



Andilaye Impact Evaluation: The impact of enhanced, demand-side sanitation and hygiene promotion on sustained behavior change and health in Amhara, Ethiopia

May 19, 2020

The Andilaye intervention was developed by Emory University, in partnership with the Federal Ministry of Health, the Amhara Regional Health Bureau, and relevant Zonal Health Departments and Woreda Health Offices. Funding for the project is provided by World Bank, International Initiative for Impact Evaluation (3ie), and Children's Investment Fund Foundation (CIFF)



Principal Investigator: Matthew C. Freeman, PhD, MPH

Local Co-PI: Abebe Gebremariam Gobezeayehu, MD

Study Managers: Maryann G. Delea, PhD, MPH
Jedidiah S. Snyder, MPH

Study Team: Mulusew Belew, MSc
Resom Berhe, MPH
Tina Braccio, MPH
Frederick Goddard, MS
Molly Linabarger, MPH
Katie Micek, MPH
Siraj Muhammed, MPH
Ashlin Rakhra, MPH
Gloria Sclar, MPH
Hiwote Solomon, MPH
Yihenew Tesfaye, MS
Mulat Woreta, BA, PGD
Melkamu Zegeye, MPH
Kassahun Zewudie, MPH

Co-Investigators: Tenagnework Antefe, BSc
Bethany Caruso, PhD, MPH
Thomas Clasen, PhD, JD
Maryann Delea, PhD, MPH
Joshua V. Garn, PhD, MS
Craig Hadley, PhD
Kenneth Maes, PhD

Enumerators: Yeworkwuha Abay, Mantegbosh Abebe, Selamawit Abebe, Tigist Abebe, Rosa Abesha, Mahider Adamu, Balemlaye Addisu, Adanech Admasu, Tirusew Alayu, Adisalem Arega, Yalemwork Asaye, Destaw Asnakew, Yalemwork Ayanew, Ayalnesh Belay, Asayech Bimrew, Tigist Bitew, Tiruzer Engidaw, Yeserash Gashaw, Natsenat Gebretsadkan, Woyneshet Genetu, Tewodaj Gizachew, Tibeltalech Mihiret, Yehizbalem Minale, Senait Mulualem, Eleni Nebiyu, Elsabet Seyoum, Mulubirhan Shitu, Sewunet Tadesse, Beza Tesfaye, Rahel Tsegaye, Sintayehu Wasihun, Maritu Yibrie

Artist: Yemi Yetneberk

Acknowledgments: Funding for the project was provided by World Bank, International Initiative for Impact Evaluation (3ie), 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 the Federal Democratic Republic of Ethiopia Ministry of Health's NTD Coordinator (including former Coordinator Biruk Kebede); 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.

Project websites: <http://www.freemanresearchgroup.org/andilaye>
[World Bank project profile](#)

Table of contents

Table of contents	i
Table of figures	i
Table of tables	i
List of acronyms	ii
Executive summary	iii
1. Introduction	1
1.1. Study rationale	1
1.2. Ethiopia context	2
1.3. Study aim	2
1.4. Research question and hypothesis	2
1.5. Objectives	2
1.6. Policy context and implications for policy and practice	3
1.7. Key partners	3
2. Study overview	5
2.1. Study design	5
2.2. Impact evaluation	6
2.3. Process evaluation	7
2.4. Timeline	7
2.5. Study setting	8
2.6. Context	8
2.7. Funding	8
2.8. Ethics approval	9
3. Formative research and intervention design	12
3.1. Formative research	12
3.2. Intervention design	12
3.3. Description of intervention	13
3.4. Target audiences	14
3.5. Intervention activities	14
3.6. Key behavior change tools utilized by the <i>Andilaye</i> intervention	15
3.7. Intervention delivery	22
4. Impact evaluation: methods	26
4.1. Sample size	26
4.2. Eligibility criteria	26
4.3. Recruitment	27
4.4. Random allocation	28
4.5. Data collection and data management methods	29
4.6. Human subjects and ethics	30
4.7. Balance of treatment arms at baseline	31
4.8. Analytical methods	31
5. Intervention implementation and process evaluation	36
5.1. Implementation	36
5.2. Process data collection results	36
5.3. Implementation of district-level activities	37
5.4. Implementation of community-level activities	38
5.5. Implementation of group-level activities	40

5.6.	Implementation of household-level activities	40
5.7.	Qualitative support of implementation challenges and successes	41
5.8.	Photos of <i>Andilaye</i> implementation	49
6.	Impact evaluation: results.....	54
6.1.	Survey results.....	54
6.2.	WASH coverage.....	54
6.3.	WASH-related behaviors.....	55
6.4.	Respondent-reported diarrhea.....	56
6.5.	Mental well-being	56
6.6.	Heterogeneity across sub-groups	57
7.	Discussion	63
7.1.	Interpretation of study findings.....	63
7.2.	Holistic WASH programming and the HEP	64
7.3.	Mental well-being	64
7.4.	WASH behavior change.....	65
7.5.	Intervention design and implementation	66
7.6.	Study strengths	67
7.7.	Study limitations	68
7.8.	Policy influence	69
8.	Policy-relevant findings and recommendations	71
	References.....	74
	Appendices.....	80
	Links to resources found in report.....	80
	Appendix A – Supplementary resources.....	80
	Appendix B – Supplementary process evaluation results.....	80
	Appendix C – Supplementary impact evaluation results	80

Table of figures

Figure 1. <i>Andilaye</i> timeline	10
Figure 2. Map of study <i>woredas</i> : Bahir Dar Zuria [1], Fogera [2], and Farta [3].....	11
Figure 3: <i>Andilaye</i> logic model, summarizing the theory of change on which the <i>Andilaye</i> intervention is grounded	23
Figure 4: Diagram summarizing the <i>Andilaye</i> intervention	24
Figure 5: Study flow diagram	35
Figure 6. <i>Andilaye</i> intervention compliance for key behavior change activities reported by household respondents from intervention <i>kebeles</i>	44
Figure 7. Endline distribution of select WASH indicators	62

Table of tables

Table 1. <i>Andilaye</i> intervention behavioral change catalyzing and maintenance activities.....	25
Table 2. Baseline data.....	33
Table 3. Summary of process data	46
Table 4. Endline levels of the WASH indicators of primary interest.....	58
Table 5. Endline respondent-reported diarrheal outcomes.....	59
Table 6. Endline respondent-reported sanitation insecurity	60
Table 7. Endline respondent-reported water insecurity	60
Table 8. Endline respondent-reported mental health outcomes.....	61

List of acronyms

3ie	International Initiative for Impact Evaluation
ARHB	Amhara Regional Health Bureau
CHC	Cluster Health Center
CIFF	Children’s Investment Fund Foundation
CLTSH	Community-led total sanitation and hygiene
CRT	Cluster-randomized trial
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
HIP	Hygiene Improvement Program
HoH	Head of household
IRB	Institutional Review Board
JMP	WHO/UNICEF Joint Monitoring Programme
MDA	Mass drug administration
NTD	Neglected tropical disease
ODF	Open defecation-free
RANAS	The Risks, Attitudes, Norms, Abilities, Self-Regulation approach for systematic behavior change
RCT	Randomized, controlled trial
SIEF	Strategic Impact Evaluation Fund
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

Introduction: Poor water, sanitation, and hygiene (WASH) behaviors are key drivers of infectious disease transmission and adverse mental well-being. While WASH is seen as a critical enabler of health, there are important knowledge gaps related to the content and delivery of effective, holistic WASH programming. Corresponding impacts of WASH on mental well-being are also underexplored. There is a need for more robust implementation research that yields information regarding whether and how community-based, demand-side interventions facilitate progressive and sustained adoption of improved sanitation and hygiene behaviors and downstream health impacts. The aim of the *Andilaye* Trial - Amharic for “togetherness/integration” - was to use behavioral theory and evidence from formative research to inform the design of a novel, holistic, community-based WASH intervention (i.e., “*Andilaye*”) and evaluate its impact on sustained behavior change and mental well-being.

Study overview: Emory University and its consortium partners conducted a three-year study in rural and peri-urban Amhara, Ethiopia. The *Andilaye* Trial consisted of three phases: (1) formative research and intervention design, (2) intervention implementation and process evaluation, and (3) impact evaluation. The study’s primary research question was: To what extent can *Andilaye*’s enhanced, demand-side sanitation and hygiene intervention impact sustained behavior change and mental well-being, above and beyond community-led total sanitation and hygiene (CLTSH)? We hypothesized that individuals in communities randomized to receive the *Andilaye* intervention are more likely to sustainably adopt improved sanitation and hygiene behaviors and demonstrate gains in mental well-being compared to individuals in communities randomized to the standard of care CLTSH programming. The study was funded by the World Bank’s Strategic Impact Evaluation Fund (SIEF), the International Initiative for Impact Evaluation (3ie), and the Children’s Investment Fund Foundation (CIFF). The trial was registered on [clinicaltrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT03075436) (NCT03075436).

Formative research and intervention design: We developed a theoretically-informed, evidence-based behavioral intervention called *Andilaye*. The purpose was to develop an intervention that incorporated holistic WASH messages, was demand-side focuses, and could be delivered at scale within the Ethiopian Health Extension Programme (HEP). The intervention focused on behavioral maintenance (not only behavior change), and was designed to complement prevailing programs, specifically CLTSH, and messages from the HEP. We developed the intervention to be delivered through the government structures. *Andilaye*’s intervention motto - “*Together we can be a strong, caring, healthy community*” - and related intervention components offered aspirational messages that emphasized the need for collective action to make positive change in the community. The intervention focused on three behavioral themes, informed by formative research: (1) sanitation, (2) personal hygiene, and (3) household environmental sanitation. Within these themes were eleven specific behaviors and practices that were targeted by the intervention. Intervention activities operated at four levels: (1) district, (2) community, (3) group, and (4) household. Key activities, amongst others, included community mobilization and commitment events, community conversations, and household counseling visits with caregivers – all of which were guided by behavior change tools that were informed by our formative research with illustrations produced by an artist based in Ethiopia.

Impact evaluation: methods: We randomly selected and assigned 50 sub-districts (*kebeles*) from three purposively selected districts (*woredas*); half to receive the *Andilaye* intervention, and half the standard of care sanitation and hygiene programming (i.e., CLTSH). During baseline (March-April, 2017), 1,589 eligible households (at least one child aged 1-9 years and consent provided to participate in the study) were randomly selected and enrolled into the cluster-randomized trial from study *kebeles*. At baseline and follow-up household visits, we collected data on an array of behavioral factors, potential moderators (e.g., water and sanitation insecurity, collective efficacy), and our primary study outcomes: sanitation and hygiene behaviors and mental well-being.

Intervention implementation and process evaluation: The fidelity of action planning workshops and trainings at the district and community-levels were high – as these activities were facilitated or co-facilitated by the *Andilaye* team. Women’s Development Army Leaders (WDALs) and their supervisors (i.e., Health Extension Workers [HEWs]) were trained to facilitate monthly counseling visits with households in their catchment area. However, at endline, only 59% of caregivers from study-enrolled households allocated to the intervention arm reported receiving *Andilaye* counseling visits. Over the course of the *Andilaye* Trial, we observed indications that additional resource considerations need to be addressed when determining whether this intervention approach can and should go to scale. The *Andilaye* intervention was implemented in collaboration with the *Woreda* Health Offices, HEWs, WDALs, and other community change agents in the 25 *kebeles* randomly selected to the *Andilaye* intervention arm. While supportive supervision considerations were acknowledged and incorporated into the design of the *Andilaye* intervention, these requirements did not go above and beyond what is expected of the HEP. However, due to a number of reasons, and as verified in our process evaluation, there was limited support extended to HEWs as designed in the HEP. The *Andilaye* team increased its role in facilitating and monitoring the implementation of activities, as needed, when local government officials did not take on this role. However, this limited integration into the HEP may have resulted in low participation, adherence, and compliance of *kebele* and *woreda* level stakeholders to the *Andilaye* intervention.

Impact evaluation: results: At endline, 62% of the 1,472 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 endline (prevalence ratio [PR]=0.99; 95% CI: 0.82, 1.21). There was no difference in the prevalence of improved latrines (PR=1.14; 95% CI: 0.82, 1.60) or in the prevalence of households with fully constructed latrines (PR=1.15; 95% CI: 0.87, 1.53) when comparing intervention and control arms. All measures of latrine utilization were similar as well. This includes indicators of urination, defecation (both for respondents and other household members), disposal of child feces, and sanitation sharing. Overall, 40% of respondents’ primary place of defecation during the last two days was in the open, and only 46% 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, which was similar between the intervention and control arms. All other environmental sanitation measures were also similar between the intervention and control arms.

Presence of hand and facewashing stations were reported and observed in 98% of households, although presence of water, and presence of soap were observed in only 20% and 2.2% of stations, respectively. Respondent-reported handwashing was similar in the intervention arm compared to the control arm. Respondents reported washing their hands with soap/ash/soapy water 44% of the time;

they reported washing their hands after defecation 49% of the time; and they reported washing their hands before food preparation 51% of the time.

The prevalence of anxiety, depression, and emotional distress among respondents was lower at endline 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. The intervention did not reduce diarrhea in the intervention arm compared to the control arm. Reported blood in the stool over the last seven days was, however, lower among all children aged 0-10 years in the intervention arm compared to the control arm (PR=0.47; 95% CI: 0.23, 0.93).

Discussion: We did not find that our intervention led to improvements in our primary outcomes, including mental well-being and sustained behavior change. However, the potential effectiveness of the intervention could not be evaluated as fidelity was not maintained after training was delivered by the *Andilaye* team. Few rigorously evaluated WASH interventions are delivered directly through government structures, and these limited results of poor intervention fidelity are consistent with several other large-scale WASH effectiveness studies. Findings from our trial may not only shed light on some of the factors contributing to poor dose delivery in *Andilaye* intervention communities, but also poor dose delivery of the HEP and CLTSH programming more broadly. While CLTSH has fostered sanitation and hygiene improvements in Ethiopia, evidence of behavioral slippage, or regression to unimproved practices in communities previously declared open defecation free, exists. Other limitations of CLTSH, such as its focus on disgust, poor triggering, and over-saturation of HEWs, have been documented. We employed rigorous formative research and social and behavioral theory to develop *Andilaye*, a scalable intervention designed to address these issues and complement existing service delivery within Ethiopia's HEP. Limited integration of *Andilaye* activities into the HEP likely resulted in minimal impact on sustained behavior change and mental well-being. However, evidence from this trial may help address knowledge gaps related to scalable alternatives to CLTSH and inform sanitation and hygiene programming and policy in Ethiopia and beyond.

Key outcomes:

	Overall	Intervention	Control
Intervention	%	%	%
Communities receiving CLTSH triggering or retriggering	----	0	0
Communities receiving the <i>Andilaye</i> intervention	----	100	0
<i>Andilaye</i> compliance	%	%	%
Reported awareness of community mobilization and commitment	----	22	----
Reported attendance of community mobilization and commitment	----	18	----
Reported awareness of community conversation	----	46	----
Reported attendance of community conversation	----	28	----
Reported receiving initial counseling visit	----	59	----
Reported receiving follow-up counseling visit	----	43	----
Attrition	%	%	%
Loss to follow up	7.4	6.3	8.4
Sanitation	%	%	%
Households with at least one latrine	61.6	61.2	62.0
Households with improved latrine	32.5	34.6	30.6
Households with fully constructed latrine	30.9	33.0	28.7
Respondent's primary place of defecation was OD during last 2 days	39.5	40.2	38.8
Respondent defecated in any latrine during last 2 days	45.5	45.8	45.3
Personal hygiene	%	%	%
Reported HH hand or facewashing station(s)	98.0	98.3	97.7
The last time the respondent washed he/she used soap/ash/soapy water	44.0	46.0	42.0
The last time the respondent defecated, he/she cleaned hands with water and soap, substitute	49.0	51.9	46.1
The last time the respondent prepared food, he/she cleaned hands with water and soap, substitute before beginning food preparations	51.0	53.6	48.5
Household environmental sanitation	%	%	%
Animal feces/waste not observed out in open in compound	53.8	56.4	51.2
Mental well-being	%	%	%
High Anxiety	23.5	22.2	24.8
High Depression	15.4	14.0	16.9
High Emotional distress	15.2	14.0	16.4
Poor well-being	26.4	25.2	27.8

Policy-relevant findings:

- **Theory to practice:** The use of behavioral theory and a structured approach to intervention development – that incorporates stakeholder feedback – yields important guidance in selecting behaviors to target, identifying leading indicators of behavior change, and informing the program’s approach.
- **Behavioral maintenance:** To prevent behavioral slippage, there is a need for approaches that complement CLTSH and focus on behavioral maintenance.
- **WASH-NTD programming:** 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.
- **Incremental improvements:** A focus on small, incremental improvements in WASH practices and facilities may be viewed as more achievable by program participants, particularly in low resource settings, and as such, may garner greater success.
- **Positive motivators:** Negative affective motivators may not be the most appropriate or effective drivers of change (especially in the Ethiopian context), and may actually erode mental well-being.
- **Over-extension of HEWs:** Although over-extension of HEWs was addressed through the engagement of additional community change agents for group-level activities and household-level activities, there is a need for more integration of activities and support to HEWs by Woreda Health Offices and CHCs for *all* health programs.
- **Gaps in supportive supervision to HEWs and WDALs:** Findings from our trial may not only shed light on some of the factors contributing to poor dose delivery in *Andilaye* intervention communities, but also poor dose delivery of the HEP and CLTSH programming more broadly.
- **Strengthening the WDAL structure:** While *Andilaye* demonstrated progress in improving the activity, function, and capacity of WDALs to act as model households, there is a need to further strengthen the WDAL system and explain what duties and responsibilities they should have as a volunteer.
- **Mental well-being and insecurity:** Quantifying the relationships between WASH improvements and these non-traditional outcomes and impacts can inform programs and policies that facilitate health equity.
- **Collective efficacy:** The development and application of collective efficacy scales allow implementers to better design and target community-level interventions, and examine the role of collective efficacy in the effectiveness of community-based WASH programming.
- **Behavioral antecedents:** The development and application of indicators addressing intermediate behavioral antecedents may yield usable tools that could be used for future WASH evaluations.

1. Introduction

1.1. Study rationale

Poor water, sanitation, and hygiene (WASH) are key drivers of infectious disease transmission and result in adverse mental and social well-being (Freeman et al., 2017, Sclar et al., 2017, Sclar et al., 2018, Wolf et al., 2018). Diarrhea accounts for an estimated 1.4 million deaths annually (Lozano et al., 2012, Pruss-Ustun et al., 2014) and nearly 20% of all under-5 deaths in low-income settings (Boschi-Pinto et al., 2008). Deficiencies in WASH are also a major contributor of neglected tropical diseases (NTDs) (Freeman et al., 2013, WHO, 2019). Over one billion people are at risk of soil-transmitted helminthiasis, which leads to nearly five million disability adjusted life years (DALYs), and schistosomiasis leads to a further two million DALYs (Pullan et al., 2014, Murray et al., 2013). Trachoma, the leading infectious cause of blindness (Resnikoff et al., 2004), is precipitated by repeated infection with bacteria, which is often perpetuated by poor hygiene (Stocks et al., 2014). These infections are environmentally mediated (Prüss-Ustün et al., 2016), and are largely attributed to inadequate WASH (Strunz et al., 2014, Grimes et al., 2014).

Many WASH programs focus only on measures of infectious diseases or growth of young children to assess programmatic impact and public health relevance. This narrow focus is at odds with the World Health Organization (WHO)'s definition of health as "a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity" (WHO, 2002). Poor WASH conditions are associated with adverse mental health outcomes (Sclar et al., 2018). Stress and depression, particularly amongst women, are linked with poor sanitation access (Caruso et al., 2018). Improvements in women's mental health, however, would likely require more than physical access to sanitation facilities, and may also entail changes in gendered norms that contribute to women's sanitation insecurity (Caruso et al., 2017b). Evidence also suggests associations between water and sanitation insecurity and mental well-being (Stevenson et al., 2016, Caruso et al., 2018), though most data focusing on the role of WASH in improving social and mental well-being is observational (Sclar et al., 2018). Quantifying the relationships between WASH improvements and these non-traditional outcomes and impacts can inform programs and policies that facilitate health equity.

While WASH-related research generally focuses on the individual, communities are often a more critical point of investigation. Due to the dynamics of infectious disease transmission, the effects of WASH interventions often depend on community level coverage and uptake (Garn et al., 2017a). Financing and maintenance of certain water and sanitation interventions may also depend on minimum levels of social capital and social cohesion (Cameron et al., 2019, Person et al., 2017, Alam et al., 2017, Bisung et al., 2014). Evidence suggests that correct, consistent, and sustained uptake of WASH interventions—other critical drivers of protective effects—depend in part on a variety of social constructs, such as collective efficacy, descriptive and social norms, and social identity (Harter et al., 2018, Delea, 2019). Nevertheless, research on these inter-personal behavioral factors, their influence on collective WASH behaviors, and health and development effects are comparatively limited. Consequently, there is inadequate evidence about how these behavioral factors should be considered in developing, implementing, and evaluating WASH programming and policy.

1.2. Ethiopia context

Despite steady reductions in open defecation since 2000, Ethiopians have some of the poorest access to basic water and sanitation, and one of the highest levels of access inequity (UNICEF/WHO, 2017). The Ethiopia Federal Ministry of Health's (FMOH) HEP, and its accompanying community-led total sanitation and hygiene (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 to 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 Services Package being delivered via the HEP, and its accompanying CLTSH module, are currently being scaled throughout Ethiopia (Wang et al., 2016).

1.3. Study aim

The aim of the *Andilaye* Trial - Amharic for “togetherness/integration” - was to use behavioral theory and evidence from formative research to inform the design of a novel holistic community-based WASH intervention (i.e., “*Andilaye*”) and evaluate its impact on sustained behavior change and mental well-being.

1.4. Research question and hypothesis

The study's primary research question was: **To what extent can *Andilaye*'s enhanced, demand-side sanitation and hygiene intervention impact sustained behavior change and mental well-being, above and beyond CLTSH?** We hypothesized that individuals in communities randomized to receive the *Andilaye* intervention were more likely to sustainably adopt improved sanitation and hygiene behaviors and demonstrate gains in mental well-being compared to individuals in communities randomized to the standard of care CLTSH programming.

1.5. Objectives

As outlined below, there are several documented limitations of CLTSH in Ethiopia, and WASH programming more broadly. The objective of the *Andilaye* Trial was to generate evidence to fill gaps in knowledge related to demand-side sanitation and hygiene programming.

1. CLTSH, the key approach to improving sanitation coverage and utilization in Ethiopia, has facilitated considerable changes in coverage of basic sanitation. However, some of these gains have not been sustained, and incremental improvements to improve sanitation (i.e., progress up the sanitation ladder) have not been widely promoted or achieved (UNICEF/WHO, 2017).
2. WASH programs, more broadly, have focused on catalyzing initial behavior change, and have placed little, if any emphasis on the habituation of improved behaviors and behavioral maintenance. Such approaches have fostered behavioral slippage, or regression back to unimproved behaviors and practices, and poor sustainability of behavioral outcomes and potential health impacts (Vaz Nery et al., 2019).

3. CLTSH largely focuses on leveraging shame to change norms around open defecation, but these negative affective motivators may not be the most appropriate or effective drivers of change, and may actually erode mental well-being. As indicated by our formative research, focus on negative affective motivators, poor facilitation of initial triggering, and a lack of post-triggering follow-up has left many communities with negative impressions of CLTSH initiatives.
4. Behaviors and facilities promoted by existing programs are aspirational but require considerable effort and/or capital investment to achieve. A focus on small, incremental improvements in WASH practices and facilities may be viewed as more achievable by program participants, particularly in low resource settings, and as such, may garner greater success.
5. HEWs charged with implementing CLTSH have many responsibilities, few tools, and little capacity to continually reinforce messages (Snel and Jacimovic, 2014). Although Cluster Health Centers 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 and development sectors, namely WASH and those vertical programs involved in the control and elimination of NTDs, prevent the integration and harmonization of NTD and WASH behavior change initiatives (Waite et al., 2016). For instance, FMOH's current CLTSH programming focuses on handwashing with soap at key times, yet overlooks the opportunity of promoting routine facewashing despite the high prevalence of trachoma in the country.
7. While the focus on diarrheal disease prevention and growth faltering have driven investments in WASH, recent evidence suggests that, in sub-Saharan Africa, basic improvements may not be enough to impact these health outcomes (Null et al., 2018, Humphrey et al., 2019). However, factors contributing to the influence of water, sanitation, and hygiene on other important health outcomes, such as mental well-being, remain under-studied.

