

1 **The impact of a demand-side sanitation and hygiene promotion intervention on sustained behavior**  
2 **change and health in Amhara, Ethiopia: a cluster-randomized trial**

3

4 Matthew C. Freeman<sup>1\*</sup>

5 Maryann G. Delea<sup>1</sup>

6 Jedidiah S. Snyder<sup>1</sup>

7 Joshua V. Garn<sup>2</sup>

8 Mulusew Belew<sup>3</sup>

9 Bethany A. Caruso<sup>4</sup>

10 Thomas F. Clasen<sup>1</sup>

11 Gloria D. Sclar<sup>1</sup>

12 Yihenew Tesfaye<sup>5</sup>

13 Mulat Woreta<sup>3</sup>

14 Kassahun Zewudie<sup>3</sup>

15 Abebe Gebremariam<sup>3,6</sup>

16

17

18 <sup>1</sup> Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University,

19 Atlanta, GA, USA; \*[matthew.freeman@emory.edu](mailto:matthew.freeman@emory.edu) (MCF); [mdelea@emory.edu](mailto:mdelea@emory.edu) (MGD);

20 [jedidiah.snyder@emory.edu](mailto:jedidiah.snyder@emory.edu) (JSS); [gloria.sclar@emory.edu](mailto:gloria.sclar@emory.edu) (GDS); [thomas.f.clasen@emory.edu](mailto:thomas.f.clasen@emory.edu) (TFC)

21 <sup>2</sup> School of Community Health Sciences, University of Nevada, Reno, NV, USA [jgarn@unr.edu](mailto:jgarn@unr.edu) (JVG)

22 <sup>3</sup> Emory Ethiopia, Bahir Dar and Addis Ababa, Ethiopia; [lijmulusewb@yahoo.co.uk](mailto:lijmulusewb@yahoo.co.uk) (MB);

23 [mulatworeta@ymail.com](mailto:mulatworeta@ymail.com) (MW); [lekassahun4@gmail.com](mailto:lekassahun4@gmail.com) (KZ); [abebegmariam@yahoo.com](mailto:abebegmariam@yahoo.com) (AG);

24 <sup>4</sup> Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA,

25 USA; [bcaruso@emory.edu](mailto:bcaruso@emory.edu) (BAC)

26 <sup>5</sup> Department of Social Anthropology, Bahir Dar University, Bahir Dar, Ethiopia;

27 [yihewtesfaye@gmail.com](mailto:yihewtesfaye@gmail.com) (YT)

28 <sup>6</sup> School of Nursing, Emory University; [abebegmariam@yahoo.com](mailto:abebegmariam@yahoo.com) (AG)

29

30 \*Corresponding author: Matthew C. Freeman - [matthew.freeman@emory.edu](mailto:matthew.freeman@emory.edu)

31 **ABSTRACT**

32 **Background:** Behaviors related to water, sanitation, and hygiene (WASH) are key drivers of infectious  
33 disease transmission, and experiences of WASH are potential influencers of mental well-being.  
34 Important knowledge gaps exist related to the content and delivery of effective WASH programs and  
35 their associated health impacts, particularly within the contexts of government programs implemented  
36 at scale.

37 **Methods:** We developed and tested a demand-side intervention called *Andilaye*, which aimed to change  
38 behaviors related to sanitation, personal hygiene, and household environmental sanitation. This theory-  
39 informed intervention was delivered through the existing Ethiopian Health Extension Programme (HEP).  
40 It was a multilevel intervention with a catalyzing event at community level and behavior change  
41 activities at group and household level. We randomly selected and assigned 50 *kebeles* (sub-districts)  
42 from three *woredas* (districts), half to receive the *Andilaye* intervention, and half the standard of care  
43 sanitation and hygiene programming (i.e., community-led total sanitation and hygiene [CLTSH]). We  
44 collected data on WASH access, behavioral outcomes, and mental well-being.

45 **Results:** A total of 1,589 households were enrolled into the study at baseline; 1,472 households (94%)  
46 participated in an endline assessment two years after baseline, and approximately 14 months after the  
47 initiation of a multi-level intervention. The intervention did not improve construction of latrines  
48 (prevalence ratio [PR] 0.99; 95% CI: 0.82, 1.21) or handwashing stations with water (PR: 0.96; 95% CI:  
49 0.72, 1.26), or the removal of animal feces from the compound (PR=1.10; 95% CI: 0.95, 1.28). Nor did it  
50 impact anxiety (PR 0.90; 95% CI: 0.72, 1.11), depression (PR: 0.83; 95% CI: 0.64, 1.07), emotional distress  
51 (PR: 0.86; 95% CI: 0.67, 1.09) or well-being (PR: 0.90; 95% CI: 0.74, 1.10) scores.

52 **Discussion:** We report limited impact of the intervention, as delivered, on changes in behavior and  
53 mental well-being. The effectiveness of the intervention was limited by poor intervention fidelity. While  
54 sanitation and hygiene improvements have been documented in Ethiopia, behavioral slippage, or

55 regression to unimproved practices, in communities previously declared open defecation free is  
56 widespread. Evidence from this trial may help address knowledge gaps related to challenges associated  
57 with scalable alternatives to CLTSH and inform sanitation and hygiene programming and policy in  
58 Ethiopia and beyond.

59

60 **Trial registration:** This trial was registered with [clinicaltrials.gov](https://clinicaltrials.gov) (NCT03075436) on March 9, 2017.

## 61 INTRODUCTION

62 Inadequate water, sanitation, and hygiene (WASH) are key drivers of infectious disease transmission  
63 (Freeman et al., 2017; Wolf et al., 2018). Diarrhea accounts for an estimated 1.65 million deaths  
64 annually [5] and nearly 10% of all under-5 deaths in low-income settings [6]. Deficiencies in WASH are  
65 also a major contributor of neglected tropical diseases (NTDs) [7,8]. Over one billion people are at risk of  
66 soil-transmitted helminthiasis, which leads to nearly five million disability adjusted life years (DALYs),  
67 and schistosomiasis leads to two million DALYs [9,10]. Trachoma, the leading infectious cause of  
68 blindness [11], is precipitated by repeat infections with *Chlamydia trachomatis* bacteria, which are often  
69 perpetuated by poor hygiene [12]. These infections are environmentally mediated [13], and are largely  
70 attributed to inadequate WASH [14,15].

71 While WASH studies have primarily focused on infectious diseases or anthropometric measures of  
72 growth amongst young children, this narrow focus does not fully encapsulate the World Health  
73 Organization (WHO)'s definition of health as “a state of complete physical, mental, and social well-being  
74 and not merely the absence of disease or infirmity” [2]. A growing body of research has identified  
75 linkages between water and sanitation and mental health outcomes [2,16–18]. For example, extensive  
76 qualitative and quantitative research has demonstrated how water insecurity can influence mental  
77 health, particularly among women [19–22]. Research on sanitation and mental health is emergent, and  
78 predominantly qualitative [23–25,25]. A cross-sectional study in Odisha, India, found women’s sanitation  
79 insecurity—their negative sanitation experiences and concerns—to be associated with stress,  
80 depression, distress, and impaired general well-being, even among those with access to a sanitation  
81 facility [26]. Further, a systematic review of sanitation and well-being found open defecation and use of  
82 sanitation facilities can negatively influence mental and social well-being for women and girls, especially  
83 when they experience or perceive a lack of privacy and safety [2]. As such, improvements in women’s  
84 mental health, likely require more than physical access to sanitation facilities, but also gender-sensitive

85 modifications to facilities and shifts in gender norms to improve women’s experiences of sanitation  
86 [24,26]. To date, limited research has assessed the impact of water interventions on mental health  
87 outcomes [27], and few studies have assessed the impact of sanitation interventions on mental health  
88 outcomes [2].

89 Despite the urgent need to improve sanitation and hygiene — including the target of universal basic  
90 access to sanitation as part of Sustainable Development Goal target 6.2 — many large-scale sanitation  
91 interventions have shown poor uptake and sustainability [28], as well as mixed impacts on health [1].  
92 Without sustained sanitation and hygiene behavior change, health gains are unlikely.

93 Community-led total sanitation (CLTS) has been heralded as a low-cost approach to improve community  
94 coverage of sanitation [29]. CLTS uses a demand-side approach—promoting the demand to execute  
95 improved sanitation behaviors, rather than supply-side provision of infrastructure—that involves  
96 engaging the community, typically via an initial “triggering” event, to become open defecation free  
97 (ODF) through community-activities and local champions. Rigorous evaluations of CLTS, like those of  
98 other sanitation interventions, have yielded mixed health effects [30–32]. To date, there is mixed  
99 evidence on the potential of CLTS to achieve and sustain changes to WASH coverage and access [33–36].  
100 Engaging local champions in CLTS delivery may yield beneficial results. Program delivery through Health  
101 Extension Workers (HEW) and the engagement of teachers both led to substantial improvements in  
102 sanitation coverage and use [34], although less than when delivered by trained natural leaders in Ghana  
103 [33]; yet these gains were not well sustained [37].

104 There are several documented limitations of community-led total sanitation and hygiene (CLTSH), a  
105 variation of CLTS that incorporates hygiene-related interventions. HEWs charged with implementing  
106 CLTSH have many responsibilities, limited incentives and motivations, few tools, and little capacity to  
107 continually reinforce messages [38]. The use of negative affective motivators employed by CLTS(H) may

108 not be culturally appropriate or the most effective drivers of sanitation and hygiene behavior change  
109 [39], and may erode mental well-being. Together, the focus on negative affective motivators, poor  
110 facilitation of initial triggering, and a lack of follow-up, has left many communities with negative  
111 impressions of CLTSH initiatives [40].

112 In Ethiopia, which has been implemented widely through the Ethiopian Health Extension Programme  
113 (HEP), relies chiefly on negative affective motives (e.g., shame, disgust) to drive open defecation  
114 cessation. However, like prior evaluations of CLTS, evidence suggests that CLTSH is largely ineffective,  
115 with one out of six Ethiopian households continuing to practice open defecation after their respective  
116 villages were certified as ODF [41].

