., Stewart, C.P., Pickering, A.J., Dentz, H.N., Arnold, B.F., Arnold, C.D., Benjamin-Chung, J., Clasen, T., Dewey, K.G., Fernald, L.C.H., Hubbard, A.E., Kariger, P., Lin, A., Luby, S.P., Mertens, A., Njenga, S.M., Nyambane, G., Ram, P.K., Colford, J.M., 2018. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Kenya: a cluster-randomised controlled trial. The Lancet Global Health 6, e316–e329. https://doi.org/10.1016/S2214-109X(18)30005-6
- Penakalapati, G., Swarthout, J., Delahoy, M.J., McAliley, L., Wodnik, B., Levy, K., Freeman, M.C., 2017. Exposure to Animal Feces and Human Health: A Systematic Review and Proposed Research Priorities. Environ. Sci. Technol. 51, 11537–11552. https://doi.org/10.1021/acs.est.7b02811
- Prochaska, J.O., DiClemente, C.C., 1982. Transtheoretical therapy: Toward a more integrative model of change. Psychotherapy: Theory, Research & Practice 19, 276–288. https://doi.org/10.1037/h0088437
- Rabbi, S.E., Dey, N.C., 2013. Exploring the gap between hand washing knowledge and practices in Bangladesh: a cross-sectional comparative study. BMC Public Health 13, 89. https://doi.org/10.1186/1471-2458-13-89
- Stevenson, E., Yohannes, B., 2014. Food security in the context of a pastoralist resettlement program in the Lower Omo valley, Ethiopia, Culture, technology, and development: Proceedings of the first national conference of the School of Behavioral Sciences. Hawassa, Ethiopia.
- Webb, T.L., Sheeran, P., 2006. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull 132, 249–268. https://doi.org/10.1037/0033-2909.132.2.249
- Wood, W., Neal, D.T., 2016. Healthy through habit: Interventions for initiating & maintaining health behavior change. Behavioral Science & Policy 2, 71–83. https://doi.org/10.1353/bsp.2016.0008
- World Health Organization and UNICEF, 2013. Progress on sanitation and drinking-water.
- Zou, G., 2004. A modified poisson regression approach to prospective studies with binary data. Am. J. Epidemiol. 159, 702–706.

Appendix A. Summary documents of *Andilaye* formative research and intervention design

- Project report: Andilaye formative research findings see here
- **Research note:** Exploring the barriers and facilitators to improved WASH behaviors in Amhara, Ethiopia see <u>here</u>
- Intervention design: Problem and solution trees see here
- Meeting report: Andilaye project intervention design workshop see here
- **Project summary:** *Andilaye*: the impact of enhanced, demand-side sanitation and hygiene promotion on sustained behavior change and health in Ethiopia see <u>here</u>

Appendix B. Summary of behavioral themes and domains addressed in formative research, in addition to behavior change techniques

Behavioral theme	Behavioral domains to address	Behavior change techniques
	(based on formative research around barriers & facilitators to behavior change)	
Psychological and Physical	l Capability	
Sanitation	 Latrine construction, maintenance, and upgrades HEWs and HHs know where to access materials for different latrine designs HH knowledgeable about how to progress along sanitation ladder Accurate perceptions regarding what constitutes an improved latrine HH has capacity to construct latrine from sustainable materials (e.g. cement, tin) Perception latrines can withstand environmental conditions Knowledge about how to prevent cattle/oxen from destroying latrines Knowledge about types of sanitation technology for different hydrogeological conditions Knowledge on how and where to construct latrine Latrine use Latrine is clean/does not smell Latrine pit not used for solid waste disposal Latrine is long-lasting Knowledgeable regarding latrine O&M 	 Education Training Skill demonstration Provide instruction Modelling Set graded tasks Enactive mastery experiences Verbal persuasion Reattribution training Provide contingent rewards Goal setting Action knowledge Action planning
Personal Hygiene	 Perceived ability to get clean with water Feasible methods available to filter water and address turbidity Know when and how facewashing (FW) / handwashing (HW) Know to wash hands with soap/soap substitute Know to wash face whenever dirty Perception FW with water alone can improve facial cleanliness 	
Household Environmental Sanitation	 Perceive have capacity to clean up animal feces and complete household chores/farming tasks Perceive have capacity to construct separate enclosure for animals Knowledge on how and where to construct separate enclosure for animals Knowledge on how to build chicken coops 	

Risk Factors		
Sanitation	 Latrine use Knowledge about health benefits of latrine use Internalize health risk associated with OD 	 ✓ Education/Knowledge transfer ✓ Provide information on consequences ✓ Coercion
Personal Hygiene	 Accurate perception of health risk related to <u>not</u> FW/HW Internalize health risks related to not FW/HW View all HW moments as important to health Recognition that previous generations did not practice FW/HW and thus experienced more disease Understand health benefits of FW/HW 	 ✓ Fear arousal ✓ Scenario-based risk information ✓ Framing ✓ Environmental reevaluation ✓ Persuasive arguments ✓ Personalized risk messages
Household Environmental Sanitation	 Knowledge around health risk to different animal feces Feces from chickens seen as just as harmful as feces from dogs, cats and other animals (and feces from dogs and cats seen as just as harmful as feces from humans, etc.) Cow dung viewed as important fuel source but also harmful to health (women and girls prepare patties with bare hands but wash hands with soap/soap substitute afterwards) 	
Attitude Factors		
Sanitation	 Latrine construction, maintenance, and upgrades Latrine construction prioritized Latrine use Knowledge around non-health benefits Women realized they can defecate whenever they would like (no need to follow "program") Perception of women that they can break "program" when have access to latrine People consider the benefits of using latrine during rainy season (avoid mud) People like using latrines People prefer latrine use over OD Perception OD does not provide privacy Do not feel comfortable ODing/urinating around HH compound 	 Persuasion Modelling Self-reevaluation Persuasive arguments Persuasive peripheral cues Affective persuasion De-conditioning Cue altering Implementation intentions Showing scenarios Prompt barrier identification
Personal Hygiene	 Perception water should be allocated for FW/HW Perception FW/HW requires little water (i.e., only requires a manageable amount of water) Perception that FW/HW does not take much time Soap/soap substitute is prioritized for HW Soap/soap substitute perceived to be affordable 	_
Household Environmental Sanitation	 Perceived as important to clean up animal feces in compound Do not fear animals will be stolen if in separate enclosure (e.g. confidence in dog guard, lock) People prioritize health risk over need for closeness with livestock Perceive chicken coops to be important/necessary Perception that clean-up of animal feces is the duty of both men and women in the HH 	_

Self-Regulation Factors		
Sanitation	 Latrine construction, maintenance, and upgrades HH plans to repair or upgrade latrine with strong/sustainable material Latrines reconstructed when destroyed Perceived capacity to keep rebuilding Knowledge regarding O&M Skills in place to carry out O&M 	 Planning coping responses Contingency management Coping with relapse Self-monitoring Conditioning Cue altering Stimulus control
Personal Hygiene	 Established routine for FW/HW Habituation of FW/HW Commitment to sustaining improved practices Recovery from personal setbacks to FW/HW and recovering from external obstacles hindering FW/HW Plan for addressing obstacles that arise Perceived ability to recover practice after experiencing setbacks Perceived ability to continue practice as obstacles arise Sufficient cues to action for FW/HW Cues to action for HW before and eating and food prep Cues to action for HW after defecation (e.g. tippy tap) 	 ✓ Barrier planning ✓ Reattribution training
Social Opportunity		
Sanitation	 Latrine use Latrine use is common practice Part of culture for community always use latrine Social pressure to exclusively use latrine Embarrassing to OD Perception that rural farmers use latrines Latrines give people status/prestige Acceptable to use friends' and neighbors" latrines (part of culture) 	 Restrictions Modelling Incentivizing Provide information about others' approval Implementation intentions Public commitment Stimulate communication to mobilize social support Anticipated regret
Social Opportunity (continue	ed)	
Personal Hygiene	 FW/HW is a (collective) habit HW after defecation is a habit Face cleaning with water is current habit FW/HW is a norm Perception daily face washing is done by all farmers Women who maintain cleanliness are healthy and good Farmers have social norm to be clean (view dirt under their fingernails as harmful to health) 	 ✓ Entertainment/edutainment ✓ Mobilizing social networks ✓ Highlighting norms
Household Environmental Sanitation	 Social pressure to keep HH compounds clean Norms around HH compound cleanliness 	

Physical Opportunity		
Sanitation	 Latrine construction, maintenance, and upgrades Sanitation marketing in place to drive down cost of building materials Perception there is capacity to construct latrine Physically able to construct (or have support) Local installation capacity (i.e. skilled masons) Perception there is adequate land/space for latrine Cooperation from <i>kebele</i> leaders to allocate land for latrines Supportive neighbors Latrines perceived as affordable Opportunity cost to construct is perceived to be low (able to spend time away from field, perceive won't take much time to construct) No conflicting financial interests Latrine design makes it easy for elderly, persons living with disability, and small children to use Functional latrine available at HH Water available for latrine use needs (anal cleansing, cleaning latrine, handwashing) 	 ✓ Plan social support or social change ✓ Technical assistance ✓ Environmental restructuring ✓ Infrastructure
Personal Hygiene	 Sufficient quantity of water available at home for FW/HW Water quality is good Rainy season/dry season water appears clean Soap sold at nearby shops Purchase soap/soap substitute Can afford enough soap/soap substitute for all HH members 	-
Household Environmental Sanitation	 Perceive construction of separate enclosure is affordable Opportunity cost to construct is perceived to be low (able to spend time away from field, perceive will not take much time to construct 	

Appendix C. Sample size calculation

	Unexpected prevalence (unexposed) P	# in cluster m	# clusters g	CI ta .05	Error tb .80	ICC	Magnitude of difference
0.25 baseline	0.25	20	25	2	0.848	0.05	44%
	0.25	25	25	2	0.848	0.05	41%
	0.25	30	25	2	0.848	0.05	40%
	0.25	35	25	2	0.848	0.05	39%
	0.25	40	25	2	0.848	0.05	38%
0.30 baseline	0.30	20	25	2	0.848	0.05	38%
	0.30	25	25	2	0.848	0.05	37%
	0.30	30	25	2	0.848	0.05	35%
	0.30	35	25	2	0.848	0.05	34%
	0.30	40	25	2	0.848	0.05	33%
0.35 baseline	0.35	20	25	2	0.848	0.05	34%
	0.35	25	25	2	0.848	0.05	33%
	0.35	30	25	2	0.848	0.05	31%
	0.35	35	25	2	0.848	0.05	30%
	0.35	40	25	2	0.848	0.05	30%

Appendix D. Outcome indicators and related prompts

Table D.1. Behavioral outcome indicators and related prompts

Related survey prompt	Type of data	Granularity of data (level)	Data captured on
Latrine coverage			
How many household latrines does your household have access to?	Reported & Observed	HH-level	All HHs
OBSERVE: Type of latrine.	Observed	HH-level	
OBSERVE: Is the latrine floor a smooth and cleanable surface?	Observed	HH-level	All nn launnes
Do you consider your latrine facility, as it is today, fully constructed?	Reported	HH-level	
Latrine characteristics		r	ſ
Presence of stagnant water (water log) over the floor, latrine slab?	Observed	HH-level	
Evidence of discoloration of pan, slab, or floor (e.g., yellow, green)?	Observed	HH-level	
Presence of flies in latrine?	Observed	HH-level	
Presence of drop hole cover in the latrine?	Observed	HH-level	
Drop hole cover currently situated over drop hole (i.e., fully covering drop hole)?	Observed	HH-level	
Presence of cleaning agents for washing latrine (inside or near the latrine)?	Observed	HH-level	
Presence of feces on floor/slab or other place in the latrine aside from the pit?	Observed	HH-level	
Evidence latrine is used for storage or other non-sanitation-related purpose?	Observed	HH-level	
Presence of well-worn path to latrine?	Observed	HH-level	All HH latrines
Presence of fresh feces on/in the pit or pan?	Observed	HH-level	
Is the pit that is in use full or close to being full?	Observed	HH-level	
Presence of anal cleansing item in, near latrine (e.g., paper, straw, wood, water)?	Observed	HH-level	
Presence of odor from stools or urine in the latrine?	Observed	HH-level	
Presence of leaves, spider webs, rubbish, other dirt in latrine?	Observed	HH-level	
Wet latrine floor?	Observed	HH-level	
Presence of water available near or inside latrine for hand washing?	Observed	HH-level	
Presence of cleansing agent near or inside latrine for hand washing?	Observed	HH-level	
Water available inside or near latrine for flushing or self-cleansing?	Observed	HH-level	
Facilities operation & maintenance		l	
During the last 7 days, including today, how many times was this latrine cleaned?	Reported	HH-level	
Have you added or improved anything on this latrine to upgrade it since its original construction (i.e., has the latrine been upgraded or improved upon since its original construction)?	Reported	HH-level	
Have you ever fixed anything that became broken, damaged, our worn out on this latrine since its original construction (i.e., has the latrine been repaired since its original construction)?	Reported	HH-level	
Is your latrine facility working (operating) correctly now?	Reported	HH-level	All HH latrines
OBSERVE: Does the facility require any obvious repairs?	Observed	HH-level	
OBSERVE: I latrine "serviceable", meaning the pit is not overflowing, and the floor provides a solid foundation over the latrine pit?	Observed	HH-level	
OBSERVE: Presence of cleaning agents for cleaning latrine (inside/near latrine)?	Observed	HH-level	
OBSERVE: Presence of feces on floor/slab, other place in the latrine aside from pit?	Observed	HH-level	
Latrine utilization		L	L
During the last 2 days, where was (your [respondent], HoH's, HH member's [those aged 4-17]) primary place of defecation?	Reported	Individual-level	
Does the primary place of defecation change over the course of the year?	Reported	Individual-level	Respondent, HoH,
During the last 2 days, did (you, HoH, HH members) defecate in any latrine?	Reported	Individual-level	all HH members
During the last 7 days, including today, did (you, HoH, HH member [those	D a n a nta d	la di data la cal	aged 4-17
aged 4-17]) always exclusively use a latrine for defecation?	Reported	individual-level	
Has any member of your household used this latrine for 3 or more days during the LAST 7 days, including today?	Reported	HH-level	
During the LAST 7 days, how many households have USED this household latrine?	Reported	Inter- compound- level	All nn latrines