1.6. Policy context and implications for policy and practice

Data from the *Andilaye* Trial was used to: (1) identify ways in which WASH-related behavior change components can be mainstreamed into the FMOH's HEP; (2) determine the effectiveness of an enhanced holistic demand-side sanitation and hygiene intervention that promotes NTD-preventive improved WASH behaviors; (3) investigate whether changes in personal hygiene and sanitation behaviors are sustained, and which factors contribute to this; (4) document the cost-effectiveness of integrated, holistic WASH behavior change promotion; and (5) assess whether collective efficacy and water and sanitation insecurity modify intervention effectiveness.

1.7. Key partners

World Bank, International Initiative for Impact Evaluation (3ie), the Water Supply and Sanitation Collaborative Council (WSSCC), Children's Investment Fund Foundation (CIFF), and Schistosomiasis Control Initiative (SCI): The World Bank's Strategic Impact Evaluation Fund (SIEF), Development Impact Evaluation Group (DIME), and WSP; 3ie; WSSCC; and CIFF provided funding, subject matter expertise, technical inputs, and strategic oversight for this study.

Federal Ministry of Health: The FMOH NTD Coordinator and other Government of Ethiopia (GoE) officials were engaged 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.

Amhara Regional Health Bureau: The ARHB supported 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 helped to collaboratively implement the *Andilaye* intervention.

Emory University, Emory Ethiopia: Emory University, the prime recipient of these research grants, provided research oversight and subject matter expertise related to WASH and NTDs, social and behavioral science, project costing, randomized trials, and study design. Emory Ethiopia is a certified 501(c)3 university-derived non-profit organization, registered in Ethiopia with a permanent research team in place. Emory Ethiopia has several full-time staff in Ethiopia that led the project, including Dr. Abebe Gebremariam (Co-PI and Director of Emory Ethiopia), Mulusew Belew (Regional Manager), Kassahun Zewudie (Study Manager), Mulat Woreta and Melkamu Zegeye (Monitoring and Evaluation Officers), and Siraj Muhammed (Behavior Change and Communications Specialist).

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 were members of the *Andilaye* team. These investigators played 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 guided the assessment of water insecurity and further refinement of existing metrics.

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 engaged other key WASH and NTD stakeholders working in the sector as part of the knowledge dissemination and curriculum development processes.

Study team: The principal investigators of the study were Matthew C. Freeman, PhD, MPH and Abebe Gebremariam Gobeze, MD. The study managers were Maryann G. Delea, PhD, MPH and Jedidiah S. Snyder, MPH. The study team consisted of: Mulusew Belew, MSc; Resom Berhe, MPH; Tina Braccio, MPH; Frederick Goddard, MS; Molly Linabarger, MPH; Katie Micek, MPH; Siraj Muhammed, MPH; Ashlin Rakhra, MPH; Gloria Sclar, MPH; Hiwote Solomon, MPH; Yihenew Tesfaye, MS; Mulat Woreta, BA, PGD; Melkamu Zegeye, MPH; and Kassahun Zewudie, MPH. 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.

2. Study overview

2.1. Study design

Emory University and its consortium partners conducted a two-year impact evaluation, designed as an ex-ante two-arm, parallel group cluster-randomized trial (CRT). Clusters were defined as rural or peri-urban *kebeles* from three *woredas* in West Gojjam and South Gondar Zones, Amhara National Regional State, Ethiopia. As indicated in further detail below (*see section 4.4. Random allocation*), we randomly selected and assigned clusters; half to receive the *Andilaye* intervention, half to receive the current standard of care sanitation and hygiene programming (i.e., interventions related to FMOH's existing CLTSH programming). The intervention arm received *Andilaye*, a demand-side sanitation and hygiene intervention informed by social and behavioral theory and empirical evidence, particularly evidence generated during *Andilaye's* formative research phase. The *Andilaye* intervention promotes selected, improved WASH behaviors and constituent practices deemed locally appropriate for inclusion in a demand-side intervention. The intervention focused on positive, community-oriented motivators of behavioral change, promoted achievable incremental improvements, and incorporated strategies that facilitated behavioral maintenance (i.e., prevention of behavioral slippage or relapse back to unimproved behaviors). We designed the *Andilaye* intervention to integrate behaviors that could contribute to improvements in mental and social well-being and the control of infectious diseases, including trachoma and soil-transmitted helminthiasis. The control arm (counterfactual comparator) received FMOH's existing CLTSH programming, and no attempt was made to modify the government's roll-out of these interventions or the Health Extension Services Package. The study team worked with government partners to minimize contamination from other WASH interventions in our study sites, to the greatest extent possible.

Overview of CLTSH programming and the *Andilaye* intervention:

FMOH's existing CLTSH programming*	<i>Andilaye</i> intervention
<ul style="list-style-type: none"> Nationally scaled program with a primary goal the achievement of open defecation free (ODF) status Based on the basic principles of Community-Led Total Sanitation Incorporated an added hygiene component (e.g., hand hygiene and hygienic handling/storage of water) Involved shame and disgust about one's own open defecation behaviors 	<ul style="list-style-type: none"> A scalable demand-side sanitation and hygiene intervention designed to complement existing service delivery within Ethiopia's HEP Employed rigorous formative research and practically applied social and behavioral theory Targeted three behavioral themes, informed by formative research: sanitation, personal hygiene, and household environmental sanitation Focused on positive, community-oriented motivators of behavioral change and promotes achievable incremental improvements

- | | |
|--|--|
| <ul style="list-style-type: none"> • The heart of the program is about “triggering” or “igniting” communities to change their hygiene and sanitation habits, namely by constructing and using latrines instead of defecating in the open • Implemented nationwide by government (e.g., HEWs) and non-governmental actors | <ul style="list-style-type: none"> • Incorporated strategies that facilitated behavioral maintenance along with behavior change catalyzing activities • Addressed issues related to over-extension of HEWs, through the engagement of additional community change agents as mechanisms for intervention delivery |
|--|--|

* (Federal Democratic Republic of Ethiopia, Ministry of Health, 2011)

2.2. Impact evaluation

Primary outcomes of interest

1. Sustainability of WASH-related behaviors, as measured through the proportion of individuals and households consistently practicing improved targeted WASH behaviors, including: (1) sanitation, (2) personal hygiene, and (3) household environmental sanitation, more specifically, the 11 constituent practices (*see section 3.3. Description of intervention*); and
2. Mental health was measured using validated scales. The WHO-5 (Bech, 2004) assesses overall well-being by asking participants to respond to how frequently on a five-point scale (‘at no time’ [1] to ‘all the time’ [5]) they have related to each of five statements in the previous two weeks. Scores can range from 5-25, with higher scores indicating better well-being. Scores below 13 are considered poor well-being. The Hopkins Symptom Checklist (HSCL) (Derogatis et al., 1974) is a non-diagnostic tool that includes 25 items to assess symptoms of anxiety (items 1-10), depression (items 11-25) and overall emotional distress (all 25 items). We omitted two items from the depression set. An item on sexual desire was deemed inappropriate for unmarried women. Another item on suicide ideation was deleted as we had no ability to provide clinical recourse if needed. Participants indicate how much symptoms bothered them in the previous week (‘not at all’ [1] to ‘Extremely’ [4]). The final score for each state is a mean of responses for each of the relevant items (range from 1 to 4). Scores of 1.75 or higher indicate that the condition could be present while lower scores are an indication of lower anxiety, depression, or distress.

Secondary outcomes of interest

- Short-term behavioral outcomes, measured as the proportion of households with improved (private or shared) latrines and washing facilities that are functional and available for use; proportion of households using functional latrines and washing facilities; proportion of households with all members exclusively using a latrine for defecation; proportion of households disposing of child feces in an improved latrine; proportion of households with all children in the household with a clean face and hands;
- Intermediate behavioral antecedents, measured as the proportion of households with improved knowledge regarding the implications of improved WASH practices (i.e.,

perceptions regarding negative externalities); proportion of households that indicate positive attitudes, perceptions toward improved sanitation and good hygiene practices; change in normative expectations related to open defecation, exclusive latrine use for defecation, and personal hygiene practices;

- Diarrhea period prevalence, as measured through caregiver self-report;
- Sanitation insecurity, as measured through changes in sanitation insecurity scale scores (Caruso et al., 2017b);
- Collective efficacy measures, as measured through changes in collective efficacy scale scores (Delea et al., 2018); and
- Water insecurity, as measured through changes in water insecurity scale scores (Hadley and Freeman, 2016).

2.3. Process evaluation

We conducted a process evaluation alongside our impact evaluation to assess intervention implementation fidelity, participation and dose response, and contextual factors. The purpose of the process evaluation was to determine the quality and integrity of the *Andilaye* intervention, as implemented; the extent to which participant engagement and dose response are associated with the dose delivered; and the social, political, and economic factors that may influence intervention implementation. We took a mixed methodological approach to collect process data over the *Andilaye* study period. We collected quantitative data via structured household-level surveys, activities observations, post-training assessments, and systematic audits of records. We collected 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. These data allowed us to better interpret our impact evaluation data and determine how and why change did or did not occur (e.g., via the identification of potential barriers to and facilitators of intervention uptake).

2.4. Timeline

The *Andilaye* Trial consisted of three major phases: (1) formative research and intervention design, (2) intervention implementation and process evaluation, and (3) impact evaluation (Figure 1). Formative research and intervention design were conducted during September 2016 to February 2017. *Kebele* and household enrollment took place during baseline data collection – March to April 2017. Implementation of *Andilaye* intervention activities began in September 2017 and continued, with a focus on behavior change catalyzing activities, through midline data collection – March to April 2018. Intervention activities transitioned to activities focused on behavioral maintenance, as dictated by community, group, and household-level progress, through to endline evaluation – March to May 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. While these conditions 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, no major unanticipated or unexpected events negatively influenced intervention implementation or impact evaluation.

2.5. Study setting

The *Andilaye* Trial was carried out in Amhara National Regional State, a region of Ethiopia in which WASH conditions are poor (Central Statistical Agency - CSA/Ethiopia and ICF, 2017), behavioral slippage has been documented, and several NTDs are hyperendemic (International Trachoma Initiative, 2019). As with the rest of Ethiopia, where CLTSH is being scaled nationally, study communities have either been triggered with CLTSH or are scheduled for triggering in the near future. Three districts - Bahir Dar Zuria *Woreda* in West Gojjam Zone and Fogera and Farta *Woredas* in South Gondar Zone - were targeted for this study given they represent a range of the topographical conditions present in Amhara, and Ethiopia in general (Figure 2). The selection of study districts was made in partnership with FMOH, ARHB and One WASH National Program representatives. Farta is a mountainous area that is more rural than Fogera, a low-land, marshy area in close proximity to Lake Tana. A large dam is located in Farta. Bahir Dar Zuria is a relatively flat area on the rural outskirts of Bahir Dar town.

2.6. Context

As this study operated in an area where CLTSH is being rolled out nationally, we did not interfere with the established CLTSH roll-out and implementation protocols. However, any such further implementation of the current CLTSH interventions would only bias effect estimates toward the null. Our study was designed and executed at a time when GoE and FMOH are critically evaluating the nationally scaled HEP. As a result, our intervention design considered demand-side 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 Services 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 explored the engagement and potential of alternative community change agents for intervention delivery at the community level.

The *Andilaye Impact Evaluation* did not include a mass drug administration (MDA) component, and did not seek to influence the timing of such activities in study communities. While MDA campaigns are active in Amhara, specifically the administration of Zithromax® for the treatment and prevention of active trachoma, no parasitological health outcomes were included in our evaluation. Thus, the presence of MDA in the study area did 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.

2.7. Funding

This work was supported by the World Bank Group's Strategic Impact Evaluation Fund (SIEF, 7175829), Children's Investment Fund Foundation (CIFF, 1606-01334), and the International Initiative for Impact Evaluation (3ie, TW11.1016).

2.8. Ethics approval

Ethical approval for the *Andilaye* Trial 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).

Figure 1. Andilaye timeline

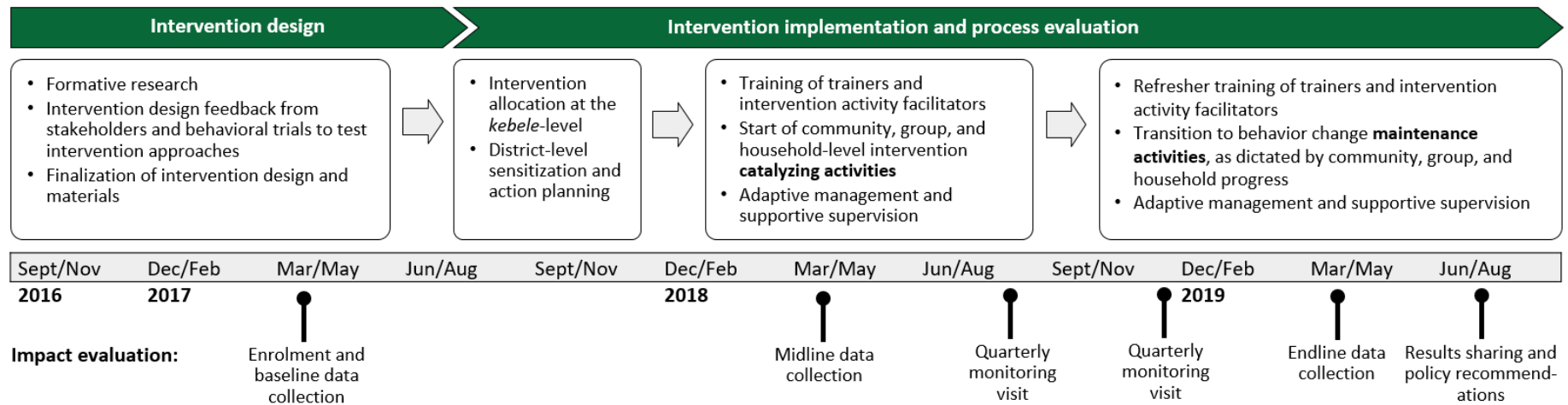
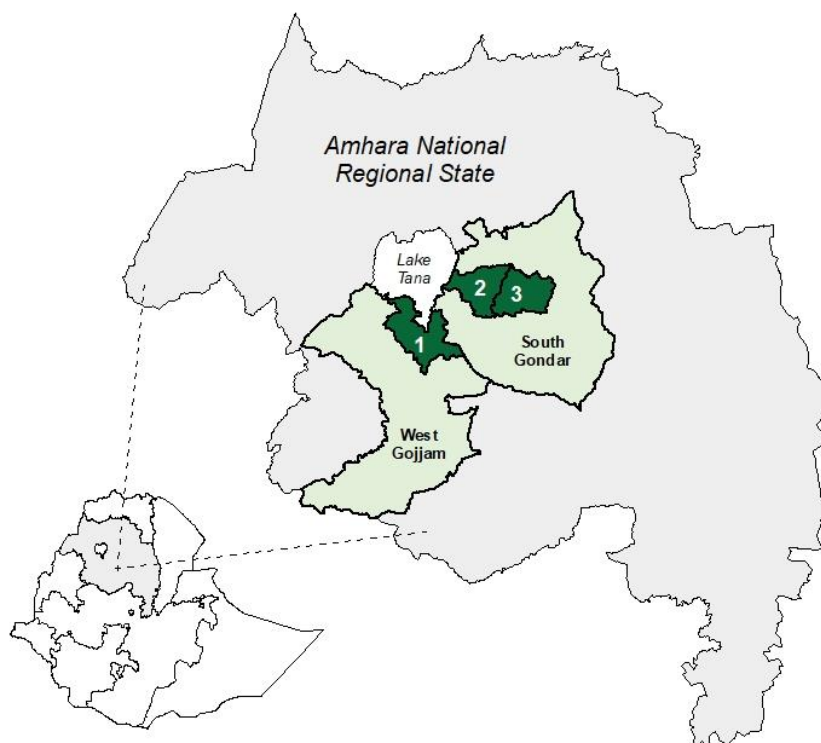


Figure 2. Map of study *woredas*: Bahir Dar Zuria [1], Fogera [2], and Farta [3]



3. Formative research and intervention design

3.1. Formative research

The *Andilaye* team executed formative research during September 2016 – February 2017, the end of rainy and beginning of dry seasons in Ethiopia. This research was grounded in several behavioral theories and frameworks, including the *Capability, Opportunity, Motivation, and Behavior (COM-B)* model (Michie et al., 2011), the *Theory of Triadic Influence (Flay and Petraitis, 1994)*, and the *RANAS (Risks, Attitudes, Norms, Abilities, and Self-regulation)* framework (Mosler, 2012). During the formative research phase, the 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 [here](#). 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 *kebeles* (two intervention, one control) in which only two or fewer 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.

3.2. Intervention design

In accordance with USAID's Technical and Operational Performance Support (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 for select behavioral themes targeted for intervention (sanitation, personal hygiene, and household environmental sanitation). These trees formed the foundation of an intervention mapping process. 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 [here](#)). 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. Problem and solution trees resulting from this process can be found [here](#). Additional details on drivers for each behavioral theme can be found in Appendix A – Supplemental Table 1.

After the intervention design meeting, the *Andilaye* team completed formal intervention mapping. The first step of this intervention mapping process involved leveraging 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.¹ The team then returned to formative research communities to conduct a series of behavioral trials, during which various intervention activities, tools, and approaches were tested among formative research households and with potential community change agents. These trials gave community members and potential change agents 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.

3.3. Description of intervention

Andilaye's intervention motto, ***“Together we can be a strong, caring, healthy community”***, and related intervention components offered aspirational messages that emphasize the need for collective action to make positive change in the community and use verbal persuasion to enhance collective efficacy perceptions. The intervention focused on three behavioral themes, informed by formative research: (1) sanitation, (2) personal hygiene, and (3) household environmental sanitation. Within these themes are 11 constituent practices targeted by the intervention (see below); these practices were identified through formative research and the intervention design process as ones that could be targeted using demand-side approaches, and were seen as achievable, per stakeholder feedback. We emphasized that behaviors usually represent a constellation of practices (Cohn, 2014), and while the *Andilaye* intervention promoted 11 constituent practices of interest, these practices represented only three improved WASH behaviors. As such, our intervention actually focused on fewer practices than many WASH interventions while also clearly specifying all necessary practices required to adopt the related improved behaviors.

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

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

Household environmental sanitation

- Keep all animals separated from the house

¹ The *Andilaye* behavioral framework reflects a compilation and adaptation of the *COM-B* and *RANAS* frameworks.

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

The logic model presented in Figure 3 reflects the summarized theory of change in which the *Andilaye* intervention is grounded. *Andilaye* intervention activities operated at four levels – district, community, group, and household (Table 1, Figure 4) – and employed a variety of behavior change and maintenance techniques. Related intervention components leveraged several motives to address behavioral antecedents and determinants at various levels of influence. Behavioral antecedents are precursors that need to be addressed before behavioral change and maintenance can occur. These include psychosocial factors such as attitudes and normative beliefs regarding improved practices, perceived and actual abilities to perform improved practices, self-regulation, and intentions to initiate and maintain the adoption of improved practices. Behavioral determinants reflect physical and contextual conditions, such as water availability and facilities access, which mediate the adoption and translation of behaviors into the execution of improved practices.

3.4. Target audiences

While primary caregivers of the study's index children (i.e., youngest child in the study household aged 1-9 years at baseline) comprised the intervention's primary target audience, we designed the intervention to address intra- and inter-personal behavioral factors and promote behavior change among all household members and the community at large.

3.5. Intervention activities

The intervention addressed issues related to over-extension of HEWs, and over-saturation of messaging via the HEP's Health Extension Services Package through the engagement of additional community change agents as mechanisms for intervention delivery. Prior to activity roll out in the community, the *Andilaye* intervention commenced with district-level capacity building activities, such as action planning and training of trainers and intervention activity facilitators (Table 1, Figure 4). Further, district-level refresher trainings and adaptive management activities were conducted to reinforce previously acquired knowledge and skills, address trainer/facilitator turnover, and review successes and address challenges faced in implementing group and household level activities. Community-level activities included the 'whole system in the room' (Federal Democratic Republic of Ethiopia Ministry of Health, 2012), community mobilization and commitment events, and cross-fertilization visits. These activities were intended to engage community stakeholders in action planning, create an enabling environment in which change may occur, and address inter-personal factors related to public commitment, social norms, and social support related to improved practices, among others. Group-level activities such as structured community conversations, provided further opportunity for peer-to-peer counselling and support. These activities served to address action knowledge and capacity, enhance barrier identification and planning, shift perceptions regarding empirical expectations, and improve perceptions regarding individual and community capabilities (e.g., self-and collective efficacy appraisals). Household-level counselling visits by WDALs provided personalized counselling to caregivers to equip them with the knowledge, skills, and motivation necessary (e.g., individual and household goal setting and monitoring [self-regulation], self-efficacy, tailored barrier identification and planning) to adopt and maintain improved WASH practices.

Our “*Andilaye* Intervention Manual” can be found [here](#).

- **The purpose of the manual is to provide a detailed outlined of each *Andilaye* intervention activity and its behavior change strategy so:** (1) key stakeholders can implement the *Andilaye* intervention with fidelity, and (2) public health practitioners, government officials, and other relevant stakeholders are able to effectively adapt the intervention or scale it in other regions of Ethiopia and perhaps beyond.
- **The manual provides a comprehensive instructional guide for all *Andilaye* intervention activities** including: (1) activity summary (aim, timeline, facilitators, target audience, tools, duration, and frequency), (2) purpose, (3) objectives, (4) detailed description and protocols, (5) checklists, and (6) process indicators.

Example activity summary:

Skills-based training of the trainers for HEWs, CHC HEWs Supervisors, Woreda Officials

Activity summary	
Activity aim	To provide skills-based training to HEWs/CHC HEWs Supervisors/ <i>Woreda</i> Officials on household-level intervention activities, supportive supervision, and on-the-job-training so HEWs can, in turn, effectively train WDALs on the implementation of household-level activities and provide supportive supervision
Activity level	District catalyzing
Timeline	After district-level Sensitizing and action planning workshop and before community-level Skills-based training of Women’s Development Army Leaders
Facilitators	<ul style="list-style-type: none"> • <i>Andilaye</i> team
Target audience	<ul style="list-style-type: none"> • All HEWs from intervention <i>kebeles</i> • At least two CHC HEWs Supervisors from all CHCs • <i>Woreda</i> Hygiene and Sanitation Officer • <i>Woreda</i> Health Extension Program Officer • <i>Woreda</i> Health Office Head
Essential tools	<ul style="list-style-type: none"> • Training Summary for Household Counselling Visits (English / Amharic) • <i>Andilaye</i> ‘Gobez!’ Flipbook (English / Amharic) • <i>Andilaye</i> Household Goal Card (English / Amharic) • <i>Andilaye</i> Household Monitoring Matrix Card (English / Amharic) • HEW Supportive Supervision Tool (English / Amharic) • <i>Woreda</i> Official Supportive Supervision Tool (English / Amharic) • Training of Trainers Action Planning Tool (English and Amharic) • CHC Action Plan Monitoring Tool (English and Amharic) • <i>Woreda</i> Action Plan Monitoring Tool (English and Amharic)
Complementary tools	<ul style="list-style-type: none"> • <i>Andilaye</i> CC Facilitator Flipbook (English / Amharic) • <i>Andilaye</i> Overview Presentation (English)
Duration	2 full-day workshop
Frequency	1x per intervention district

3.6. Key behavior change tools utilized by the *Andilaye* intervention

Illustrations developed for *Andilaye* intervention’s behavior change tools were produced by an artist based in Ethiopia and were informed by our formative research. We employed a cognitive validation process, through which we obtained feedback from potential participants and implementers (e.g., WDALs, HEWs) regarding comprehension and ease of use of the materials, to ensure the behavior change tools were locally appropriate and acceptable. Emory University oversaw the development of all behavior change tools used for the *Andilaye* intervention. The description and function of key behavior change tools developed for community, group, and household-level *Andilaye* intervention activities are provided below.

Community-level: *Andilaye* Community Commitment Banner

This banner was used during the *Andilaye* community mobilization and commitment event to guide a discussion on and facilitate community commitment for the adoption of improved WASH behaviors and practices aligning with each *Andilaye* intervention behavioral theme (sanitation, personal hygiene,

and household environmental sanitation). Community members willing to commit to community goals and cooperate with *Andilaye* community change agents (e.g., WDALs, HEWs, community conversation facilitators) signified their commitment by leaving their mark (e.g., ink thumb print or maker signature) on the empty space of the banner. See [Amharic](#) and [English](#) versions.

Group-level: Andilaye Community Conversations Flipbook

This illustrative flipbook was used by selected *gott/kebele* stakeholders to guide *Andilaye* community conversations and follow-up community conversations. The flipbook contained information regarding the ideal way to perform the 11 targeted WASH practices promoted by the *Andilaye* intervention. Through probing questions, the flipbook guided facilitators to encourage participants to reflect on their perceptions, experiences, and challenges and solutions regarding the adoption of the targeted practices. Changes in attitudes toward these improved practices were facilitated by discussing the benefits of practicing the target WASH practices, and an attempt was made to dispel misconceptions through demonstrations. These demonstrations also provided an opportunity for participants to obtain skills and action knowledge related to the execution of the improved WASH practices promoted by the *Andilaye* intervention. Lastly, the flipbook prompted facilitators to (1) encourage participants of community conversations to reach a consensus about the adoption of the improved behaviors targeted by the *Andilaye* intervention, and (2) commit to practicing the improved behaviors and encouraging others to do so as well. The flipbook was color coordinated into sections that align with each behavioral theme targeted by the *Andilaye* intervention (i.e., sanitation [green], personal hygiene [blue], and household environmental sanitation [yellow]). See [Amharic](#) and [English](#) versions.