117 We designed a study to generate evidence to address knowledge gaps related to demand-side  
118 sanitation and hygiene programming and examine less studied, yet critical, inter-personal factors related  
119 to sustained behavioral adoption and downstream health impacts [42]. Specifically, we conducted a  
120 cluster-randomized trial (CRT) to test whether an intervention delivered at scale within the existing  
121 Ethiopian HEP would lead to sustained WASH behavior change and improved mental health. Leveraging  
122 feedback received from community members and key stakeholders, we designed a theoretically-  
123 informed [43–45] and evidence-based demand-side sanitation and hygiene intervention called  
124 *Andilaye*—Amharic for “togetherness/integration.” The intervention takes a positive, encouragement  
125 approach to behavior change by promoting incremental improvements in behavior and incorporating  
126 behavioral maintenance strategies to foster sustained behavior change.

127

## 128 **METHODS**

129 The study’s primary aim was to determine whether a demand-side sanitation and hygiene intervention  
130 (*Andilaye*) impacted WASH behavior change and mental health, specifically general well-being and

131 symptoms of anxiety, depression, and non-specific emotional distress. A protocol detailing the methods,  
132 intervention, and baseline results are published elsewhere [42].

133

#### 134 ***Ethics and trial registration***

135 Ethical approval for the *Andilaye* Trial was provided by Emory University (IRB00076141), the London  
136 School of Hygiene & Tropical Medicine (9595), and locally by the Amhara Regional Health Bureau  
137 (HRTT0135909). The trial was registered with clinicaltrials.gov (NCT03075436) on March 9, 2017.

138

#### 139 ***Study design***

140 This parallel CRT was conducted in West Gojjam and South Gondar Zones of the Amhara National  
141 Regional State, a region of Ethiopia in which WASH conditions are inadequate [46], slippage in sanitation  
142 coverage and improved sanitation behaviors has been documented, and several NTDs (e.g., soil-  
143 transmitted helminths, trachoma) are hyperendemic [47]. The specific behaviors targeted by the  
144 *Andilaye* intervention and their respective key outcome indicators are detailed in **Table 1**. We sought to  
145 investigate whether any changes in WASH behaviors targeted by the *Andilaye* intervention were  
146 sustained, and we tracked intervention fidelity through a process evaluation.



147 **Table 1.** Key outcome indicators for WASH behavioral themes and constituent practices of interest of the *Andilaye* intervention at endline  
148

Indicators	Intervention		Control		PR (95% CI) <sup>a</sup>	PD (95% CI) <sup>b</sup>
	N	%	N	%		
<b>Sanitation (S)</b>						
<b>S1: Construct a long-lasting latrine that is comfortable and hygienic</b>						
- Households with access to at least one household latrine	743	61.2	729	62.0	0.99 (0.82, 1.21)	-0.004 (-0.125, 0.118)
- Households with access to an improved household latrine <sup>c</sup>	741	34.6	726	30.6	1.13 (0.81, 1.59)	0.41 (-0.070, 0.151)
- Households with access to a fully constructed household latrine	742	33.0	729	28.7	1.15 (0.86, 1.54)	0.044 (-0.462, 0.134)
<b>S2: Repair your latrine whenever it is damaged</b>						
- Facility observed to require obvious repair	455	70.1	451	80.5	<b>0.88 (0.78, 0.99)</b>	<b>-0.0985 (-0.19, -0.01)</b>
<b>S3: Upgrade your latrine so it becomes more long lasting, comfortable, and hygienic</b>						
- Household has added or improved anything on the latrine since its original construction	453	17.2	446	15.7	1.08 (0.71, 1.65)	0.012 (-0.056, 0.080)
- Households with latrine with smooth and cleanable slab/floor	743	16.3	728	13.3	1.19 (0.70, 2.03)	0.0263 (-0.05, 0.11)
- Presence of drop hole cover in the latrine	455	18.2	451	10.0	<b>1.77 (1.19, 2.63)</b>	<b>0.0789 (0.02, 0.14)</b>
<b>S4: Close your pit when it becomes full and reconstruct a new latrine</b>						
- Is the pit that is in use full or close to being full	454	11.7	451	12.6	0.92 (0.57, 1.49)	-0.0098 (-0.68, 0.05)
<b>S5: All household members use a latrine every time they defecate</b>						
- Respondent always exclusively used a latrine for defecation during last 7 days	743	53.2	729	54.1	0.99 (0.79, 1.24)	-0.0044 (-0.12, 0.12)
- Head of household always exclusively used a latrine for defecation during last 7 days	529	36.5	473	33.0	1.07 (0.79, 1.47)	0.0270 (-0.09, 0.15)
- Ages 4-17 always exclusively used a latrine for defecation during last 7 days	1447	42.6	1385	35.0	1.15 (0.89, 1.50)	0.0573 (-0.05, 0.16)
<b>S6: Immediately dispose of children's feces into the latrine</b>						
- Child feces were safely disposed of during the last 2 days	401	36.7	376	41.2	0.96 (0.69, 1.32)	-0.017 (-0.145, 0.111)
<b>Personal hygiene (PH)</b>						
<b>PH1: All household members wash their hands with water and soap or soap substitute AFTER handling animal and human feces, even children's feces</b>						
- Household hand or facewashing station(s)	743	98.3	729	97.7	1.01 (0.99, 1.02)	0.008 (-0.008, 0.023)
- The last time the respondent defecated, s/he cleaned hands with water and soap, substitute	738	51.9	725	46.1	1.13 (0.94, 1.35)	0.058 (-0.032, 0.149)
- The last time the index child defecated, s/he cleaned hands with water and soap, substitute	713	43.9	697	39.6	1.12 (0.92, 1.35)	0.0461 (-0.04, 0.13)
<b>PH2: All household members wash their hands with water and soap or soap substitute BEFORE handling food</b>						
- The last time the respondent prepared food, s/he cleaned hands with water and soap, substitute before beginning food preparations	700	53.6	703	48.5	1.11 (0.95, 1.29)	0.051 (-0.027, 0.130)
<b>PH3: All household members wash their faces with water whenever they are dirty and use soap when it is available</b>						

- Ocular discharge present among children aged 1-9 years	822	26.9	874	30.4	0.88 (0.68, 1.15)	-0.037 (-0.114, 0.040)
- Wet nasal discharge present among children aged 1-9 years	822	37.0	874	39.4	0.94 (0.78, 1.13)	-0.024 (-0.094, 0.045)
- Dry nasal discharge present among children aged 1-9 years	822	42.7	874	45.2	0.97 (0.81, 1.16)	-0.016 (-0.095, 0.063)
- Dirt/dust/other debris present among children aged 1-9 years	822	50.5	874	49.5	1.03 (0.89, 1.20)	0.016 (-0.058, 0.090)

### Household Environmental Sanitation (HES)

**HES1:** Keep all animals separated from the house

- Observed animal feces present in the compound	743	82.2	729	82.4	1.01 (0.92, 1.11)	0.0071 (-0.07, 0.08)
-------------------------------------------------	-----	------	-----	------	-------------------	----------------------

**HES2:** Keep the household compound clean by disposing of all animal feces and other waste on a DAILY basis

- Animal feces/waste not left out in open in compound	743	56.4	729	51.2	1.10 (0.95, 1.28)	0.052 (-0.029, 0.132)
- Solid waste was not observed to have been left out in the open	743	34.6	729	27.6	1.26 (0.93, 1.69)	0.0705 (-0.02, 0.17)

**Notes.** <sup>a</sup> We used log-linear binomial regression models to compare the prevalence of the outcomes between the intervention and control arms. Models accounted the stratified design by including woreda indicator variables [48], and accounted for clustering within kebeles by using generalized estimating equations with robust standard errors. <sup>b</sup> Prevalence differences (PD) were calculated using post-estimation commands to estimate the average marginal effects. <sup>c</sup> "Improved" was defined based on the WHO/UNICEF Joint Monitoring Programme (JMP) for Water Supply and Sanitation definition.

149

150

151

152 Fifty *kebeles*—the smallest government administrative unit in Ethiopia—in Amhara Region were  
153 selected and randomly assigned to either the *Andilaye* intervention or counterfactual (i.e., control  
154 comparison group). *Kebele* clusters, the randomization unit, were rural or peri-urban sub-districts from  
155 three districts (*woredas*)—Bahir Dar Zuria, Fogera, and Farta. These areas were targeted because they  
156 represented different topographical conditions present in Amhara. Following baseline data collection,  
157 we randomly selected and assigned clusters using stratified sampling by district. Randomization was  
158 done using a computer-based random number generator. To secure balance across three key potential  
159 confounders (i.e., latrine coverage, washing station with soap coverage, and head of household  
160 education), we established *a priori* that the intervention and control mean values for these three  
161 variables, using baseline survey data, should be within two standard deviations of the overall mean. As  
162 such, the randomization process was repeated twice using replacement rerandomization [49] to achieve  
163 balance according to that *a priori* criterion. **Figure 2** provides further details in a flow diagram.

164

### 165 ***“Andilaye” Intervention***

166 The *Andilay* intervention motto was *“Together we can be a strong, caring, healthy community”*.  
167 Intervention activities offered aspirational messages that emphasized the need for collective action to  
168 make positive change in the community and used verbal persuasion to enhance collective efficacy  
169 perceptions [50]. The *Andilaye* intervention focused on three WASH-related behavioral themes,  
170 informed by formative research: (1) sanitation, (2) personal hygiene, and (3) household environmental  
171 sanitation. Within these themes were 11 constituent practices targeted by the intervention (**Table 1**);  
172 these practices were identified through formative research as ones that could be targeted using  
173 demand-side approaches, and were seen as achievable, per stakeholder feedback [42].