During the LAST 7 days, NOT INCLUDING your household members, how		Inter-	
many people used this household latrine?		compound-	
Open defecation		levei	l
During the last 2 days where was (your [respondent] HoH HH member's			
[those aged 4-17]) primary place of defecation?	Reported	Individual-level	
Does the primary place of defecation change over the course of the year?	Reported	Individual-level	
During the last two days, did (you, HoH, HH member) openly defecate?	Reported	Individual-level	Respondent, HoH,
During the past 2 days, did (you, HoH, HH member) OD in or near a surface	Departed	Individual laval	all HH members
water source (e.g., pond, river, or lake)?	Reported	Individual-level	ageu 4-17
During the last 7 days, including today, did (you, HoH, HH member [those	Reported	Individual-level	
aged 4-17]) always exclusively use a latrine for defecation?	Reported		
Evidence of open defecation (i.e., human feces) in or near HH compound	Observed	HH-level	All HH compounds
The last time (the child - HH member under 4 years of age) passed stools,	Reported	Individual-level	All HH members
What happened to the stools?			<= 3 y/o
Officiation			Respondent HoH
During the past 2 days, did (you, HoH, HH member) urinate in or near a	Reported	Individual-level	all HH members
surface water source (e.g., pond, river, or lake)?	Ropolitou		aged 4-17
Animal husbandry/hygiene			
Deep this near a have any animal handing as other eximal hyphendry			Respondent, HoH,
Does this person have any animal herding or other animal husbandry	Poportod	Individual loval	all HH members
chickens)	Reported	individual-level	aged > 3 and < 17
			years
Last night, where did you keep your animals?	Observed	HH-level	
Ask: How do you dispose of the animal feces/waste in your compound?	Reported	HH-level	
OBSERVE: How does the household appear to be disposing of the animal	Observed	HH-level	All HHs
teces/waste?	0		
OBSERVE: is there animal feces present in the compound?	Observed	HH-level	
Solid waste management	Poportod		
OBSERVE: How does the HH appear to be disposing of solid waste?	Observed		All HHs
Observe. How does the first appear to be disposing of solid waste:	Obscived		
FSM			
FSM			HHs reporting a
FSM What was done with the sludge?	Reported	HH-level	HHs reporting a HH latrine pit had
FSM What was done with the sludge?	Reported	HH-level	HHs reporting a HH latrine pit had filled in the past
FSM What was done with the sludge? Face cleaning practices	Reported	HH-level	HHs reporting a HH latrine pit had filled in the past
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or	Reported	HH-level	HHs reporting a HH latrine pit had filled in the past
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else?	Reported Reported	HH-level	HHs reporting a HH latrine pit had filled in the past
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, often index child's face get cleaned?	Reported Reported Reported	HH-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cleate such as a towel or approp?	Reported Reported Reported Reported	HH-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOUL clean YOUR face?	Reported Reported Reported Reported	HH-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, how did you clean your face?	Reported Reported Reported Reported Reported	HH-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth	Reported Reported Reported Reported Reported Reported	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron?	Reported Reported Reported Reported Reported Reported Reported	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, now did you clean your face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron?	Reported Reported Reported Reported Reported Reported Reported	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Facial cleanliness	Reported Reported Reported Reported Reported Reported Reported	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, how did you clean your face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Facial cleanliness OBSERVE whether ocular discharge is absent or present	Reported Reported Reported Reported Reported Reported Reported Observed	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, how did you clean your face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Facial cleanliness OBSERVE whether ocular discharge is absent or present OBSERVE whether WET nasal discharge is absent or present	Reported Reported Reported Reported Reported Reported Reported Observed Observed	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, how did you clean your face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Facial cleanliness OBSERVE whether ocular discharge is absent or present OBSERVE whether WET nasal discharge is absent or present OBSERVE whether DRY nasal discharge is absent or present	Reported Reported Reported Reported Reported Reported Reported Observed Observed Observed	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent HH members aged 1-9 yrs
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, how did you clean your face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Facial cleanliness OBSERVE whether ocular discharge is absent or present OBSERVE whether DRY nasal discharge is absent or present OBSERVE whether DRY nasal discharge is absent or present OBSERVE whether other dirt/dust/other debris is absent or present	Reported Reported Reported Reported Reported Reported Reported Observed Observed Observed	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent HH members aged 1-9 yrs
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Facial cleanliness OBSERVE whether ocular discharge is absent or present OBSERVE whether DRY nasal discharge is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present	Reported Reported Reported Reported Reported Reported Reported Observed Observed Observed Observed	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent HH members aged 1-9 yrs
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Facial cleanliness OBSERVE whether ocular discharge is absent or present OBSERVE whether ORY nasal discharge is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirthe spectration.	Reported Reported Reported Reported Reported Reported Reported Observed Observed Observed Observed	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent HH members aged 1-9 yrs
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Facial cleanliness OBSERVE whether ocular discharge is absent or present OBSERVE whether WET nasal discharge is absent or present OBSERVE whether ORY nasal discharge is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present DBSERVE whether other dirt/dust/other debris is absent or present DBSERVE whether other dirt/dust/other debris is absent or present Enter the number of times you observed a fly land on the index child's face during a 1 minute observation. Enter the number of the color on the scale that most closely matches the darkest part of the eswere but the start of the ave struct the struct the struct of the ave struct the struct of the	Reported Reported Reported Reported Reported Reported Reported Observed Observed Observed Observed Observed	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent HH members aged 1-9 yrs Index child
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Facial cleanliness OBSERVE whether ocular discharge is absent or present OBSERVE whether DRY nasal discharge is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present DBSERVE whether other dirt/dust/other debris is absent or present Enter the number of times you observed a fly land on the index child's face during a 1 minute observation. Enter the number of the color on the scale that most closely matches the darkest part of the eye swab. Handwashing practices	Reported Reported Reported Reported Reported Reported Reported Observed Observed Observed Observed Observed	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent HH members aged 1-9 yrs Index child
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Facial cleanliness OBSERVE whether ocular discharge is absent or present OBSERVE whether DRY nasal discharge is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE at 1 minute observation. Enter the number of times you observed a fly land on the index child's face during a 1 minute observation. Enter the number of the eolor on the scale that most closely matches the darkest part of the eye swab. Handwashing practices YESTERDAY, did index child's HANDS get washed (by you, the child or <td>Reported Reported Reported Reported Reported Reported Reported Observed Observed Observed Observed Observed</td> <td>HH-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level</td> <td>HHs reporting a HH latrine pit had filled in the past Index child Respondent HH members aged 1-9 yrs Index child</td>	Reported Reported Reported Reported Reported Reported Reported Observed Observed Observed Observed Observed	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent HH members aged 1-9 yrs Index child
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Facial cleanliness OBSERVE whether ocular discharge is absent or present OBSERVE whether ORY nasal discharge is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE at minute observation. Enter the number of the color on the scale that most closely matches the darkest part of the eye swab. Handwashing practices YESTERDAY, did index child's HANDS get washed (by you, the child, or anyone else)?	Reported Reported Reported Reported Reported Reported Reported Observed Observed Observed Observed Observed Observed	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent HH members aged 1-9 yrs Index child
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Facial cleanliness OBSERVE whether ocular discharge is absent or present OBSERVE whether ORY nasal discharge is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE and the every on the scale that most closely matches the darkest part of the color on the scale that most closely matches the darkest part of the every swab. Handwashing practices YESTERDAY, did index child's HANDS get washed (by you, the child, or anyone else)? YESTERDAY when index child's	Reported Reported Reported Reported Reported Reported Reported Observed Observed Observed Observed Observed Observed	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent HH members aged 1-9 yrs Index child
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Facial cleanliness OBSERVE whether ocular discharge is absent or present OBSERVE whether URT nasal discharge is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other color on the scale that most closely matches the darkest part of the eye swab. Handwashing practices YESTERDAY, did index child's HANDS get washed (by you, the child, or anyone else)? YESTERDAY when index child's hands were washed, was soap/ash/or soapy water used?	Reported Reported Reported Reported Reported Reported Reported Observed Observed Observed Observed Observed Observed Reported Reported	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent HH members aged 1-9 yrs Index child
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, after you washed YOUR face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Facial cleanliness OBSERVE whether ocular discharge is absent or present OBSERVE whether ORY nasal discharge is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether of times you observed a fly land on the index child's face during a 1 minute observation. Enter the number of the color on the scale that most closely matches the darkest part of the eye swab. Handwashing practices YESTERDAY, did index child's hands were washed, was soap/ash/or soapy water used? The last time the index child defecated, did s/he clean his/her hands with	Reported Reported Reported Reported Reported Reported Reported Observed Observed Observed Observed Observed Observed Reported Reported	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent HH members aged 1-9 yrs Index child
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, after you washed YOUR face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Facial cleanliness OBSERVE whether ocular discharge is absent or present OBSERVE whether WET nasal discharge is absent or present OBSERVE whether DRY nasal discharge is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether of times you observed a fly land on the index child's face during a 1 minute observation. Enter the number of the color on the scale that most closely matches the darkest part of the eye swab. Handwashing practices YESTERDAY, did index child's hands were washed, was soap/ash/or soapy water used? The last time the index child defecated, did s/he clean his/her hands with water and soap, soapy water, or ash?	Reported Reported Reported Reported Reported Reported Reported Observed Observed Observed Observed Observed Observed Reported Reported Reported	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent HH members aged 1-9 yrs Index child
FSM What was done with the sludge? Face cleaning practices YESTERDAY, did index child's FACE get cleaned by you, the child, or anyone else? YESTERDAY, how did index child's face get cleaned? Yesterday, after index child's FACE was washed, was it wiped dry with a cloth such as a towel or apron? YESTERDAY, did YOU clean YOUR face? Yesterday, how did you clean your face? Yesterday, after you washed YOUR FACE, did you wipe it dry with a cloth such as a towel or apron? Facial cleanliness OBSERVE whether ocular discharge is absent or present OBSERVE whether ORY nasal discharge is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present DBSERVE whether other dirt/dust/other debris is absent or present OBSERVE whether other dirt/dust/other debris is absent or present DBSERVE whether other dirt/dust/other debris is absent or present DBSERVE	Reported Reported Reported Reported Reported Reported Reported Observed Observed Observed Observed Observed Observed Reported Reported Reported Reported	HH-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level Individual-level	HHs reporting a HH latrine pit had filled in the past Index child Respondent HH members aged 1-9 yrs Index child Index child