Household-level: Andilaye Gobezi! (Good job!) Flipbook

This illustrative flipbook was used by WDALs to guide *Andilaye* counseling visits with caregivers. The first pages of the flipbook highlighted the five steps of a counseling visit: (1) greeting household members and building rapport with the caregiver, (2) conducting a transect walk of the entire household compound to get a better understanding of the household's current WASH practices and conditions (including a detailed illustration on what to look for), (3) setting goals using the *Andilaye* Household Goal Card, (4) performing tailored (behavioral theme and need-specific) inter-personal counselling using the following sections of the flipbook, and (5) setting an appointment for the next follow up visit and concluding remarks. The inter-personal counseling pages of the flipbook contained information regarding: (1) ideal ways to perform the 11 improved WASH practices targeted by the *Andilaye* intervention, (2) benefits related to the improved practices, (3) barrier identification (specific personal and household barriers are discussed, but the flipbook contains a list of common barriers, identified during formative research, to help facilitate the conversation), and (4) information regarding how to plan for, cope with, and overcome barriers (e.g., personal setbacks or shocks to the system [e.g., floods, drought]). The flipbook was organized into sections and color coordinated to align with each behavioral theme targeted by the *Andilaye* intervention (sanitation [green], personal hygiene [blue], and household environmental sanitation [yellow]). See [Amharic](#) and [English](#) versions.

Household-level: Andilaye Household Goal Card

This goal card provided intervention households several options for committing to and prioritizing incremental changes each month. These incremental changes represented progressive steps toward aspirational goals that reflect the 11 improved WASH practices promoted by the *Andilaye* intervention. Allowing households to (1) identify the incremental changes they think are most feasible

at a given time-point, and (2) select the timing of their own behavioral goals facilitated a tailored approach to behavioral adoption at the household level. Such an approach provided each household with the opportunity to consider its own situation and select the goals it felt were most important and attainable to focus on each month. Mastery experiences obtained through the attainment of small, incremental goals also served to enhance and reinforce efficacy appraisals (i.e., perceptions regarding the ability and autonomy to pursue and achieve tasks and/or goals). The goal card hung in intervention households so all members could see and track their progress toward the goals. Progress made on the attainment of household goals were reviewed by the caregiver and the WDAL during follow-up *Andilaye* counseling visits. Like other *Andilaye* intervention tools, the *Andilaye Household Goal Card* was color coordinated to align with each behavioral theme targeted by the intervention (sanitation [green], personal hygiene [blue], and household environmental sanitation [yellow]). See [Amharic](#) and [English](#) versions.

“Together we can be a strong, caring, healthy community!”



Construct a long lasting and comfortable latrine and encourage all family members to use exclusively for defecation and urination



Promote your personal hygiene by washing your hands and face several times a day



Ensure the sanitation of your household compound by cleaning up animal and child feces and separating animals from the main living room



EMORY
UNIVERSITY



The Andilaye intervention was developed by Emory University, in partnership with the Federal Ministry of Health. Funding for the project is provided by World Bank, SIE, and CDF.

*The **Andilaye Community Commitment Banner** had summary illustrations and messages representing each behavioral theme and space to the right for community by-laws to be documented and for community members to mark their commitment.*



SESSION 1 Sanitation (Latrine construction & use)

DISCUSSION QUESTIONS

1. What comes to mind when you are asked to "construct a long-lasting latrine that is comfortable and hygienic"?
 - How many of you have constructed a household latrine like this in your compound?
 - What are the benefits, challenges, solutions?
2. What is your family's experience with having all members of the family only using a latrine every time they defecate?
 - What are the benefits, challenges, solutions?
 - Are there women's challenges to defecate during the day? Solutions?
 - Are there challenges to immediately disposing of children's feces into the latrine? Solutions?
3. What comes to mind when you are asked to repair and upgrade your latrine?
 - What are the benefits, challenges, solutions?
4. What comes to mind when you are asked to close your latrine's pit when it becomes full and reconstruct a new latrine?
 - What are the benefits, challenges, solutions?

DEMONSTRATIONS PROTOCOLS Sanitation (Latrine construction & use)

VISIT TO MODEL HOUSEHOLD

1. In advance of the community conversation, the facilitator should identify a household modelling improved sanitation practices (e.g., durable, clean, hygienic latrine kept in good condition), and ask the head of household whether the participants can visit his/her home (toward the end of the community conversation)
2. Together with the model household, immediately after the community conversation, the facilitator will take the participants to the model household
3. The facilitator and the model household will ask the participants who are resistant to constructing/upgrading/repairing their household latrine about their specific barriers to doing so
4. Then the model household will share how their household prioritized these practices, and what the household has gotten due to constructing their household latrine, upgrading and/repairing it regularly, and exclusively using the latrine for defecation
5. The model household members and the facilitator should encourage the late adopters, and tell them that they can adopt these practices – they have the ability and community support to do so, and that by adopting these behaviors in their household, they are contributing to a stronger, healthier, caring community
6. The facilitator will strengthen what the model household said
7. Then the facilitator will get the HEW, and tell them the name and home of the people who are late adopters, or are resisting latrine construction, upgrade, repair
8. The HEW, and if possible the facilitator, will go with the WDAL to that household during her next Andilaye household counseling visit to support the WDAL in advising the household

PARTICIPANTS' DECISION AND PLEDGE Sanitation (latrine construction & use)

FACILITATOR INSTRUCTIONS

1. Ask participants to summarize what was discussed, agreed to, and decided on during Session 1. Then summarize their responses and anything else they missed
2. Allow participants to ask if there is something not clear on the summary and try to give an answer for the questions
3. Show the participants the Andilaye Household Goal Card, and explain how the messages on the card, the household visit by WDALs, and the community conversation are interconnected
4. Inform participants to share the information and skills covered during this community conversation with their family members
5. Ask participants to show that they agree and demonstrate their commitment by going around the group and having each person separately state their name, and that they agree and commit to practicing the improved behaviors discussed during the CC and encouraging others in their household and community to do the same
6. Finally, close the session by recognizing and thanking all group members for their active participation and remind them the date, time and place of Session 2

The cover of the **Andilaye Community Conversations Flipbook** (top left) and an example of discussion questions (top right), demonstration protocol (bottom left) and facilitator instructions to guide participants' decision and pledge for the sanitation behavioral theme (bottom right). Between each of these pages were illustrative examples of the benefits, barriers and solutions related to each targeted practice (the same illustrations included in the household-level Andilaye Gobezi! (Good job!) Flipbook and Andilaye Household Goal Card).



1 Construct a long lasting and comfortable pit latrine

Benefits!

- You will not touch feces when doing gardening activities and no longer get infection.
- Your children will be happy and do well in school because they will not get sick from feces in compound.

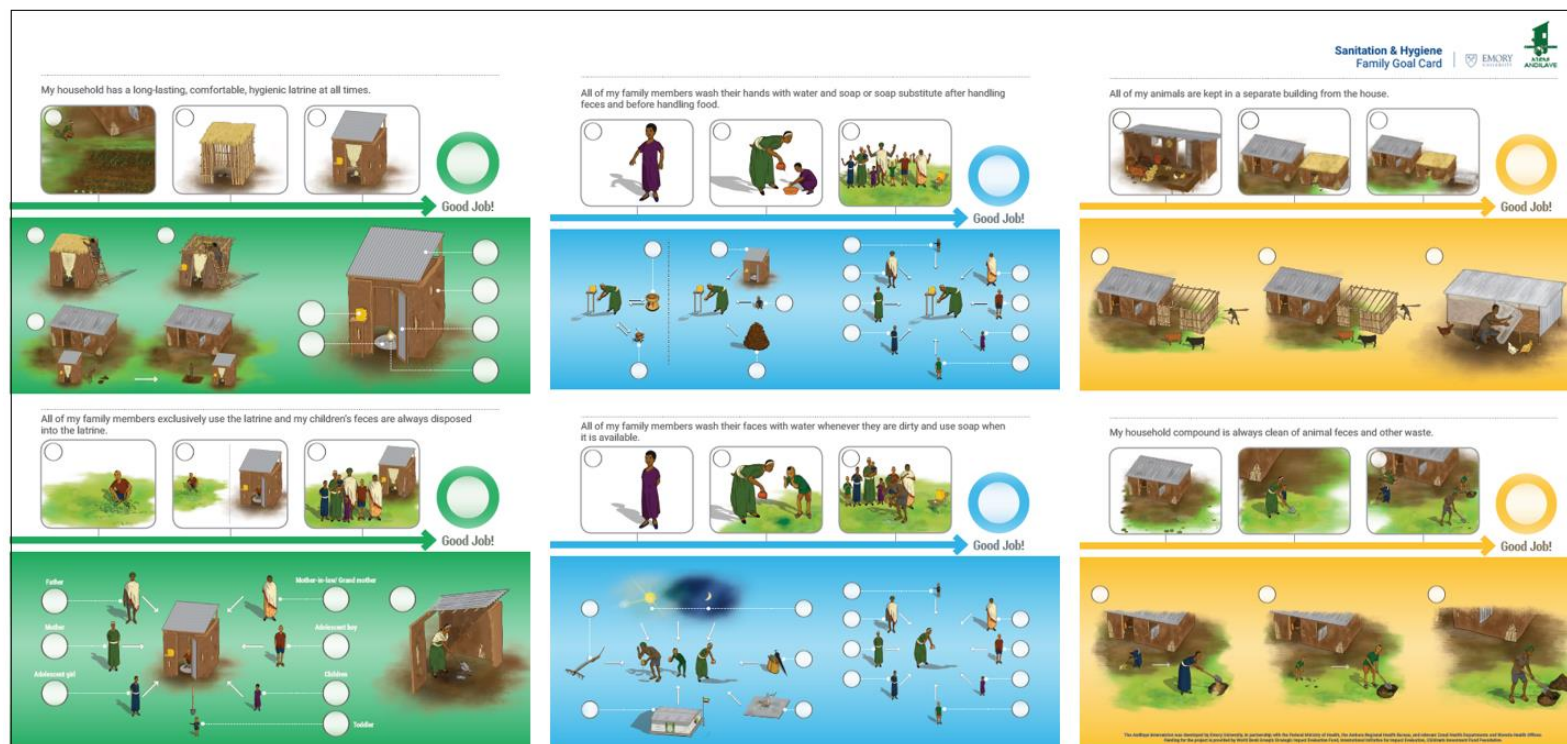
1 Construct a long lasting and comfortable pit latrine

Barrier and Solution

I do not have time to construct a latrine. I am too busy working in my field.

A strong and smart farmer makes time for latrine construction so he stays healthy and can continue to work in his field.

The cover of the **Andilaye Gobezi! (Good job!) Flipbook** (top left) and illustrative summary of the five steps of a counseling visit an example of discussion questions (top right). Each behavioral theme contained illustrative examples of the benefits of the improved practices (bottom left) and barriers and solutions related to each practice promoted within the respective behavioral theme (bottom right) to prompt the WDAL during inter-personal counseling. Between these pages were full page illustrative examples (shown to the caregiver) of the ideal ways to perform the 11 improved WASH practices targeted by the Andilaye intervention. The Flipbook also contained prompts that WDALs can use to counsel caregivers regarding how to plan for, cope with, and overcome barriers.



The Andilaye Household Goal Card offered the caregiver options for household-level goal selection related to each of the behavioral themes targeted by the Andilaye intervention. Large colored circles on the goal card indicated and celebrated aspirational Gobezi! (Good job!) targets.

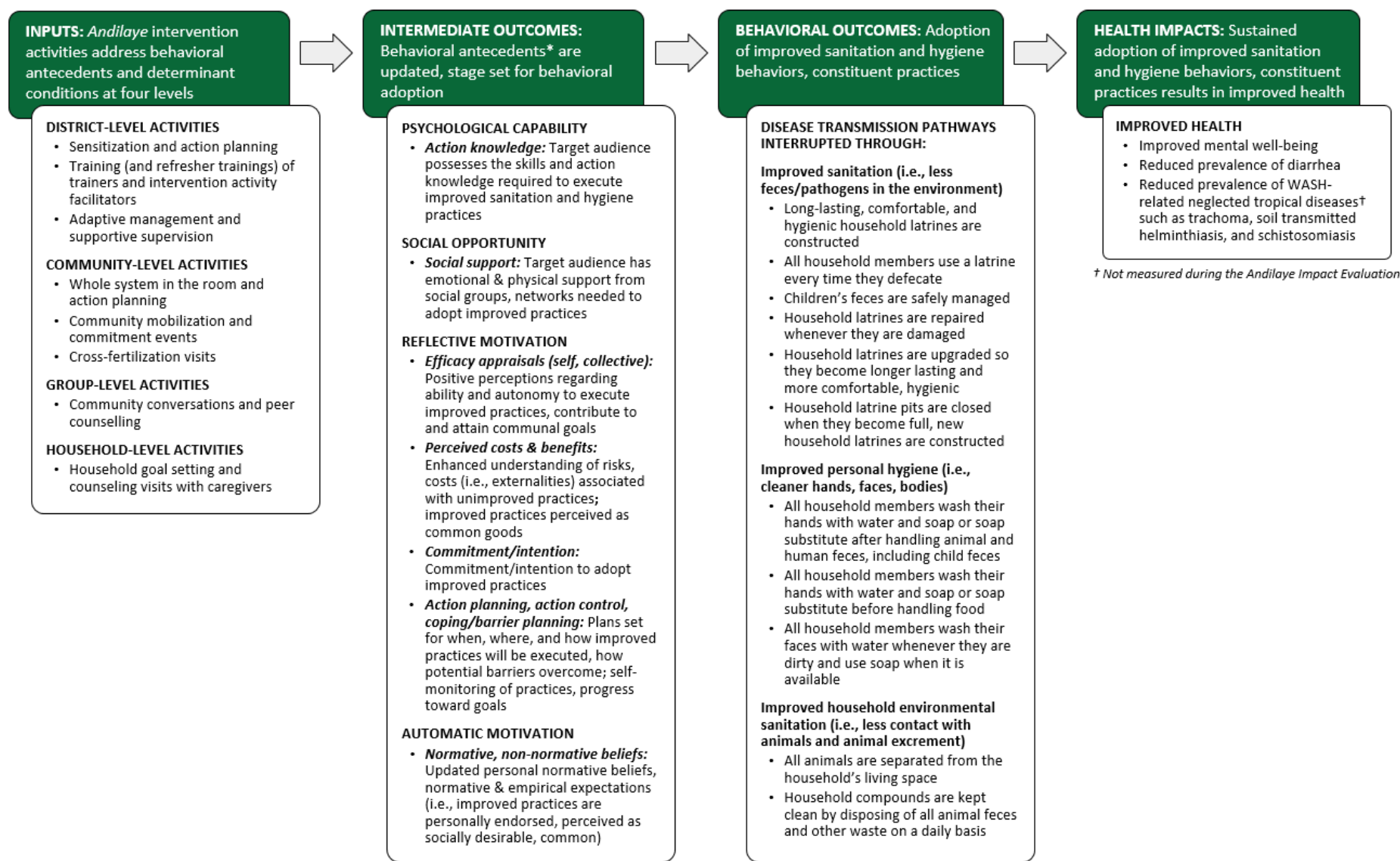
3.7. Intervention delivery

The original conceptualization of this study reflected an impact evaluation of a government-led intervention. However, this plan became problematic as it became evident that a government-led roll out of the *Andilaye* intervention in a select number of *kebeles* was both politically and logistically infeasible due to a lack of resources available to do so. Specific information provided from government officials regarding these limitations indicated that a lack of per diem and transportation limited their ability to get out to communities to routinely conduct supportive supervision, on-the-job training and implementation support. Other limitations included competing job and program priorities. Our alternative approach, to develop an intervention that complemented existing government programming (i.e., HEP), yielded important policy and programmatic findings. However, this modified approach did result in a delayed implementation timeline. See Appendix A – Supplemental Table 2 for a summary of relevant roles and responsibilities related to HEP and the *Andilaye* Trial that demonstrates the alignment of these initiatives.

The implementation and overall supervision of the *Andilaye* intervention at each level was led by a qualified Ethiopian-based *Andilaye* team (Emory Ethiopia). 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 could not assist with implementation facilitation roles. The *Andilaye* team utilized detailed field protocols and agendas included in the *Andilaye Intervention Manual* to guide the roll out of intervention activities and maximize quality and fidelity of *Andilaye* intervention trainings and workshops. The *Andilaye* team developed standardized field protocols, tools, and materials for each intervention activity that incorporated comprehensive feedback from key stakeholders and study participants (e.g., WDALs, HEWs, community members) during intervention design. The employment of such materials facilitated intervention implementation fidelity and minimized the opportunity for “on-the-spot” modifications from the per protocol design of the intervention.

Per protocol participation, adherence, and compliance of *kebele* and *woreda* level stakeholders (e.g., WDALs, HEWs, CHC HEWs Supervisors, *woreda* officials) was facilitated through supportive supervision which cascaded from the *Andilaye* team to Woreda Health Office officials and on to CHC HEW supervisors and HEWs. 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. *Kebele* and *woreda* level stakeholders were incentivized to partake in *Andilaye* trainings and workshops by receiving training material and a regional-standard per diem to compensate them for travel and accommodation to the events. WDALs and HEWs were provided in-kind motivators (i.e., *Andilaye umbrella*) to support the direct implementation and supportive supervision of intervention activities, respectively, and encourage adherence to intervention activities and targeted behaviors amongst community members in intervention clusters after training. 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. At the end of the *Andilaye* Trial, all study households were supplied a bar of soap as a small gesture of our appreciation for their participation in our survey.

Figure 3: Andilaye logic model, summarizing the theory of change on which the Andilaye intervention is grounded



* Targeted behavioral antecedents summarized here

Figure 4: Diagram summarizing the *Andilaye* intervention

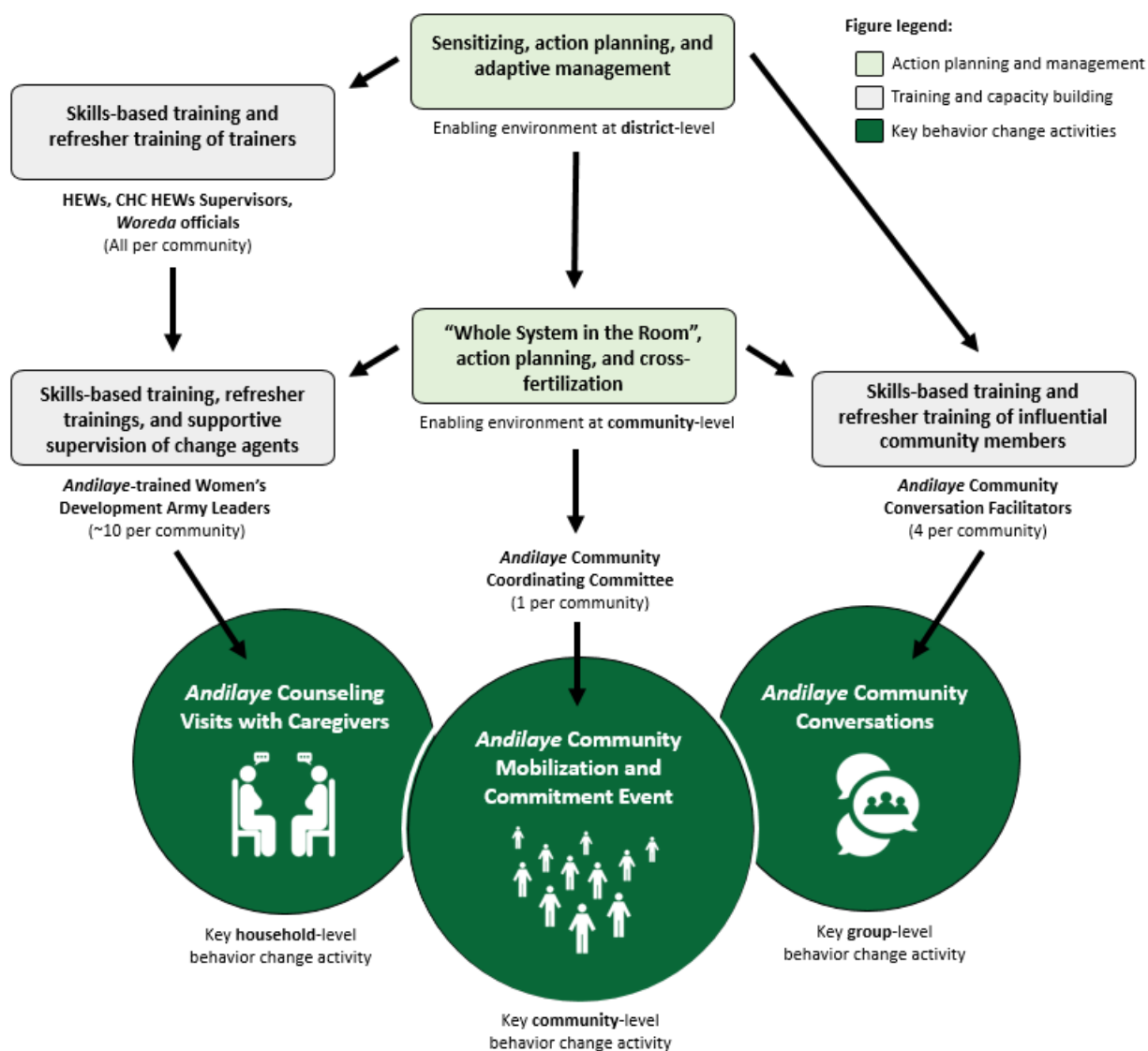


Table 1. *Andilaye* intervention behavioral change catalyzing and maintenance activities

	Activity	Aim
District	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.
	Skills-based training of the trainers for HEWs, CHC HEWs Supervisors, <i>Woreda</i> officials	To provide skills-based training to HEWs/CHC HEWs Supervisors/ <i>Woreda</i> officials on household (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 them with supportive supervision and on-the-job training.
	Training of community conversation facilitators	To provide comprehensive facilitator training to selected <i>gott</i> and <i>kebele</i> stakeholders on the ‘community conversations’ group-level intervention activity.
	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.
	Adaptive management workshops	To leverage monitoring data to facilitate 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.
Community	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 community-level 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 “ <i>strong, caring, healthy community</i> .”
	Skills-based training of Women’s Development Army Leaders	To provide skills-based training to WDALs on household-level intervention activities, as detailed in the training of the trainers for HEWs, CHC HEWs Supervisors, and <i>Woreda</i> officials.
	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.
	Community mobilization and commitment event	To improve action knowledge, barrier identification and planning, and attitudes regarding targeted NTD-preventive WASH behaviors through a form of contextually appropriate and interactive edutainment. To initiate the process of shifting social norms through community-generated and managed by-laws and sanctions and public commitment thereof.
	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.
Group	Community conversations	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 away from unimproved practices and adopting improved sanitation and hygiene practices.
	Follow-up community conversations	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.
Household	Counseling visits with caregivers	To provide personalized counseling to caregivers to equip them with the knowledge, skills, and motivation necessary to adopt improved WASH practices. To foster action capacity, self-efficacy, and barrier planning so caregivers maintain the improved WASH practices.
	Follow-up barrier planning counseling visits with caregivers	To provide continuous follow-up to households such that the house graduates from counseling related to initial adoption of improved practices to counseling related to behavioral maintenance skills. These visits will progressively focus more and more on specific barrier identification and planning skills so the caregiver can maintain his/her improved WASH practices, especially as personal setbacks, systemic shocks, and other obstacles arise.

4. Impact evaluation: methods

4.1. Sample size

We powered this study on mental well-being outcomes, as measure by the HSCL (Derogatis et al., 1974), which reflected the most restrictive primary outcomes 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 (Hadley and Patil, 2006, Hadley et al., 2008b). Drawing on two studies that have used the HSCL in East Africa (Derogatis et al., 1974), we estimated that average scores generated by this tool would be around 1.5 (standard deviation [SD] 0.5). Using unpublished data from a large on-going study of young people in Ethiopia (Hadley et al., 2008a), we estimated the intra-class correlation coefficient (ICC) for a measure of mental health was approximately 0.05, although we suspected this may be low given the sample consisted of young people, and the measure captured more severe mental health symptoms. Unpublished data from rural households in South Omo, Ethiopia are approximately 33% higher on a common mental health scale; suggesting that water insecurity is linked to mental health, and the effect is appreciable (Stevenson EGJ et al., 2014). There are limited data on the impact of changes in WASH on mental well-being, so we used a similar difference to estimate our impact. Our sample size determination (Appendix A – Supplemental Table 3) indicated that we should recruit and enroll a total of 25 households from each of 50 study clusters (25 clusters per study arm). We increased our final sample size to accommodate for 20% of households being lost to follow-up. Our target sample, therefore, included 30 households in each *kebele* study cluster, or 1,500 households in total (i.e., 750 per study arm).