174 Intervention activities and behavior change tools were informed by our formative research and  
175 specifically designed to incorporate techniques that addressed behavioral factors such as action

176 knowledge, perceived personal and household barriers to behavioral adoption, identification and  
177 planning, and behavioral control perceptions amongst others [42]. Activities occurred at four levels—  
178 district, community, group, and household ([Supplemental Figure 1 and Supplemental Table 1](#)). Key  
179 activities included community mobilization and commitment events, community conversations, and  
180 household counseling visits with caregivers—all of which were guided by behavior change tools (e.g.,  
181 community commitment banner, community conversations flipbook, and household counseling flipbook  
182 and goal cards) with illustrations produced by an artist based in Ethiopia.

183 The *Andilaye* intervention commenced with district-level capacity building activities, such as action  
184 planning and training of trainers who would facilitate intervention activities. Further, district-level  
185 refresher trainings and adaptive management activities were conducted to reinforce previously acquired  
186 knowledge and skills, address trainer/facilitator turnover, and review successes and address challenges  
187 faced in implementing group and household level activities. Community-level activities included the  
188 ‘Whole System in the Room’ [51], community mobilization and commitment events, and cross-  
189 fertilization visits. During household-level counselling visits, trained Women’s Development Army  
190 Leaders (WDALs) provided personalized counselling to caregivers to equip them with the knowledge,  
191 skills, and motivation necessary to adopt and maintain improved WASH practices. Structured  
192 community conversations, implemented by trained community facilitators, provided further opportunity  
193 for group-level counselling and support.

194

195 **Figure 1.** The *Andilaye* Trial consists of three major phases: (1) formative research and intervention  
196 design, (2) intervention implementation and process evaluation, and (3) impact evaluation. *Kebele* and  
197 household enrollment took place during baseline data collection (March to April 2017). Implementation  
198 of *Andilaye* intervention activities began in September 2017 and continued through midline data  
199 collection (March to April 2018), quarterly monitoring (June to July and November to December 2018),  
200 and endline evaluation (March to May 2019). See [Supplemental Table 2](#) for specific dates of the delivery  
201 of intervention activities. Midline data reflected at least 2 months since the start of household-level  
202 behavior change activities and 3 weeks since the completion of a catalyzing community-level

203 mobilization and commitment event. Our endline data reflected the implementation of 14-15 months of  
204 household-level behavior change activities and 6-7 months of group-level behavior change activities (as  
205 dictated by trained activity facilitators) and 13-14 months since the community mobilization and  
206 commitment events.

207  
208 The *Andilaye* intervention was delivered through Ethiopia’s HEP, via trained government-salaried  
209 *Woreda* (district) Health Office officials, HEWs, and volunteer WDALs (Figure 1; see Supplemental Table  
210 2 for an alignment of relevant roles and responsibilities of the HEP and *Andilaye* Trial). Implementation  
211 of the *Andilaye* intervention was overseen by an Ethiopian-based study team. The Health Extension  
212 Services Package, and its accompanying CLTSH module delivered via the HEP, were being scaled  
213 throughout Ethiopia [52] and reflected the existing government-supported demand-side sanitation and  
214 hygiene approach. No attempt was made to modify the roll out of this standard care of WASH  
215 programming in any of our study clusters. Results from our equivalence analyses indicated balance in  
216 the number of previously CLTSH-triggered and ODF certified *kebeles*, between study arms (Figure 2).  
217 Baseline statistics, along with the fact that 39 of 50 *kebele* clusters randomly selected for inclusion in the  
218 *Andilaye* Trial had been triggered with CLTSH, and certified ODF, provided strong evidence that  
219 behavioral slippage was, indeed, an issue that needed to be addressed in Amhara and perhaps  
220 elsewhere in Ethiopia.

221  
222 [Figure 2. Study flow diagram]  
223

224  
225 ***Outcomes of interest***

226 Survey instruments administered for our impact evaluation collected data on key outcomes through  
227 self-reports from respondents and other household members. Primary outcomes of interest included  
228 mental health and three targeted WASH behavioral themes (1) sanitation, (2) personal hygiene, and (3)

229 household environmental sanitation behaviors, consisting of 11 constituent practices (Table 1). To  
230 measures WASH outcomes, we pulled from standard WASH indicators, and leveraged formative  
231 research data to contextually adapt survey prompts and answer choices (shown in Supplemental Table 4  
232 of Delea et al., 2019). Sustainability of WASH-related behaviors was measured through the proportion of  
233 individuals and households consistently practicing target behaviors at midline and endline.

234 For mental health, we assessed subjective well-being using the validated WHO's Well-Being Index  
235 (WHO-5) [53] and symptoms of anxiety, depression, and non-specific emotional distress using the  
236 Hopkins Symptom Checklist (HSCL) [54]. The WHO-5 asks the respondent to indicate how frequently  
237 they relate to each of five statements in the previous two weeks using a five-point Likert scale. Higher  
238 scores are better (range: 0-25) with scores below 13 indicating poor well-being. The HSCL is a non-  
239 diagnostic tool that includes 25 items to assess symptoms of anxiety (items 1-10), depression (items 11-  
240 25) and overall emotional distress (all 25 items). We omitted two items from the depression set: an item  
241 on sexual desire, which was deemed inappropriate for unmarried women, and an item on suicide  
242 ideation, because we were unable to provide clinical recourse if needed. Participants indicated how  
243 much symptoms bothered them in the previous week ('not at all' [1] to 'Extremely' [4]). The final score  
244 for each state is a mean of responses for each of the relevant items (range from 1 to 4). Scores of 1.75 or  
245 higher indicate that the condition could be present while lower scores are an indication of lower anxiety,  
246 depression, or distress.

247 Secondary outcomes included diarrhea period prevalence, measured through caregiver report of the  
248 index child (i.e., youngest child in the household aged one to nine years at baseline), sanitation  
249 insecurity, and water insecurity. For sanitation insecurity, we asked respondents to indicate how often  
250 (never, sometimes, often, always) they felt one of seven different forms of sanitation insecurity (i.e., 7  
251 factors). Scores were means of all items in the factor. A higher score represents higher sanitation

252 insecurity. The factors were predesignated, and based on a validation that was done in another study  
253 [24]. Water insecurity was measured through the Household Water Insecurity Experiences (HWISE) scale  
254 [55]. HWISE includes 12 items with four response categories (never, rarely, sometimes, often/always).  
255 The score is the sum of responses, ranging from 0–36. A higher score indicates greater household water  
256 insecurity, and we considered households to be water insecure with a score of 12 or more, as described  
257 elsewhere [55]. Measures of collective efficacy [50] and intermediate behavioral antecedents were also  
258 collected, but are not reported here.

259

#### 260 ***Process evaluation***

261 We conducted a process evaluation alongside our impact evaluation to describe and analyze key aspects  
262 of the *Andilaye's* implementation and provide insights and understanding of program impacts. We  
263 defined fidelity as the degree to which the intervention or program was delivered as intended [56].  
264 Quantitative process data on dose delivered, participation, and dose received were collected through  
265 the direct observation of all district and community-level activities—these activities being facilitated or  
266 co-facilitated by the Ethiopian-based study team. Questions were incorporated into our survey  
267 instruments administered during quarterly monitoring and endline data collection to capture exposure  
268 to key *Andilaye* behavior change activities by respondents from study-enrolled households in  
269 intervention *kebeles*. This included self-reported awareness of and attendance at the community  
270 mobilization and commitment event and community conversations, and the number of household  
271 counseling visits received from WDALs. Per protocol, all community members were targeted to attend  
272 the community mobilization and commitment event; routine (1-2 per month) community conversations  
273 primarily focused on influential community members (e.g., male heads of households, religious leaders,  
274 mother-in-laws) targeted in the ‘Whole System in the Room’; and caregivers were to receive monthly

275 household counseling visits (each visit lasting approximately 30 minutes) following the typical structure  
276 for the WDAL and in accordance the HEP.

277

### 278 ***Sample size and power***

279 A detailed description of sample size considerations is published in the study protocol [42]. Briefly, we  
280 powered this study on mental well-being outcomes, as measure by the HSCL [54], utilizing data from  
281 Ethiopia and East Africa suggesting that approximately 20-35% of rural women experience elevated  
282 symptoms of common mental disorders such as anxiety and depression [57,58]. Our sample size  
283 determination indicated we should recruit and enroll 25 households from each of 50 *kebele* study  
284 clusters, with 25 *kebeles* per study arm, targeting one index child—the youngest child aged one to nine  
285 years at baseline—per household. We increased our final sample size to accommodate for 20% of  
286 households being lost to follow-up. Our target sample, therefore, included 30 households in each *kebele*  
287 study cluster, or 1,500 households in total (i.e., 750 per study arm).

288

### 289 ***Data collection***

290 Data were collected via structured household interviews and observations by trained enumerators  
291 during rounds of data collection. Surveys were collected using mobile phones equipped with the freely  
292 available Open Data Kit (<http://opendatakit.org/>). We interviewed the same households at each round  
293 of data collection. **Figure 1** and **Supplemental Table 2** provide a summary of the timeline of intervention  
294 implementation in relation to points of data collection for our impact evaluation.

295

### 296 ***Analytical methods***

297 We followed a pre-analysis plan developed following baseline data collection [42]. The primary analysis  
298 method was an “intention-to-treat” analysis, which compares the intervention arm to the



299 counterfactual arm without regard to intervention fidelity or compliance. The majority of our primary  
300 and secondary outcomes were binary variables, and for these we used log-linear binomial regression  
301 models and report the prevalence ratio (PR). For these binary outcomes we also present prevalence  
302 differences (PD), which were calculated using post-estimation commands to estimate the average  
303 marginal effects. For continuous outcomes, such as the WHO-5 and HSCL scores, we used linear  
304 regression models. All models included an intervention variable as a fixed effect, account for the  
305 stratified design through the inclusion of the *woreda* indicator variable [59], and incorporated  
306 generalized estimating equations with robust standard errors to account for the clustering of  
307 observations within *kebeles*. For each of our primary outcomes of interest, we assessed if there was  
308 interaction across various sub-groups, including exposure to previous CLTSH triggering and sex of the  
309 index child. We also assessed if water insecurity modified the effectiveness on hygiene behaviors. For all  
310 of these analyses, we included interaction terms to test if effect modification was present (i.e., the  
311 interaction term had a p-value <0.05).