		1		
The last time you defecated, did you clean your hands with water and soap, soapy water, or ash?	Reported	Individual-level		
The last time you prepared food, did you clean your hands with water and	Reported	Individual-level		
Hand cleanliness				
Finger nail cleanliness: Left hand	Observed	Individual-level		
Finger nail cleanliness: Eich hand	Observed		Personalant	
Finger half cleanliness: Light hand	Observed		nrocont UU	
Finger pad cleanliness: Left hand	Observed		members aged 1-	
Pringer pau cleaniness. Right fiand	Observed		10 vrs	
Palm cleanliness. Leit fidiu	Observed		10 y13	
Pathing prostions	Observed	Individual-level		
During the last 7 days, including today, how many times did the index shild				
bathe or get bathed by another?	Reported	Individual-level	Index child	
During the last 7 days, including today, how many times did YOU bathe your body?	Reported	Individual-level	Respondent	
During the last 7 days, including today, was surface water (e.g., pond, river or	Reported	HH-level	Respondent and	
lake) used for bathing?	Reported		index child	
Fomite washing				
During the last 7 days, including today:				
How many times did you wash the index child's clothes	Reported	Individual-level	Index child	
How many times did you wash YOUR own clothes	Reported	Individual-level	Respondent	
How many times did you wash towels	Reported	HH-level	rtoopondont	
How many times did you wash bedsheets	Reported	HH-level		
How many times did you wash baby carriers	Reported	HH-level	All HHs	
How many times did you wash children's toys	Reported	HH-level		
Washing station coverage	Roportou			
Does your household have any hand or facewashing stations?	Reported	HH-level		
OBSERVE: What are the components of the washing station:	Observed	HH-level		
OBSERVE: Location of the handwashing/facewashing station	Observed	HH-level		
OBSERVE: Presence/absence of water at washing stations	Observed	HH-level	All HHs	
OBSERVE: Presence/absence of soap or soap substitute (e.g. soapy water	00001100			
ash) at the washing stations	Observed	HH-level		
Shoe wearing				
			HoH all other HH	
If present, OBSERVE whether (HoH, HH member) is wearing shoes.	Observed	Individual-level	members aged <	
OBSERVE: Did the respondent wear shoes to walk to the latrine?	Observed	Individual-level	Respondent	
Access to clean water (drinking, domestic use)	00001100	individual lovel	rtoopondont	
What is the main source of DRINKING WATER for members of your				
household?	Reported	HH-level		
Is this main source of DRINKING WATER located within your own household compound?	Reported	HH-level		
What is the main source of water used by your household for other purposes, such as cooking and handwashing?	Reported	HH-level	Airtins	
Is this main source of water used for other purposes located within your own	Observed	HH-level	HH-level	
household compound?	00001100			
Water storage		r		
Please show me all of the containers your household uses to STORE water (enumerator enters number of containers)	Observed	HH-level	All HHs	
Number of small-necked containers used to store water.	Observed	HH-level		
Water treatment				
Did you do anything to treat THIS water?	Reported	HH-level		
What method of treatment did you use to treat THIS water (i.e., the water currently in the container, or the remnants thereof)?	Reported	HH-level	All HHs	

Table D.2. Health outcome indicators and related prompts

Outcome	Related survey prompt	Data captured on			
Enteric infection					
	In the LAST 2 days, has (HH member [0-10 years of age]) had three or more loose stools per day (diarrhea)?				
Diarrhea	During the last 7 days, including today, has (HH member [0-10 years of age]) had three or more loose stools per day (diarrhea)?	Children ages 0 - 10 years			
	During the last 7 days, including today, has (HH member [0-10 years of age]) had blood in the stool (dysentery)?				
	Mental well-being				
	Suddenly scared for no reason - Read response options.				
	Feeling fearful - Read response options.				
	Faintness, dizziness or weakness - Read response options.				
	Nervousness or shakiness inside - Read response options.				
Anxiety	Heart pounding or racing - Read response options.				
symptoms	Trembling - Read response options.				
	Feeling tense or keyed up - Read response options.				
	Headaches - Read response options.				
	Spell of terror or panic - Read response options.				
	Feeling restless or can't sit still - Read response options.				
	Feeling low in energy, slowed down - Read response options.				
	Blaming yourself for things - Read response options.				
	Crying easily - Read response options.				
	Poor appetite - Read response options.				
	Difficulty falling asleep, staying asleep - Read response options.	Respondent			
	Feeling hopeless about future - Read response options.				
Depression	Feeling blue - Read response options.				
symptoms	Feeling lonely - Read response options.				
	Feeling of being trapped or caught - Read response options.				
	Worry too much about things - Read response options.				
	Feeling no interest in things - Read response options.				
	Feeling everything is an effort - Read response options.				
	Feeling of worthlessness - Read response options.				
	I have felt cheerful and in good spirits - Read response options.				
	I have felt calm and relaxed - Read response options.				
WHO 5 well-	I have felt active and vigorous - Read response options.				
being index	I woke up feeling fresh and rested - Read response options.				
	My daily life has been filled with things that interest me - Read response options.				
	Are you currently ill?				

* All indicators are respondent-reported, and generate individual-level data

Appendix E. Random allocation of study clusters

Figure E.1. Flow chart indicating *kebele* eligibility, randomization, allocation, and follow-up



Kabala	Intervention assignment	Data oprollod	Enrolled HH	CLTSH status
Robir Dor Zurio (N=16)	intervention assignment	Date enfoneu	Enroneu AA	
	O - m true l	0/0/0047	00	Trianand and ODE Qurtified
Andasa/Yigoma	Control	3/8/2017	36	Triggered, and ODF Certified
		3/8/2017	32	Triggered, and ODF Certified
Debranta	Intervention	4/9/2017	32	Triggered, but not yet ODF Certified
Dehina Mariam	Intervention	4/5/2017	33	Triggered, and ODF Certified
Fereswoga	Intervention	4/7/2017	31	Triggered, and ODF Certified
Gonbat	Intervention	3/21/2017	32	Triggered, and ODF Certified
Kimbaba	Control	4/9/2017	32	Triggered, but not yet ODF Certified
Lata	Intervention	4/9/2017	35	Triggered, and ODF Certified
Maqual	Control	3/8/2017	33	Triggered, and ODF Certified
Robit	Control	3/21/2017	30	Triggered, and ODF Certified
Sebatamit	Control	3/7/2017	36	Non-triggered
Sekelet	Control	4/5/2017	35	Non-triggered
Tentalaguna	Control	3/21/2017	30	Triggered, and ODF Certified
Wogelsa	Intervention	4/5/2017	32	Triggered, and ODF Certified
Wondatta	Intervention	4/7/2017	32	Triggered, but not yet ODF Certified
Yiqodi	Control	4/7/2017	34	Triggered, and ODF Certified
Farta (N=22)			-	
Arga	Intervention	3/23/2017	30	Triggered and ODE Certified
Ariendo	Control	3/18/2017	30	Triggered, and ODF Certified
Ata	Control	3/25/2017	30	Triggered, and ODE Certified
Avidde	Intervention	3/21/2017	33	Triggered, and ODF Certified
	Intervention	J/J/2017 1/2/2017	30	Triggered, and ODF Certified
	Intervention	2/20/2017	32	Triggered, and ODE Certified
Azawule		3/29/2017	3Z 20	Triggered, and ODF Certified
Dulo Debretebore Evenue	Control	3/10/2017	30	Triggered, and ODF Certified
Cana Machawasha	Control	3/27/2017	33	Inggered, and ODF Certilled
		3/31/2017	34	Non-Inggered
GIRIDI	Intervention	3/23/2017	31	Triggered, and ODF Certified
Hiruy Aba Aregay	Intervention	3/29/2017	34	Triggered, and ODF Certified
Kanat	Intervention	3/29/2017	33	Triggered, and ODF Certified
Kolay	Control	3/18/2017	31	Iriggered, and ODF Certified
Mahidere Mariam	Control	4/2/2017	34	Triggered, and ODF Certified
Mayinet	Intervention	3/25/2017	32	Triggered, and ODF Certified
Megendi	Control	3/31/2017	31	Triggered, and ODF Certified
Sahirna	Intervention	3/27/2017	31	Triggered, and ODF Certified
Selamko	Control	3/23/2017	30	Non-triggered
Tsegur	Control	3/18/2017	30	Triggered, and ODF Certified
Wewa	Intervention	4/2/2017	33	Triggered, and ODF Certified
Worken	Intervention	3/25/2017	30	Triggered, and ODF Certified
Wukro	Control	3/27/2017	33	Triggered, and ODF Certified
Fogera (N=12)				
Adis Betekristian	Intervention	3/12/2017	30	Triggered, but not vet ODF Certified
Alembur Zuria	Control	4/11/2017	32	Triggered, and ODF Certified
Anguko	Intervention	3/14/2017	30	Triggered and ODF Certified
Hagere Selam	Control	3/16/2017	30	Triggered, and ODF Certified
Kubar Abo	Control	3/10/2017	31	Triggered, and ODF Certified
Kuhar Micheal	Intervention	3/12/2017	33	Triggered, but not vet ODF Certified
Makisegnit Ketema	Control	3/12/2017	31	Triggered, and ODF Certified
Rib Gabrial	Control	1/12/2017	31	Triggered, but not yet ODE Certified
	Intervention	3/16/2017	30	Triggered, but not yet ODF Certified
N/aii Awuramba		3/10/2017	20	Triggered, but not yet ODF Certified
Woroto Zuria	Control	3/10/2017	30	Triggered, and ODF Certified
vvoreta Zuria		3/10/2017	<u>ა</u> ს აე	Triggered, and ODF Certified
zeng	intervention	3/14/2017	J∠	Inggerea, and ODF Certified

Table E.1. Intervention assignment of study clusters
Table E.2. Characteristics of previous and current models of CLTSH and the Andilaye intervention

	Previous demand-side models (e.g., CLTS, CLTBCHS)	Current CLTSH model (comparison arm)	Andilaye intervention (intervention arm)
Emphasis	Enabling community behavior change and triggering the desire for an ODF community rather than subsidies for demonstration of household toilets	Intensive interpersonal communication to cultivate improvements in sanitation and hygiene practices through community problem-solving and action	Integrating facial hygiene and NTD prevention messaging into the current CLTSH model to optimize potential health outcomes attributable to CLTSH
Messaging	 Improved sanitation to promote ODF communities Did not include positive motivators promoting trachoma-specific prevention and control behaviors 	 Improved sanitation and hygiene (primarily hand hygiene and safe water handling practices) to promote ODF communities, and peripherally attend to "E" component of the SAFE strategy for trachoma control Does not include positive motivators promoting trachoma-specific prevention and control behaviors 	 Improved sanitation and hygiene to promote ODF communities and attend to NTD control integrated, NTD-specific prevention and control behaviors (e.g., integration of NTD prevention messaging amongst the entire intervention group, integrated facial hygiene promotion into personal hygiene promotion in a sub-set of the intervention arm) The overarching intervention motto and related intervention functions work to improve individual and community-level agency (i.e., enhance self- and collective efficacy) Focuses on positive, community-oriented motivators of behavioral change, promotes achievable incremental improvements, and incorporates strategies that facilitate behavioral maintenance (i.e., prevention of behavioral slippage or relapse back to unimproved behaviors).
Implementation	 GoE implementation: Previous localized implementation in Ethiopia Discontinued, but serves as the foundation of the current CLTSH model 	 GoE implementation: FMoH scaling nationally, with phased roll-out and triggering first in communities GoE determines are "ripe" for change Local government and NGO partners follow communities until they achieve ODF, little (if any) subsequent follow-up to determine sustainability of results (access, utilization, and progression along the sanitation ladder) Study implementation: "Standard of care" comparison arm implementation, roll-out per GoE implementation plan 	 GoE implementation: The study team will work with GoE to test the <i>Andilaye</i> intervention in a sub-set of <i>gotts</i> and <i>kebeles</i> in Amhara (i.e. <i>Andilaye Impact Evaluation</i>) Longitudinal follow-up beyond ODF certification to gather data for sustainability analyses Study implementation: Intervention arm implementation, per study design Implementation and fidelity assessed at midterm, end line, and during longitudinal surveillance visits
Major outcomes	Absence of feces in the openODF communities	 Absence of feces in the open ODF communities Improved hand hygiene Increase in hygienic water handling/storage practices 	 Absence of feces in the open ODF communities Improved hand hygiene Increase in hygienic water handling/storage practices Improved understanding of NTD prevention Improved facial cleanliness Sustained latrine infrastructure quality and use
Major inputs	Software trainingCapacity building	 Whole system in the room and action planning (WSR) Software training Capacity building 	 WSR Software training Capacity building Enhanced behavior change communication messaging resources that integrates WASH- related, NTD-preventive behavior change components to that of a holistic WASH approach

Appendix F. Implementation of *Andilay*e intervention behavioral change catalyzing and maintenance