Following baseline, we conducted an ex-post power calculation on our main outcome. Given the prevalence of poor well-being (33% - Table 2), and a calculated ICC of 0.026, lower than expected, we determined that we were well powered (>99%) to detect our original expected 37% relative reduction (an absolute difference of 12% points). Assuming 80% power, the absolute detectable difference is 8% points (a 24% relative reduction). For our other health outcomes, we conducted ex-post power calculations, and determined that we are able to detect a: 7%-point reduction in anxiety, with baseline of 30% (ICC: 0.007); 6%-point reduction in depression, with baseline of 21% (ICC: 0.01); and 5%-point reduction in emotional distress, with baseline of 17% (ICC: 0.005). Given the low prevalence of diarrhea (Table 2) and recent large-scale studies that showed mixed effects of the impact of WASH on diarrhea (Luby et al., 2018b, Null et al., 2018), we considered this a secondary outcome. Given a baseline reported diarrheal prevalence of 9% for the index child during the past 7 days (ICC: 0.077), we were powered to see a 6%-point reduction in diarrhea outcomes.

4.2. Eligibility criteria

Cluster-level criteria

Rural and peri-urban *kebeles* within Bahir Dar Zuria, Fogera, and Farta *Woredas* that were accessible throughout the course of the year were eligible for inclusion in the *Andilaye* Trial. It was determined that those inaccessible areas would likely not receive any such intervention at scale given the challenges of access. As such, our study was externally valid to areas that would likely be included in

any such approach implemented at scale. 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 relevant *Woreda* Health Offices helped study staff identify *kebeles* that met these eligibility criteria. While sanitation coverage and utilization were originally incorporated as inclusion criteria, the veracity of existing data were not verifiable in many *kebeles* in which initial visits were made. For example, only one latrine was observed in a community in which sanitation coverage was reportedly over 80%, and the community reported this being the case for as long as people could recall. Due to uncertainty regarding sanitation coverage and utilization data, and the challenge of behavioral slippage, even in *kebeles* previously declared as open defecation free (ODF), those criteria were dropped from inclusion requirements prior to baseline data collection. As such, any *kebele* meeting the previously stated eligibility criteria was eligible, regardless of sanitation coverage and utilization of previous CLTSH triggering or ODF verification status.

We employed a structured sampling strategy to randomly select eligible *kebeles* within the sampling frame. The primary sampling unit for this study was the *kebele*. The secondary sampling unit for this study was the household; specifically, any household residing in a targeted, sentinel village (*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’ (Hayes and Moulton, 2009) approach to purposively select one to two *gotts* that were either situated in or near the center of the *kebele* (if there were centric *gotts*) or were not adjacent to any other study *kebele* (in the event there are no centric *gotts*). This approach minimized spill-over of intervention effects and other externalities associated with the research between intervention and control clusters, especially those adjacent to each other. The number of targeted *gotts* depended only on the number of eligible households identified in *gott* census books.

Household-level criteria

Study household inclusion criteria for the *Andilaye* Trial included any household randomly selected from the *gott* census book residing in the target *gott(s)* that: (1) had at least one child aged 1-9 years at baseline (i.e., the study’s index children to assess behavioral outcomes) and consented to allowing study staff to observe the children, specifically their faces and hands, and (2) provided consent to participate in the study, with at least one adult household member consenting to serve as the primary survey respondent.

4.3. Recruitment

Cluster-level recruitment

From the *kebele* sampling frame, we employed a random number generator to identify 50 eligible *kebeles* clusters from across the three *woredas* targeted for our study. Given each of the three *woredas* vary with regard to their hydrogeological conditions and the size and number of *kebeles*, we used a stratified selection approach (at the *woreda* level). 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 (Figure 2). Our baseline results indicated that 78% (39 of 50) of *kebele* clusters randomly selected for inclusion in the *Andilaye* Trial had been triggered with CLTSH and certified as ODF according to *Woreda* Health Office records. Another 14% (7 of 50) of study clusters had been triggered with CLTSH but had not yet been certified ODF, and the remaining 8% (4 of 50) had not been triggered with CLTSH, as of the commencement of this study, though they were slated for triggering.

Household-level recruitment

Recruitment took place within the home compounds during baseline data collection. Enumerators contacted adult members of the household; they 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 using a structured consent form. Enumerators assessed household level eligibility by asking potential survey respondents a series of questions that lead to a determination of eligibility, and enrolled eligible and consenting households into the trial. During the consent process, enumerators informed respondents 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. Enumerators targeted primary survey respondents, based on the following order of priority: (1) the primary female caregiver of the index child, (2) any female household member who serves as a caregiver, (3) any male household member who serves as a caregiver, and (4) any household member over 18 year of age. If households were absent after three visits, no eligible adult respondent was available or refused to consent, or upon further conversation with the household, it became apparent that it was not eligible (e.g., no child aged 1-9 years), the enumerator recorded the information and notified the field supervisor prior to replacing that household with the next randomly selected household on the study roster. Of 1,691 surveys initiated during baseline, 1,589 (94%) households met all inclusion criteria and were enrolled into the study – 89 households more than the targeted sample size. Targeted households that did not meet inclusion criteria were excluded for the following reasons: 81 households did not have a household member between 1-9 years of age, 17 households had no eligible respondent available, three surveys were initiated but not fully completed, and one household did not consent to participate.

4.4. Random allocation

Following baseline data collection, we used a stratified random design to assign study *kebeles* to either the *Andilaye* intervention or control arm (CLTSH). Within each stratum (*woreda*), researchers from Emory University used a computer-based random number generator to generate a random number between zero and one for each *kebele* cluster, and then placed the clusters in ascending order by their randomly generated numbers. 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 re-randomization (Lachin, 1988) to secure balance across three key variables (latrine coverage, washing station with soap coverage, and head of household education), as indicated by our baseline survey. CRTs, particularly trials with a small number of clusters, often have individual-level imbalances between study 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 priori* criterion. Results from our equivalence analyses indicated balance in the number of previously triggered and ODF certified *kebeles*, between study arms, with 20 triggered and ODF *kebeles* in the intervention arm (80%, 20 of 25 study clusters) and 19 in the control arm (76%, 19 of 25 study clusters). Figure 5 provides further details in a flow diagram.

4.5. Data collection and data management methods

Survey instruments

Survey instruments administered at baseline and follow-up data collection rounds (midline and endline) consisted of several modules aimed at collecting data on key outcome indicators through reports from respondents and other household members (Appendix A – Supplemental Table 4). When developing these tools, we pulled from a reserve of existing WASH indicators, and leveraged formative research data to contextually adapt survey prompts and answer choices. To the greatest extent possible, we included validated metrics for assessment. Prior to enumerator training, the survey instrument was translated into Amharic, and back-translated by two independent Amharic speakers. The study team discussed and reconciled any discrepancies noted between the intended English prompts and the Amharic translations (identified via the back-translations). 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 ‘think-aloud’ and verbal probing techniques, we 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 administered it during 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 were incorporated, as appropriate. At the end of the piloting period, key data were checked and analyzed, and further modifications were made to the tool prior to finalizing the instrument and supervisor and enumerator trainings. Select finalized survey prompts and answer choices can be found [here](#).

Training

To the greatest extent possible, the same group of field supervisors and enumerators were re-hired and re-trained for data collection activities (i.e., baseline, midline, quarterly monitoring, endline). Prior to engaging with the full team of enumerators, senior technical research staff from Emory University and Emory Ethiopia oriented field supervisors to the *Andilaye* Trial, 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). Senior research staff then conducted a training with field supervisors and 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 pairs of two to practice administering the survey. Upon the conclusion of the office-based practice session, the enumerators 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. Pilot testing of the instrument ensured enumerators were familiar and comfortable with the survey prior to official data collection.

Quality assurance

The *Andilaye* Trial survey was administered electronically on password-protected mobile phones by enumerators and field supervisors (e.g., completing validation surveys) to improve accuracy of data entry and enable immediate review of results. Data is stored securely using the freely available Open Data Kit (<http://opendatakit.org/>). 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, but independent of enumerator assessment for validation surveys) at 10% of households in each study *kebele*. Supervisor and enumerator data captured from this sub-set of households was compared to determine inter-rater reliability of related metrics.

All enumerators and field supervisors were external evaluators who were hired for discrete data collection activities (i.e., baseline, midline, quarterly monitoring, endline). Study staff did not disclose study cluster treatment allocation to field supervisors or enumerators. However, given the nature of the intervention, they may have observed intervention materials in *Andilaye* intervention clusters, which may signal treatment allocation.

Loss to follow up

We measured and tracked individual, household, and community-level changes in key outcomes over multiple time points (i.e., baseline, midline, quarterly monitoring, endline). Households enrolled in the study are lost to follow-up if: (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 midline) or aged 1-11 years (at endline) resides in the household any longer. Households with no eligible respondent available after three attempts during midline data collection continued to be enrolled in the study, and were visited for subsequent quarterly monitoring and endline data collection.

4.6. 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) on March 9, 2017 under the name “The impact of enhanced, demand-side sanitation and hygiene promotion on sustained behavior change in Ethiopia.” We provided study participants with full details regarding the study as well as their rights as participants of the study prior to inquiring about consent to participate. Consents were

obtained orally due to low literacy rates of the population and concerns about historically coercive practices which including obtaining signatures. Oral consent was approved by all ethics boards. The consent process was conducted in the local Amharic language. We took appropriate steps to ensure confidentiality for all study participants.

4.7. Balance of treatment arms at baseline

Our baseline analyses established reference measures for follow-up analyses. We conducted bivariate assessments of balance between study arms at baseline (Table 2). As previously mentioned, 1,589 households were enrolled into the study at baseline. Given we prioritized targeting of the primary female caregiver of the index child 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. Demographic variables were balanced across study arms, with no meaningful differences in the prevalence of key demographic variables between arms (Table 2). Similarly, key outcomes of interests were also balanced between the intervention and control arms at baseline. Importantly, facial cleanliness observations of the 1,385 index children were similar to all other children aged 1-9 years in the study households (data not shown). This suggests that our index children serve as acceptable sentinels of behavioral outcomes for children of similar ages within the larger household.

At baseline, sanitation and hygiene conditions were found to be generally poor (Table 2). Only 40% of all households had a sanitation facility that was classified as improved based on the WHO/UNICEF Joint Monitoring Programme (JMP) for Water Supply and Sanitation definition. Further, 38% of respondents' primary place of defecation during the two days prior to survey administration was in the open, and only 46% of respondents had defecated in any latrine during the two-day reporting period. These statistics, along with the fact that 39 of 50 *kebele* clusters randomly selected for inclusion in the *Andilaye* Trial have been triggered with CLTSH, and certified ODF, provide strong evidence that behavioral slippage is, indeed, an issue that needs to be addressed in Amhara and perhaps elsewhere in Ethiopia.

Our key health outcome, mental well-being, was shown to be poor, per the WHO-5 scale, amongst approximately one-third of respondents at baseline (Table 2). The baseline prevalence of anxiety, depression, and emotional distress amongst respondents was 30%, 21%, and 17%, respectively. The distributions of these scores were generally balanced, when comparing the intervention and control arms. Finally, the prevalence of diarrhea during the week (i.e., seven days) preceding the survey was 9% among index children, and was also generally balanced across study arms (Table 2).

4.8. Analytical methods

The primary method of analysis for all primary and secondary outcomes followed an “intention-to-treat” analysis, which compares the intervention arm to the control arm, without regard to intervention fidelity or compliance. For binary outcomes, such as our targeted WASH behaviors, we preferentially used log-linear binomial regression models. For continuous outcomes, such as the mental well-being scale scores, we used linear regression models. All models included an intervention variable as a fixed effect, and account for the stratified design through the inclusion of the *woreda*

indicator variable (Kahan and Morris, 2012), and incorporated generalized estimating equations with robust standard errors to account for the clustering of observations within *kebeles*.

The majority of our primary and secondary outcomes (Table 2) are binary variables, and for these we preferentially used log-linear binomial models; however, log-linear binomial models are known to have difficulty converging, and so we may instead use modified Poisson regression if we encounter problems with convergence (Zou, 2004). There is often interest in showing an absolute measure along with a relative measure (e.g., a prevalence ratio), so we also presented prevalence differences. We used the same models as described above (e.g., log-linear binomial models), but used post-estimation commands to estimate the average marginal effects.

Given no imbalances were detected between study arms at baseline for any of the primary variables of interest (Table 2), 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 was 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 used interaction terms and/or stratification, and we presented the impact of the intervention at each level of the sub-group variable (e.g., separately for boys and girls).

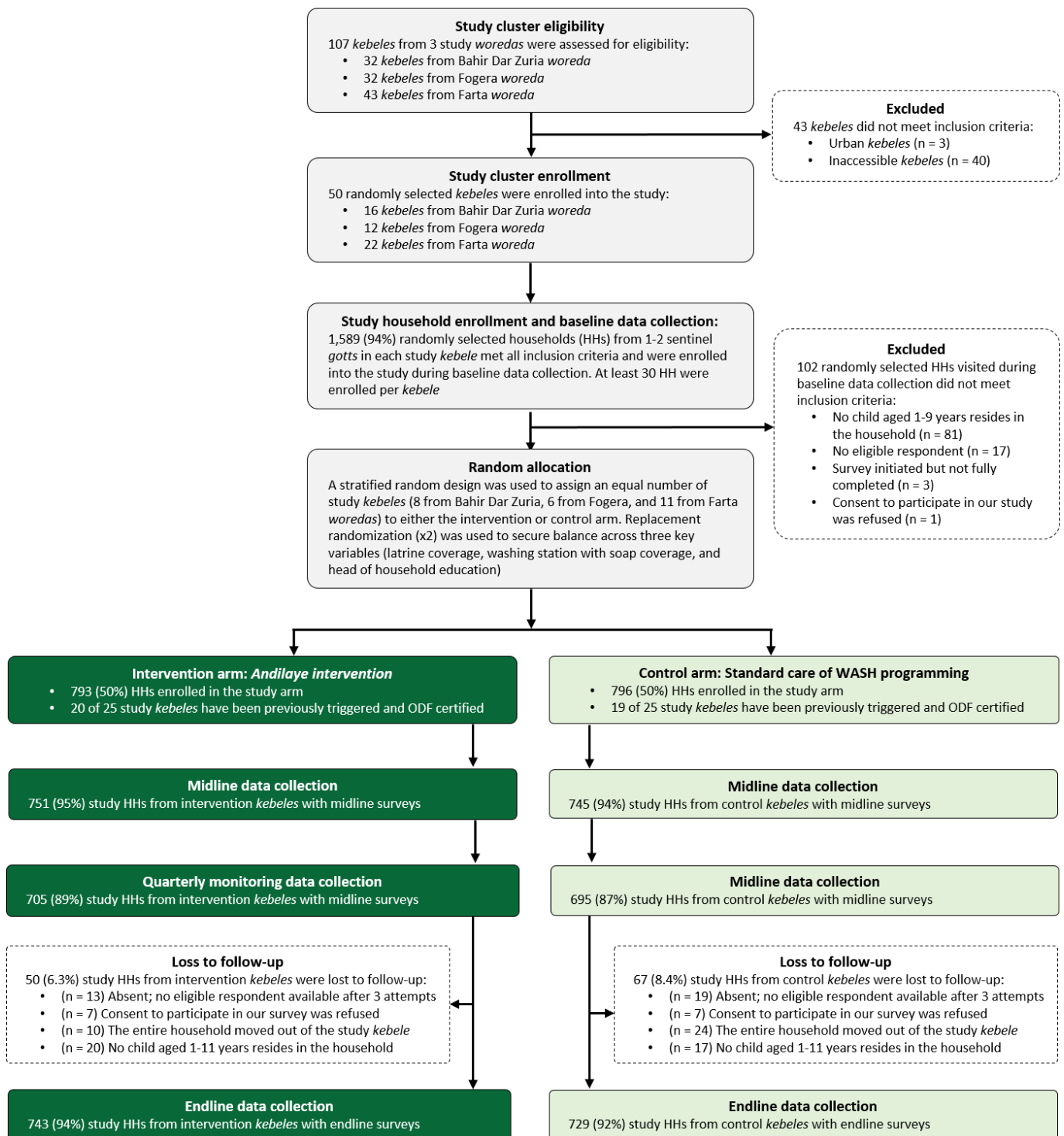
Table 2. Baseline data

Indicator	Overall		Intervention		Control	
Demographic information	N	%	N	%	N	%
Respondent 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
Primary caregiver/mother has at least secondary education	1589	12.8	793	11.9	796	13.8
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
Demographic information	N	mean (SE)	N	mean (SE)	N	mean (SE)
Respondent's age	1589	33.5 (0.38)	793	33.7 (0.52)	796	33.3 (0.55)
Head of household age	1589	41.3 (0.46)	793	41.6 (0.54)	796	41.1 (0.73)
Number of household members	1589	5.3 (0.08)	793	5.3 (0.10)	796	5.3 (0.12)
Household latrine coverage	N	%	N	%	N	%
Households with access to at least one household latrine	1589	65.5	793	64.1	796	66.8
Households with access to an improved household latrine ^a	1553	39.8	775	39.9	778	39.7
Households with access to a fully constructed household latrine	1583	30.7	792	29.6	791	31.9
Household latrine operation and maintenance	N	%	N	%	N	%
Household has added or improved something on the latrine since its original construction	1028	12.9	504	11.9	524	13.9
Household latrine characteristics	N	%	N	%	N	%
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 practices	N	%	N	%	N	%
Respondent's primary place of defecation was in the open 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
Child feces were safely disposed of during the last 2 days	961	40.2	463	38.9	498	41.4
Sharing of household latrine facilities	N	mean (SE)	N	mean (SE)	N	mean (SE)
Given household has a household latrine, number of people from another household who used this latrine during last 7 days, exclusive of household members	1037	0.94 (0.12)	506	1.08 (0.18)	530	0.79 (0.14)
Animal husbandry / other household sanitation practices	N	%	N	%	N	%
Animal feces/waste not left out in open in compound	1589	44.1	793	42.0	796	46.2
Facial cleanliness among children aged 1-9 years	N	%	N	%	N	%
Ocular discharge present	1944	40.3	932	42.2	1012	38.6
Wet nasal discharge present	1944	47.3	932	47.6	1012	46.9
Dry nasal discharge present	1944	65.4	932	64.7	1012	66.1

Dirt/dust/other debris present	1944	69.9	932	68.7	1012	71.0
Fly observations	N	mean (SE)	N	mean (SE)	N	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)
Household washing station coverage	N	%	N	%	N	%
Household hand or facewashing station(s)	1589	78.9	793	77.1	796	78.8
Handwashing practices	N	%	N	%	N	%
The last time the respondent washed, s/he used soap/ash/soapy water	1588	36.4	793	35.1	795	37.7
The last time the respondent defecated, s/he cleaned hands with water and soap, substitute	1585	37.2	791	36.3	794	38.0
The last time the respondent prepared food, s/he cleaned hands with water and soap, substitute before beginning food preparations	1586	39.7	791	41.0	795	38.5
Diarrhea among index children	N	%	N	%	N	%
During the last 2 days, including today, index 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, index child had three or more loose stools per day	1575	9.1	778	8.1	797	10.0
Mental health scores	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Anxiety score ^b	1584	1.56 (0.62)	790	1.56 (0.60)	794	1.56 (0.63)
Depression score ^b	1588	1.46 (0.52)	793	1.45 (0.51)	795	1.46 (0.52)
Emotional distress score ^b	1583	1.38 (0.48)	790	1.38 (0.47)	793	1.38 (0.48)
WHO-5 well-being score ^c	1586	16.0 (7.0)	792	15.6 (7.1)	794	16.3 (6.8)
Score above cutoff indicating poor mental health	N	%	N	%	N	%
High Anxiety ^d	1584	29.7	790	29.6	794	29.7
High Depression ^d	1588	20.8	793	21.3	795	20.4
High Emotional distress ^d	1583	17.3	790	17.2	793	17.4
Poor well-being ^e	1586	33.2	792	31.1	794	35.3

Notes. ^a "Improved" based on the WHO/UNICEF Joint Monitoring Programme (JMP) for Water Supply and Sanitation definition. ^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. ^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. ^d Each of these scores was dichotomized, with scores greater than 1.75 indicating a positive status for any of the three outcomes. ^e Each of these scores was dichotomized with scores below 13 indicating poor well-being.

Figure 5: Study flow diagram



5. Intervention implementation and process evaluation

5.1. Implementation

Control

Reports from *Woreda* Health Offices collected at the end of the trial (May 2019) indicated that none of the 50 study *kebeles* (control and intervention) received CLTSH triggering or re-triggering during the course of the *Andilaye* Trial.

Intervention

The implementation of *Andilaye* intervention activities began in September 2017 and continued, with a focus on behavior change catalyzing activities, through midline data collection – March to April 2018. Intervention activities transitioned to activities focused on behavioral maintenance, as dictated by community, group, and household-level progress, through to endline evaluation – March to May 2019. A complete list of the dates of implementation of *Andilaye* intervention activities can be found in Appendix B – Supplemental Table 1.

Costs related to the implementation of *Andilaye* activities are indicated in Appendix B – Supplemental Table 2. Implementer economic cost data for all intervention components were collected from monthly financial records reported from the implementing organization (Emory Ethiopia) and verified through interviews of project coordinators and accountants. Costs were collected in the local currency of the Ethiopian birr (ETB) and converted to US dollars (USD), using the official exchange rate of 28 ETB to 1 USD. The implementation of *Andilaye* intervention components were 2,326 USD per intervention community, plus 687 USD per intervention community for per diem and transportation for the *Andilaye* team to facilitate or co-facilitate activities.

The extent to which the intervention was delivered, as designed (i.e., dose delivered); and the extent of engagement of participants with the intervention (i.e., participation and dose received), and environmental factors which may have influenced intervention implementation (i.e., contextual factors) are discussed below. Complete process evaluation results for all intervention levels (district, community, group, and household) can be found in Appendix B – Supplemental Tables 3-6. These data are summarized in Table 3.

5.2. Process data collection results

Quantitative process data on dose delivered, participation, and dose received were collected through the direct observation of all district and community-level activities – these activities being facilitated or co-facilitated by the *Andilaye* team. At the group-level, records from community conversation facilitators were periodically audited during quarterly monitoring, refresher trainings, and endline to document items such as the number of community conversation sessions conducted, behavioral themes addressed, number of community members and target groups that participated, and demonstration activities conducted (see “Tracking Form” on pages 49-52 of the *Andilaye Community Conversations Facilitator Flipbook*). Similarly, one randomly selected HEW (N=25) and four randomly selected WDAs (N=100) from each intervention *kebele* were surveyed during community visits for

midline, quarterly monitoring, and endline to document the extent to which they have received supportive supervision and on-the-job training for the counseling visits with caregivers from CHC HEWs Supervisors and HEWs, respectively. Lastly, caregivers from study-enrolled households in intervention *gottts/kebeles* (N=793) were surveyed on their participation and engagement (i.e., compliance) in the community mobilization and commitment event, community conversations, and counseling visits with caregivers during quarterly monitoring and endline visits. A total of 703 (89%) and 708 (89%) surveys with caregivers were completed from the study-enrolled households in intervention *kebeles* for quarterly monitoring and endline, respectively. These compliance data are summarized in Figure 6, which provides a graphical representation of *Andilaye* intervention compliance reported by household respondents from *kebeles* in Bahir Dar Zuria (N=8), Fogera (N=6), Farta (N=11), and overall (N=25) for key behavior change activities.

Qualitative process data were collected via semi-structured interviews and informal discussions with key informants including government stakeholders (N=12), HEWs (N=21), WDALs (N=39), and community members at large such as caregivers (N=49). Participant observation was also completed during relevant trainings and intervention activities at the beginning of implementation.

5.3. Implementation of district-level activities

Dose delivered

As per intervention design, all study districts (N=3) were supposed to receive district-level catalyzing and maintenance activities. Overall, there were no major changes to the delivery of the district-level intervention activities compared to what was outlined in the *Andilaye Intervention Manual* and accompanying field protocols (Table 3). All study districts received the sensitization and action planning workshop, trainings of trainers and intervention activity facilitators (including one round of refresher trainings) and an adaptive management workshop. The expectation of payment (per diem) associated with district-level activities did create challenges in implementation. *Kebele* and *woreda* level stakeholders were incentivized to partake in trainings and workshops by receiving training material and a regional-standard per diem to compensate them for travel and accommodation to attend the trainings. When scheduling trainings and workshops, the *Andilaye* team was often faced with negotiations from the *woreda* to conduct the trainings out of town, the reason being that this would result in an increase in per diem. These negotiations and other district-level health program priorities (e.g., community-based health insurance, vaccine campaigns, etc.) often resulted in delayed implementation of intervention activities. Considerable delays were faced in scheduling the skills-based refresher training of community conversation facilitators and the adaptive management workshop in Bahir Dar Zuria compared to other districts. As a result, these activities were implemented nearly two months after the implementation in the other two districts and nearly one month prior to endline.