312  
313 To assess whether any improvements in WASH behaviors were sustained between the midterm and the  
314 follow-up period, we compared the prevalence of key targeted sanitation, hygiene, household  
315 environmental sanitation indicators between the demand-side intervention arm and the counterfactual  
316 arm group using the baseline, midline, and endline data (Figure 1). Based on the endline data, at the  
317 time of data collection household-level activities had taken place for the past 14-15 months, group-level  
318 activities for the past 6-7 months, and community-level mobilization and commitment events were  
319 completed 13-14 months prior. Elements of the intervention were still ongoing during the time of our  
320 endline data collection, and we did not collect further data after the endline visit.

321

322 **RESULTS:**

323 **Survey results**

324 Our baseline assessment showed balance in terms of our primary outcomes of interest and  
325 demographic variables [42]. Our endline results reflect complete data from 1,472 (93%) of 1,589  
326 households enrolled in the study at baseline, and was well within our sample size requirement of 1,250  
327 households. Of the 793 enrolled intervention households and 796 enrolled counterfactual households,  
328 retention was similar in both arms, at 94% and 92%, respectively (Figure 2). A large majority (90%) of the  
329 respondents were female, by design. Of these 1,472 respondents, 85% were the mother of the index  
330 child.

331

332 **Process evaluation**

333 Reports from *Woreda* Health Offices collected at endline indicated that none of the 50 *kebele* study  
334 clusters (intervention or counterfactual) received additional CLTSH triggering or re-triggering during the  
335 course of the *Andilaye* Trial. The process evaluation survey was conducted at 703 randomly selected  
336 study-enrolled households in the months following the event (i.e., quarterly monitoring visits).

337

338 For the *Andilaye* intervention, the fidelity of action planning workshops and trainings at the district and  
339 community-levels were high (See Supplemental Table 3 for a summary of process data on dose  
340 delivered, participation, and dose received). In each district (*woredas*), we conducted a sensitization and  
341 action planning workshop, trainings of trainers and intervention activity facilitators (including one round  
342 of refresher trainings) and an adaptive management workshop. All intervention *kebeles* (n=25) had  
343 delivery of the 'Whole System in the Room' and action planning activity, community mobilization and  
344 commitment event, and skill-based training of WDALs (including two rounds of review meetings and  
345 refresher trainings).

346

347 Respondent-reported exposure of key behavior change activities varied, but was generally suboptimal  
348 (Figure 3). Overall, only 18% of respondents reported attending the community mobilization and  
349 commitment event, and 22% reported being aware of the activity, according to data collected from  
350 intervention household respondents via process evaluation surveys administered months following  
351 quarterly monitoring visits. WDALs and their supervisors (i.e., HEWs) were trained to facilitate monthly  
352 counseling visits with households in their catchment area. However, among intervention household  
353 respondents of process evaluation surveys administered at endline (n=665), only 59% reported receiving  
354 a counseling visit and 43% reported receiving at least one follow-up visit. No intervention *kebeles* had  
355 WDALs conducting counseling visits monthly. Among household respondents reporting at least one  
356 household visit (n=391), the average number of visits reported was 2-3 during the 14-15-month  
357 implementation period (Supplemental Table 3). Among respondents of the group-level process  
358 evaluation surveys administered at endline (n=707), 28% reported attending at least one community  
359 conversation, and 46% reported being aware of the activity.

360

361 **[Figure 3.** Respondent-reported exposure of key behavior change activities of the *Andilaye* intervention.  
362 Respondents from study-enrolled households in intervention *kebeles* (n=793) were surveyed on their  
363 awareness and attendance in the community-level mobilization and commitment event during quarterly  
364 monitoring visits and awareness and attendance of group-level community conversations and frequency  
365 household-level counseling visits received by endline visits. A total of 703 (89%) and 707 (89%) surveys  
366 with responses to process evaluation prompts were completed from quarterly monitoring and endline,  
367 respectively. Reported frequency of counseling visits reflected a total of 665 survey responses as these  
368 questions were not relevant for study-enrolled households that were residents of caregivers who were  
369 trained as Women's Development Army Leaders (WDALs) responsible for conducting the *Andilaye*  
370 counselling visits (n=42).]  
371

### 372 ***Impacts on sanitation, personal hygiene, and household environmental sanitation***

373 The intervention did not increase latrine access. At endline, 62% of both intervention and control  
374 households had at least one latrine (prevalence ratio [PR] 0.99; 95% CI: 0.82, 1.21) (Table 1). There was  
375 no difference in the prevalence of improved latrines (PR 1.13; 95% CI: 0.81, 1.59) or in fully constructed

376 latrines (PR 1.15; 95% CI: 0.86, 1.54). Although there were improvements in many latrine characteristics  
377 in the intervention arm compared to the counterfactual arm (e.g., presence of water available or  
378 cleansing agent near or inside the latrine for handwashing, and water available for flushing or self-  
379 cleansing), the conditions (e.g., presence of feces on floor) of latrines in the intervention arm were often  
380 poor (Supplemental Table 4).

381

382 The intervention did not impact defecation practices. Overall, 40% of respondents reported practicing  
383 open defecation during the previous two days; only 46% of respondents had defecated in any latrine  
384 during the previous two days (Supplemental Table 4). All measures of latrine utilization and non-  
385 utilization were similar across intervention and counterfactual arms. This includes indicators of  
386 respondent open defecation (PR: 1.05; 95% CI: 0.76, 1.45), safe disposal of child feces (PR: 0.96; 95% CI:  
387 0.69, 1.32), and number of people from another household who used a latrine during last seven days  
388 (difference: -0.40; 95% CI: -0.85, 0.05).

389

390 The intervention did not impact personal hygiene behaviors. The prevalence of washing stations with  
391 water (PR: 0.96; 95% CI: 0.72, 1.26) was similar between the intervention and counterfactual arms.  
392 Presence of hand or face washing stations were observed in 98% of households (Table 1), although  
393 water and soap were observed in only 20% of handwashing stations and 2% of facewashing stations  
394 (Supplemental Table 4). The prevalence of stations with soap was higher in the intervention arm,  
395 although only 3% of households in this arm had a washing station with soap present. Among all children  
396 aged 1-9 years at endline, observations of facial cleanliness indicated 29% had ocular discharge, 38% had  
397 wet nasal discharge, 44% had dry nasal discharge, and 50% had dust, dirt, or debris on their faces (Table  
398 1). There were no meaningful differences between the study arms for any of these facial cleanliness  
399 measures.

400

401 We found no evidence that the *Andilaye* intervention impacted household environmental sanitation.  
402 Across both arms, the majority of respondents and heads of household had animal herding  
403 responsibilities (88% overall), and animal feces were present in the compound in 82% household  
404 compounds (Supplemental Table 4). A similar proportion of households in intervention and control kept  
405 animals separate from the house (PR=1.01, 95% CI 0.91, 1.11). About half of households did not leave  
406 animal feces/waste in the open (Table 1); this was similar between the intervention and counterfactual  
407 arms (PR=1.10; 95% CI: 0.95, 1.28).

408

#### 409 *Sustained changes of key indicators*

410 There was little difference in the sustainability of key targeted indicators — assessed by changes  
411 between midline and endline — on sanitation access and practices, personal hygiene access and  
412 practices, and household environmental sanitation over the course of follow up (Figure 4). At midline,  
413 most variables continued to show little difference between the intervention and counterfactual arms,  
414 although the prevalence of drop hole covers in latrines and the prevalence of appropriate hygiene  
415 behaviors were more common in the intervention arm. At the endline visit, the prevalence of drop hole  
416 cover was largely sustained in the intervention arm, while the prevalence of drop hole covers decreased  
417 in the control arm (PR=1.77' 95% CI: 1.19, 2.63). All other variables at endline had similar prevalence  
418 levels when comparing the two arms. While the prevalence of hand hygiene behaviors was maintained  
419 at levels similar to the midline visit, increases in hand hygiene behaviors in the counterfactual arm  
420 narrowed the difference between the intervention and control arms at endline. Over the two follow-up  
421 surveys, there was an increase in the prevalence of household hand or facewashing stations that  
422 appeared among study arms.

423

424 **[Figure 4. The prevalence of key sanitation, hygiene, and environmental sanitation indicators over time.]**

425

426 **Impacts on mental health**

427 There was no difference between study arms in the scores for anxiety, depression, emotional distress or

428 general well-being (Table 2). There was also no difference between the intervention and counterfactual

429 arms in the prevalence of each mental health condition: anxiety (PR= 0.90; 95% CI: 0.72, 1.11),

430 depression (PR: 0.83; 95% CI: 0.64, 1.07), emotional distress (PR: 0.86; 95% CI: 0.67, 1.09) and poor well-

431 being (PR: 0.90; 95% CI: 0.74, 1.10) (Table 2).

432

433

434

435 **Table 2. Mental well-being outcomes at endline**

436

Indicator	Cronbach Alpha	Intervention		Control		PR (95% CI) <sup>d</sup>	PD (95% CI) <sup>e</sup>
		N	Mean (SD)	N	Mean (SD)		
<b>Scores</b>							<b>difference (95% CI)<sup>ae</sup></b>
Anxiety score <sup>b</sup>	0.89	742	1.46 (0.61)	729	1.52 (0.64)	-	-0.059 (-0.136, 0.018)
Depression score <sup>b</sup>	0.87	742	1.35 (0.48)	728	1.39 (0.52)	-	-0.036 (-0.081, 0.010)
Emotional distress score <sup>b</sup>	0.93	741	1.29 (0.46)	728	1.33 (0.49)	-	-0.042 (-0.093, 0.008)
Well-being score <sup>c</sup>	0.97	749	17.6 (6.8)	728	17.0 (6.7)	-	0.545 (-0.187, 1.277)
<b>Prevalence</b>		<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>		
High Anxiety <sup>f</sup>	-	742	22.2	729	24.8	0.90 (0.72, 1.11)	-0.026 (-0.076, 0.024)
High Depression <sup>f</sup>	-	742	14.0	728	16.9	0.83 (0.64, 1.07)	-0.029 (-0.067, 0.010)
High Emotional distress <sup>f</sup>	-	741	14.0	728	16.4	0.86 (0.67, 1.09)	-0.023 (-0.060, 0.014)
Poor well-being <sup>g</sup>	-	749	25.2	728	27.8	0.90 (0.74, 1.10)	-0.028 (-0.079, 0.024)

**Notes.** <sup>a</sup>We used linear regression models to estimate the difference 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. <sup>b</sup>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. <sup>c</sup>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. <sup>d</sup>We used similar log-linear binomial regression models to compare the prevalence of the outcomes between the intervention and control arms. <sup>e</sup>Prevalence differences (PD) were calculated using post-estimation commands to estimate the average marginal effects. <sup>f</sup>Each of the above scores was dichotomized, with scores greater than 1.75 indicating a positive status for any of the three outcomes. <sup>g</sup>The above score was dichotomized with scores below 13 indicating poor well-being.