Activity	Aim	Facilitator	Target audience		Essential tools	Complementary tools	Duration
District-level ca	talyzing activities						
Sensitizing and action planning workshop	To orient key stakeholders to the <i>Andilaye</i> intervention and engage them in intervention action planning so as to generate buy-in and foster an enabling environment in which the intervention can be implemented.	• <i>Andilaye</i> team - Emory Ethiopia	 Relevant woreda officers Regional COWASH representative Woreda COWASH representative ARHB core process owner Kebele administrators Representatives from other agencies/organizations working on sanitation and hygiene study kebeles 	✓ ✓	Overview of <i>Andilaye</i> presentation <i>Woreda</i> action planning tool	 ✓ Andilaye 'Gobez!' Flipbook ✓ Andilaye Household Goal Card ✓ Community mobilization and commitment event – video and photo examples 	1 day (1x)
Skills-based training of the trainers for HEWs, CHC HEWs supervisors, Woreda officials	To provide skills-based training to HEWs/CHC HEWs supervisors/ <i>Woreda</i> officials on HH-level intervention activities, supportive supervision, and on-the-job- training so HEWs can, in turn, effectively train WDALs on the implementation of HH- level activities and provide supportive supervision.	• <i>Andilaye</i> team - Emory Ethiopia	 Woreda hygiene and sanitation officer Woreda health officer HEWs CHC HEWs supervisors 		Training guide Andilaye 'Gobez!' Flipbook Andilaye Household Goal Card Andilaye Household Monitoring Matrix Card Training handout Supportive supervision checklists HEW action planning tool	 ✓ Community mobilization and commitment event – video and photo examples 	2 days (1x)
Training of community conversation facilitators District-level ma	To provide comprehensive facilitator training to selected <i>gott</i> and <i>kebele</i> stakeholders on the 'community conversations' group-level intervention activity.	 Andilaye team - Emory Ethiopia Woreda hygiene and sanitation officer Woreda health officer 	• Selected <i>gott/kebele</i> stakeholders (see activity 'Whole system in the room and action planning')	✓ ✓ ✓	Training guide Andilaye CC Facilitator Flipbook Kebele action planning tool	 ✓ Andilaye Household Goal Card ✓ Andilaye 'Gobez!' Flipbook 	2 days (1x)
Adaptive	To leverage monitoring data to facilitate	Andilave team - Emory	Woreda bygiene and	~	Supportive supervision	✓ Andilave 'Gobez!'	1 day (1x)
management workshops	evidence-based, controlled, and documented operational-specific modifications during critical program moments (i.e., "change gates"). To improve intervention outcomes and resource management by learning from monitored program outcomes.	Ethiopia	 sanitation officer <i>Woreda</i> health officer CHC HEWs supervisors 	V	checklists Monitoring data	Flipbook ✓ <i>Andilaye</i> Household Goal Card	

Activity	Aim	Facilitator	Target audience	Essential tools	Complementary tools	Duration
District-level ma	aintenance activities (continued)					
Skills-based refresher training for supervisors and facilitators	To reinforce previously acquired knowledge and skills and address trainer/facilitator turnover. Prior experience indicates that such trainings serve to sustain actor motivation and further strengthen capacity.	• <i>Andilaye</i> team - Emory Ethiopia	 <i>Woreda</i> hygiene and sanitation officer <i>Woreda</i> health officer HEWs CHC HEWs supervisors 	 ✓ Training guide ✓ Andilaye 'Gobez!' Flipbook ✓ Andilaye Household Goal Card ✓ Andilaye Household Monitoring Matrix Card ✓ Training handout ✓ HEW action planning tool 	 ✓ Monitoring data 	2 days (1x)
Community-leve	el <u>catalyzing</u> activities					
Skills-based training of Women's Development Army Leaders	To provide skills-based training to WDALs on HH-level intervention activities, as detailed in the training of the trainers for HEWs, CHC HEWs supervisors, <i>Woreda</i> officials.	• HEWs	• WDALs	 ✓ Training handout ✓ Andilaye 'Gobez!' Flipbook ✓ Andilaye Household Goal Card ✓ Andilaye Household Monitoring Matrix Card ✓ HEW action planning tool 	 Community mobilization and commitment event video and photo examples 	2 days (1x)
Whole system in the room and action planning	To engage key community stakeholders, orient them to the <i>Andilaye</i> intervention, and facilitate their involvement in intervention action planning. This participatory approach aims to generate buy-in and foster an enabling environment (i.e., social opportunity) in which the <i>Andilaye</i> intervention can be supported and effectively implemented for a "strong, caring, healthy community."	 <i>Kebele</i> administrator Key influencers from the study <i>gott</i> Relevant <i>woreda</i> officer <i>Andilaye</i> team – Emory Ethiopia 	 <i>Kebele</i> manager HEWs Agricultural extension workers School directors WDAL Religious leaders Influential elders and other influential people from the <i>gott</i> Other potential change agents, as identified by the HEWs and <i>woreda</i> officials 	 ✓ Kebele action planning tool 	 ✓ Andilaye 'Gobez!' Flipbook ✓ Andilaye Household Goal Card ✓ Andilaye CC Facilitator Flipbook ✓ Community mobilization and commitment event – video and photo examples 	0.5 day (1x)

Activity	Aim	Facilitator	Target Audience	Essential Tools	Complementary Tools	Duration
Community-leve	el catalyzing activities (continued)					
Community mobilization and commitment event	To shift social norms (including community by-laws and sanctions), and improve action knowledge, barrier identification and planning, and attitudes regarding targeted WASH behaviors through a form of contextually appropriate and interactive edutainment.	 <i>Kebele</i> administrator Other <i>kebele</i> Stakeholders (identified during 'Whole system in the room and action planning') Cultural performance group HEWs, CHC HEWs supervisors Relevant <i>woreda</i> officers 	Community members	 ✓ Playbill ✓ WASH role play scripts ✓ Andilaye commitment banner 		0.5 day (1x)
Community-leve	el <u>maintenance</u> activities					
Skills-based review meetings and refresher trainings for Women's Development Army Leaders	To reinforce previously acquired knowledge and skills, address WDAL turnover, and review successes and address challenges faced in implementing counseling visits with caregivers. Prior experience indicates that such trainings serve to sustain actor motivation and further strengthen capacity.	• HEWs	• WDALs	 ✓ Monitoring data ✓ Andilaye 'Gobez!' Flipbook ✓ Andilaye Household Goal Card ✓ Andilaye Household Monitoring Matrix Card ✓ Supportive supervision checklist 		2 days (1x)
Household graduation and maintenance celebration events	To hold a celebration to reward households/communities and to motivate one another to sustain well-earned gains. Celebrations foster motivation and also help reinforce improved behaviors and promote healthy competition among communities.	 <i>Kebele</i> administrator HEWs WDALs 	Community members		 ✓ Andilaye commitment banner ✓ Andilaye Household Goal Card 	0.5 days (1x)
Cross- fertilization visits	To provide an opportunity to share experiences across different intervention communities – to address common implementation bottlenecks, propose solutions, and share perspectives on preliminary behavior change and health outcomes.	CHC HEWs supervisors	• WDALs • HEWs		 ✓ Andilaye commitment banner 	0.5 days (1x)

Activity	Aim	Facilitator	Target audience	Essential tools	Complementary tools	Duration
Group-level cata	alyzing activity					
Community conversation s	To change factual beliefs and attitudes, enhance action knowledge, improve perceptions of capability, identify and make plans to overcome barriers, and shift social norms regarding targeted behaviors through community group dialogue. To carry out demonstrations that address key factors associated with both breaking unimproved practices and adopting improved ones.	HEWsCC facilitators	 Heads of HH (i.e. men) Religious leaders Mother-in-laws WDALs 1 to 5 Other groups identified during 'Whole system in the room and action planning' activity 	 ✓ Andilaye CC Facilitator Flipbook 	 ✓ Andilaye Household Goal Card 	1 hour /session* (1x/2 weeks) *3 sessions per CC group
Group-level mai	intenance activity					
Community conversation s	To generate community-level dialogue regarding nuanced issues associated with maintenance of improved practices and barriers thereof through a follow-up round of community group dialog. To carry-out demonstrations related to behavioral maintenance issues.	HEWsCC facilitators	 Heads of HH (i.e. men) Religious leaders Mother-in-laws WDALs 1 to 5 Other groups identified during 'Whole system in the room and action planning' activity 	 ✓ Andilaye CC Facilitator Flipbook 	 ✓ Andilaye Household Goal Card 	1 hour /session* (1x/2 weeks) *3 sessions per CC group
Household-leve	I <u>catalyzing</u> activity					
Counseling visits with caregivers	To provide personalized counseling to caregivers to equip them with the knowledge, skills, and motivation necessary to develop improved WASH practices. To foster action capacity, self- efficacy, and barrier planning so caregivers maintain the improved WASH practices.	• WDALs	 Primary audience: Primary caregiver of HH Secondary audience: Other HH members present during the HH visits 	 ✓ Andilaye Household Goal Card ✓ Andilaye 'Gobez!' Flipbook ✓ Andilaye Household Monitoring Matrix Card 		30 min per HH (1x/month)
Household-leve	l <u>maintenance</u> activity					
Focused behavioral maintenance counseling visits	To provide continuous follow-up to households such that the house graduates from counseling related to initial adoption of improved behaviors to counseling related to behavioral maintenance skills. These visits will progressively focus on specific barrier identification and planning skills so the caregiver can maintain his/her improved WASH practices, especially as personal setbacks, systematic shocks, and other obstacles arise.	• WDALs	 Primary audience: Primary caregiver of HH Secondary audience: Other HH members present during the HH visits 	 ✓ Andilaye Household Goal Card ✓ Andilaye 'Gobez!' Flipbook ✓ Andilaye Household Monitoring Matrix Card 		30 min per HH (1x/month)

Appendix G. Cost of implemented intervention components

Intervention Components	Avg. Unit Cost		Unit Definition	Quantity	Total Co	ost
Andilaye tools and materials	ETB	USD		#	ETB	USD
Andilaye illustration development	972.22	42.27	Illustration	54	52,500.05	2,282.61
Printing Andilaye 'Gobez!' Flipbook	260.00	11.30	Flipbook	400	104,000.00	4,521.74
Printing of Andilaye Household Goal Card	46.01	2.00	Goal card	3,583	164,869.36	7,168.23
Printing of Andilaye Monitoring Matrix Card	25.47	1.11	Matrix card	3,436	87,532.00	3,805.74
Printing of Andilaye CC Facilitator Flipbook*	-	-	Flipbook	-	-	-
In kind motivators for Andilaye implementers*	-	-	Item	-	-	-
Woreda level planning workshops						
Woreda sensitizing and action planning workshops	9,006.39	391.58	Workshop	3	27,019.16	1,174.75
Woreda adaptive management workshop (round 1)*	-	-	Workshop	-	-	-
Woreda adaptive management workshop (round 2)*	-	-	Workshop	-	-	-
Kebele level planning workshops						
Whole system in the room and action planning	172.74	7.51	Workshop	25	4,318.44	187.76
Andilaye community events						
Community mobilization and commitment event	8,659.03	376.48	Event	25	216,475.75	9,411.99
Training of the trainers for Andilaye counseling caregiver visits						
Training of the trainers (HEWs/CHCs/Woreda Officials)	38,539.10	1,675.61	Workshop	4	154,156.41	6,702.45
Training of the trainers (HEWs/CHCs/Woreda Officials) mop up	25,000.13	1,086.96	Workshop	1	25,000.13	1,086.96
Refresher training for HEWs/CHCs*	-	-	Workshop	-	-	-
Training of the facilitators for Andilaye counseling caregiver visits						
Training of WDALs	3,432.74	149.25	Workshop	25	85,818.59	3,731.24
Review meeting and refresher training with WDALs (round 1)*	-	-	Workshop	-	-	-
Review meeting and refresher training with WDALs (round 2)*	-	-	Workshop	-	-	-
Training of the facilitators for Andilaye community conversations						
Training of the CC facilitators*	-	-	Workshop	-	-	-
Review meeting and refresher training with CC facilitators*	-	-	Workshop	-	-	-
Field travel for Andilaye team facilitators						
Transportation for Andilaye team facilitators	2,469.60	107.37	Day	75	185,220.20	8,053.05
Per diem for Andilaye team facilitators	339.11	14.74	Day	264	89,526.00	3,892.43
Total cost of implemented intervention components, as of July 2018					1,196,436.09	52,018.96

* Intervention activities that are scheduled for implementation after reporting date (i.e. August 2018-February 2019)

Appendix H. Process data on the mplementation of *Andilaye* intervention activities to date