Participation and dose received

The overall aims of district-level activities were to create an enabling environment through the trainings of trainers and intervention activity facilitators and development of action plans for the implementation of subsequent activities. Importantly, these activities were designed to equip government health workers and community change agents from all intervention *kebeles* (N=25) with the necessary skills to implement community, group, and household-level *Andilaye* activities. This

included targeting Health Office Heads and Hygiene and Sanitation Officers from all *woredas*, at least two HEWs Supervisors from each CHC, and all HEWs and four community conversation facilitators from intervention *kebeles*. Overall, participation of district-level activities was as planned. A “mop up” training was conducted to assure all target population were trained on *Andilaye*. Further, multiple trainings were conducted in districts to assure less than 30 trainees per session. As a result, an average of 2-3 HEWs per *kebele* (N=56), 18 CHC HEWs Supervisors, three *Woreda* Health Office Heads, and three Hygiene and Sanitation Officers were trained and retrained on *Andilaye* counseling visits with caregivers; and four community conversation facilitators were trained and retrained per *kebele* (N=100). Ensuring participation of some targeted representatives from ARHB and *Woreda* Health Offices, however, proved challenging for the sensitization and action planning workshop and refresher trainings (Appendix B – Supplemental Table 3).

5.4. Implementation of community-level activities

Dose delivered

All intervention *kebeles* were designed to receive community-level catalyzing and maintenance activities. The *Andilaye* team increased its role in facilitating when local government officials (e.g., CHC HEWs Supervisors, *Kebele* Administrators, relevant *woreda* officials) could not assist with their implementation role at the community-level. Overall, with this additional support, there were no major changes to the delivery of the community-level intervention activities compared to what was outlined in the *Andilaye Intervention Manual* and accompanying field protocols (Table 3). All intervention *kebeles* received the ‘whole system in the room’ and action planning activity, community mobilization and commitment event, and skill-based training of WDALs (including two rounds of review meetings and refresher trainings). Cross-fertilization visits were conducted in 1-2 ‘model’ *kebeles* per study district as part of the district-level adaptive management workshops where *Andilaye* activity facilitators from all intervention *kebeles* observed model counseling visits with caregivers and community conversations. The engagement of government health workers (e.g. CHC HEWs Supervisors) did create challenges in implementing WDAL trainings. Training of WDALs were planned to be facilitated by HEWs with help from CHC HEWs Supervisors, but the involvement of CHC HEWs Supervisors was minimal in some *kebeles*. Thus, the *Andilaye* team trained WDALs together with HEWs to assure the activity was implemented appropriately. Similarly, the facilitation of the ‘whole system in the room’ and action planning activity was planned to be conducted by *Kebele* Administrators and relevant *woreda* officials, but one *kebele* had an administrator not facilitating the workshop and only 8% (2/25) of intervention *kebeles* had a representative from the *woreda* in attendance.

Participation and dose received

The overall aim of the ‘whole system in the room’ and action planning activity was to create an enabling environment at the community-level. Here, key community-level stakeholders such as *Kebele* Managers, HEWs, Agricultural Extension Workers, school directors, WDALs, religious leaders, influential elders and other influential people were targeted for participation. On average, *kebeles* had 8-9 key community-level stakeholders that participated in the activity. Similar to district-level activities the expectation of payment (per diem) associated with the activity caused challenges in engaging participants. Although this was a one to two hour activity located in the community and government policies were followed, participants seemed to expect some kind of per diem (perhaps because some had to walk up to 3 hours round trip), and when they found out that there was no per diem (at the

end of the activity), some participants expressed their disappointment. These challenges may have been avoided with increased facilitation of relevant *woreda* officials, as planned. That said, facilitators were able to lead community stakeholders to complete the action plan.

Mobilizing community members to participate in the community mobilization and commitment event presented a challenge in terms of participation. The coordinating committee identified during the 'whole system in the room' and action planning activity were tasked with mobilization and overall planning and coordination of the community event. The payment (per diem) issue discussed above may have influenced the motivation of some coordinating committee members to follow through with activities identified in the action plan, as only 36% (9/25) of intervention *kebeles* had all coordinating committee in attendance during the event. While, on average, each community had around 300 adult community members in attendance during the event, only 18% (130/703) of caregivers surveyed from study-enrolled households in intervention *kebeles* during quarterly monitoring reported attending the event (Figure 6). Ideally, all community events would have taken place on religious holidays and days that did not conflict with social gatherings or religious activities (e.g., funerals, church) to maximize community mobilization. However, the requirement to implement the community events in a relatively short amount of time (less than one month) as part of the timeline of the trial may have added to the challenges in mobilizing the community, as some events were required to be conducted on working days. Overall, the events were successful in engaging the participants as planned (i.e., to determine practices no longer deemed acceptable by the community and establish community by-laws and monitoring thereof). However, some *kebeles* did not conduct the commitment ceremony at the end of the event due to concern with the political situation (recent declarations of states of emergency) and public protests in the region in the last few years. Further, community leaders in some *kebeles* felt that it was best to discuss community by-laws for practices no longer deemed acceptable by the community after the event.

Overall, the participation corresponded with the execution of WDALs trained for counseling visits with caregivers. We planned to train 10 WDALs per community to facilitate monthly visits with households in their catchment area. On average, intervention *kebeles* had 9-10 WDALs trained and retrained on the household-level activities. That said, considerable gaps in implementation were associated with the engagement of participants in these training activities. Per protocol, and in accordance to HEP roles and responsibilities (Appendix A – Supplemental Table 2), HEWs were trained and provided materials to offer supportive supervision and on-the-job-training to each WDAL in her catchment area, as the WDALs conduct initial and monthly follow-up household counselling visits. Similarly, CHC HEWs Supervisors were provided supervisory checklists and trained to perform supportive supervision and on-the-job training with HEWs for a minimum of one round of initial household counselling visits and at least once per month for subsequent monthly visits. However, only 43% (43/100) of WDALs and 20% (5/25) of HEWs surveyed during endline reported receiving supportive supervision from relevant government officials (e.g., HEWs and CHC HEWs Supervisors, respectively) during the implementation period. This lack of supportive supervision likely negatively influenced both motivations and capabilities of WDALs to conduct household counselling visits, as indicated per protocol.

5.5. Implementation of group-level activities

Dose delivered

During district-level trainings, community conversation facilitators in each intervention *kebele* identified key populations to be targeted for community conversations. Facilitators planned to conduct community conversation sessions for the three behavioral themes with all individuals within the target population who were residing in intervention communities (i.e., sentinel *gotts*). As the community conversations were designed to target 20-25 community members at a time, pairs of facilitators in each *kebele* were encouraged to implement the activity on a monthly basis. Overall, there were gaps in delivery of these group-level activities. Despite the provision of trainings and in-kind motivators, two *kebeles* in Bahir Dar Zuria reported that they failed to implement the activity within the eight months since the initial training (October, 2018) (Table 3). Of *kebeles* with activity implemented (N=23), the average number of community conversation sessions implemented were 7-8 per *kebele* (about one per month). Thus, suggesting that *both* pairs of facilitators were not conducting monthly community conversation sessions, as planned. Seasonality may have affected mobilization of community members to attend the community conversations. The frequency of reported community conversations were lowest in the months of November, December, and January (harvesting season). Overall, facilitators reported that punctuality and motivation of participants as a challenge in conducting community conversations. During refresher trainings, many facilitators suggested encouraging participation within the community by-laws as well as working with *kebele* officials to mobilize households.

Participation and dose received

The target audience for community conversations included a diverse group of community members, such as husbands, fathers, community or religious leaders, elders, youths, students, *kebele* officials, administrators, and health development leaders. Overall, 1,211 community members were reported by facilitators to have attended at least one community conversation session. On average, this was about 53 community members per *kebele*. During endline data collection, 28% (199/708) of caregivers surveyed from study-enrolled households in intervention *kebeles* reported attending at least one community conversation (Figure 6). Interestingly, 46% (329/708) of these respondents reported that they had heard about the community conversations, suggesting an engagement in community group dialogue of those that did participate. Further, demonstration activities to dispel misconceptions of WASH practices were conducted in 80% (132/165) of all sessions reported, suggesting the activity outcome to improve perceptions regarding individual and community capabilities to adopt and maintain improved WASH practices were addressed as planned in a majority of the community conversations.

5.6. Implementation of household-level activities

Household level

Per protocol, WDALs were to act as the primary counsellor and visit each household in her women's development army network to conduct an *Andilaye* household counselling visit about once per month, with each visit lasting around 30 minutes. WDALs from all intervention *kebeles* implemented *Andilaye* counseling visits with caregivers (Table 3). However, no *kebele* had WDALs conducting monthly counseling visits according to caregivers from study-enrolled households in intervention *kebeles*

surveyed during endline. Of households reporting at least one visit (N=390)², the average number of visits was 2-3 during the 14-15 months of implementation (i.e., since the initial trainings in January-February 2018). This suggests visits were likely only conducted following each round of WDAL training and refresher training (N=3) for a majority of WDALs.

Participation and dose received

All households with children in the community were targeted to receive the household-level intervention activity. While *Andilaye Household Goal Cards* were observed in 65% of study-enrolled households in intervention *kebeles* during endline, only 59% (390/665) of caregivers reported that they received a counseling visit (Figure 6). All behavior change techniques designed into the counseling visits with caregivers (e.g., inter-personal counseling on action planning, barrier identification and planning; goal setting, commitment, and self-regulation), were critical to the *Andilaye* intervention (Appendix A – Supplemental Table 1). However, only 63% (419/665) of households were able to identify the WDALs responsible for conducting their counseling visits – suggesting that the rapport required for tailored inter-personal counseling was not achieved in the other 35% of households. Of the households reporting at least one counseling visit, 72% reported that they set household goals or incremental improvements, and two-thirds reported that they identified barriers, and their WDAL provided counseling on how to plan for, cope with, and overcome barriers – suggesting that some households were not exposed to the intended intervention, as designed.

5.7. Qualitative support of implementation challenges and successes

Gaps in supportive supervision to HEWs

To the greatest degree possible, the *Andilaye* team encouraged CHC HEWs Supervisors to incorporate *Andilaye*-related supportive supervision to HEWs within their existing activities. Limited support from government officials seemed to impact the motivation of HEWs:

“So far no one has provided support to us from the health center. It is the health-center head and the supervisor who got training with us but no one has provided support for us so far...It would be nice if they come and see and provide comments for us.” – [Key informant interview with HEW from intervention kebele]

Gaps in supportive supervision to WDALs

Within the HEP, HEWs are to provide supportive supervision and evaluation of WDAL teams and conduct biweekly meetings to evaluate the performance of the development teams. Process data and feedback from WDALs suggest that this *Andilaye*-related supportive supervision and on-the-job training were limited in some *kebeles*:

“The support we got is just one day [during the training], and we are doing the work by ourselves and no one from anywhere has come and supports us.” – [Key informant interview with WDAL from intervention kebele]

² Excluding 43 study-enrolled households who happen to have caregivers who were trained as WDALs responsible for conducting *Andilaye* counselling visits (Bahir Dar Zuria, N=5; Fogera, N=16, Farta, N=22).

“The extension worker encourages us when she find us otherwise there is no one who provide support to us it may be because they didn't find me at home but I have never seen anyone here who came and supported me. I don't think there's a special attention given for this project but they trained us and we are trying even if we have our issues.” – [Key informant interview with WDAL from intervention kebele]

Gaps within the WDAL structure

Early on, it became evident that although Woreda Health Offices and CHC HEWs Supervisors stated that the WDAL structure should be in place throughout the region, these structures were not well established in some areas. Of the 25 intervention *kebeles*, the WDALs structure was fully functional in less than half. This limited definition of the duties and responsibilities WDALs in some communities, along with lack of supportive supervision, likely hindered the ability of WDALs to act as the primary counsellor for *Andilaye*. In some cases WDALs experience opposition and insufficient cooperation from families and communities as a result of the HEWs and/or CHC HEWs Supervisors not accompanying WDALs during the initial round of household counselling visits:

“I think it is good if we get more support and there are more people working with us. People [community members] may not listen to us since we are farmers [like them] so it is good if workers from the health office or the health extension worker work with us [support us]. People may say that we are visiting them because we are paid but if there are other people [HEW or from the health center] going house-to-house to with us then it would be good.” – [Key informant interview with WDAL from intervention kebele]

Further, despite training efforts, qualitative feedback from caregivers suggests that the fidelity of the WDALs to act as primary counselors may have not been successful in some cases:

“The WDAL just came and gave the tools and left the house without counseling.” – [Key informant interview with caregiver from intervention kebele]

Strengthening the WDAL structure

Trainings may have improved motivations of WDALs through enhancing their capacity by providing counselling training and counseling tools, and small per diems during training days. Many WDALs reported their satisfaction in working on *Andilaye* as a way to increase their ability to be a model household:

“Yes, I'm happy it is because I'm working for the community to be clean and to accept the cleanness message and to clean our household compound. In earlier time when we were harvesting cabbage we would picking cabbage from places [household garden] where there is no human feces [since human feces were everywhere in the garden] but now since people are using toilet and no one openly defecates we can harvest cabbage without worry and easily and that is why I am happy.” – [Key informant interview with WDAL from intervention kebele]

“I worked as Andilaye WDAL since its start and have 16 households under my supervision. I constructed latrine which is better than I had before and face washing has become a common

practice in my family. I separated animals, including chicken, from my house.” – [Key informant interview with WDAL from intervention kebele]

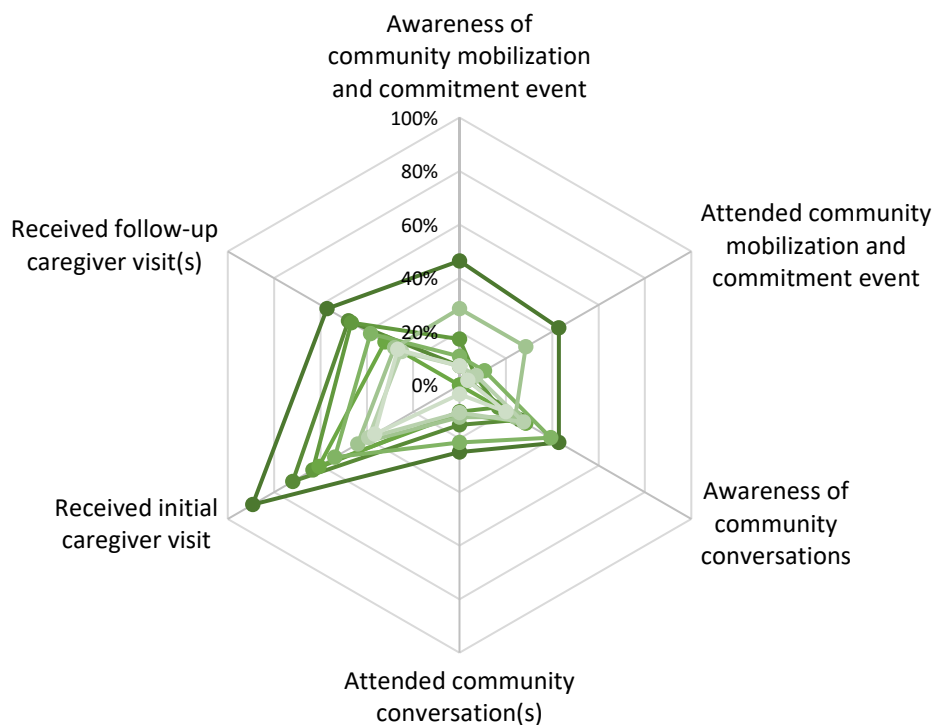
Over-extension of HEWs

The *Andilaye* intervention was designed to address the over-extension of HEWs through the engagement of additional community change agents for group-level activities (i.e., community conversation facilitators) and household-level activities (i.e., WDALs). This was often recognized by the HEWs:

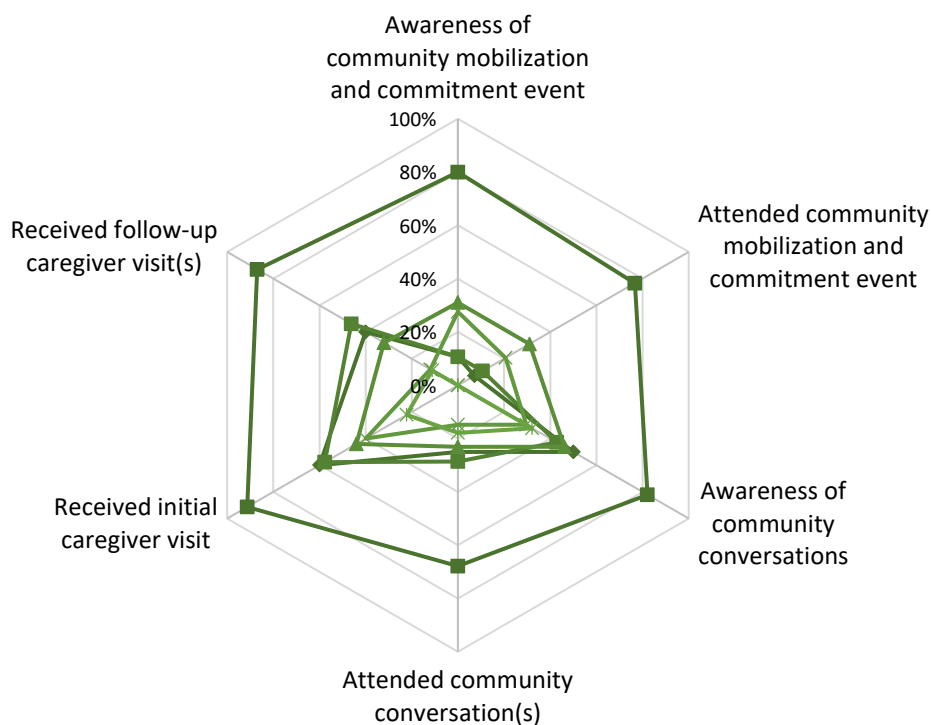
“Personally, I am happy. Because first the project help us to train development group leaders [WDALs], second the project whether it is small or large but has some motivation for the [WDALs]... From the 34 [WDALs] that we have in our Kebele 10 of them were selected and got the training.... The women have got awareness and they have better understanding about the project and they started working. So having the WDALs trained in the project is a support for us, it is one of the good input to succeed our work. The activities in the project are our normal duties and when we have a support for our duty it is an input for us and I am happy about that.” – [Key informant interview with HEW from intervention kebele]

Figure 6. *Andilaye* intervention compliance for key behavior change activities reported by household respondents from intervention *kebeles*

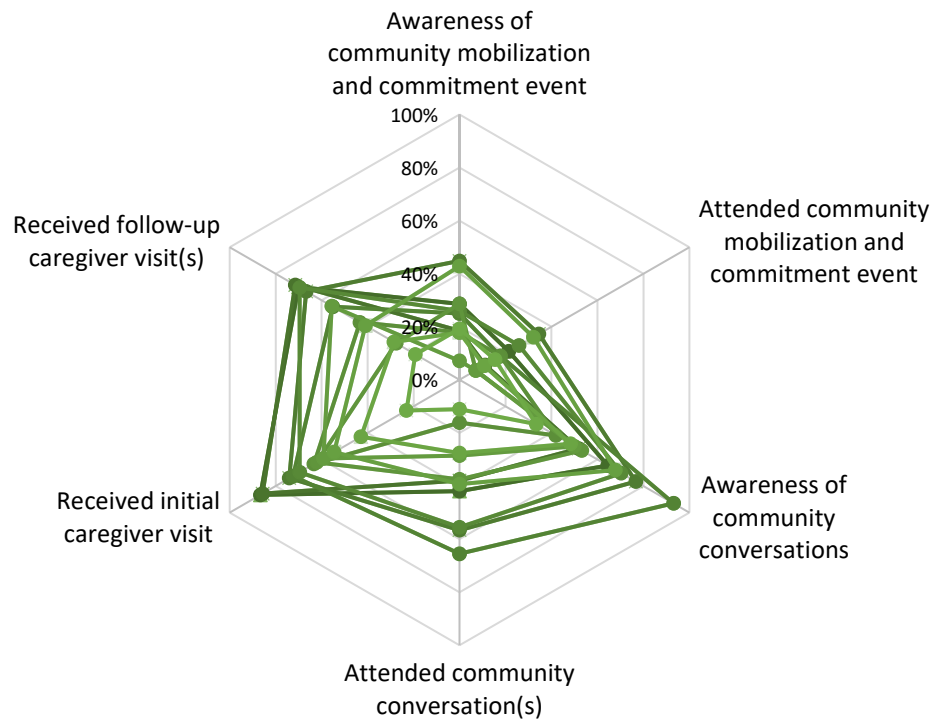
A. Bahir Dar Zuria intervention *kebeles* (N=8)



B. Fogera intervention *kebeles* (N=6)



C. Farta intervention *kebeles* (N=11)



D. Overall

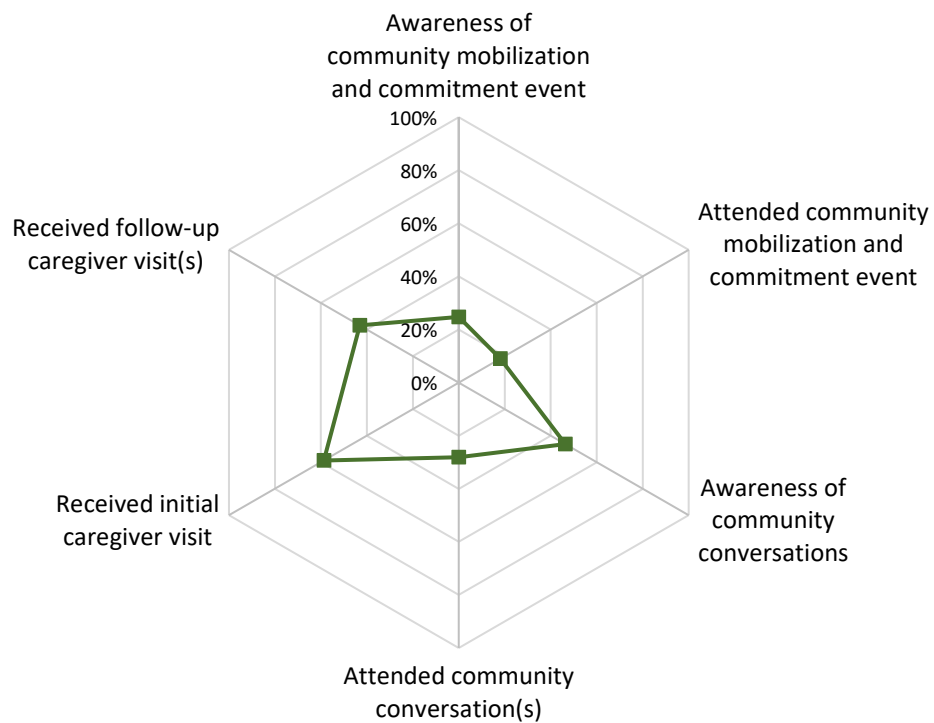


Table 3. Summary of process data

Key: ✓ Implemented as planned, ▼ Challenges in implementation, ✗ Gaps in implementation. (1), (2), and (3) used to identify contextual details for dose delivered, participation, and dose received, respectively.