437

438 **Secondary health outcomes**

439 *Reported diarrhea*

440 Diarrhea prevalence during the last seven days among index children was similar in the intervention  
441 (7%) and counterfactual (6%) arms (PR: 1.20; 95% CI: 0.74, 1.93; **Table 3**). Among index children, there  
442 were also similarities comparing study arms in diarrhea prevalence when assessing episodes over the  
443 last two days (PR: 1.25; 95% CI: 0.71, 2.22).

444

445 **Table 3.** Secondary health outcomes at endline

446

Indicator	Cronbach Alpha	Intervention		Control		PR (95% CI) <sup>a</sup>	PD (95% CI) <sup>b</sup>
		N	%	N	%		
<b>Diarrhea</b>							
During the last 2 days, index child had three or more loose stools per day	-	730	5.9	720	4.7	1.62 (0.71, 2.22)	0.012 (-0.018, 0.042)
During the last 7 days, index child had three or more loose stools per day	-	731	7.1	721	6.0	1.20 (0.74, 1.93)	0.012 (-0.019, 0.043)
<b>Water and sanitation insecurity scores</b>		<b>N</b>	<b>mean (SE)</b>	<b>N</b>	<b>mean (SE)</b>	-	<b>difference (95% CI)<sup>c</sup></b>
Water-HWISE Scale <sup>d</sup>	0.96	565	1.7 (0.37)	388	2.7 (0.89)	-	-1.3 (-3.2, 0.61)
Sanitation-Potential harms <sup>d</sup>	0.85	365	0.46 (0.026)	327	0.50 (0.033)	-	-0.050 (-0.130, 0.030)
Sanitation-Social expectations resultant repercussions <sup>d</sup>	0.79	366	0.28 (0.025)	327	0.30 (0.022)	-	-0.029 (-0.089, 0.030)
Sanitation-Physical exertion or strain <sup>d</sup>	0.57	366	0.42 (0.046)	328	0.40 (0.043)	-	0.010 (-0.110, 0.129)
Sanitation-Night concerns <sup>d</sup>	0.56	366	0.32 (0.022)	328	0.37 (0.027)	-	-0.050 (-0.119, 0.019)
Sanitation-Social support <sup>d</sup>	0.88	366	0.10 (0.021)	328	0.20 (0.023)	-	<b>-0.103 (-0.162, -0.433)</b>
Sanitation-Physical agility <sup>d</sup>	0.56	366	0.14 (0.017)	328	0.14 (0.020)	-	0.000 (-0.050, 0.051)
Sanitation-Defecation place <sup>d</sup>	0.81	366	0.35 (0.038)	327	0.32 (0.028)	-	0.024 (-0.058, 0.106)
<b>Water insecurity prevalence</b>		<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>PR (95% CI)<sup>a</sup></b>	<b>PD (95% CI)<sup>b</sup></b>
Water insecure (HWISE score 12 or more) <sup>d</sup>	-	565	5.7	388	8.8	0.50 (0.21, 1.2)	-0.049 (-0.12, 0.026)

**Notes.** <sup>a</sup>We used log-linear binomial regression models to compare the prevalence of 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. <sup>b</sup>Prevalence differences (PD) were calculated using post-estimation commands to estimate the average marginal effects. <sup>c</sup>We used similar linear regression models to estimate the difference in the outcomes comparing the intervention and control arms. <sup>d</sup>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 pre-designated, and based on a validation that was done in another study (Caruso et al., 2017a). A higher score represents higher sanitation insecurity. <sup>d</sup>We used similar linear regression models to estimate difference comparing the outcomes between the intervention and control arms. This used a 12-item scale with four response categories (never, rarely, sometimes, often/always), and a total summed score of those response categories ranging from 0–36. A higher score indicates greater household water insecurity. We considered water insecure as a score of 12 or more, as described elsewhere scale [55].

447

448



449 *Water and Sanitation insecurity*

450 The intervention did not reduce water insecurity prevalence between intervention (5.7%) and  
451 counterfactual (8.8%) arms (PR: 0.50 (95% CI: .21, 1.2); Table 3). At endline, sanitation insecurity scores  
452 related to social support were statistically lower (i.e., better) in the intervention arm than in the control  
453 arm (score difference: -0.10, 95% CI: -0.16, -0.43), indicating a reduced frequency of experiencing the  
454 circumstances in the social support domain (e.g., trouble finding support to watch dependents during  
455 urination, worry about dependents when going to defecate, had to leave dependents alone to urinate,  
456 etc.). Other sanitation insecurity measures were similar between arms.

457

458 **Interaction and effect modification**

459 There was no interaction of the intervention by previous CLTSH triggering for any of the primary  
460 outcome variables of interest. We did not detect effect measure modification by sex for any of the four  
461 mental health outcomes. Similarly, we did not detect interaction by child's sex for any of these  
462 outcomes. We also did not detect interaction between the intervention and water insecurity on any of  
463 the primary handwashing or face washing variables.

464

465 **DISCUSSION:**

466 The *Andilaye* intervention generally did not improve WASH conditions or outcomes. Without sustained  
467 changes to these WASH conditions and behaviors, changes in well-being were not likely, and indeed,  
468 were not detected. Improving sanitation and hygiene behaviors in rural communities remains a  
469 considerable challenge, especially in regions with poor water access and high levels of WASH-related  
470 NTD endemicity. Most studies designed to change sanitation and hygiene behavior are *efficacy* studies  
471 [28] – meant to assess changes under controlled conditions; ours was an *effectiveness* study, designed  
472 to measure changes in a real-world context. We believe that poor fidelity of intervention delivery played

473 a considerable role in uptake of the intervention, pointing to challenges in delivering demand-side  
474 sanitation and hygiene interventions at scale and through existing community-based models.

475 We did not find statistical differences between study arms at endline and few promising trends in  
476 intervention communities for some of the targeted behaviors. Changes were much lower than with  
477 approaches found by Crocker et al. elsewhere in Ethiopia [34,37]. Similarly, Apanga et al. found that  
478 implementation of the Rural Sustainable Sanitation and Hygiene for All (SSH4A) approach that employs a  
479 multidimensional intervention led to a large increase of 77 percentage points in sanitation coverage in  
480 Ethiopia and also coverage gains in many other countries under study [60], although slippage did occur  
481 after conclusion of program activities, whereas many other countries sustained their previous sanitation  
482 coverage gains [61].

483 Our *Andilaye* intervention did not impact validated mental health measures. Few studies have measured  
484 the impact of a sanitation intervention on mental health outcomes despite calls for broader  
485 investigations of sanitation-related health impacts [2,62,63]. We assessed if changes to sanitation access  
486 and sanitation insecurity — changes that we anticipated would be generated by this intervention —  
487 would lead to improved mental health states, including improved well-being and reduction in symptoms  
488 associated with anxiety, depression, distress and general wellbeing. In rural India, women’s experiences  
489 of sanitation, as measured by a validated sanitation insecurity measure, were associated with well-  
490 being, anxiety, depression, and distress, even when women had access to a facility [26]. Similarly, in  
491 urban Mozambique, latrine location and neighborhood violence were important determinants of safety  
492 perceptions and corresponding psychosocial stress [64]. These findings highlight the need for  
493 interventions to consider the experience of sanitation beyond access to a facility alone. The intention-to-  
494 treat analysis did not detect changes to mental well-being scores or to sanitation insecurity scores,  
495 which was perhaps limited by our evaluation period (**Figure 1**). We believe further studies are warranted

496 to test the hypothesis that improved sanitation would impact mental well-being, as our intervention did  
497 not change sanitation behaviors, quality, or access.

498 The purpose of this study was to develop and test an intervention that could be scaled within the  
499 existing Ethiopian HEP. The intervention was designed to be incorporated into prevailing programs (e.g.,  
500 HEP) to demonstrate potential for scale-up, and did not succeed in this regard. For example, despite  
501 strong attendance of trainings, action planning workshops, provision of supportive supervision and on-  
502 the-job-training tools, WDALs and HEWs did not conduct the necessary household visits. They reported  
503 that they did not receive supportive supervision from relevant government officials in accordance with  
504 their action plans (Supplemental Table 3). While supportive supervision considerations were  
505 acknowledged and incorporated into the design of the *Andilaye* intervention, these requirements did  
506 not go above and beyond what is expected of the HEP [65]. These delivery challenges are consistent  
507 with those associated with CLTSH programming and HEP more broadly [38,66]. Additionally, a majority  
508 of intervention *kebeles* had non-active WDALs at the start of implementation, as identified by our  
509 Ethiopian-based study team during initial recruitment of activity facilitators. A cross-sectional study in  
510 four regions of Ethiopia found similar trends in varying levels of WDAL strategy implementation strength  
511 among 423 *kebeles* [67]. Importantly, findings from Damtew et al. suggest HEP outreach activities were  
512 higher in *kebeles* where active WDAL density was higher (i.e., fewer households per active WDAL).  
513 Although HEWs were paid health workers, WDALs were not. This has brought questions of ethics and  
514 sustainability as WDALs are increasingly asked to provide more and more services. Recent qualitative  
515 and quantitative studies suggest that unpaid WDALs are actually worse off than their peers and makes  
516 women, especially unmarried women, vulnerable to negative gossip and psychological distress [68,69].  
517 Although this point goes much deeper into the political economy, it is an important gap to bring up in  
518 the context of empowering women volunteers to enact positive change in their communities [70].  
519 Further, when community health workers are paid to deliver the intervention, there is evidence of

520 successful delivery [71]. Together, these findings raise questions about the possibility of bringing new  
521 programs and approaches to the HEP without adequate support.