Activity	Intervention implementation fidelity To document the extent to which the intervention was delivered as planned	Participation and dose response To assess the extent of engagement of participants with the intervention	Context To understand which environmental factors may influence intervention implementation
District-level catalyzing activi	ities		
Sensitizing and action planning workshop	 Dates of implementation: September – October 2017 Proportion of study <i>woredas</i> with workshop: 100% (3/3) Proportion of study <i>woredas</i> with workshops agenda completed as planned: 100% (3/3) Proportion of study <i>woredas</i> with a completed action planning tool: 100% (3/3) 	 Participation: Total number of participants: 47 Proportion of workshops with at least one representative from the <i>Woreda</i> health office in attendance: 100% (3/3) Proportion of workshops with <i>Woreda</i> hygiene and sanitation officer in attendance: 100% (3/3) Proportion of intervention <i>kebeles</i> with administrator in attendance: 96% (24/25) COWASH representative attended at least one workshop: Yes Amhara Regional Health Bureau representative attended at least one workshop: Yes Proportion of participants that endorsed completed action plan: 89% (42/47) 	 Context: A <i>kebele</i> administrator from one non-intervention <i>kebele (wonjeta)</i> was invited by mistake instead a <i>kebele</i> administrator from an intervention <i>kebele (wondata)</i> was missed (not invited). The names of these two <i>kebeles</i> sound very similar and perhaps similarity of names may have created the mismatch while extending invitations. The sensitization workshop that was completed in <i>Bahir Dar</i> city (the regional capital city) was the one with the highest number of stakeholders' turnout including regional level government and non-government organizations (NGOs) stakeholders. A NGO stakeholder invited in <i>Farta woreda</i> did not attend the workshop. Some workshop participants (mainly regional representatives and <i>woreda</i> administrators) from each of the 3 <i>woredas</i> did not attend the afternoon session of the workshop; this might have impacted the number of action plan endorsing stakeholders.
Skills-based training of the trainers for HEWs, CHC HEWs supervisors, Woreda officials	 Dates of implementation: December 2017 – January 2018 Proportion of study woredas with training: 100% (3/3) Proportion of study woredas with training agenda completed as planned: 100% (3/3) Percent of study woredas with a completed action planning tool: 100% (3/3) 	 Participation: Total number of participants: 96 Proportion of trainings with at least one representative from the <i>Woreda</i> health office in attendance: 100% (4/4) Proportion of trainings with <i>Woreda</i> hygiene and sanitation officer in attendances: 75% (3/4) Proportion of intervention <i>kebele</i> health centers with CHC HEWs supervisors in attendance: 100% (18/18) Proportion of intervention <i>kebeles</i> with at least one HEW in attendance: 100% (25/25) Dose response: Percentage of participants to reiterate the motto of the <i>Andilaye</i> intervention correctly in post assessment: 65% Percentage of participants to list the 3 targeted behavioral domains of the <i>Andilaye</i> intervention correctly in post assessment: 72% Percentage of participants to list the 11 targeted WASH practices of the <i>Andilaye</i> intervention correctly in post assessment: 70% Proportion of intervention <i>kebeles</i> with representatives that endorsed completed action plan: 100% (25/25) 	 Context: Four skills-based training of the trainers (ToT) were completed in the three study <i>woredas</i>. Two ToTs were completed in Farta <i>woreda</i>, the <i>woreda</i> with the highest number of intervention kebeles. Since there is only one <i>Woreda</i> hygiene and sanitation officer in one woreda, the maximum number of <i>Woreda</i> hygiene and sanitation officer who can attend the four ToTs are three. The Farta <i>Woreda</i> hygiene and sanitation officer only attended one of the two ToTs in that <i>woreda</i>. Participants were encouraged to refer to their personal note when responding "Dose response" questions.
Training of community conversation facilitators	Scheduled for August 2018		

District-level maintenance act	ivities		
Adaptive management workshops	Scheduled for November 2018		
Skills-based refresher training for supervisors and facilitators	Scheduled for September 2018		
Community-level catalyzing ac	ctivities		
Skills-based training of Women's Development Army Leaders	 Dates of implementation: January – February 2018 Proportion of study <i>kebeles</i> with training: 100% (25/25) Proportion of study <i>kebeles</i> with training agenda completed as planned: 100% (25/25) 	 Participation: Proportion of intervention <i>kebele</i> WDALs in attendance: 96% (212/220) Dose response: Percentage of WDALs (in groups) to reiterate the motto of the <i>Andilaye</i> intervention correctly in post assessment: 80% Percentage of WDALs (in groups) to list the 3 targeted behavioral domains of the <i>Andilaye</i> intervention correctly in post assessment: 97% Percentage of WDALs (in groups) to list the 11 targeted WASH practices of the <i>Andilaye</i> intervention correctly in post assessment: 97% 	 Context: Training of WDALs was planned to primarily done by HEWs with help from CHC HEWs supervisors and CHC heads, but the involvement of CHC HEWs supervisors and CHC heads was minimal in some kebeles. In some of the Kebeles, Emory Bahirdar staff members had acted as trainers and trained WDALs together with HEWs.
Whole system in the room and action planning	 Dates of implementation: January – March 2018 Proportion of study <i>kebeles</i> with workshop: 100% (25/25) Proportion of study <i>kebeles</i> with workshop completed as planned: 100% (25/25) Percent of study <i>kebeles</i> with a completed action planning tool: 100% (25/25) 	 Post assessment. 95% Participation: Proportion of workshops with <i>kebele</i> administrator facilitating: 96% (24/25) Proportion of workshops with a <i>Woreda</i> official facilitating: 8% (2/25) Proportion of workshops with a key community-level influencers facilitating: 44% (11/25) Proportion of workshops with the <i>kebele</i> manager in attendance: 88% (22/25) Proportion of workshops with the <i>kebele</i> HEWs in attendance: 100% (25/25) Proportion of workshops with the <i>kebele</i> school directors in attendance: 60% (15/25) Proportion of workshops with the <i>kebele</i> wDALs in attendance: 44% (11/25) Dose response: Proportion of study <i>kebeles</i> with community conversation facilitators identified in the action plan: 100% (25/25) Proportion of study <i>kebeles</i> with a master of ceremony for the 'Community mobilization and commitment event' identified in the action plan: 100% (25/25) 	 Context: Expectation of payment (per diem) associated with whole system in the room and action planning (WSR) activity was the main issue and in some cases a clear disappointment for participants of WSR activities. Although the WSR is one to two hours activity, participants (a composition of community members and salaried government workers) seem to expect some kind of payment (per diem) and when, at the end of the WSR, they found out that there is no payment, some participants in most <i>kebeles</i> expressed their clear disappointment. HEWs, who invited WSR participants, were outspoken in expressing their dissatisfaction. Money was allotted to invite participants to coffee and tea during WSR activities, however, since the time of the WSR was fasting season (Easter fasting season) most participants mentioned that they do not eat or drink before noon, therefore, the WSR money assigned for coffee/tea could not be used to buy coffee/tea. In some <i>kebeles</i>, at the end of the WSR activities, the WSR facilitator (Emory Ethiopia) would give the money allotted for coffee/tea to one of the participants break fasting, but in some <i>kebeles</i>, when participants see the amount of money (150 to 200 birr for 12 people) they would make fun of it and in few cases openly expressed their disappointment. WSR participants' expectation of payment (per diem) appeared to impact Emory Ethiopia staff members' interest to facilitate WSR activities, especially WSR completed at the end.

Community mobilization and commitment event	 Dates of implementation: March – April 2018 Proportion of study <i>kebeles</i> with event completed: 100% (25/25) Proportion of study <i>kebeles</i> with event completed as planned: 100% (25/25) 	 Participation: Average estimated attendance of community members per event: 325 Proportion of events with coordinating committee in attendance: 36% (9/25) Proportion of events with <i>kebele</i> HEW in attendance: 96% (24/25) Proportion of events with <i>kebele</i> WDALs in attendance: 92% (23/25) Dose response: Proportion of study <i>kebeles</i> that determined practices no longer deemed to be acceptable by the community at the end of the event: 96% (24/25) Proportion of study <i>kebeles</i> that determined improved behaviors at the end of the event: 92% (23/25) Proportion of study <i>kebeles</i> that determined improved behaviors at the end of the event: 96% (24/25) 	 Context: Setting community bylaws and, most importantly, determining regulations (deciding penalties) for people who violated the bylaws were the most challenging activities of the event. In few Kebeles HEWs expressed their concern on determining regulations (setting penalties) considering the intense political atmosphere and public protests in the region in the last few years. Mobilizing community members to come to the event was another challenge. The payment (per diem) issue associated with WSR activities might have impacted the involvement of some event coordinating committee members and masters of ceremonies who were identified during WSR activities. Since four performance groups were hired to perform in the 25 intervention <i>kebeles</i>, there was a clear difference in the capability of the groups to complete their job as planned. The two performance groups hired from Farta <i>woreda</i> and performed in <i>kebeles in Farta</i> and Fogera <i>woredas</i> were, by far, better than the two performance groups who were hired from Bahir Dar. 			
		(24/25)	performing their duties as planned. Since similar payment was decided to be paid for performance groups hired from Farta woreda and Bahir Dar Zuria woreda, it was not possible to get best performance groups in Bahir Dar Zuria <i>woreda</i> that would agree to work with the already established payment because of more opportunities for performance groups in Bahir Dar than in Farta.			
Community-level maintenance	e activities					
Skills-based review meetings and refresher trainings for Women Development Army Leaders	Scheduled for October 2018					
Household graduation and maintenance celebration events	Ongoing (following Skills-based refresher training for supervisors and facilitators)					
Cross-fertilization visits	Ongoing (following Skills-based refresher tra	aining for supervisors and facilitators)				
Group-level catalyzing activity	/					
Community conversations	Ongoing (following Training of community co	onversation facilitators)				
Group-level maintenance activ	vity					
Community conversations	Ongoing (following Training of community co	onversation facilitators)				
Household-level catalyzing ac	tivity					

Counseling visits with caregivers	 Dates of implementation: January 2018 ongoing Proportion of study <i>kebeles</i> with at least WDALs conducting at least on visit: 100% (25/25) 	 Participation: Proportion of study intervention households with an <i>Andilaye</i> Household Goal Card observed during follow-up 1: 54% (419/780) Dose response: See 'Impact analysis and results of the key evaluation questions' section 	 Context: Gaps were observed in some of the intervention <i>kebeles</i> related to fidelity of household level <i>counseling visits with caregivers</i>, these includes: In some <i>kebeles</i>, most WDALs performed few household counselling visits (frequency) and there are cases where WDALs performed counselling visits only to few households (not to all household members in their development network), and indications for possible reductions in household visits because of the coming of the rainy season and the associated field work (farming). Very limited or in some <i>kebeles</i> complete absence of supportive supervision and on the job training given to WDALs by HEWs. In most <i>kebeles</i>, absence of support and supervision to HEWs by CHC HEWs supervisors and/or CHC heads. 			
Household-level maintenance	activity					
Focused behavioral maintenance counseling visits Ongoing (following Skills-based review meetings and refresher trainings for Women Development Army Leaders)						

Appendix I. Dates of key activities and one year follow-up data collection

	(1) Date of skills-based training of	(2) Date of community mobilization and	(3) Date of follow-up 1 data	Days between	Days between (2)
Kebele	WDALs	commitment event	collection	(1) and (3)	and (3)
Bahir Dar Zuria (N=8)				Avg. 66	Avg. 22
Debranta	1/26/18	3/14/2018	4/3/2018	67	20
Fereswoga	1/24/18	3/14/2018	4/3/2018	69	20
Gonbat	1/26/18	3/14/2018	4/3/2018	67	20
Wondatta	1/24/18	3/14/2018	4/3/2018	69	20
Dehina Mariam	2/23/18	4/1/2018	4/25/2018	61	24
Chenta	2/20/18	4/3/2018	4/25/2018	64	22
Lata	2/23/18	4/3/2018	4/25/2018	61	22
Wogelsa	2/21/18	4/5/2018	4/25/2018	63	20
Farta (N=11)				Avg. 68	Avg. 24
Wewa	2/1/18	3/16/2018	4/13/2018	71	28
Azawure	2/2/18	3/18/2018	4/13/2018	70	26
Mayinet	2/2/18	3/18/2018	4/16/2018	73	29
Sahirna	2/2/18	3/21/2018	4/15/2018	72	25
Ayiva Niva	2/2/18	3/25/2018	4/15/2018	72	21
Giribi	2/14/18	3/25/2018	4/16/2018	61	22
Ayidde	2/12/18	3/28/2018	4/18/2018	65	21
Kanat	2/1/18	3/28/2018	4/18/2018	76	21
Worken	2/16/18	3/28/2018	4/18/2018	61	21
Arga	2/16/18	3/30/2018	4/20/2018	63	21
Hiruy Aba Aregay	2/16/18	3/30/2018	4/20/2018	63	21
Fogera (N=6)				Avg. 71	Avg. 24
Adis Betekristian	1/23/18	3/6/2018	3/28/2018	64	22
Zeng	1/23/18	3/6/2018	3/28/2018	64	22
Tiwa	1/25/18	3/8/2018	3/30/2018	64	22
Woji Awuramba	1/25/18	3/8/2018	3/30/2018	64	22
Kuhar Micheal	1/12/18	3/8/2018	4/1/2018	79	24
Anguko	1/12/18	3/16/2018	4/13/2018	91	28
All kebeles (N=25)				Avg. 68	Avg. 23