Activity	Dose delivered (1)		Participation (2)		Dose received (3)		Context
	Delivery of activity	...as planned	Participation of target population	...as planned	Outcome/engagement of activity	...as planned	Details
Sensitization and action planning	All study districts	✓	47 regional, <i>woreda</i> , and <i>kebele</i> stakeholders	▼	Action plan completed for district-level activities	✓	▼ (2) One intervention <i>kebele</i> had an administrator not in attendance; ARHB representative in 2/3 workshops
Skills-based training and refresher of trainers	All study districts	✓	Average of 2-3 HEWs per <i>kebele</i> (N=56), 18 CHC HEW Supervisors, 3 Health Office Heads, and 3 Hygiene and Sanitation Officers	✓	Action plan completed for household-level activities	✓	✓ (2) A “mop up” training was conducted to assure all target population were trained; multiple trainings were conducted in districts to assure less than 30 trainees per session
Skills-based training and refresher training of community conversation facilitators	All study districts	▼	4 facilitators per <i>kebele</i> trained (N=100)	✓	Action plan completed for group-level activities	✓	▼ (1) Delays in scheduling refresher training in one study district (conducted one month prior to endline) ✓ (2) Multiple trainings were conducted in districts to assure less than 30 trainees per session
Adaptive management workshop	All study districts	▼	71 <i>woreda</i> and <i>kebele</i> stakeholders	✓	Action plan for ways forward in overcoming challenges in district, community, group, and household-level activities	✓	▼ (1) Delays in scheduling workshop in one study district (conducted one month prior to endline)

Activity	Dose delivered		Participation		Dose received		Context
Community-level	Delivery of activity	...as planned	Participation of target population	...as planned	Outcome/engagement of activity	...as planned	Details
Whole system in the room and action planning	All intervention <i>kebeles</i>	✓	Average of 8-9 stakeholders per <i>kebele</i> (N=233)	✓	Action plan completed for selection of community coordinating committee, community conversation facilitators, and WDALS	✓	<p>✓ (1) Facilitation of the workshop was planned to be conducted by <i>Kebele</i> Administrators and relevant <i>woreda</i> officials, but one <i>kebele</i> had an administrator not facilitating the workshop and only 2/25 <i>kebeles</i> had a representative from the <i>woreda</i> in attendance. Thus, the <i>Andilaye</i> team increased its role in facilitating when local government officials could not take on this role</p> <p>✓ (3) 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 complete the action plan</p>
Skills-based training and refresher training of WDALS	All intervention <i>kebeles</i>	✓	Average of 9-10 WDALS per <i>kebele</i> (N=236)	✓	HEWs and WDALS received routine supportive supervision and on-the-job training to conduct counseling visits with caregivers	✗	<p>✓ (1) Training of WDALS were planned to be facilitated by HEWs with help from CHC HEWs Supervisors, but the involvement of CHC HEWs Supervisors was minimal in some <i>kebeles</i>. Thus, the <i>Andilaye</i> team trained WDALS together with HEWs</p> <p>✗ (3) At endline, only 20% (5/25) of HEWs interviewed reported receiving supportive supervision from CHC HEWs Supervisors and only 43% (43/100) of WDALS interviewed reported receiving supportive supervision from HEWs</p>
Community mobilization and commitment event	All intervention <i>kebeles</i>	✓	Average of 309 community members per <i>kebele</i>	✗	Public commitment of improved WASH practices and regulations for monitoring by-laws	✗	<p>✗ (2) Only 9/25 <i>kebeles</i> had all coordinating committee in attendance during the event; mobilizing community members (tasked to the committee) presented a challenge in participation. During quarterly monitoring visits, only 18% (130/703) of household respondents reported that they attended the event; conducting events on working days may have decreased mobilization</p> <p>✗ (3) Four <i>kebeles</i> did not conduct the commitment ceremony at the end of the event due to concern with the political situation or decision of community leaders to agree on by-laws later</p>

Activity	Dose delivered		Participation		Dose received		Context
Group-level	Delivery of activity	...as planned	Participation of target population	...as planned	Outcome/engagement of activity	...as planned	Details
Community conversations	23/25 intervention <i>kebeles</i>	▼	Average of 53 community members per <i>kebele</i>	▼	Perceptions improved regarding individual and community capabilities to adopt and maintain improved WASH practices	▼	<p>▼ (1) Despite the provision of trainings and in-kind motivators, two <i>kebeles</i> failed to implement the activity before endline. Of <i>kebeles</i> with activity implemented, the average number of community conversation sessions was 7-8 (about one per month – half of what was planned)</p> <p>▼ (2) Overall, facilitators reported that punctuality and motivation of participants as a challenge in conducting community conversations. During endline visits, 28% (199/708) of household respondents reported attending at least one community conversation</p> <p>▼ (3) 80% (132/165) of community conversations had demonstration activities conducted to dispel misconceptions of improved WASH practices; during endline visits, 46% (329/708) of household respondents reported hearing about the community conversations, suggesting an engagement in community group dialogue of those that did participate</p>
Household-level	Delivery of activity	...as planned	Participation of target population	...as planned	Outcome/engagement of activity	...as planned	Details
Counseling visits with caregivers	All intervention <i>kebeles</i>	✗	59% of caregivers from study households reported counseling visits	✗	Counseling provided to caregivers to equip them with the knowledge, skills, motivation, barrier planning, and self-regulation necessary to adopt and maintain improved WASH practices	▼	<p>✗ (1) No <i>kebeles</i> had WDALs conducting monthly counseling visits. Of households with at least one visit, the average number of visits was 2-3 (one fifth of what was planned)</p> <p>✗ (2) During endline, only 59% of household respondents reported that they received a counseling visit; only 63% (419/665) of households were able to identify the WDALs responsible for conducting counseling visits</p> <p>▼ (3) Of households reporting at least one visit 72% reported that they set household goals or incremental improvements, and two-thirds reported that they identified barriers</p>

5.8. Photos of *Andilaye* implementation



Community members gather underneath a tree for an Andilaye community mobilization and commitment event.



Members of a performance group act out role play skits during an Andilaye community mobilization and commitment event. Performed skits align with each Andilaye intervention behavioral domain.



By signing the Andilaye banner, community members demonstrate their willingness to commit to actively aspiring toward goals of improved sanitation and hygiene practices, and cooperating with community change agents such as WDALs, HEWs, community conversation facilitators, and other key stakeholders.



Emory Ethiopia staff member Mulat Woreta supervises a WDAL training on Andilaye counseling visits with caregivers. Here, two HEWs role play the tailored (behavioral theme and need-specific) inter-personal counselling using the Andilaye Gobezi! (Good job!) Flipbook.



In-kind umbrellas (modeled by Emory Ethiopia staff: Mulat Woreta, Tigist Bahiru, and Kassahun Zewudie) were distributed to HEWs, WDAs, and community conversation facilitators to help shade them from the sun as they conduct Andilaye group and household-level activities.



Anthropologist Yihenew Tesfaye from Oregon State University sits down with a group of WDAs trained on Andilaye to discuss the intervention. Here, the group can be seen with their Andilaye Gobez! (good job!) flipbooks as they partake in an Andilaye community mobilization and commitment event.



Each intervention community has four community conversation facilitators trained on the group-level Andilaye intervention activity. Here, Emory Ethiopia staff member Mulat Woreta interviews a Kebele Police Officer trained on Andilaye community conversations in front of the Andilaye banner at the community's health post.



A group of HEWs discuss upcoming plans for Andilaye intervention activities.



Trained Andilaye community conversation facilitators discuss challenges, successes, and next steps in implementing Andilaye community conversations during a refresher training.



The team of enumerators, field supervisors, and Emory Ethiopia staff which led data collection in the study enrolled households for process evaluation and impact evaluation.

6. Impact evaluation: results

6.1. Survey results

A total of 1,589 households were enrolled into the study at baseline. At midline, 27 (2%) of these households were lost-to-follow up (i.e., did not have a member between 1-10 years of age or moved away from the *kebele*) and were therefore no longer eligible for the study. Thus, at endline, we visited a total of 1,562 households; 16 households did not have a member between 1-11 years of age, 40 had no eligible respondent available, 14 did not consent to take the survey, and 20 had moved away from the *kebele*. The endline results reflect complete data from the remaining 1,472 households (94%) (Figure 5). Given we prioritized targeting of the primary female caregiver of the index child for endline survey administration, a large majority (90%) of the respondents were female, by design. Of these 1,472 respondents, 85% were the mother of the index child – identical to baseline (Table 2).

6.2. WASH coverage

Impacts on sanitation coverage

At endline, we observed that 62% of households had at least one latrine (Table 4), which was similar to baseline (66%; Table 2). Latrine coverage was also similar, when comparing intervention and control arms at midline (PR=0.99; 95% CI: 0.82, 1.21). 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, 33% met criteria of an improved facility based on the JMP definition (World Health Organization and UNICEF, 2013), which was seven percentage points lower than at baseline. There was no difference in the prevalence of improved latrines, when comparing intervention and control arms (PR=1.14; 95% CI: 0.82, 1.60). Similarly, there was no difference in the prevalence of households with fully constructed latrines (PR=1.15; 95% CI: 0.87, 1.53). As for secondary outcomes, there were no differences in the prevalence of latrines with smooth, cleanable surfaces, when comparing intervention to control, with only 15% 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 7A. 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.

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=906; Table 4). 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=1.81; 95% CI: 1.21, 2.71). Intervention households were more likely to have water available for handwashing near or inside the latrine (PR=2.28; 95% CI: 1.08, 4.81), cleansing agents for handwashing near or inside the latrines (PR=21.7; 95% CI: 2.17, 216), and water available inside or near the latrines for flushing or self-cleansing (PR=2.35; 95% CI: 1.05, 5.28). 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, 82% of intervention latrines still had flies present, only 18% had a drop hole cover, and 50% 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.

Impacts on washing station coverage

Presence of hand and facewashing stations were reported and observed in 98% of households, although presence of water, and presence of soap were observed in only 20% and 2% of stations, respectively (Table 4). The prevalence of hand and facewashing stations was similar between the two study arms (PR=1.01; 95% CI: 0.99, 1.02), and the prevalence of stations with water was also similar when comparing the two arms (PR=0.96; 95% CI: 0.72, 1.26). The prevalence of stations with soap was higher in the intervention arm, although only 3% of households in this arm had a hand or facewashing station with soap present.

6.3. WASH-related behaviors

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 respondent urination near surface water (PR=0.82; 95% CI: 0.46, 1.48), respondent open defecation (PR=1.05; 95% CI: 0.76, 1.45), head of household open defecation (PR=0.98; 95% CI: 0.80, 1.21), children openly defecating (PR=0.95, 95% CI: 0.77, 1.16), disposal of child feces (PR=0.96; 95% CI: 0.69, 1.32; Figure 7B), and number of households sharing a sanitation facility (difference=0.11; 95% CI: -0.42, 0.19; Appendix C – Supplemental Table 3). Overall, 40% of respondents' primary place of defecation during the last two days was in the open, and only 46% of respondents had defecated in any latrine during the last two days (Table 4).

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. Unlike midline (data not shown), this finding was not statistically significant across the multiple measures of handwashing (Table 4). When reporting on their last handwashing experience, respondents reported washing their hands with soap/ash/soapy water 44% of the time (Figure 7C); they reported washing their hands after defecation 49% of the time; and they reported washing their hands before food preparation 51% of the time. 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 31% of the time (Figure 7D). The prevalence of these reported practices were similar in both study arms. At midline, 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) (data not shown). Although there was a slight increase in children reportedly had their hands washed after defecation overall at endline (37% to 42%), this reported practice was no longer statistically different at endline, when comparing intervention and control arms.

There were 1,696 children aged 1-9 years whose facial cleanliness was observed. Observations of facial cleanliness showed ocular discharge among 29% of children, wet nasal discharge among 40%, dry nasal discharge among 46%, and dust, dirt or other debris on 53% of children. There were no meaningful differences between the intervention and control arm for any of these facial cleanliness measures. When restricting these analyses to just the 1,024 index children who were present for endline observation, facial cleanliness results were similar to those provided by all children aged 1-9 years. Overall, 99% 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 3, and was similar in both the intervention and control groups (difference=-0.15; 95% CI: -0.90, 0.60).

The intervention had little measurable 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 (see Appendix C – Supplemental Table 6).

Impacts on household environmental sanitation

Across both arms, the majority of both respondents and heads of household had animal herding responsibilities (88% overall), and animal feces were present in the compound in 82% of the households (Appendix C – Supplemental Table 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 4). This was similar between the intervention and control arms (PR=1.10; 95% CI: 0.95, 1.28). All other environmental sanitation measures were quite similar, when comparing intervention and control arms.

6.4. Respondent-reported diarrhea

Diarrhea prevalence (7% overall) was similar in the intervention and control arms (PR=1.20; 95% CI: 0.74, 1.93) 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 5). Among index children, there were also similarities comparing intervention and control arms in diarrhea prevalence when assessing episodes over the last two days (5% overall) (PR=1.25; 95% CI: 0.71, 2.22) and the prevalence of diarrhea with blood in the stool (2% overall) (PR=0.67; 95% CI: 0.26, 1.71). Blood in the stool over the last seven days was, however, statistically lower among all HH members age 0-10 years in the intervention arm compared to the control arm (PR=0.47; 95% CI: 0.23, 0.93).

6.5. Mental well-being

Mental health scores

The scores for anxiety, depression, and emotional distress among respondents were all lower at endline than baseline among the overall population, indicating better overall status of each condition assessed (Table 2, Table 8). Using the intention-to-treat analysis on our endline data, the mean scores were similar between the intervention and control arm for the anxiety score (difference= -0.026, 95% CI: -0.08, 0.02), depression score (difference= -0.029, 95% CI: -0.07, 0.01), and the emotional distress score (difference= -0.023, 95% CI: -0.06, 0.01). However, each of the estimates trended in the

preventive direction (Table 8). Box and whisker plots indicate similar distributions between the two arms (Appendix C – Supplemental Figure 1). The WHO-5 well-being scores were also similar, when comparing intervention and control arms (difference= -0.028, 95% CI: -0.08, 0.02). Similarly, scores at endline trended towards better mental health compared to baseline scores (Table 9, Appendix C – Supplemental Figure 1).

When assessing the prevalence (rather than a score) of anxiety, depression, emotional distress and well-being, there were no differences in the prevalence of any of these variables comparing the intervention and control arms (Table 8).

Water and sanitation insecurity

Sanitation insecurity scores related to social support were lower in the intervention arm than in the control arm (-0.10, 95% CI: -0.16, -0.43), indicating a reduced frequency of experiencing the circumstances in the social support domain (e.g. trouble finding support to watch dependents during urination, worry about dependents when going to defecate, Had to leave dependents alone to urinate, etc.) but all of the other sanitation insecurity and water insecurity measures were similar between arms (Table 6; Table 7).

6.6. 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 CLTSH 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 if there was effect modification of the intervention by respondent sex for each of the mental health outcomes, and we did not detect statistically significant effect measure modification by sex for any of the four mental health outcomes. Similarly, when assessing whether the child's sex modified the effect of the intervention on our primary outcome variables related to children, we did not detect interaction by sex for any of these outcomes. We also did not detect interaction between the intervention and water insecurity on any of the primary handwashing or facewashing variables.

Table 4. Endline levels of the WASH indicators of primary interest

	Overall		Intervention		Control			
	N	%	N	%	N	%	PR (95% CI) ^a	PD (95% CI) ^b
Latrine coverage observations								
Households with at least one latrine (%)	1472	61.6	743	61.2	729	62.0	0.99 (0.82, 1.21)	-0.004 (-0.125, 0.118)
Households with improved latrine (%) ^c	1467	32.5	741	34.6	726	30.6	1.13 (0.81, 1.59)	0.041 (-0.070, 0.151)
Households with fully constructed latrine (%)	1471	30.9	742	33.0	729	28.7	1.15 (0.86, 1.54)	0.044 (-0.462, 0.134)
Sanitation facility operation and maintenance								
HH has reported adding or improving anything on this latrine since its original construction (%)	899	16.5	453	17.2	446	15.7	1.08 (0.71, 1.65)	0.012 (-0.056, 0.080)
Observed latrine characteristics								
Presence of drop hole cover in the latrine (%)	906	14.1	455	18.2	451	10.0	1.77 (1.19, 2.63)	0.079 (0.020, 0.137)
Among those with a drop hole, a cover was situated over drop hole (%)	130	67.7	85	70.6	45	62.2	1.07 (0.86, 1.34) ^d	0.045 (-0.097, 0.187)
Respondent reported defecation								
Respondent's primary place of defecation was OD during last 2 days (%)	1472	39.5	743	40.2	729	38.8	1.05 (0.76, 1.45)	0.020 (-0.106, 0.145)
Respondent defecated in any latrine during last 2 days (%)	1469	45.5	740	45.8	729	45.3	1.01 (0.79, 1.29)	0.003 (-0.110, 0.115)
Safely disposed of child feces (%)	777	38.9	401	36.7	376	41.2	0.96 (0.69, 1.32)	-0.017 (-0.145, 0.111)
Respondent reported latrine sharing								
Given HH has a latrine, number of people who used this latrine from ANOTHER HH during last 7 days, not including your HH members	902	mean (SE)	454	mean (SE)	448	mean (SE)	-	difference (95% CI) ^e
	902	0.91 (0.14)	454	0.71 (0.16)	448	1.11 (0.22)	-	-0.400 (-0.847, 0.049)
Animal husbandry/other HH sanitation								
Animal feces/waste not observed out in open in compound (%)	1472	53.8	743	56.4	729	51.2	1.10 (0.95, 1.28)	0.052 (-0.029, 0.132)
Observed facial cleanliness among children ages 1-9								
Clean face, defined as absence of nasal and ocular discharge (wet or dry) from the face (%)	1696	43.2	822	40.7	874	30.4	1.07 (0.88, 1.29)	0.028 (-0.052, 0.108)
Ocular discharge is present (%)	1696	28.7	822	26.9	874	30.4	0.88 (0.68, 1.15)	-0.037 (-0.114, 0.040)
Wet nasal discharge is present (%)	1696	38.2	822	37.0	874	39.4	0.94 (0.78, 1.13)	-0.024 (-0.094, 0.045)
Dry nasal discharge is present (%)	1696	44.0	822	42.7	874	45.2	0.97 (0.81, 1.16)	-0.016 (-0.095, 0.063)
Dirt/dust/other debris is present (%)	1696	50.0	822	50.5	874	49.5	1.03 (0.89, 1.20)	0.016 (-0.058, 0.090)
Objective hygiene metric among children ages 1-9								
Facial cleanliness, as measured via a novel personal hygiene metric (i.e., 11-point brown scale)	1010	mean (SE)	490	mean (SE)	520	mean (SE)	-	difference (95% CI) ^e
	1010	5.2 (0.09)	490	5.2 (0.12)	520	5.2 (0.12)	-	-0.018 (-0.339, 0.304)
Hand cleanliness, as measured via a novel personal hygiene metric (i.e., 11-point brown scale)	1011	mean (SE)	490	mean (SE)	521	mean (SE)	-	difference (95% CI) ^e
	1011	2.7 (0.09)	490	2.6 (0.16)	521	2.7 (0.09)	-	-0.068 (-0.407, 0.272)

Fly observations	N	mean (SE)	N	mean (SE)	N	mean (SE)	-	difference (95% CI) ^e
Number of times flies landed on index child's face during minute observation	1024	3.3 (0.21)	502	3.2 (0.33)	522	3.4 (0.26)	-	-0.150 (-0.900, 0.603)
Washing station coverage	N	%	N	%	N	%	PR (95% CI) ^a	PD (95% CI) ^b
Reported HH hand or facewashing station(s) (%)	1472	98.0	743	98.3	729	97.7	1.01 (0.99, 1.02)	0.008 (-0.008, 0.023)
Respondent reported handwashing practices	N	%	N	%	N	%	PR (95% CI) ^a	PD (95% CI) ^b
The last time the respondent washed he/she used soap/ash/soapy water (%)	1468	44.0	740	46.0	728	42.0	1.09 (0.91, 1.31)	0.039 (-0.042, 0.120)
The last time the respondent defecated, he/she cleaned hands with water and soap, substitute (%)	1463	49.0	738	51.9	725	46.1	1.13 (0.94, 1.35)	0.058 (-0.032, 0.149)
The last time the respondent prepared food, he/she cleaned hands with water and soap, substitute before beginning food preparations (%)	1403	51.0	700	53.6	703	48.5	1.11 (0.95, 1.29)	0.051 (-0.027, 0.130)

Notes. ^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 7A for all latrine type categories. ^d Robust standard error not used for this model to allow for model convergence. ^e We used similar linear regression models to estimate difference comparing the outcomes between the intervention and control arms

Table 5. Endline respondent-reported diarrheal outcomes

	Overall		Intervention		Control			
Among all HH members 0-10 years	N	%	N	%	N	%	PR (95% CI) ^a	PD (95% CI) ^b
In the LAST 2 days, HH member had three or more loose stools per day (%)	3017	6.8	1507	6.9	1510	6.6	1.04 (0.69, 1.58)	0.003 (-0.024, 0.030)
During the last 7 days, including today, HH member had three or more loose stools per day (%)	3017	8.1	1506	8.2	1511	8.1	1.01 (0.71, 1.44)	0.001 (-0.027, 0.029)
During the last 7 days, including today, HH member had blood in the stool (%)	3012	2.4	1503	1.5	1509	3.3	0.47 (0.23, 0.93)	-0.018 (-0.038, 0.002)
Among index children	N	%	N	%	N	%	PR (95% CI) ^a	PD (95% CI) ^b
In the LAST 2 days, child had three or more loose stools per day (%)	1450	5.3	730	5.9	720	4.7	1.62 (0.71, 2.22)	0.012 (-0.018, 0.042)
During the last 7 days, including today, child had three or more loose stools per day (%)	1452	6.5	731	7.1	721	6.0	1.20 (0.74, 1.93)	0.012 (-0.019, 0.043)
During the last 7 days, including today, child had blood in the stool (%)	1449	2.1	730	1.6	719	2.5	0.67 (0.26, 1.71)	-0.008 (-0.031, 0.014)

Notes. ^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 6. Endline respondent-reported sanitation insecurity

Sanitation insecurity^b	Overall		Intervention		Control		difference (95% CI)^a
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	
Factor 1-Potential harms	692	0.48 (0.021)	365	0.46 (0.026)	327	0.50 (0.033)	-0.050 (-0.130, 0.030)
Factor 2-Social expectations resultant repercussions	694	0.29 (0.017)	366	0.28 (0.025)	327	0.30 (0.022)	-0.029 (-0.089, 0.030)
Factor 3-Physical exertion or strain	693	0.41 (0.031)	366	0.42 (0.046)	328	0.40 (0.043)	0.010 (-0.110, 0.129)
Factor 4-Night concerns	694	0.35 (0.018)	366	0.32 (0.022)	328	0.37 (0.027)	-0.050 (-0.119, 0.019)
Factor 5-Social support	694	0.15 (0.017)	366	0.10 (0.021)	328	0.20 (0.023)	-0.103 (-0.162, -0.433)
Factor 6-Physical agility	694	0.14 (0.013)	366	0.14 (0.017)	328	0.14 (0.020)	0.000 (-0.050, 0.051)
Factor 7-Defecation place	693	0.34 (0.024)	366	0.35 (0.038)	327	0.32 (0.028)	0.024 (-0.058, 0.106)

Notes. ^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., 2017a). A higher score represents higher sanitation insecurity.