522 Evidence suggests that it is important to move away from information-based interventions to address  
523 the array of behavioral factors and determinants that operate at various levels of influence [72–77]. Few  
524 sanitation and hygiene interventions employ behavioral theory to locally adapt messages [37,78].  
525 Exceptions include the *SuperAmma* intervention, which was developed and implemented in India, and  
526 found substantial gains in handwashing with soap [79] and studies that examined the effectiveness of  
527 the risk, attitudes, norms, abilities and self-regulation (RANAS) behavioral model to intervention design  
528 and showed positive impacts on a variety of WASH behaviors including safe water consumption, solar  
529 water disinfection, handwashing, and cleaning of shared sanitation [80–86]. Our intervention aimed to  
530 focus on a variety of contextually appropriate behavioral factors rather than knowledge alone. Given the  
531 low fidelity of the intervention delivery, further capacity building of federal, regional, and local-level  
532 government officials as well as community-level change agents may be necessary for the successful  
533 implementation of approaches that move beyond dissemination of information and messages [87–89].

534

### 535 ***Study strengths***

536 Our intervention was theory-informed and included an extensive intervention design process during  
537 which we emphasized the solicitation and incorporation of feedback from key stakeholders at regional,  
538 zonal, *woreda*, and community-levels. It was designed to be delivered at scale within the Ethiopia HEP.  
539 We utilized a randomized study design, in which intervention and counterfactual communities were  
540 allocated to treatment arms randomly. While cluster randomized trials tend to emphasize internal  
541 validity, we made considerable effort to enhance external validity. Our study was spread over three  
542 *woredas* in two zones, yielding a heterogeneous mix of contexts. To improve internal validity, we used a

543 'fried egg' [90] approach to select kebeles and minimize spillover. We targeted both rural and peri-urban  
544 communities and collected behavioral outcome data on a variety of household members (e.g., primary  
545 female caregiver of index child, head of household, all children aged 0-17 years).

546

547 ***Study limitations***

548 The study faced significant delays in gaining local ethical approval to start the project which led to  
549 truncated implementation and follow-up periods. Key government actors were less involved than  
550 planned, which may have led to sub-optimal fidelity. The integration of *Andilaye* intervention activities  
551 into non-Ethiopian HEP delivery structures (e.g., hired independent community implementers) may yield  
552 further investigations into the effectiveness of the intervention on sustained behavior change and  
553 mental well-being. Several of our behavioral outcomes were reported (vs. observed), and these types of  
554 outcomes may be prone to reporting biases, indicated by differences in our reported and observed  
555 measures.

556

557 **CONCLUSIONS:**

558 We did not find that the *Andilaye* intervention yielded changes in behaviors and conditions related to  
559 sanitation, personal hygiene, or household environmental sanitation; nor did it impact mental health  
560 outcomes. Limited integration of *Andilaye* activities into the HEP likely explains the minimal impact  
561 observed and points to considerable challenges related to implementing demand-side interventions at  
562 scale in Ethiopia. Evidence from this trial may help address knowledge gaps related to scalable  
563 alternatives to CLTSH and inform sanitation and hygiene programming and policy in Ethiopia and  
564 beyond. A greater emphasis on implementation research in WASH delivery would support tools and  
565 approaches for developing, testing, and adapting scalable best-practice interventions [91].

566

567

568

569

570 **Acknowledgements**

571 The authors would like to thank our *Andilaye* study participants, who generously gave their time to  
572 participate in our formative work, behavioral trials, and summative assessment surveys. We are grateful  
573 for the support we receive from numerous partners at the Democratic Republic of Ethiopia's Federal  
574 Ministry of Health; the Amhara Regional Health Bureau; South Gondar and West Gojjam Zonal Health  
575 Departments; and the Bahir Dar Zuria, Farta, and Fogera *Woreda* Health Offices. We would also like to  
576 acknowledge support provided by the Health Extension Workers, Women's Development Army Leaders,  
577 and Health Development Army members from our study *kebeles*. We thank the cadre of field  
578 supervisors and enumerators who captured these data (Yeworkwuha Abay, Mantegbosh Abebe,  
579 Selamawit Abebe, Tigist Abebe, Rosa Abesha, Mahider Adamu, Balemlaye Addisu, Adanech Admasu,  
580 Tirusew Alayu, Adisalem Arega, Yalemwork Asaye, Destaw Asnakew, Yalemwork Ayanew, Ayalnesh  
581 Belay, Asayech Bimrew, Tigist Bitew, Tiruzer Engidaw, Yeserash Gashaw, Natsenat Gebretsadkan,  
582 Woyneshet Genetu, Tewodaj Gizachew, Tibeltalech Mihiret, Yehizbalem Minale, Senait Mulualem, Eleni  
583 Nebiyu, Elsabet Seyoum, Mulubirhan Shitu, Sewunet Tadesse, Beza Tesfaye, Rahel Tsegaye, Sintayehu  
584 Wasihun, and Maritu Yibrie).

585

586

587    **REFERENCES**

- 588    1.    Freeman MC, Garn JV, Sclar GD, Boisson S, Medlicott K, Alexander KT, et al. The impact of  
589        sanitation on infectious disease and nutritional status: A systematic review and meta-  
590        analysis. *Int J Hyg Environ Health*. 2017;220: 928–949. doi:10.1016/j.ijheh.2017.05.007
- 591    2.    Sclar GD, Penakalapati G, Caruso BA, Rehfuess EA, Garn JV, Alexander KT, et al. Exploring  
592        the relationship between sanitation and mental and social well-being: A systematic review  
593        and qualitative synthesis. *Soc Sci Med*. 2018;217: 121–134.  
594        doi:10.1016/j.socscimed.2018.09.016
- 595    3.    Sclar GD, Garn JV, Penakalapati G, Alexander KT, Krauss J, Freeman MC, et al. Effects of  
596        sanitation on cognitive development and school absence: A systematic review. *Int J Hyg  
597        Environ Health*. 2017;220: 917–927. doi:10.1016/j.ijheh.2017.06.010
- 598    4.    Wolf J, Hunter PR, Freeman MC, Cumming O, Clasen T, Bartram J, et al. Impact of drinking  
599        water, sanitation and handwashing with soap on childhood diarrhoeal disease: updated  
600        meta-analysis and meta-regression. *Trop Med Int Health*. 2018;23: 508–525.
- 601    5.    Troeger C, Blacker BF, Khalil IA, Rao PC, Cao S, Zimsen SR, et al. Estimates of the global,  
602        regional, and national morbidity, mortality, and aetiologies of diarrhoea in 195 countries: a  
603        systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect Dis*.  
604        2018;18: 1211–1228. doi:10.1016/S1473-3099(18)30362-1
- 605    6.    Roth GA, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and  
606        national age-sex-specific mortality for 282 causes of death in 195 countries and territories,  
607        1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*.  
608        2018;392: 1736–1788. doi:10.1016/S0140-6736(18)32203-7
- 609    7.    Freeman MC, Ogden S, Jacobson J, Abbott D, Addiss DG, Amnie AG, et al. Integration of  
610        water, sanitation, and hygiene for the prevention and control of neglected tropical  
611        diseases: a rationale for inter-sectoral collaboration. *PLoS Negl Trop Dis*. 2013;7: e2439.  
612        doi:10.1371/journal.pntd.0002439
- 613    8.    WHO. WASH and health working together: A “how to” guide for neglected tropical disease  
614        programmes. Geneva: World Health Organization & the Neglected Tropical Disease NGO  
615        Network; 2019.
- 616    9.    Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted  
617        life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic  
618        analysis for the Global Burden of Disease Study 2010. *Lancet*. 2013;380: 2197–2223.
- 619    10.   Pullan RL, Smith JL, Jirasaria R, Brooker SJ. Global numbers of infection and disease  
620        burden of soil transmitted helminth infections in 2010. *Parasit Vectors*. 2014;7: 1.