Appendix J. Supplemental results

Table J.1. Baseline characteristics

	Overall		Intervention		Coi	ntrol
Latrine coverage	Ν	%	Ν	%	Ν	%
Households with latrine with smooth and cleanable slab/floor (%)	1582	15.0	790	11.9	792	18.1
% of present latrines that were fully constructed	1034	47.0	507	46.2	527	47.8
Sanitation facility operation and maintenance	Ν	%	Ν	%	Ν	%
Latrine cleaned during the last seven days (%)	1036	22.7	506	23.7	530	21.7
Have ever fixed anything that became broken, damaged, our worn out on this latrine since its original construction (%)	1027	21.5	504	19.6	523	23.3
Respondent reports latrine facility is correctly working (operating) (%)	1037	93.3	507	92.9	530	93.6
Facility observed to require obvious repair (%)	1033	74.6	505	76.2	528	73.1
The latrine was observed to be serviceable ^a (%)	1034	84.7	506	83.4	528	86.0
Cleaning agents for washing latrine were observed inside or near the latrine (%)	1033	2.42	505	3.0	528	1.9
Feces observed on floor/slab, or other place in the latrine aside from pit (%)	1033	52.2	505	54.7	528	49.8
Latrine characteristics	Ν	%	Ν	%	Ν	%
Stagnant water present over the floor / latrine slab (%)	1033	5.2	505	5.9	528	4.6
Pan, slab, or floor was discolored (e.g., yellow, green) (%)	1033	56.3	505	55.1	528	57.6
Presence of flies in latrine (%)	1033	80.3	505	82.0	528	78.6
Presence of cleaning agents for washing latrine (%)	1033	2.4	505	3.0	528	1.90
Presence of feces on floor/slab or other place in the latrine (%)	1033	52.2	505	54.7	528	49.8
Evidence latrine is used for storage or other non-sanitation-related purpose (%)	1033	8.3	505	10.5	528	6.25
Presence of well-worn path to latrine (%)	1033	92.3	505	92.3	528	92.2
Presence of fresh feces on/in the pit or pan (%)	1029	80.8	505	81.7	528	79.9
Is the pit that is in use full or close to being full (%)	1029	20.5	505	21.3	528	19.8
Presence of anal cleansing item in, near latrine (%)	1031	29.2	504	28.8	527	29.6
Presence of odor from stool or urine in the latrine (%)	1033	81.1	503	82.2	526	80.1
Presence of leaves, spider webs, rubbish, other dirt in latrine (%)	1033	22.2	503	23.8	526	20.6
Wet latrine floor (%)	1033	32.8	505	29.7	528	35.8
Presence of water available near or inside latrine for hand washing (%)	1032	2.71	505	4.0	527	1.5
Presence of cleansing agent near or inside latrine for hand washing (%)	1033	2.2	505	3.2	528	1.3
Water available inside or near latrine for flushing or self-cleansing (%)	1032	3.9	504	4.6	528	3.2
Latrine utilization	Ν	%	Ν	%	Ν	%
Urination	Ν	%	Ν	%	Ν	%
Respondent urinated in/or near surface water	1588	5.5	792	4.4	796	6.5
Head of household urinated in/near surface water (%)	1001	5.6	496	6.1	505	5.2
Ages 1-14 urinated in/near surface water (%)	2746	10.1	1390	8.0	1356	12.3
Defecation	N	%	Ν	%	Ν	%
Respondent openly defecated during last 2 days (%)	1589	46.5	793	47.8	796	45.2
% of OD that was near surface water	738	5.4	378	4.8	360	6.1
Respondent always exclusively used a latrine for defecation during last 7 days (%)	1588	47.9	793	46.3	795	49.6
Respondent's primary place of defecation changes over the course of the year (%)	1587	26.9	791	24.9	796	28.9

HoH's primary place of defecation was OD during last 2 days (%)	1481	39.2	746	41.6	735	36.9
HoH defecated in any latrine during last 2 days (%)	1278	53.9	640	56.1	638	51.7
HoH openly defecated during last 2 days (%)	1204	61.4	596	63.9	608	58.9
% of OD that was near surface water	572	6.3	300	6.7	272	5.9
HoH always exclusively used a latrine for defecation during last 7 days (%)	1237	35.9	612	32.5	625	39.2
HoH's primary place of defecation changes over the course of the year (%)	1542	24.9	771	23.2	771	2636
Ages 4-17 primary place of defecation was OD during last 2 days (%)	3633	43.6	1826	44.6	1807	42.6
Ages 4-17 defecated in any latrine during last 2 days (%)	3281	51.5	1663	51.1	1618	51.9
Ages 4-17 openly defecated during last 2 days (%)	3093	62.0	1549	62.0	1544	62.1
% of OD that was near surface water	1614	11.7	825	9.5	789	14.1
Ages 4-17 always exclusively used a latrine for defecation during last 7 days (%)	3069	34.0	1519	33.8	1550	34.1
Ages 4-17 primary place of defecation changes over the course of the year (%)	3633	23.1	1845	20.3	1818	25.9
Evidence of open defecation (i.e., human feces) in/near HH compound (%)	1589	57.3	793	60.2	796	54.4
Given you have a latrine, HH members have used this latrine for 3 or more days during the LAST 7 days (%)	1037	86.4	507	85.8	530	87.0
	N	mean	N	mean	N	mean
	IN	(SE)	IN	(SE)	IN	(SE)
Given you have a latrine, number of HHs that USED	1037	1.52	507	1.40	530	1.62
THIS HH latrine in last 7 days mean (SE)	1007	(0.84)	507	(0.11)	550	(0.12)
Environmental sanitation	Ν	%	Ν	%	Ν	%
Animal husbandry/other HH sanitation	Ν	%	Ν	%	Ν	%
Respondent has animal herding or other animal husbandry responsibilities ^b (%)	1589	89.4	793	90.8	701	88.1
Head of household has animal herding or other animal husbandry responsibilities ${}^{\mathfrak{b}}$ (%)	1589	89.2	793	90.8	796	87.7
Observed animal feces present in the compound (%)	1589	84.5	793	87.3	796	81.8
Solid waste management	Ν	%	Ν	%	Ν	%
Solid waste was not observed to have been left out in the open (%)	1589	26.8	793	26.0	796	27.6
Personal hygiene practices	Ν	%	Ν	%	Ν	%
Face cleaning practices	Ν	%	Ν	%	Ν	%
Yesterday the index child's face was cleaned (%)	1554	95.7	774	96.0	780	95.4
Yesterday, after index child's FACE was washed it was it wiped dry with a cloth (%)	1458	34.2	730	36.0	728	32.3
Yesterday, the respondent cleaned his/her face (%)	1586	98.3	792	98.2	794	98.4
Yesterday, the respondent wiped his/her face dry with a cloth such as a towel or apron (%)	1526	36.2	762	39.2	764	33.1
Facial cleanliness	Ν	%	Ν	%	Ν	%
Among index children ages 1-9 years	Ν	%	Ν	%	Ν	%
Ocular discharge is present (%)	1385	42.3	670	44.0	715	40.8
Wet nasal discharge is present (%)	1385	50.5	670	49.6	715	51.5
Dry nasal discharge is present (%)	1385	67.4	670	67.5	715	67.3
Dirt/dust/other debris is present (%)	1385	71.3	670	70.5	715	72.0
Handwashing practices	Ν	%	Ν	%	Ν	%
Yesterday, the index child's HANDS were washed (%)	1553	97.6	772	98.7	781	96.5
Yesterday, the index child's were washed with soap/ash/or soapy water used (%)	1516	38.5	761	35.9	755	41.1
The last time the index child defecated, he/she cleaned hands with water and soap, soapy water, or ash (%)	1553	26.0	773	23.2	780	26.7
Yesterday, the respondent washed his/her hands with water (%)	1587	97.3	793	96.5	794	98.1
Hand cleanliness	N	%	Ν	%	Ν	%
Respondent's finger nails clean on left hand (%)	1589	81.8	793	83.2	796	80.3

Respondent's finger nail clean on right hand (%)	1589	81.7	793	83.1	796	79.7
Respondent's finger pads clean on left hand (%)	1589	75.3	793	76.2	796	74.4
Respondent's Finger pads clean on right hand (%)	1589	75.1	793	75.8	796	74.4
Respondent's palm clean on left hand (%)	1589	68.2	793	68.9	796	67.6
Respondent's palm clean on right hand (%)	1589	68.6	793	69.2	796	68.0
Index child's finger nails clean on left hand (%)	1385	91.8	670	93.0	715	90.6
Index child's finger nail clean on right hand (%)	1385	90.8	670	92.5	715	89.2
Index child's finger pads clean on left hand (%)	1385	86.4	670	86.7	715	86.2
Index child's finger pads clean on right hand (%)	1385	86.3	670	86.9	715	85.7
Index child's palm clean on left hand (%)	1385	85.6	670	85.7	715	85.5
Index child's palm clean on right hand (%)	1385	85.6	670	85.2	715	85.9
Shoe wearing	Ν	%	Ν	%	Ν	%
Respondent wore shoes to walk to the latrine (%)	1040	50.1	508	46.5	532	53.6
Head of household is wearing shoes (%)	582	56.2	280	54.3	302	58.0
Washing station coverage	Ν	%	Ν	%	Ν	%
Observed water in at least one washing stations (%)	1244	18.3	617	15.9	627	20.7
Observed soap in at least some of the HW stations (%)	1244	4.3	617	4.7	627	3.8
Respondent-reported fomite and bathing practices	N	%	Ν	%	Ν	%
Fomite washing	Ν	%	Ν	%	Ν	%
Respondent washed the index child's clothes at least once during the last 7 days (%)	1585	86.6	791	86.5	794	86.8
Respondent washed their own clothes at least once during the last 7 days (%)	1587	78.7	793	80.1	794	77.3
Respondent washed towels at least once during the last 7 days (%)	1587	65.1	792	67.9	795	62.3
Respondent washed bedsheets at least once during the last 7 days (%)	1586	30.2	791	30.7	795	29.7
Respondent washed baby carrier at least once during the last 7 days (%)	1585	28.0	791	27.9	794	28.1
Respondent washed children's toys at least once during the last 7 days (%)	1546	1.8	778	1.2	768	2.3
Bathing practices	N	%	Ν	%	Ν	%
Index child bathed at least once during the last 7 days (%)	1572	76.3	783	76.0	789	76.7
Respondent bathed their body at least once during last 7 days (%)	1587	56.6	791	58.5	796	54.7
Respondent used surface water at least once for bathing during the last 7 days (%)	1587	56.8	791	58.5	796	54.7

^a Meaning the pit is not overflowing, and the floor provides a solid foundation over the latrine pit

Table J.2. Midline latrine coverage (key indicators in blue)

	Ov	verall In		Intervention		ntrol	PR (95% CI) ^a	PD (95% CI) ^b
	Ν	%	Ν	%	Ν	%		
Households with at least one latrine (%)	1496	67.7	751	67.8	745	67.5	1.01 (0.85, 1.22)	0.0098 (-0.11, 0.13)
Households with improved latrine (%) ^c	1485	29.8	743	31.0	742	28.4	1.09 (0.77, 1.55)	0.0254 (-0.08, 0.13)
Households with latrine with smooth and cleanable slab/floor (%)	1495	12.4	750	13.3	745	12.4	1.07 (0.60, 1.94)	0.0105 (-0.07, 0.09)
Households with fully constructed latrine (%)	1490	31.9	749	33.8	741	30.0	1.14 (0.86, 1.50)	0.0405 (-0.05, 0.13)
% of present latrines that were fully constructed	1006	47.2	507	49.9	499	44.3	1.12 (0.95, 1.32)	0.0529 (-0.03, 0.13)

^a We used log-linear binomial models to estimate prevalence ratios comparing the outcomes between the intervention and control arms. Models accounted the stratified design by including woreda indicator variables, (Kahan and Morris, 2012) and accounted for clustering within kebeles by using generalized estimating equations with robust standard errors. ^b We used similar linear regression models to estimate difference comparing the outcomes between the intervention and control arms. ^c Improved based on JMP definition; (World Health Organization and UNICEF, 2013) see Figure 4A for all latrine type categories.