Table 7. Endline respondent-reported water insecurity

Respondent reported water insecurity	Overall		Intervention		Control		PR (95% CI)^a	PD (95% CI)^b
	N	%	N	%	N	%		
Water insecure for any of the four reported insecurity measures (%)	1385	25.6	698	21.1	687	30.3	0.71 (0.49, 1.02)	-0.091 (-0.186, 0.004)

^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

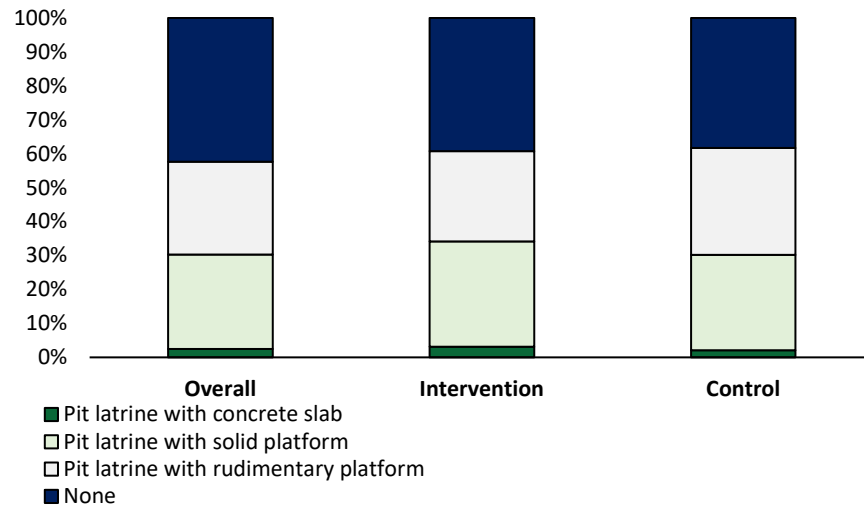
Table 8. Endline respondent-reported mental health outcomes

	Overall		Intervention		Control		Relative measure	Absolute measure
Mental health scores	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)		difference (95% CI) ^a
Anxiety score ^b	1471	1.49 (.63)	742	1.46 (.61)	729	1.52 (.64)	-	-0.059 (-0.136, 0.018)
Depression score ^b	1470	1.37 (.50)	742	1.35 (.48)	728	1.39 (.52)	-	-0.036 (-0.081, 0.010)
Emotional distress score ^b	1469	1.31 (.48)	741	1.29 (.46)	728	1.33 (.49)	-	-0.042 (-0.093, 0.008)
WHO-5 well-being score ^c	1471	17.3 (6.8)	749	17.6 (6.8)	728	17.0 (6.7)	-	-0.545 (-0.187, 1.277)
Score above cutoff indicating poor mental health	N	%	N	%	N	%	PR (95% CI) ^d	PD (95% CI) ^e
High Anxiety ^f	1471	23.5	742	22.2	729	24.8	0.90 (0.72, 1.11)	-0.026 (-0.076, 0.024)
High Depression ^f	1470	15.4	742	14.0	728	16.9	0.83 (0.64, 1.07)	-0.029 (-0.067, 0.010)
High Emotional distress ^f	1469	15.2	741	14.0	728	16.4	0.86 (0.67, 1.09)	-0.023 (-0.060, 0.014)
Poor well-being ^g	1471	26.4	749	25.2	728	27.8	0.90 (0.74, 1.10)	-0.028 (-0.079, 0.024)

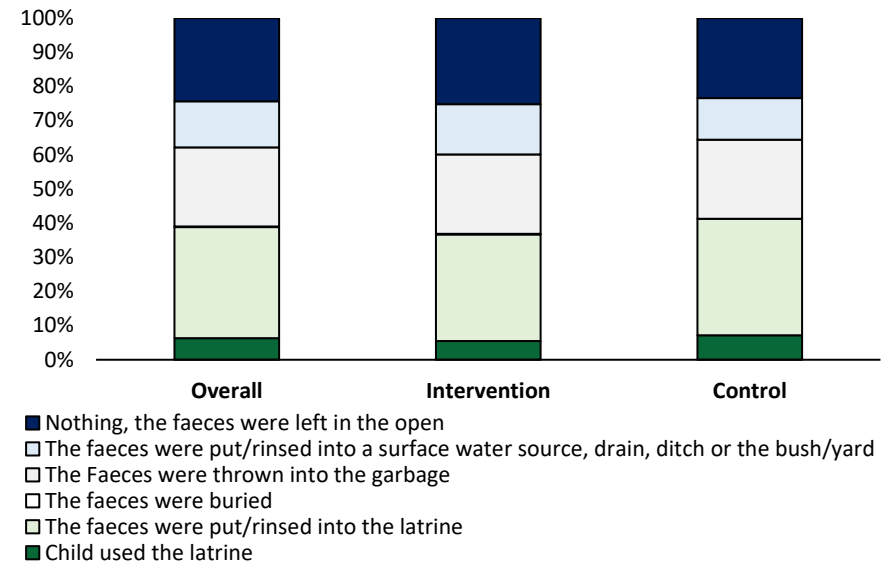
Notes. ^a We used linear regression models to estimate the change in the outcomes comparing the intervention and control arms. Models accounted the stratified design by including woreda indicator variables, (Kahan and Morris, 2012) and accounted for clustering within kebeles by using generalized estimating 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. ^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. ^d 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. ^e We used similar linear regression models to estimate difference comparing the outcomes between the intervention and control arms ^f Each of these scores was dichotomized, with scores greater than 1.75 indicating a positive status for any of the three outcomes. ^g Each of these scores was dichotomized with scores below 13 indicating poor well-being.

Figure 7. Endline distribution of select WASH indicators

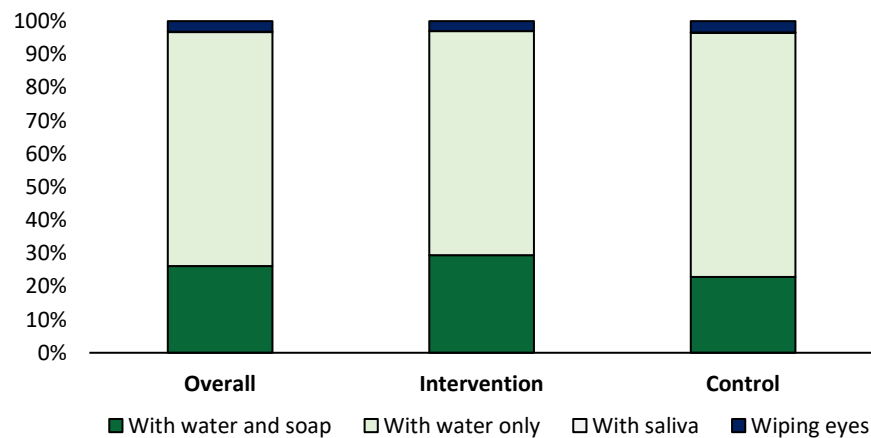
A. Latrine type distribution



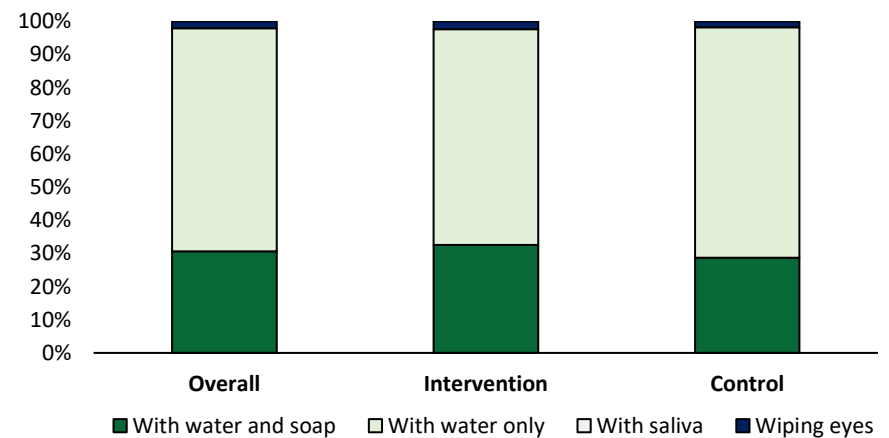
B. Disposal of child feces for children under 4 years of age (n=777)



C. Method respondent used to clean face during the day preceding the survey



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



7. Discussion

7.1. Interpretation of study findings

Few of our primary and secondary outcomes were impacted by our intervention, including sanitation and hygiene behaviors such as latrine use, improvements to latrine conditions, facewashing and handwashing, and compound cleanliness, which was perhaps limited by our evaluation period (14-15 months). Without changes to these WASH conditions and behaviors, changes in mental well-being are not likely, and indeed, were not detected here. The *Andilaye* intervention employed behavioral theory and utilized extensive formative research and piloting. However, the potential effectiveness of the intervention could not be evaluated as fidelity - defined as “the degree to which the intervention or program was delivered as intended (Carroll et al., 2007)” - was not maintained after training was delivered by the *Andilaye* team. Few rigorously evaluated WASH interventions are delivered directly through government structures, and these limited results of poor intervention fidelity are consistent with several other large-scale WASH effectiveness studies (Sinha et al., 2017, Clasen 2014, Patil 2014). A notable intervention is the *SuperAmmu* study in India which used behavioral theory to rollout an at-scale intervention that showed substantial behavior change (Biran et al., 2014). Studies from Bangladesh, Ethiopia, Uganda, Kenya, and Sierra Leone (and our formative research) have shown that lack of affordability of resources for latrine construction or handwashing is the top reason why behavior slippage occurs after CLTS programs are implemented (Tyndale-Biscoe, 2013, Hanchett, 2011). These studies also showed that security of land or land-ownership status also affected whether people would invest in improving their latrines. Perhaps these constraints, along with our evaluation period, may have affected the magnitude of behavior change and maintenance observed in our study.

Several recent high-profile studies have sought to evaluate the effects of single and combined WASH and nutritional interventions on child illness and growth. Even with high fidelity and adequate behavior change, these studies revealed inconsistent impact on infectious disease outcomes including growth and diarrhea (Luby et al., 2018a, Null et al., 2018, Humphrey et al., 2019), though results on soil-transmitted helminth infections were more encouraging (Pickering et al., 2019a, Ercumen et al., 2019). In all of these studies, the WASH interventions (water chlorination; sanitation; handwashing with soap) were limited to the household or compound of the index child. Considering these results, a recent call for future research in the WASH sector has been made to focus on developing and evaluating interventions that are “radically more effective in reducing fecal contamination in the domestic environment than the interventions implemented in these trials” (Pickering et al., 2019b). Although poor intervention fidelity has limited our ability to determine the effectiveness of the intervention, our novel holistic community-based WASH intervention embodies several elements of “Transformative WASH” presented in this call, including: community coverage of improved sanitation facilities; complete separation of animal feces from people's living environments; reductions in fecal contamination on surfaces where young children crawl and play; and different modalities of behavior change.

Most notably, to our knowledge, no previous study has measured the impact of a sanitation and hygiene intervention on mental well-being using an experimental design. Our study proposes that quantifying the relationships between WASH improvements and these non-traditional outcomes and

impacts (e.g., mental well-being and insecurity) can inform programs and policies that facilitate health equity.

7.2. Holistic WASH programming and the HEP

The purpose of our intervention was to determine if we could design an effective intervention in the context of Ethiopia's HEP. This integration was advocated from key stakeholders at the national and regional level^{3,4}. We could not determine effectiveness, as fidelity was too low to show changes in behavior at the household. Findings from our trial may reveal factors contributing to poor dose delivery in *Andilaye* intervention communities, but also poor dose delivery of the HEP and CLTSH programming more broadly (Assefa et al., 2019, Gebremariam et al., 2018, Gebremariam and Tsehaye, 2019, Snel and Jacimovic, 2014) (Assefa et al., 2019; Gebremariam et al., 2018; Gebremariam and Tsehaye, 2019; Snel and Jacimovic, 2014). While supportive supervision considerations were acknowledged and incorporated into the design of the *Andilaye* intervention, these requirements did not go above and beyond what is expected of the HEP (Workie, 2013). When community health workers are paid to deliver the intervention, there is evidence of successful delivery (Fankhauser et al., 2019). However, due to a number of reasons, and as verified in our trial, there is limited support extended to HEWs as designed in the HEP. Further, a majority of intervention *kebeles* had non-functional and non-active WDALs at the start of implementation. While higher WDA strategy implementation strength, measured through WDAL density, has been shown to be associated with improved HEP outreach activities, it is also suggested to be pro poor (Damtew et al., 2018). The intervention addressed barriers in WDAL illiteracy, and lack assertiveness and confidence through on-the-job training and the use of illustrative behavior change tools to guide them in their activities. However, outside of the provision of in-kind motivators and per diem for trainings, the intervention did not incorporate investments in pay for these lower-level health workers – in accordance with government policies. Although HEWs were paid health workers, WDALs were not. This has brought questions of ethics and sustainability as WDALs are increasingly asked to provide more and more services. Recent qualitative and quantitative studies suggest that unpaid WDALs are actually worse off than their peers and makes women, especially unmarried women, vulnerable to negative gossip and psychological distress (Maes et al., 2019, Maes et al., 2018). Although this point goes much deeper into the political economy, it is an important gap to bring up in the context of empowering women volunteers to enact positive change in her community (Closser et al., 2019). Together, these findings raise questions about the possibility of bringing new programs and approaches to the HEP without adequate support.

7.3. Mental well-being

We added validated mental well-being measure to this WASH impact evaluation to determine if changes to sanitation access and sanitation security that we anticipated to be generated by this intervention would lead to improved mental health states, including improved well-being and reduction in symptoms associated with anxiety, depression, and distress. A cross-sectional study in rural India found that women's experiences of sanitation, as measures by a validated sanitation

³ 2016 [letter](#) from FMOH NTD focal point for the Disease Control Directorate to ARHB

⁴ 2016 [letter](#) from ARHB Core Process Owner of Community Health Programs to Farta, Fogera, and Bahir Dar Zuria Health Offices

insecurity measure, were associated with well-being, anxiety, depression, and distress, even when they had access to a facility (Caruso et al., 2018). These findings highlighted the need for interventions and programs to consider experience of sanitation beyond access to a facility alone. This study is one of the first experimental studies to assess the impact of sanitation and hygiene improvements on sanitation insecurity and on mental well-being. However, our intention to treat analysis did not result in changes to sanitation insecurity scores or to mental health scores. We cannot weigh in on our hypothesis that sanitation would impact sanitation insecurity and mental well-being as our intervention did not change sanitation coverage, quality, or access. As such, we should not expect to see a change in sanitation insecurity or mental well-being. Overall, the scores for anxiety, depression, and emotional distress were lower, though not significantly, at endline than baseline among all survey respondents in both arms, demonstrating improvement. The mental well-being scores were similar between intervention and control arms, although each of the estimates trended toward an improved direction. Improvement in scores for anxiety, depression and emotional distress in the overall population from baseline may be attributed to variation in non-controllable factors related to year on year changes, such as drought, food security, etc.

7.4. WASH behavior change

Observed and reported sanitation coverage and utilization were similar between intervention and control communities at endline. These findings are not surprising given the gaps in intervention fidelity previously discussed. Access to any household latrine was lower than baseline, with only 62% of all study households having access. This indicates slight behavioral slippage in household latrine coverage from baseline (66%). Only 46% of endline respondents had defecated in any latrine during the last two days. The conditions of the household latrine in intervention communities are still inadequate. All measures of latrine utilization were similar in intervention and control communities. Amongst households with any latrine, only 33% met criteria of an improved facility, per the JMP definition – this was seven percent lower than baseline. There was also no difference in the proportion of household latrines with a smooth and cleanable platform surface between baseline and endline. There were, however, increases 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 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 had a higher prevalence of water and soap present.

Personal hygiene facilities coverage and practices were similar between intervention and control communities at endline. While presence of washings station was observed to be high (98%), 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 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%. 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 29% of 1,696 children aged 1-9 years, wet nasal discharge present amongst 38%, and dry nasal discharge amongst 44%.

Household environmental sanitation indicators were also similar between intervention and control communities, signaling the need for further improvement. Animal feces were present in the compound in 83% of the households. However, 54% of households managed the feces in some way (i.e., did not leave it out in the open).

We measured and tracked individual, household, and community-level changes in key outcomes over multiple time points (i.e., baseline, midline, quarterly monitoring, and endline). Although there were few differences between intervention and control communities at endline, average changes over time for primary outcomes targeted by the *Andilaye* intervention do suggest promising trends in intervention communities, given the challenges faced in implementation. Within the sanitation behavioral theme, percent differences between the intervention and control communities over time were greatest among households with a fully constructed latrine (+7%) and presence and use of a drop hole cover (+10% and +10%). Within personal hygiene, percent differences were greatest among children ages 1-9 with ocular discharge present (-7%) and among respondents washing their hands with soap or soap substitute in general and after defecating (+7% and +8%). Lastly, percent differences in the absence of animal feces from the household compound (i.e., no animal feces left out in the open) was +9%. We plan to conduct further analysis of these trends, along with our measurements in precursors needed to be addressed before behavior change and maintenance can occur (i.e., behavioral antecedents and determinants), which may yield important findings on mediators and moderators of sanitation and hygiene behavior in the rural Ethiopian context.

7.5. Intervention design and implementation

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, soil transmitted helminthiasis, schistosomiasis). 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 Services Package messaging. The overarching intervention approach was designed to be incorporated into prevailing programs (e.g., HEP) to demonstrate potential for scale-up.

Over the course of the *Andilaye* Trial, we observed 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-job-training tools, *Woreda* Health Office officials, CHC HEWs Supervisors, and HEWs themselves did not make supportive supervisory visits to WDALs in accordance with their action plans. Even a limited number of supportive supervisory visits seem to be prohibitive. Limited integration of *Andilaye* activities within the larger HEP during the *Andilaye* Trial may have been further compromised as funds and resources allocated for ONE WASH and other government programs were almost exclusively for water provision and did not align directly with *Andilaye* activities.

While gaps in intervention fidelity were documented overall, process data does suggest higher quality implementation of group and household-level activities in some intervention *kebeles* compared to others (Figure 5). For example, in one intervention *kebele*, more than 91% of study households reporting caregiver visits (compared to the overall average of 59%) and 68% of respondents from study households reported attending community conversations (compared to the overall average of 28%). Further, process data at the district-level suggest ownership and integration of the *Andilaye* intervention into the HEP was more successful in Farta than the other two study districts (Figure 5). We plan to further investigate our process evaluation data to identify how and why these *kebeles*/districts were more successful in implementing *Andilaye*, as designed.

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.

7.6. Study 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* (Flay and Petraitis, 1994), 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 CRTs 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 *gottis* 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). The process elements of the study help provide additional context on potential institutional factors related to the implementation of program service delivery and findings are consistent with other research in Amhara and Ethiopia.

We have collected extensive 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 allowed us to monitor seasonal trends, and therefore improve the precision of our inferences and external validity across time. We conducted 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.

7.7. Study limitations

The study has several notable limitations. First, 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. The implementation of the project faced significant delays in gaining 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.

Another limitation is that key actors were less involved than planned (e.g., *Woreda* Health Office officials, CHC HEWs Supervisors, HEWs did not conduct 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 are currently discussing the possibility to integrate *Andilaye* intervention activities into an NGO-based WASH-NTD program in Ethiopia whose delivery structures (e.g. hired independent community implementers) may yield further investigations into the effectiveness of the intervention on sustained behavior change and mental well-being. Given heterogeneity in intervention fidelity, we also plan to supplement our primary intention-to-treat analyses, presented in this report, with an analysis that attempts to assess the impact of adherence to the *Andilaye* intervention on our primary outcomes of interest (e.g., as-treated analysis). We have performed such analyses in other studies (Garn et al., 2016, Garn et al., 2017b).

The results presented here should be interpreted cautiously as there were ~30 outcomes of interest, and therefore a high possibility of some false positives (e.g., ~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 been met.

7.8. Policy influence

Throughout the *Andilaye* Trial, we worked with stakeholders at the national, regional, and local levels to disseminate key learnings of the study, including the 2018 NTD Annual Review Meeting and Research Symposium in Hawassa, Ethiopia. Our team proposed the initial idea to hold the 3rd WASH-NTD Roundtable in Addis Ababa, Ethiopia in 2018, sponsored by the FMOH. Dr. Freeman was on the coordinating committee for the meeting and presented our *Andilaye* findings at this international forum, giving us an important policy influencing platform. These findings have also been presented at relevant conferences such as UNC Water and Health, Coalition for Operational Research on NTDs (COR-NTD), Water, Engineering and Development Centre (WEDC) International Conference, and the Neglected Tropical Disease NGO Network (NNN) Annual Conference. Our overall policy-relevant findings and recommendations from the *Andilaye* Trial (see section 8) were presented to stakeholders during *woreda* and regional-level results dissemination and review meetings in June-July 2019 and to national and regional-level NTD focal points in Addis Ababa, Ethiopia in June 2019.

We anticipate the publication of several manuscripts addressing knowledge gaps related to scalable alternatives to CLTSH and sanitation and hygiene programming that we have documented during the *Andilaye* Trial. Five of which have already been published and discuss: (1) the design of the cluster-randomized trial (Delea et al., 2019), (2) innovations in demand-side sanitation and hygiene

interventions (Berhe et al., 2018), (3) the application of discrete choice experiments to elicit stated preferences for latrine use and construction (Goddard et al., 2018), (4) the development and application of scales to examine the role of collective efficacy in the effectiveness of community-based programming (Delea et al., 2018), and (5) the development and reliability of a quantitative personal hygiene assessment tool (Delea et al., 2020).

8. Policy-relevant findings and recommendations

We employed rigorous formative research and practically applied social and behavioral theory to develop *Andilaye*, a scalable intervention designed to address these issues and complement existing service delivery within Ethiopia's HEP. Limited integration of *Andilaye* activities into the HEP likely resulted in minimal impact on sustained behavior change and mental well-being. However, evidence from this trial may help address knowledge gaps related to scalable alternatives to CLTSH and inform sanitation and hygiene programming and policy in Ethiopia and beyond.

Designing “holistic” hygiene and sanitation programming:

- CLTSH has facilitated considerable changes in coverage of basic sanitation. Some of these gains have not been sustained (as indicated by our impact evaluation), and incremental improvements to improve sanitation have not been widely promoted or achieved.
- WASH programs, more broadly, have focused on catalyzing initial behavior change, and have placed little, if any emphasis on the habituation of improved behaviors and behavioral maintenance.
- Siloed approaches within the health and development sectors, namely WASH and those vertical programs involved in the control and elimination of NTDs, prevent the integration and harmonization of NTD and WASH behavior change initiatives.
- Behaviors and facilities promoted by existing programs are aspirational but require considerable effort and/or capital investment to achieve. As indicated in our impact evaluation, such approaches have fostered behavioral slippage, or regression back to unimproved behaviors and practices, and poor sustainability of behavioral outcomes and potential health impacts.
- CLTSH largely focuses on leveraging shame to change norms around open defecation. As indicated by our formative research, focus on negative affective motivators, poor facilitation of initial triggering, and a lack of post-triggering follow-up has left many communities with negative impressions of CLTSH initiatives.

As demonstrated by the Andilaye Trial...

- **Theory to practice:** The use of behavioral theory and a structured approach to intervention development – that incorporates stakeholder feedback – yields important guidance in selecting behaviors to target, identifying leading indicators of behavior change, and informing the program's approach.
- **Behavioral maintenance:** To prevent behavioral slippage, there is a need for approaches that complement CLTSH and focus on behavioral maintenance.
- **WASH-NTD programming:** 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.
- **Incremental improvements:** A focus on small, incremental improvements in WASH practices and facilities may be viewed as more achievable by program participants, particularly in low resource settings, and as such, may garner greater success.

- **Positive motivators:** Negative affective motivators may not be the most appropriate or effective drivers of change (especially in the Ethiopian context), and may actually erode mental well-being.

Addressing “implementation gaps” of hygiene and sanitation programming:

- HEWs charged with implementing CLTSH have many responsibilities, few tools, and little capacity to continually reinforce messages. Although *Woreda* Health Offices and CHCs are expected to closely support and monitor HEWs, due to a number of reasons, and as verified in our process evaluation, there is limited support extended to them as designed in the HEP.
- While supportive supervision considerations were acknowledged and incorporated into the design of the *Andilaye* intervention, these requirements do not go above and beyond what is expected of the HEP. In addition, the *Andilaye* intervention provides actual tools and protocols for the execution of supportive supervision. However, even a limited number of supportive supervisory visits from the *Woreda* Health Offices and CHCs seem to be prohibitive.
- *Woreda* Health Offices and CHCs state that though the WDAL system should be in place throughout the region (as outlined in the HEP), the system is not well established in some areas. Consequently, a majority of *Andilaye* communities had non-functional and non-active WDALs at the start of implementation. This is not unique for *Andilaye* and is an issue for all types of community health delivery modalities.

As demonstrated by the Andilaye Trial...

- **Over-extension of HEWs:** Although over-extension of HEWs was addressed through the engagement of additional community change agents for group-level activities and household-level activities, there is a need for more integration of activities and support to HEWs by *Woreda* Health Offices and CHCs for *all* health programs.
- **Gaps in supportive supervision to HEWs and WDALs:** Findings from our trial may not only shed light on some of the factors contributing to poor dose delivery in *Andilaye* intervention communities, but also poor dose delivery of the HEP and CLTSH programming more broadly.
- **Strengthening the WDAL structure:** While *Andilaye* demonstrated progress in improving the activity, function, and capacity of WDALs to act as model households, there is a need to further strengthen the WDAL system and explain what duties and responsibilities they should have as a volunteer.

Measuring “impact” of hygiene and sanitation programming:

- While the focus on diarrheal disease prevention and growth faltering have driven investments in WASH, recent evidence suggests that, in sub-Saharan Africa, basic improvements may not be enough to impact these health outcomes. However, factors contributing to the influence of WASH on other important health outcomes, such as mental well-being and water and sanitation insecurity, remain under-studied.
- In some cases WASH intervention demonstrate lower than expected health gains due to low uptake and sustained adoption of interventions at a community level. These findings represent common challenges for public health and development programs relying on collective action. One possible explanation may be low collective efficacy – perceptions regarding a group’s ability to execute actions related to a common goal.

- As with the traditional focus of WASH programs on catalyzing initial behavior change, few tools have been developed to measure precursors that need to be addressed before behavioral change and maintenance can occur (i.e., behavioral antecedents). These include psychosocial factors such as attitudes and normative beliefs regarding improved practices, perceived and actual abilities to perform improved practices, self-regulation, and intentions to initiate and maintain the adoption of improved practices.

As demonstrated by the Andilaye Trial...

- **Mental well-being and insecurity:** Quantifying the relationships between WASH improvements and these non-traditional outcomes and impacts can inform programs and policies that facilitate health equity.
- **Collective efficacy:** The development and application of collective efficacy scales allow implementers to better design and target community-level interventions, and examine the role of collective efficacy in the effectiveness of community-based WASH programming.
- **Behavioral antecedents:** The development and application of indicators addressing intermediate behavioral antecedents may yield usable tools that could be used for future WASH evaluations.