- 621 11. Resnikoff S, Pascolini D, Etya'ale D, Kocur I, Pararajasegaram R, Pokharel GP, et al. Global  
622 data on visual impairment in the year 2002. *Bull World Health Organ.* 2004;82: 844–51.
- 623 12. Stocks ME, Ogden S, Haddad D, Addiss DG, McGuire C, Freeman MC. Effect of water,  
624 sanitation, and hygiene on the prevention of trachoma: a systematic review and meta-  
625 analysis. *PLoS Med.* 2014;11: e1001605.
- 626 13. Prüss-Ustün A, Wolf J, Corvalán C, Bos R, Neira M. Preventing disease through healthy  
627 environments: A global assessment of the burden of disease through environmental risks.  
628 Geneva: WHO; 2016.
- 629 14. Grimes JE, Croll D, Harrison WE, Utzinger J, Freeman MC, Templeton MR. The relationship  
630 between water, sanitation and schistosomiasis: a systematic review and meta-analysis.  
631 *PLoS Negl Trop Dis.* 2014;8: e3296.
- 632 15. Strunz EC, Addiss DG, Stocks ME, Ogden S, Utzinger J, Freeman MC. Water, sanitation,  
633 hygiene, and soil-transmitted helminth infection: a systematic review and meta-analysis.  
634 *PLoS Med.* 2014;11: e1001620.
- 635 16. Bisung E, Elliott SJ. Psychosocial impacts of the lack of access to water and sanitation in  
636 low-and middle-income countries: a scoping review. *J Water Health.* 2017;15: 17–30.
- 637 17. Kangmennaang J, Elliott SJ. Linking water (in)security and wellbeing in low-and middle-  
638 income countries. *Water Secur.* 2021;13: 100089. doi:10.1016/j.wasec.2021.100089
- 639 18. Wutich A, Brewis A, Tsai A. Water and mental health. *WIREs Water.* 2020;7.  
640 doi:10.1002/wat2.1461
- 641 19. Brewis A, Choudhary N, Wutich A. Household water insecurity may influence common  
642 mental disorders directly and indirectly through multiple pathways: Evidence from Haiti.  
643 *Soc Sci Med.* 2019;238: 112520. doi:10.1016/j.socscimed.2019.112520
- 644 20. Collins SM, Mbullo Owuor P, Miller JD, Boateng GO, Wekesa P, Onono M, et al. 'I know  
645 how stressful it is to lack water!' Exploring the lived experiences of household water  
646 insecurity among pregnant and postpartum women in western Kenya. *Glob Public Health.*  
647 2019;14: 649–662. doi:10.1080/17441692.2018.1521861
- 648 21. Mushavi RC, Burns BFO, Kakuhikire B, Owembabazi M, Vořechovská D, McDonough AQ, et  
649 al. "When you have no water, it means you have no peace": A mixed-methods, whole-  
650 population study of water insecurity and depression in rural Uganda. *Soc Sci Med.*  
651 2020;245: 112561. doi:10.1016/j.socscimed.2019.112561
- 652 22. Stevenson EG, Greene LE, Maes KC, Ambelu A, Tesfaye YA, Rheingans R, et al. Water  
653 insecurity in 3 dimensions: An anthropological perspective on water and women's  
654 psychological distress in Ethiopia. *Soc Sci Med.* 2012;75: 392–400.  
655 doi:10.1016/j.socscimed.2012.03.022

- 656 23. Biswas D, Joshi S. Sanitation and gendered psychosocial stress in peri-urban Bangalore.  
657 Dev Pract. 2021;31: 548–557. doi:10.1080/09614524.2020.1862762
- 658 24. Caruso BA, Clasen TF, Hadley C, Yount KM, Haardörfer R, Rout M, et al. Understanding and  
659 defining sanitation insecurity: Women’s gendered experiences of urination, defecation  
660 and menstruation in rural Odisha, India. *BMJ Glob Health*. 2017;2: e000414.
- 661 25. Sahoo KC, Hulland KRS, Caruso BA, Swain R, Freeman MC, Panigrahi P, et al. Sanitation-  
662 related psychosocial stress: A grounded theory study of women across the life-course in  
663 Odisha, India. *Soc Sci Med*. 2015;139: 80–89. doi:10.1016/j.socscimed.2015.06.031
- 664 26. Caruso BA, Cooper HL, Haardörfer R, Yount KM, Routray P, Torondel B, et al. The  
665 association between women’s sanitation experiences and mental health: A cross-sectional  
666 study in Rural, Odisha India. *SSM-Popul Health*. 2018;5: 257–266.
- 667 27. Stevenson E, Ambelu A, Caruso B, Tesfaye Y, Freeman M. Community water improvement,  
668 household water insecurity, and women’s psychological distress: An intervention and  
669 control study in Ethiopia. *PLoS One*. 2016;11: e0153432.
- 670 28. Garn JV, Sclar GD, Freeman MC, Penakalapati G, Alexander KT, Brooks P, et al. The impact  
671 of sanitation interventions on latrine coverage and latrine use: A systematic review and  
672 meta-analysis. *Int J Hyg Environ Health*. 2017;220: 329–340.
- 673 29. Kar K, Chambers R. Handbook on community-led total sanitation. 2008.
- 674 30. Patil SR, Arnold BF, Salvatore AL, Briceno B, Ganguly S, Colford JM Jr, et al. The Effect of  
675 India’s Total Sanitation Campaign on Defecation Behaviors and Child Health in Rural  
676 Madhya Pradesh: A Cluster Randomized Controlled Trial. *PLOS Med*. 2014;11: e1001709.  
677 doi:10.1371/journal.pmed.1001709
- 678 31. Pattanayak SK, Yang J-C, Dickinson KL, Poulos C, Patil SR, Mallick RK, et al. Shame or  
679 subsidy revisited: social mobilization for sanitation in Orissa, India. *Bull World Health  
680 Organ*. 2009;87: 580–587.
- 681 32. Pickering AJ, Djebbari H, Lopez C, Coulibaly M, Alzua ML. Effect of a community-led  
682 sanitation intervention on child diarrhoea and child growth in rural Mali: a cluster-  
683 randomised controlled trial. *Lancet Glob Health*. 2015;3: e701–e711. doi:10.1016/S2214-  
684 109X(15)00144-8
- 685 33. Crocker J, Abodoo E, Asamani D, Domapielle W, Gyapong B, Bartram J. Impact evaluation  
686 of training natural leaders during a community-led total sanitation intervention: a cluster-  
687 randomized field trial in Ghana. *Environ Sci Technol*. 2016;50: 8867–8875.
- 688 34. Crocker J, Geremew A, Atalie F, Yetie M, Bartram J. Teachers and sanitation promotion: an  
689 assessment of community-led total sanitation in Ethiopia. *Environ Sci Technol*. 2016;50:  
690 6517–6525.

- 691 35. Guiteras R, Levinsohn J, Mobarak AM. Encouraging sanitation investment in the  
692 developing world: a cluster-randomized trial. *Science*. 2015;348: 903–906.
- 693 36. Hammer J, Spears D. Village sanitation and child health: effects and external validity in a  
694 randomized field experiment in rural India. *J Health Econ*. 2016;48: 135–148.
- 695 37. Crocker J, Saywell D, Bartram J. Sustainability of community-led total sanitation outcomes:  
696 Evidence from Ethiopia and Ghana. *Int J Hyg Environ Health*. 2017;220: 551–557.
- 697 38. Snel M, Jacimovic R. Turning CLTS Challenges into Opportunities for Success. PanAfrican  
698 CLTS Programme: Empowering Self-Help Sanitation of Rural and Peri-Urban Communities  
699 and Schools in Africa. The Hague, The Netherlands: IRC; 2014.
- 700 39. Alemu F, Kumie A, Medhin G, Gebre T, Godfrey P. A socio-ecological analysis of barriers to  
701 the adoption, sustainability and consistent use of sanitation facilities in rural Ethiopia. *BMC*  
702 *Public Health*. 2017;17: 706.
- 703 40. Berhe R, Delea M, Sclar G, Woreta M, Zewudie K, Muhammed S, et al. Maintaining  
704 behaviour change: innovations in demand-side sanitation and hygiene interventions. Shaw  
705 RJ Ed Transform Sustain Resilient WASH Serv Proc 41st WEDC Int Conf Nakuru Kenya 9-13  
706 July 2018. 2018; 6.
- 707 41. Abebe TA, Tucho GT. Open defecation-free slippage and its associated factors in Ethiopia:  
708 a systematic review. *Syst Rev*. 2020;9: 252. doi:10.1186/s13643-020-01511-6
- 709 42. Delea MG, Snyder JS, Belew M, Caruso BA, Garn JV, Sclar GD, et al. Design of a parallel  
710 cluster-randomized trial assessing the impact of a demand-side sanitation and hygiene  
711 intervention on sustained behavior change and mental well-being in rural and peri-urban  
712 Amhara, Ethiopia: Andilaye study protocol. *BMC Public Health*. 2019;19: 801.  
713 doi:10.1186/s12889-019-7040-6
- 714 43. Flay BR, Petraitis J. The Theory of Triadic Influence: A New Theory of Health Behavior With  
715 Implications for Preventive Interventions. *Adv Med Sociol*. 1994;4: 19–44.
- 716 44. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The  
717 behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques:  
718 building an international consensus for the reporting of behavior change interventions.  
719 *Ann Behav Med Publ Soc Behav Med*. 2013;46: 81–95. doi:10.1007/s12160-013-9486-6
- 720 45. Mosler HJ. A systematic approach to behavior change interventions for the water and  
721 sanitation sector in developing countries: a conceptual model, a review, and a guideline.  
722 *Int J Env Health Res*. 2012;22: 431–49. doi:10.1080/09603123.2011.650156
- 723 46. CSA-Ethiopia ICF. International: Ethiopia Demographic and Health Survey 2016. Cent Stat  
724 Agency Ethiop ICF Int Addis Ababa Ethiop Calverton Md USA. 2017.

- 725 47. Initiative International Trachoma. Global Trachoma Atlas. Trach Atlas Available. 2019.
- 726 48. Kahan BC, Morris TP. Reporting and analysis of trials using stratified randomisation in  
727 leading medical journals: review and reanalysis. *BMJ*. 2012;345: e5840.  
728 doi:10.1136/bmj.e5840
- 729 49. Lachin JM. Statistical properties of randomization in clinical trials. *Control Clin Trials*.  
730 1988;9: 289–311.
- 731 50. Delea M, Sclar G, Woreta M, Haardörfer R, Nagel C, Caruso B, et al. Collective efficacy:  
732 Development and validation of a measurement scale for use in public health and  
733 development programmes. *Int J Environ Res Public Health*. 2018;15: 2139.
- 734 51. The Federal Democratic Republic of Ethiopia Ministry of Health. Implementation Guideline  
735 for CLTSH Programming. Addis Ababa. 2012.
- 736 52. Wang H, Tesfaye R, Ramana G, Chekagn C. Ethiopia Health Extension Program: An  
737 Institutionalized Community Approach for Universal Health Coverage. Washington, D.C.:  
738 World Bank; 2016. doi:10.1596/978-1-4648-0815-9
- 739 53. Bech P. Measuring the dimension of psychological general well-being by the WHO-5. *Qual  
740 Life Newsl*. 2004; 15–16.
- 741 54. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The Hopkins Symptom Checklist  
742 (HSCL): A self-report symptom inventory. *Behav Sci*. 1974;19: 1–15.
- 743 55. Young SL, Boateng GO, Jamaluddine Z, Miller JD, Frongillo EA, Neilands TB, et al. The  
744 Household Water InSecurity Experiences (HWISE) Scale: development and validation of a  
745 household water insecurity measure for low-income and middle-income countries. *BMJ  
746 Glob Health*. 2019;4. doi:10.1136/bmjgh-2019-001750
- 747 56. Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for  
748 implementation fidelity. *Implement Sci*. 2007;2: 40. doi:10.1186/1748-5908-2-40
- 749 57. Hadley C, Tegegn A, Tessema F, Cowan JA, Asefa M, Galea S. Food insecurity, stressful life  
750 events and symptoms of anxiety and depression in east Africa: evidence from the Gilgel  
751 Gibe growth and development study. *J Epidemiol Community Health*. 2008;62: 980–986.
- 752 58. Hadley C, Patil CL. Food insecurity in rural Tanzania is associated with maternal anxiety and  
753 depression. *Am J Hum Biol*. 2006;18: 359–368.
- 754 59. Kahan BC, Morris TP. Reporting and analysis of trials using stratified randomisation in  
755 leading medical journals: review and reanalysis. *BMJ*. 2012;345: e5840.  
756 doi:10.1136/bmj.e5840