	Ove	rall	Interv	ention	Cor	ntrol	PR (95% CI) ^a	PD (95% CI) ^b
	N	%	N	%	N	%		
Stagnant water present over the floor / latrine slab (%)	1010	7.0	507	7.5	503	6.6	1.05 (0.41, 2.66)	0.0032 (-0.06, 0.06)
Pan, slab, or floor was discolored (e.g., vellow, green) (%)	1011	58.2	508	53.4	503	63.0	0.86 (0.72, 1.04)	-0.0839 (-0.19, 0.02)
Presence of flies in latrine (%)	1011	77.7	508	72.4	503	82.9	0.87 (0.76, 1.00)	-0.1062 (-0.21,0.00)
Presence of drop hole cover in the latrine (%)	1011	19.5	508	23.8	503	15.1	1.56 (1.07, 2.28)	0.0851 (0.00, 0.16)
Among those with a drop hole, a cover was situated over drop hole (%)	197	70.6	121	68.6	76	75.0	0.90 (0.76, 1.08)	-0.0727 (-0.20, 0.50)
Presence of cleaning agents for washing latrine (%)	1010	7.1	508	9.7	502	4.6	1.99 (0.68, 5.89)	0.0426 (-0.03, 0.11)
Presence of feces on floor/slab or other place in the latrine (%)	1011	57.2	508	50.8	503	63.6	0.80 (0.68, 0.93)	-0.1287 (-0.22, -0.04)
Evidence latrine is used for storage or other non-sanitation-related purpose (%)	1011	8.7	508	9.5	503	8.0	1.07 (0.51, 2.23)	0.0053 (-0.05, 0.07)
Presence of well-worn path to latrine (%)	1011	93.4	508	93.7	503	93.0	1.00 (0.96, 1.04)	-0.0031 (-0.04, 0.03)
Presence of fresh feces on/in the pit or pan (%)	1010	80.1	508	78.0	502	82.3	0.95 (0.85, 1.06)	-0.0419 (-0.12, 0.05)
Is the pit that is in use full or close to being full (%)	1009	14.7	507	15.2	502	14.1	1.09 (0.73, 1.62)	0.0123 (-0.15, 0.07)
Presence of anal cleansing item in, near latrine (%)	1011	47.3	508	48.4	503	46.1	1.03 (0.83, 1.29)	0.0154 (-0.09, 0.12)
Presence of odor from stool or urine in the latrine (%)	1011	79.8	508	75.2	503	84.5	0.88 (0.79, 0.98)	-0.9782 (-0.18, -0.02)
Presence of leaves, spider webs, rubbish, other dirt in latrine (%)	1011	24.2	508	23.0	503	25.5	0.94 (0.71, 1.25)	-0.0158 (-0.08, 0.05)
Wet latrine floor (%)	1011	37.1	508	37.0	503	37.2	0.97 (0.72, 1.31)	-0.0102 (-0.12, 0.10)
Presence of water available near or inside latrine for hand washing (%)	1011	13.8	508	20.5	503	7.0	2.57 (1.03, 6.41)	0.1119 (-0.01, 0.23)
Presence of cleansing agent near or inside latrine for hand washing (%)	1011	3.8	508	6.5	503	1.0	5.75 (1.77, 18.70)	0.0606 (0.00, 0.12)
Water available inside or near latrine for flushing or self-cleansing (%)	1010	4.7	508	6.9	502	2.4	2.67 (1.12, 6.34)	0.0429 (0.00, 0.09)
Sanitation operation and maintenance	Ν	%	Ν	%	Ν	%		
Latrine cleaned during the last seven days (%)	1004	26.9	507	29.8	497	23.9	1.18 (0.81, 1.71)	0.0420 (-0.05, 0.14)
HH has added or improved anything on this latrine since its original construction (%)	1006	24.3	506	23.3	500	25.2	0.91 (0.72, 1.15)	-0.0223 (-0.08, 0.04)
Have ever fixed anything that became broken, damaged, our worn out on this latrine since its	1007	216	506	30.2	501	32.0	0.80 (0.71 1.12)	0.0355 (0.11.0.04)
original construction (%)	1007	51.0	500	30.2	501	32.9	0.09 (0.71, 1.13)	-0.0355 (-0.11, 0.04)
Is your latrine facility working (operating) correctly now? (%)	1010	94.0	508	95.1	502	92.8	1.02 (0.99, 1.06)	0.0226 (-0.01, 0.06)
Facility observed to require obvious repair (%)	1011	76.8	508	72.1	503	81.5	0.88 (0.79, 0.99)	-0.0958 (-0.18, -0.01)
The latrine was observed to be serviceable ^b (%)	1011	83.5	508	82.3	503	84.7	0.98 (0.88, 1.10)	-0.0158 (-0.11, 0.08)
Cleaning agents for washing latrine were observed inside or near the latrine (%)	1010	7.1	508	9.7	502	4.6	1.99 (0.68, 5.89)	0.0426 (-0.03, 0.11)

Table J.3. Midline latrine characteristics (key indicators in blue, significant findings bolded)

^a We used log-linear binomial models to estimate prevalence ratios comparing the outcomes between the intervention and control arms. Models accounted the stratified design by including woreda indicator variables, (Kahan and Morris, 2012) and accounted for clustering within kebeles by using generalized estimating equations with robust standard errors. ^b We used similar linear regression models to estimate difference comparing the outcomes between the intervention and control arms.

Table J.4. Midline latrine utilization (key indicators in blue)

	Ov	verall	Interv	Intervention		ntrol	PR (95% CI)ª	PD (95% CI) ^b
Urination	Ν	%	Ν	%	Ν	%		
Respondent urinated in/or near surface water	1486	4.5	745	3.9	741	5.1	0.75 (0.36, 1.58)	-0.0128 (-0.05, 0.02)
Head of household urinated in/near surface water (%)	663	3.6	350	2.9	313	4.5	0.56 (0.21, 1.51)	-0.0213 (-0.06, 0.02)
Ages 4-17 urinated in/near surface water (%)	2521	5.5	1283	5.5	1238	5.5	0.66 (0.42, 1.04)	-0.0071 (-0.05, 0.03)
Defecation	Ν	%	Ν	%	Ν	%		
Respondent's primary place of defecation was OD during last 2 days (%)	1496	34.6	751	34.5	745	34.6	0.98 (0.69, 1.41)	-0.0065 (-0.13, 0.12)
Respondent defecated in any latrine during last 2 days (%)	1491	44.9	749	47.3	742	42.5	1.12 (0.89, 1.41)	0.0521 (-0.05, 0.16)
Respondent openly defecated during last 2 days (%)	1496	41.9	751	42.3	745	41.5	1.01 (0.76, 1.35)	0.0057 (-0.12, 0.13)
% of OD that was near surface water	623	3.5	314	3.2	309	3.9	0.77 (0.30, 2.00)	-0.0097 (-0.45, 0.26)
Respondent always exclusively used a latrine for defecation during last 7 days (%)	1496	55.2	751	54.9	745	55.6	0.99 (0.80, 1.22)	-0.0054 (-0.12, 0.11)
Respondent's primary place of defecation changes over the course of the year (%)	1496	25.0	751	26.1	745	23.9	1.08 (0.80, 1.45)	0.0189 (-0.05, 0.09)
HoH's primary place of defecation was OD during last 2 days (%)	1082	34.5	560	36.3	522	32.6	1.09 (0.73, 1.63)	0.0307 (-0.11, 0.17)
HoH defecated in any latrine during last 2 days (%)	857	58.2	454	59.5	403	56.8	1.05 (0.83, 1.32)	0.0285 (-0.10, 0.16)
HoH openly defecated during last 2 days (%)	980	58.3	501	57.1	479	59.5	0.98 (0.78, 1.22)	-0.0139 (-0.14, 0.11)
% of OD that was near surface water	289	5.2	152	4.0	137	6.6	0.48 (0.14, 1.65)	-0.0416 (-0.12, 0.04)
HoH always exclusively used a latrine for defecation during last 7 days (%)	972	39.6	505	41.2	467	37.9	1.03 (0.75, 1.43)	0.0133 (-0.12, 0.15)
HoH's primary place of defecation changes over the course of the year (%)	1173	22.4	612	23.5	561	21.2	1.15 (0.80, 1.66)	0.0308 (-0.05, 0.11)
Ages 4-17 primary place of defecation was OD during last 2 days (%)	3498	39.9	1770	38.1	1728	41.8	0.97 (0.69, 1.36)	-0.0109 (-0.14, 0.02)
Ages 4-17 defecated in any latrine during last 2 days (%)	2987	51.1	1510	53.9	1477	48.2	1.16 (0.93, 1.44)	0.0753 (-0.04, 0.19)
Ages 4-17 openly defecated during last 2 days (%)	2869	52.5	1443	50.7	1426	54.3	0.98 (0.79, 1.22)	-0.0103 (-0.12, 0.10)
% of OD that was `near surface water	1218	8.0	596	9.1	622	6.9	1.26 (0.56, 2.86)	0.0194 (-0.05, 0.09)
Ages 4-17 always exclusively used a latrine for defecation during last 7 days (%)	2815	41.0	1422	43.9	1393	38.0	1.10 (0.83, 1.45)	0.0380 (-0.08, 0.15)
Ages 4-17 primary place of defecation changes over the course of the year (%)	3537	21.8	1778	23.5	1759	20.1	1.17 (0.87, 1.57)	0.0335 (-0.03, 0.09)
Evidence of open defecation (i.e., human feces) in/near HH compound (%)	1496	53.9	751	51.9	745	56.0	0.93 (0.77, 1.11)	-0.0420 (-0.14, 0.05)
Given you have a latrine, HH members have used this latrine for 3 or more days during the	1000	85 5	507	86.8	502	84 3		0 01/11 (-0 05 0 08)
LAST 7 days (%)	1009	00.0	507	00.0	302	04.5	1.01 (0.94, 1.10)	0.0141 (-0.03, 0.00)
Safely disposed of child feces (%)	873	46.4	438	43.8	435	45.0	0.97 (0.73, 1.27)	-0.0160 (-0.14, 0.11)
Sonitation charing	N	mean	NI	mean	NI	mean		difference
Sanitation sharing	IN	(SE)	IN	(SE)	IN	(SE)	-	(95% CI) ^c
Given you have a latrine, number of HHs that USED THIS HH latrine in last 7 days mean	1005	1.78	509	1.91	496	1.65	_	0 17 (-0 27 0 62)
(SE)		(0.14)		(0.23)		(0.13)		(0.2., 0.02)
Given HH has a latrine, number of people who used this latrine from ANOTHER HH	1001	1.10	505	0.94	496	1.27	-	-0.41 (-0.88, 0.06)
during last 7 days, not including your HH members		(0.14)		(0.17)		(0.22)		(, ,

^a We used log-linear binomial models to estimate prevalence ratios comparing the outcomes between the intervention and control arms. Models accounted the stratified design by including woreda indicator variables, (Kahan and Morris, 2012) and accounted for clustering within kebeles by using generalized estimating equations with robust standard errors. ^b We used similar linear regression models to estimate difference comparing the outcomes between the intervention and control arms. ^c We used similar linear regression models to estimate difference comparing the outcomes between the intervention and control arms. ^c We used similar linear regression models to estimate difference comparing the outcomes between the intervention and control arms.

Table J.5. Midline washing station coverage (key indicators in blue)

	Overall		II Intervention		Control		PR (95% CI) ^a	PD (95% CI) ^b	
	Ν	%	Ν	%	Ν	%			
Reported HH hand or facewashing station(s) (%)	1494	95.0	750	97.3	744	92.6	1.02 (0.99, 1.04)	0.0154 (-0.01, 0.04)	
Observed HH hand or facewashing station(s) with water (%)	1494	20.0	750	18.7	744	21.2	0.89 (0.66, 1.20)	-0.0238 (-0.09, 0.04)	
Observed HH hand or facewashing station(s) with soap (%)	1494	1.9	750	2.9	744	0.9	3.09 (1.07, 8.96)	0.0219 (-006, 0.05)	

^a We used log-linear binomial models to estimate prevalence ratios comparing the outcomes between the intervention and control arms. Models accounted the stratified design by including woreda indicator variables, (Kahan and Morris, 2012) and accounted for clustering within kebeles by using generalized estimating equations with robust standard errors. ^b We used similar linear regression models to estimate difference comparing the outcomes between the intervention and control arms.