References

- ALAM, M.-U., WINCH, P. J., SAXTON, R. E., NIZAME, F. A., YEASMIN, F., NORMAN, G., MASUD, A.-A., BEGUM, F., RAHMAN, M., HOSSAIN, K., LAYDEN, A., UNICOMB, L. & LUBY, S. P. 2017. Behaviour change intervention to improve shared toilet maintenance and cleanliness in urban slums of Dhaka: a cluster-randomised controlled trial. *Tropical Medicine & International Health*, 22, 1000-1011.
- ASSEFA, Y., GELAW, Y. A., HILL, P. S., TAYE, B. W. & VAN DAMME, W. 2019. Community health extension program of Ethiopia, 2003-2018: successes and challenges toward universal coverage for primary healthcare services. *Global Health*, 15, 24.
- BECH, P. 2004. Measuring the dimension of psychological general well-being by the WHO-5. *Quality of Life newsletter*, 15-16.
- BERHE, R., DELEA, M., SCLAR, G., WORETA, M., ZEWUDIE, K., MUHAMMED, S., SNYDER, J., LINABARGER, M., GEBREMARIAM, A. & FREEMAN, M. C. 2018. Maintaining behaviour change: innovations in demand-side sanitation and hygiene interventions. *IN: Shaw, R.J. (ed). Transformation towards sustainable and resilient WASH services: Proceedings of the 41st WEDC International Conference, Nakuru, Kenya, 9-13 July 2018*, 6.
- BIRAN, A., SCHMIDT, W. P., VARADHARAJAN, K. S., RAJARAMAN, D., KUMAR, R., GREENLAND, K., GOPALAN, B., AUNGER, R. & CURTIS, V. 2014. Effect of a behaviour-change intervention on handwashing with soap in India (SuperAmma): a cluster-randomised trial. *Lancet Glob Health*, 2, e145-54.
- BISUNG, E., ELLIOTT, S. J., SCHUSTER-WALLACE, C. J., KARANJA, D. M. & BERNARD, A. 2014. Social capital, collective action and access to water in rural Kenya. *Social Science & Medicine*, 119, 147-154.
- BOSCHI-PINTO, C., VELEBIT, L. & SHIBUYA, K. 2008. Estimating child mortality due to diarrhoea in developing countries. *Bull World Health Organ*, 86, 710-717.
- CAMERON, L., OLIVIA, S. & SHAH, M. 2019. Scaling up sanitation: Evidence from an RCT in Indonesia. *Journal of Development Economics*, 138, 1-16.
- CARROLL, C., PATTERSON, M., WOOD, S., BOOTH, A., RICK, J. & BALAIN, S. 2007. A conceptual framework for implementation fidelity. *Implement Sci*, 2, 40.
- CARUSO, B. A., CLASEN, T., YOUNT, K. M., COOPER, H. L. F., HADLEY, C. & HAARDÖRFER, R. 2017a. Assessing Women's Negative Sanitation Experiences and Concerns: The Development of a Novel Sanitation Insecurity Measure. *Int J Environ Res Public Health*, 14.
- CARUSO, B. A., CLASEN, T. F., HADLEY, C., YOUNT, K. M., HAARDÖRFER, R., ROUT, M., DASMOHAPATRA, M. & COOPER, H. L. 2017b. Understanding and defining sanitation insecurity: Women's gendered experiences of urination, defecation and menstruation in rural Odisha, India. *BMJ global health*, 2, e000414.
- CARUSO, B. A., COOPER, H. L., HAARDÖRFER, R., YOUNT, K. M., ROUTRAY, P., TORONDEL, B. & CLASEN, T. 2018. The association between women's sanitation experiences and mental health: A cross-sectional study in Rural, Odisha India. *SSM-population health*, 5, 257-266.
- CENTRAL STATISTICAL AGENCY - CSA/ETHIOPIA & ICF 2017. Ethiopia Demographic and Health Survey 2016. Addis Ababa, Ethiopia: CSA and ICF.
- CLOSSER, S., NAPIER, H., MAES, K., ABESHA, R., GEBREMARIAM, H., BACKE, G., FOSSETT, S. & TESFAYE, Y. 2019. Does volunteer community health work empower women? Evidence from Ethiopia's Women's Development Army. *Health Policy Plan*.
- COHN, S. 2014. From health behaviours to health practices: an introduction. *Sociology of Health & Illness*, 36, 157-162.
- DELEA, M., SCLAR, G., WORETA, M., HAARDÖRFER, R., NAGEL, C., CARUSO, B., DREIBELBIS, R., GOBEZAYEHU, A., CLASEN, T. & FREEMAN, M. 2018. Collective efficacy: Development and validation of a measurement scale for use in public health and development programmes. *International journal of environmental research and public health*, 15, 2139.

- DELEA, M. G. 2019. *Social constructs, behaviour change, and the uptake of community-based WASH interventions: Metrics and analytical approaches for measuring collective efficacy*. London School of Hygiene & Tropical Medicine.
- DELEA, M. G., SNYDER, J. S., BELEW, M., CARUSO, B. A., GARN, J. V., SCLAR, G. D., WORETA, M., ZEWUDIE, K., GEBREMARIAM, A. & FREEMAN, M. C. 2019. Design of a parallel cluster-randomized trial assessing the impact of a demand-side sanitation and hygiene intervention on sustained behavior change and mental well-being in rural and peri-urban Amhara, Ethiopia: Andilaye study protocol. *BMC Public Health*, 19, 801.
- DELEA, M. G., SNYDER, J. S., WORETA, M., ZEWUDIE, K., SOLOMON, A. W. & FREEMAN, M. C. 2020. Development and reliability of a quantitative personal hygiene assessment tool. *Int J Hyg Environ Health*, 227, 113521.
- DEROGATIS, L. R., LIPMAN, R. S., RICKELS, K., UHLENHUTH, E. H. & COVI, L. 1974. The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. *Behavioral science*, 19, 1-15.
- DREIBELBIS, R., WINCH, P. J., LEONTSINI, E., HULLAND, K. R., RAM, P. K., UNICOMB, L. & LUBY, S. P. 2013. The Integrated Behavioural Model for Water, Sanitation, and Hygiene: a systematic review of behavioural models and a framework for designing and evaluating behaviour change interventions in infrastructure-restricted settings. *BMC Public Health*, 13, 1015.
- ERCUMEN, A., BENJAMIN-CHUNG, J., ARNOLD, B. F., LIN, A., HUBBARD, A. E., STEWART, C., RAHMAN, Z., PARVEZ, S. M., UNICOMB, L., RAHMAN, M., HAQUE, R., COLFORD, J. M., JR. & LUBY, S. P. 2019. Effects of water, sanitation, handwashing and nutritional interventions on soil-transmitted helminth infections in young children: A cluster-randomized controlled trial in rural Bangladesh. *PLoS Negl Trop Dis*, 13, e0007323.
- FANKHAUSER, K., NAGEL, C. L., BARSTOW, C. K., KIRBY, M. & THOMAS, E. A. 2019. Geospatial-temporal, demographic, and programmatic adoption characteristics of a large-scale water filter and improved cookstove intervention in Western Province, Rwanda. *Cogent Environmental Science*, 5, 1625481.
- FEDERAL DEMOCRATIC REPUBLIC OF ETHIOPIA MINISTRY OF HEALTH 2012. Implementation Guideline for CLTSH Programming. Addis Ababa.
- FLAY, B. R. & PETRAITIS, J. 1994. The Theory of Triadic Influence: A New Theory of Health Behavior With Implications for Preventive Interventions. *Advances in Medical Sociology*, 4, 19-44.
- FREEMAN, M. C., GARN, J. V., SCLAR, G. D., BOISSON, S., MEDLICOTT, K., ALEXANDER, K. T., PENAKALAPATI, G., ANDERSON, D., MAHTANI, A. G., GRIMES, J. E. T., REHFUESS, E. A. & CLASEN, T. F. 2017. The impact of sanitation on infectious disease and nutritional status: A systematic review and meta-analysis. *International Journal of Hygiene and Environmental Health*, 220, 928-949.
- FREEMAN, M. C., OGDEN, S., JACOBSON, J., ABBOTT, D., ADDISS, D. G., AMNIE, A. G., BECKWITH, C., CAIRNCROSS, S., CALLEJAS, R., COLFORD, J. M., JR., 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 Negl Trop Dis*, 7, e2439.
- 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.
- GARN, J. V., FREEMAN, M. C. & CONSORTIUM, G. 2017a. Global Trachoma Mapping Project: Sanitation Coverage Threshold Levels and Protection against Trachoma. *American Journal of Tropical Medicine and Hygiene*, 95, 563-563.
- GARN, J. V., TRINIES, V., TOUBKISS, J. & FREEMAN, M. C. 2017b. 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.

- GEBREMARIAM, B., HAGOS, G. & ABAY, M. 2018. Assessment of community led total sanitation and hygiene approach on improvement of latrine utilization in Laelay Maichew District, North Ethiopia. A comparative cross-sectional study. *PLoS One*, 13, e0203458.
- GEBREMARIAM, B. & TSEHAYE, K. 2019. Effect of community led total sanitation and hygiene (CLTSH) implementation program on latrine utilization among adult villagers of North Ethiopia: a cross-sectional study. *BMC Res Notes*, 12, 478.
- GODDARD, F. G. B., DELEA, M. G., SCLAR, G. D., WORETA, M., ZEWUDIE, K. & FREEMAN, M. C. 2018. Quantifying user preferences for sanitation construction and use: Application of discrete choice experiments in Amhara, Ethiopia. *Tropical Medicine & International Health*, 23, 1364-1373.
- GRIMES, J. E., CROLL, D., HARRISON, W. E., UTZINGER, J., FREEMAN, M. C. & TEMPLETON, M. R. 2014. The relationship between water, sanitation and schistosomiasis: a systematic review and meta-analysis. *PLoS neglected tropical diseases*, 8, e3296.
- HADLEY, C. & FREEMAN, M. C. 2016. Assessing reliability, change after intervention, and performance of a water insecurity scale in rural Ethiopia. *Food Security*, 8, 855-864.
- HADLEY, C., LINDSTROM, D., TESSEMA, F. & BELACHEW, T. 2008a. Gender bias in the food insecurity experience of Ethiopian adolescents. *Social science & medicine*, 66, 427-438.
- HADLEY, C. & PATIL, C. L. 2006. Food insecurity in rural Tanzania is associated with maternal anxiety and depression. *American Journal of Human Biology*, 18, 359-368.
- HADLEY, C., TEGEGN, A., TESSEMA, F., COWAN, J. A., ASEFA, M. & GALEA, S. 2008b. Food insecurity, stressful life events and symptoms of anxiety and depression in east Africa: evidence from the Gilgel Gibe growth and development study. *Journal of Epidemiology & Community Health*, 62, 980-986.
- HANCHETT, S. K., LAURIE; KAHN, MOHIDUL HOQUE; KULLMAN, CRAIG; AHMED, ROKEYA 2011. Long-Term Sustainability of Improved Sanitation in Rural Bangladesh. Washington, D.C.: World Bank.
- HARTER, M., MOSCH, S. & MOSLER, H.-J. 2018. How does Community-Led Total Sanitation (CLTS) affect latrine ownership? A quantitative case study from Mozambique. *BMC Public Health*, 18, 387.
- HAYES, R. J. & MOULTON, L. H. 2009. *Cluster Randomised Trials*, Taylor & Francis.
- 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.
- HUMPHREY, J. H., MBUYA, M. N. N., NTOZINI, R., MOULTON, L. H., STOLTZFUS, R. J., TAVENGWA, N. V., MUTASA, K., MAJO, F., MUTASA, B., MANGWADU, G., CHASOKELA, C. M., CHIGUMIRA, A., CHASEKWA, B., SMITH, L. E., TIELSCH, J. M., JONES, A. D., MANGES, A. R., MALUCCIO, J. A., PRENDERGAST, A. J. & SANITATION HYGIENE INFANT NUTRITION EFFICACY TRIAL, T. 2019. Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on child stunting and anaemia in rural Zimbabwe: a cluster-randomised trial. *Lancet Glob Health*, 7, e132-e147.
- INTERNATIONAL TRACHOMA INITIATIVE. 2019. *Trachoma Atlas* [Online]. Available: <http://www.trachomaatlas.org/> [Accessed].
- 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.
- LACHIN, J. M. 1988. Statistical properties of randomization in clinical trials. *Control Clin Trials*, 9, 289-311.
- LOZANO, R., NAGHAVI, M., FOREMAN, K., LIM, S., SHIBUYA, K., ABOYANS, V., ABRAHAM, J., ADAIR, T., AGGARWAL, R., AHN, S. Y., ALMAZROA, M. A., ALVARADO, M., ANDERSON, H. R., ANDERSON, L. M., ANDREWS, K. G., ATKINSON, C., BADDOUR, L. M., BARKER-COLLO, S., BARTELS, D. H., BELL, M. L., BENJAMIN, E. J., BENNETT, D., BHALLA, K., BIKBOV, B., ABDULHAK, A. B., BIRBECK, G., BLYTH, F., BOLLIGER, I., BOUFIOUS, S., BUCELLO, C., BURCH,

- M., BURNEY, P., CARAPETIS, J., CHEN, H., CHOU, D., CHUGH, S. S., COFFENG, L. E., COLAN, S. D., COLQUHOUN, S., COLSON, K. E., CONDON, J., CONNOR, M. D., COOPER, L. T., CORRIERE, M., CORTINOVIS, M., DE VACCARO, K. C., COUSER, W., COWIE, B. C., CRIQUI, M. H., CROSS, M., DABHADKAR, K. C., DAHODWALA, N., DE LEO, D., DEGENHARDT, L., DELOSSANTOS, A., DENENBERG, J., DES JARLAIS, D. C., DHARMARATNE, S. D., DORSEY, E. R., DRISCOLL, T., DUBER, H., EBEL, B., ERWIN, P. J., ESPINDOLA, P., EZZATI, M., FEIGIN, V., FLAXMAN, A. D., FOROUZANFAR, M. H., FOWKES, F. G. R., FRANKLIN, R., FRANSEN, M., FREEMAN, M. K., GABRIEL, S. E., GAKIDOU, E., GASPARI, F., GILLUM, R. F., GONZALEZ-MEDINA, D., HALASA, Y. A., HARING, D., HARRISON, J. E., HAVMOELLER, R., HAY, R. J., HOEN, B., HOTEZ, P. J., HOY, D., JACOBSEN, K. H., JAMES, S. L., JASRASARIA, R., JAYARAMAN, S., JOHNS, N., KARTHIKEYAN, G., KASSEBAUM, N., KEREN, A., KHOO, J.-P., KNOWLTON, L. M., KOBUSINGYE, O., KORANTENG, A., KRISHNAMURTHI, R., LIPNICK, M., LIPSHULTZ, S. E., et al. 2012. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, 380, 2095-2128.
- 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. 2018a. 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.
- 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., JR. 2018b. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Bangladesh: a cluster randomised controlled trial. *Lancet Glob Health*, 6, e302-e315.
- MAES, K., CLOSSER, S., TESFAYE, Y. & ABESHA, R. 2019. Psychosocial distress among unpaid community health workers in rural Ethiopia: Comparing leaders in Ethiopia's Women's Development Army to their peers. *Soc Sci Med*, 230, 138-146.
- MAES, K., CLOSSER, S., TESFAYE, Y., GILBERT, Y. & ABESHA, R. 2018. Volunteers in Ethiopia's women's development army are more deprived and distressed than their neighbors: cross-sectional survey data from rural Ethiopia. *BMC Public Health*, 18, 258.
- MURRAY, C. J., VOS, T., LOZANO, R., NAGHAVI, M., FLAXMAN, A. D., MICHAUD, C., EZZATI, M., SHIBUYA, K., SALOMON, J. A. & ABDALLA, S. 2013. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, 380, 2197-2223.
- 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. 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.
- PERSON, M. T., DELEA, M. G., GARN, J. V., ALEXANDER, K., ABAIRE, B. & FREEMAN, M. C. 2017. Assessing the influence of social capital on water point sustainability in rural Ethiopia. *Journal of Water, Sanitation and Hygiene for Development*, 7, 611-622.
- PICKERING, A. J., NJENGA, S. M., STEINBAUM, L., SWARTHOUT, J., LIN, A., ARNOLD, B. F., STEWART, C. P., DENTZ, H. N., MUREITHI, M., CHIENG, B., WOLFE, M., MAHONEY, R., KIHARA, J., BYRD, K., RAO, G., MEERKERK, T., CHERUIYOT, P., PAPAIAKOVOU, M., PILOTTE, N., WILLIAMS, S. A., COLFORD, J. M., JR. & NULL, C. 2019a. Effects of single and integrated water, sanitation, handwashing, and nutrition interventions on child soil-transmitted helminth and *Giardia* infections: A cluster-randomized controlled trial in rural Kenya. *PLoS Med*, 16, e1002841.
- PICKERING, A. J., NULL, C., WINCH, P. J., MANGWADU, G., ARNOLD, B. F., PRENDERGAST, A. J., NJENGA, S. M., RAHMAN, M., NTOZINI, R., BENJAMIN-CHUNG, J., STEWART, C. P., HUDA, T. M. N., MOULTON, L. H., COLFORD, J. M., JR., LUBY, S. P. & HUMPHREY, J. H. 2019b. The

- WASH Benefits and SHINE trials: interpretation of WASH intervention effects on linear growth and diarrhoea. *Lancet Glob Health*, 7, e1139-e1146.
- PRUSS-USTUN, A., BARTRAM, J., CLASEN, T., COLFORD, J. M., JR., CUMMING, O., CURTIS, V., BONJOUR, S., DANGOUR, A. D., DE FRANCE, J., FEWTRELL, L., FREEMAN, M. C., GORDON, B., HUNTER, P. R., JOHNSTON, R. B., MATHERS, C., MAUSEZAH, D., MEDLICOTT, K., NEIRA, M., STOCKS, M., WOLF, J. & CAIRNCROSS, S. 2014. Burden of disease from inadequate water, sanitation and hygiene in low- and middle-income settings: a retrospective analysis of data from 145 countries. *Trop Med Int Health*, 19, 894-905.
- PRÜSS-USTÜN, A., WOLF, J., CORVALÁN, C., BOS, R. & NEIRA, M. 2016. Preventing disease through healthy environments: A global assessment of the burden of disease through environmental risks. Geneva: WHO.
- PULLAN, R. L., SMITH, J. L., JASRASARIA, R. & BROOKER, S. J. 2014. Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. *Parasites & vectors*, 7, 1.
- RESNIKOFF, S., PASCOLINI, D., ETYA'ALE, D., KOCUR, I., PARARAJASEGARAM, R., POKHAREL, G. P. & MARIOTTI, S. P. 2004. Global data on visual impairment in the year 2002. *Bull World Health Organ*, 82, 844-51.
- SCLAR, G. D., GARN, J. V., PENAKALAPATI, G., ALEXANDER, K. T., KRAUSS, J., FREEMAN, M. C., BOISSON, S., MEDLICOTT, K. O. & CLASEN, T. 2017. Effects of sanitation on cognitive development and school absence: A systematic review. *International Journal of Hygiene and Environmental Health*, 220, 917-927.
- SCLAR, G. D., PENAKALAPATI, G., CARUSO, B. A., REHFUESS, E. A., GARN, J. V., ALEXANDER, K. T., FREEMAN, M. C., BOISSON, S., MEDLICOTT, K. & CLASEN, T. 2018. Exploring the relationship between sanitation and mental and social well-being: A systematic review and qualitative synthesis. *Social Science & Medicine*, 217, 121-134.
- SNEL, M. & JACIMOVIC, R. 2014. Turning CLTS Challenges into Opportunities for Success. PanAfrican CLTS Programme: Empowering Self-Help Sanitation of Rural and Peri-Urban Communities and Schools in Africa. The Hague, The Netherlands: IRC.
- STEVENSON, E., AMBELU, A., CARUSO, B., TESFAYE, Y. & FREEMAN, M. 2016. Community water improvement, household water insecurity, and women's psychological distress: An intervention and control study in Ethiopia. *PloS one*, 11, e0153432.
- STEVENSON EGJ, YOHANNES B & C., H. 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, 2014 Hawassa, Ethiopia. 123-30.
- STOCKS, M. E., OGDEN, S., HADDAD, D., ADDISS, D. G., MCGUIRE, C. & FREEMAN, M. C. 2014. Effect of water, sanitation, and hygiene on the prevention of trachoma: a systematic review and meta-analysis. *PLoS medicine*, 11, e1001605.
- STRUNZ, E. C., ADDISS, D. G., STOCKS, M. E., OGDEN, S., UTZINGER, J. & FREEMAN, M. C. 2014. Water, sanitation, hygiene, and soil-transmitted helminth infection: a systematic review and meta-analysis. *PLoS medicine*, 11, e1001620.
- TYNDALE-BISCOE, P. B., MATTHEW; KIDD, ROSS, 2013. Plan International ODF Sustainability Study. London, UK: Plan International UK.
- UNICEF/WHO 2017. Progress on Drinking Water, Sanitation and Hygiene: 2017 Update and SDG Baselines. *Joint Monitoring Programme for Water Supply, Sanitation and Hygiene (JMP)*. New York, NY: UNICEF/World Health Organization.
- VAZ NERY, S., TRAUB, R. J., MCCARTHY, J. S., CLARKE, N. E., AMARAL, S., LLEWELLYN, S., WEKING, E., RICHARDSON, A., CAMPBELL, S. J., GRAY, D. J., VALLELY, A. J., WILLIAMS, G. M., ANDREWS, R. M. & CLEMENTS, A. C. A. 2019. WASH for WORMS: A Cluster-Randomized Controlled Trial of the Impact of a Community Integrated Water, Sanitation, and Hygiene and Deworming Intervention on Soil-Transmitted Helminth Infections. *Am J Trop Med Hyg*, 100, 750-761.

- WAITE, R. C., VELLEMAN, Y., WOODS, G., CHITTY, A. & FREEMAN, M. C. 2016. Integration of water, sanitation and hygiene for the control of neglected tropical diseases: a review of progress and the way forward. *Int Health*, 8 Suppl 1, i22-7.
- WANG, H., TESFAYE, R., RAMANA, G. & CHEKAGN, C. 2016. Ethiopia Health Extension Program: An Institutionalized Community Approach for Universal Health Coverage. Washington, D.C.: World Bank.
- WHO 2002. Constitution of the World Health Organization 1946. *Bull World Health Organ*, 80, 983-984.
- WHO 2019. WASH and health working together: A “how to” guide for neglected tropical disease programmes. Geneva: World Health Organization & the Neglected Tropical Disease NGO Network.
- WOLF, J., HUNTER, P. R., FREEMAN, M. C., CUMMING, O., CLASEN, T., BARTRAM, J., HIGGINS, J. P., JOHNSTON, R., MEDLICOTT, K. & BOISSON, S. 2018. Impact of drinking water, sanitation and handwashing with soap on childhood diarrhoeal disease: updated meta-analysis and meta-regression. *Tropical medicine & international health*, 23, 508-525.
- WORLD HEALTH ORGANIZATION & UNICEF 2013. Progress on sanitation and drinking-water - 2013 update. Geneva, Switzerland.
- ZOU, G. 2004. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol*, 159, 702-6.

Appendices

Links to resources found in report

- [Andilaye Intervention Manual](#): A how-to-guide on implementing intervention activities
- [Andilaye Formative Research Note](#): Exploring the barriers and facilitators to improved WASH behaviors in Amhara, Ethiopia
- [Andilaye Meeting Report](#): Intervention design workshop
- [Andilaye Formative Research Findings](#): Problem and solution trees
- [Andilaye Survey Prompts](#): Sub-set of *Andilaye* Trial survey prompts and answer choices
- [Andilaye Community Commitment Banner](#): Amharic and English versions
- [Andilaye Community Conversations Flipbook](#): Amharic and English versions
- [Andilaye Gobezi! \(Good job!\) Flipbook](#): Amharic and English versions
- [Andilaye Household Goal Card](#): Amharic and English versions
- [Andilaye Supportive Supervision Checklists](#): Amharic and English versions

Appendix A – Supplementary resources

- [Supplemental Table 1](#). Summary of behavioral drivers for each domain and related behavior change techniques mapped out using behavioral theories and frameworks
- [Supplemental Table 2](#). Alignment of relevant roles and responsibilities for the HEP and *Andilaye* Trial
- [Supplemental Table 3](#). Sample size calculations
- [Supplemental Table 4](#). Indicator subgroups used to evaluate the impact the *Andilaye* intervention has on sustained WASH behavior change, diarrhea, and mental well-being

Appendix B – Supplementary process evaluation results

- [Supplemental Table 1](#). Dates of implementation for *Andilaye* intervention behavioral change catalyzing and maintenance activities
- [Supplemental Table 2](#). Costs of implementation for *Andilaye* intervention behavioral change catalyzing and maintenance activities
- [Supplemental Table 3-6](#). Process data for *Andilaye* intervention behavioral change catalyzing and maintenance activities

Appendix C – Supplementary impact evaluation results

- [Supplemental Table 1-8](#). Endline data for *Andilaye* impact evaluation
- [Supplemental Figure 1](#). Box and whisker plots for respondent-reported mental health outcomes