Table J.6. Midline personal hygiene practices among respondents and children (key indicators in blue, significant findings bolded)

	Ove	rall	Interve	ntion	Con	trol	PR (95% CI) ^a	PD (95% CI) ^b
Respondent reported face cleaning practices	Ν	%	Ν	%	Ν	%		
Yesterday the index child's face was cleaned (%)	1460	93.8	731	95.1	729	92.5	1.02 (0.99, 1.06)	0.0218 (-0.01, 0.05)
Yesterday, after index child's FACE was washed it was it wiped dry with a cloth (%)	1281	39.6	646	40.3	635	38.9	1.03 (0.84, 1.27)	0.0118 (-0.07, 0.09)
Yesterday, the respondent cleaned his/her face (%)	1493	95.9	750	97.1	743	94.8	1.03 (1.00, 1.06)	0.0259 (0.00, 0.05)
Yesterday, the respondent wiped his/her face dry with a cloth such as a towel or	1210	110	695	16 1	662	12 1	1 00 (0 90 1 22)	0 0271 (0 05 0 12)
apron (%)	1540	44.0	005	40.4	005	43.1	1.09 (0.09, 1.32)	0.0371 (-0.03, 0.12)
Facial cleanliness	Ν	%	Ν	%	Ν	%		
Among all children ages 1-9 years	Ν	%	Ν	%	Ν	%		
Ocular discharge is present (%)	1441	37.7	682	37.4	759	37.9	0.97 (0.81, 1.16)	-0.0122 (-0.08, 0.06)
Wet nasal discharge is present (%)	1441	45.8	682	45.0	759	46.5	0.96 (0.82, 1.12)	-0.0183 (-0.09, 0.05)
Dry nasal discharge is present (%)	1441	55.3	682	55.1	759	55.5	0.99 (0.89, 1.10)	-0.0077 (-0.07, 0.05)
Dirt/dust/other debris is present (%)	1441	53.0	682	52.2	759	53.6	0.97 (0.83, 1.13)	-0.0151 (-0.10, 0.06)
Among index children ages 1-9 years	Ν	%	Ν	%	Ν	%		
Ocular discharge is present (%)	921	39.5	444	38.5	477	40.5	0.97 (0.81, 1.16)	-0.122 (-0.08, 0.06)
Wet nasal discharge is present (%)	921	47.2	444	45.7	477	48.6	0.96 (0.82, 1.12)	-0.0183 (-0.09, 0.05)
Dry nasal discharge is present (%)	921	59.1	444	56.8	477	61.2	0.99 (0.89, 1.10)	-0.0077 (-0.07, 0.05)
Dirt/dust/other debris is present (%)	921	54.4	444	52.5	477	56.2	0.97 (0.83, 1.13)	-0.0151 (-0.10, 0.06)
	NI	mean	NI	mean		mean	difference	
	N	(SE)	N	(SE)	N	(SE)	(95% CI) ^b	
Number of times a fly land on the index child's face during a 1 minute	021	3.1	111	3.3	177	2.9	0 /1 (-0 27 1 08)	
observation	921	(0.19)	444	(0.32)	4//	(0.22)	0.41 (-0.27, 1.00)	-
Respondent reported handwashing practices	Ν	%	Ν	%	Ν	%		
Yesterday, the index child's HANDS were washed (%)	1458	98.2	729	98.4	729	98.1	1.00 (0.99, 1.02)	0.0039 (-0.01, 0.02)
The last time the index child's hands were washed, soap/ash/or soapy water used (%)	1418	45.2	713	48.4	705	42.0	1.17 (0.95, 1.42)	0.0696 (-0.02, 0.16)

The last time the index child defecated, he/she cleaned hands with water and soap,	1116	27.2	707	12.2	700	21.0	1 /1 /1 00 1 93)	0 1285 (0 03 0 23)
soapy water, or ash (%)	1410	57.2	101	45.5	109	51.0	1.41 (1.03, 1.03)	0.1203 (0.03, 0.23)
Yesterday, the respondent washed his/her hands with water (%)	1494	98.7	751	99.1	743	98.3	1.01 (1.00, 1.02)	0.0080 (0.00, 0.02)
The last time the respondent washed he/she used soap/ash/soapy water (%)	1492	44.0	750	48.1	742	39.8	1.22 (1.00, 1.47)	0.0865 (0.00, 0.17)
The last time the respondent defecated, he/she cleaned hands with water and	1106	12.2	747	40.1	720	27 /	1 30 (1 03 1 66)	0 1153 (0 01 0 22)
soap (%)	1400	43.3	141	49.1	139	37.4	1.30 (1.03, 1.00)	0.1155 (0.01, 0.22)
The last time the respondent prepared food, he/she cleaned hands with water	1407	15.2	710	50.0	COF	20.6	1 20 (1 02 1 62)	0 1120 (0 01 0 22)
and soap before beginning food preparations (%)	1407	40.5	112	50.6	095	39.0	1.29 (1.02, 1.02)	0.1139 (0.01, 0.22)
Hand cleanliness								
Respondent's finger nails clean on left hand (%)	1496	74.1	751	71.0	745	77.3	0.92 (0.84 (1.02)	-0.0600 (-0.13, 0.01)
Respondent's finger nail clean on right hand (%)	1496	73.1	751	69.6	745	76.5	0.91 (0.82, 1.01)	-0.0659 (-0.14, 0.01)
Respondent's finger pads clean on left hand (%)	1496	65.4	751	62.9	745	67.9	0.93 (0.84, 1.03)	-0.0485 (-0.12, 0.02)
Respondent's Finger pads clean on right hand (%)	1496	65.5	751	62.5	745	68.6	0.91 (0.82, 1.01)	-0.0595 (-0.13, 0.01)
Respondent's palm clean on left hand (%)	1496	56.0	751	52.9	745	59.2	0.90 (0.78, 1.03)	-0.0605 (-0.14, 0.02)
Respondent's palm clean on right hand (%)	1496	56.6	751	53.7	745	59.5	0.91 (0.79, 1.04)	-0.0549 (-0.13, 0.02
Index child's finger nails clean on left hand (%)	967	86.5	466	86.1	501	86.8	1.00 (0.94, 1.06)	0.0010 (-0.05, 0.05)
Index child's finger nail clean on right hand (%)	967	86.0	466	85.2	501	86.8	0.98 (0.92, 1.04)	-0.0197 (-0.07, 0.03)
Index child's finger pads clean on left hand (%)	967	77.0	466	74.3	501	79.6	0.96 (0.88, 1.04)	-0.0345 (-0.10, 0.03)
Index child's finger pads clean on right hand (%)	967	76.4	466	74.7	501	78.0	0.97 (0.89, 1.06)	-0.0242 (-0.09, 0.04)
Index child's palm clean on left hand (%)	967	72.2	466	70.2	501	74.1	0.95 (0.86, 1.04)	-0.0375 (-0.11, 0.03)
Index child's palm clean on right hand (%)	967	71.4	466	69.5	501	73.1	0.95 (0.86, 1.04)	-0.0387 (-0.11, 0.03)
Midline personal hygiane practices not targeted by intervention						-		
Shoe weeking								
Shoe wearing	1010	EAE	500	FO 1	502	56.0		0.0407 (0.47, 0.07)
Respondent wore shoes to wark to the latime (%)	612	54.5 60.0	209 272	52.1	340	50.9 61.2	0.91(0.73, 1.13) 0.00(0.81, 1.21)	-0.0497(-0.17, 0.07)
Children 0.17 wearing shoes	2161	50.0	1013	JO.0 10 1	11/18	51 5	0.99(0.01, 1.21) 0.07(0.86 1.10)	-0.0044 (-0.13, 0.12)
Fomite washing	2101	50.4	1015	+5.1	1140	01.0	0.37 (0.00, 1.10)	
Respondent washed the index child's clothes at least once during the last 7 days (%)	1487	85.3	747	84.3	740	86.4	0.98 (0.92, 1.05)	-0.0132 (-0.07, 0.04)
Respondent washed their own clothes at least once during the last 7 days (%)	1496	77.7	751	76.6	745	78.8	0.98(0.90, 1.06)	-0.0160 (-0.08, 0.05)
Respondent washed towels at least once during the last 7 days (%)	1492	65.8	750	63.3	742	68.3	0.93 (0.84, 1.03)	-0.0469 (-0.11, 0.02)
Respondent washed bedsheets at least once during the last 7 days (%)	1493	37.1	749	37.4	744	36.8	1.01 (0.81, 1.26)	0.0033 (-0.08, 0.09)
Respondent washed baby carrier at least once during the last 7 days (%)	1488	31.5	748	32.6	740	30.4	1.07 (0.89, 1.29)	0.0210 (-0.04, 0.08)
Respondent washed children's toys at least once during the last 7 days (%)	1492	0.9	751	0.7	741	1.2	0.55 (0.20, 1.49)	-0.0056 (-0.02, 0.00)
Bathing practices								
Index child bathed at least once during the last 7 days (%)	1466	77.6	734	76.3	732	78.8	0.97 (0.91, 1.04)	-0.0209 (-0.07, 0.03)
Respondent bathed their body at least once during last 7 days (%)	1495	60.7	751	62.6	744	58.9	1.06 (0.95, 1.18)	0.0341 (-0.03, 0.10)
Respondent used surface water at least once for bathing during the last 7 days (%)	1235	20.3	612	14.2	623	26.3	0.54 (0.31, 0.96)	-0.1244 (-0.26, 0.01)

^a We used log-linear binomial models to estimate prevalence ratios comparing the outcomes between the intervention and control arms. Models accounted the stratified design by including woreda indicator variables,(Kahan and Morris, 2012) and accounted for clustering within kebeles by using generalized estimating equations with robust standard errors. ^b We used similar linear regression models to estimate difference comparing the outcomes between the intervention and control arms.

Table J.7. Midline animal husbandry/hygiene, animal feces, and solid waste management

	Ove	Overall		ention	n Control		Control		PR (95% CI) ^a	PD (95% CI) ^b
Animal husbandry / other HH sanitation	Ν	%	Ν	%	Ν	%				
Respondent has animal herding or other animal husbandry responsibilities ^b (%)	1493	90.4	749	91.6	744	89.1	1.04 (0.98, 1.09)	0.0314 (-0.01, 0.08)		
Head of household has animal herding or other animal husbandry responsibilities ^b (%)	1232	91.0	632	92.3	600	89.7	1.05 (0.99, 1.11)	0.0437 (-0.01, 0.10)		
Observed animal feces present in the compound (%)	1496	83.4	751	82.7	745	84.0	0.99 (0.88, 1.10)	-0.0103 (-0.10, 0.08)		
Animal feces/waste not left out in open in compound (%)	1496	51.1	751	52.9	745	49.3	1.06 (0.80, 1.40)	0.03 (-0.12, 0.17)		
Solid waste management	Ν	%	Ν	%	Ν	%				
Solid waste was not observed to have been left out in the open (%)	1496	33.2	751	35.4	745	30.9	1.14 (0.79, 1.63)	0.0425 (-0.08, 0.16)		

^a We used log-linear binomial models to estimate prevalence ratios comparing the outcomes between the intervention and control arms. Models accounted the stratified design by including woreda indicator variables, (Kahan and Morris, 2012) and accounted for clustering within kebeles by using generalized estimating equations with robust standard errors. ^b E.g., taking care of cows, bulls, oxen, goats, sheep, horses, mules, donkeys, or chickens.

Table J.8. Midline water practices not targeted by intervention

	C	verall	Intervention			Control		
Water storage	Ν	mean (SE)	Ν	mean (SE)	Ν	mean (SE)	difference (95% CI) ^a	
Number of containers used to store water	1489	3.7 (0.08)	748	3.7 (0.13)	741	3.7 (0.11)	0.07 (-0.22, 0.36)	-
Number of small-necked containers used to store water	1486	2.5 (0.07)	748	2.4 (0.09)	738	2.5 (0.10)	-0.03 (-0.29, 0.23)	-
Water treatment	Ν	%	Ν	%	Ν	%	PR (95% CI) ^b	PD (95% CI) ^b
Did you do anything to treat THIS water? (%)	1495	8.3	751	9.7	744	6.9	1.36 (0.76, 2.44)	0.0256 (-0.03, 0.08)
Respondent reported water insecurity	Ν	%	Ν	%	Ν	%	PR (95% CI) ^b	PD (95% CI) ^b
Water insecure for any of the four reported insecurity measures (%)	1420	36.4	716	35.2	704	37.6	0.89 (0.64, 1.25)	0.0424 (-0.17, 0.08)

^a We used linear regression models to estimate difference comparing the outcomes between the intervention and control arms. Models accounted the stratified design by including woreda indicator variables, (Kahan and Morris, 2012) and accounted for clustering within kebeles by using generalized estimating equations with robust standard errors. ^b We used similar log-linear binomial models to estimate prevalence ratios comparing the outcomes between the intervention and control arms.

Figure J.1. Box and whisker plots for respondent-reported mental health outcomes

A. Medians and distributions of anxiety, depression, and emotional distress scores, shown by treatment arm. Scores greater than 1.75 indicate positive status for related outcomes



 B. Medians and distributions of WHO-5 mental well-being scores, shown by treatment arm.
 Scores below 13 indicate poor mental wellbeing