- 757 60. Apanga PA, Garn JV, Sakas Z, Freeman MC. Assessing the Impact and Equity of an  
758 Integrated Rural Sanitation Approach: A Longitudinal Evaluation in 11 Sub-Saharan Africa  
759 and Asian Countries. *Int J Environ Res Public Health*. 2020;17: 1808.  
760 doi:10.3390/ijerph17051808
- 761 61. Sakas Z, Apanga PA, Freeman M, Snyder JS, Garn JV. Assessing the sustainability of an  
762 integrated rural sanitation approach in 10 countries in Africa and Asia. Colorado WASH  
763 Symposium; 2021 Mar 11; University of Colorado, Boulder.
- 764 62. Caruso BA, Sevilimedu V, Fung IC-H, Patkar A, Baker KK. Gender disparities in water,  
765 sanitation, and global health. *Lancet Lond Engl*. 2015;386: 650–651. doi:10.1016/S0140-  
766 6736(15)61497-0
- 767 63. Pradyumna A, Im P, Ck G. Moving beyond sanitation’s diarrhoea fixation. *Lancet Glob*  
768 *Health*. 2015;3: e16. doi:10.1016/S2214-109X(14)70288-8
- 769 64. Shiras T, Cumming O, Brown J, Muneme B, Nala R, Dreibelbis R. Shared latrines in Maputo,  
770 Mozambique: exploring emotional well-being and psychosocial stress. *BMC Int Health*  
771 *Hum Rights*. 2018;18: 30. doi:10.1186/s12914-018-0169-z
- 772 65. Workie NW and R Gandham NV. Ethiopia - The health extension program in Ethiopia.  
773 Washington, DC: World Bank; 2013. Available:  
774 [http://documents.worldbank.org/curated/en/356621468032070256/Ethiopia-The-health-](http://documents.worldbank.org/curated/en/356621468032070256/Ethiopia-The-health-extension-program-in-Ethiopia)  
775 [extension-program-in-Ethiopia](http://documents.worldbank.org/curated/en/356621468032070256/Ethiopia-The-health-extension-program-in-Ethiopia)
- 776 66. Assefa Y, Gelaw YA, Hill PS, Taye BW, Van Damme W. Community health extension  
777 program of Ethiopia, 2003-2018: successes and challenges toward universal coverage for  
778 primary healthcare services. *Glob Health*. 2019;15: 24. doi:10.1186/s12992-019-0470-1
- 779 67. Damtew ZA, Karim AM, Chekagn CT, Fesseha Zemichael N, Yihun B, Willey BA, et al.  
780 Correlates of the Women’s Development Army strategy implementation strength with  
781 household reproductive, maternal, newborn and child healthcare practices: a cross-  
782 sectional study in four regions of Ethiopia. *BMC Pregnancy Childbirth*. 2018;18: 373.  
783 doi:10.1186/s12884-018-1975-y
- 784 68. Maes K, Closser S, Tesfaye Y, Abesha R. Psychosocial distress among unpaid community  
785 health workers in rural Ethiopia: Comparing leaders in Ethiopia’s Women’s Development  
786 Army to their peers. *Soc Sci Med*. 2019;230: 138–146.  
787 doi:10.1016/j.socscimed.2019.04.005
- 788 69. Maes K, Closser S, Tesfaye Y, Gilbert Y, Abesha R. Volunteers in Ethiopia’s women’s  
789 development army are more deprived and distressed than their neighbors: cross-sectional  
790 survey data from rural Ethiopia. *BMC Public Health*. 2018;18: 258. doi:10.1186/s12889-  
791 018-5159-5

- 792 70. Closser S, Napier H, Maes K, Abesha R, Gebremariam H, Backe G, et al. Does volunteer  
793 community health work empower women? Evidence from Ethiopia's Women's  
794 Development Army. Health Policy Plan. 2019. doi:10.1093/heapol/czz025
- 795 71. Fankhauser K, Nagel CL, Barstow CK, Kirby M, Thomas EA. Geospatial-temporal,  
796 demographic, and programmatic adoption characteristics of a large-scale water filter and  
797 improved cookstove intervention in Western Province, Rwanda. Cogent Environ Sci.  
798 2019;5: 1625481.
- 799 72. Briscoe C, Aboud F. Behaviour change communication targeting four health behaviours in  
800 developing countries: a review of change techniques. Soc Sci Med 1982. 2012;75: 612–  
801 621. doi:10.1016/j.socscimed.2012.03.016
- 802 73. Hulland K, Martin N, Dreibelbis R, DeBruicker Valliant J, Winch P. What factors affect  
803 sustained adoption of safe water, hygiene and sanitation technologies? A systematic  
804 review of literature Hulland K, Martin N, Dreibelbis R, DeBruicker Valliant J, Winch P (2015)  
805 What factors affect sustained adoption of safe water, hygiene and sanitation  
806 technologies? A systematic review of literature. London: EPPI-Centre, Social Science  
807 Research Unit, UCL Institute of Education, University College London. UCL Institute of  
808 Education, University College London: London: EPPI-Centre, Social Science Research Unit;  
809 2015 p. 169.
- 810 74. Marteau TM, Hollands GJ, Fletcher PC. Changing human behavior to prevent disease: the  
811 importance of targeting automatic processes. Science. 2012;337: 1492–1495.  
812 doi:10.1126/science.1226918
- 813 75. Rabbi SE, Dey NC. Exploring the gap between hand washing knowledge and practices in  
814 Bangladesh: a cross-sectional comparative study. BMC Public Health. 2013;13: 89.  
815 doi:10.1186/1471-2458-13-89
- 816 76. Wood W, Neal DT. Healthy through habit: Interventions for initiating & maintaining health  
817 behavior change. Behav Sci Policy. 2016;2: 71–83. doi:10.1353/bsp.2016.0008
- 818 77. Delea MG, Solomon H, Solomon AW, Freeman MC. Interventions to maximize facial  
819 cleanliness and achieve environmental improvement for trachoma elimination: A review  
820 of the grey literature. PLoS Negl Trop Dis. 2018;12: e0006178.
- 821 78. Whittington D, Radin M, Jeuland M. Evidence-based policy analysis? The strange case of  
822 the randomized controlled trials of community-led total sanitation. Oxf Rev Econ Policy.  
823 2020;36: 191–221.
- 824 79. Biran A, Schmidt WP, Varadharajan KS, Rajaraman D, Kumar R, Greenland K, et al. Effect of  
825 a behaviour-change intervention on handwashing with soap in India (SuperAmmma): a  
826 cluster-randomised trial. Lancet Glob Health. 2014;2: e145-54. doi:10.1016/S2214-  
827 109X(13)70160-8

- 828 80. Contzen N, Meili IH, Mosler H-J. Changing handwashing behaviour in southern Ethiopia: A  
829 longitudinal study on infrastructural and commitment interventions. *Soc Sci Med.*  
830 2015;124: 103–114. doi:10.1016/j.socscimed.2014.11.006
- 831 81. Friedrich MND, Kappler A, Mosler H-J. Enhancing handwashing frequency and technique of  
832 primary caregivers in Harare, Zimbabwe: A cluster-randomized controlled trial using  
833 behavioral and microbial outcomes. *Soc Sci Med.* 2018;196: 66–76.  
834 doi:10.1016/j.socscimed.2017.10.025
- 835 82. Huber AC, Tobias R, Mosler H-J. Evidence-Based Tailoring of Behavior-Change Campaigns:  
836 Increasing Fluoride-Free Water Consumption in Rural Ethiopia with Persuasion: Tailoring  
837 of Behavior-Change Campaigns. *Appl Psychol Health Well-Being.* 2014;6: 96–118.  
838 doi:10.1111/aphw.12018
- 839 83. Inauen J, Mosler H-J. Developing and testing theory-based and evidence-based  
840 interventions to promote switching to arsenic-safe wells in Bangladesh. *J Health Psychol.*  
841 2014;19: 1483–1498. doi:10.1177/1359105313493811
- 842 84. Kraemer SM, Mosler H-J. Effectiveness and Effects of Promotion Strategies for Behaviour  
843 Change: Solar Water Disinfection in Zimbabwe: PROMOTION OF BEHAVIOUR CHANGE.  
844 *Appl Psychol.* 2012;61: 392–414. doi:10.1111/j.1464-0597.2011.00475.x
- 845 85. Lilje J, Mosler H-J. Effects of a behavior change campaign on household drinking water  
846 disinfection in the Lake Chad basin using the RANAS approach. *Sci Total Environ.*  
847 2018;619–620: 1599–1607. doi:10.1016/j.scitotenv.2017.10.142
- 848 86. Tumwebaze IK, Mosler H-J. Effectiveness of group discussions and commitment in  
849 improving cleaning behaviour of shared sanitation users in Kampala, Uganda slums. *Soc Sci*  
850 *Med.* 2015;147: 72–79. doi:10.1016/j.socscimed.2015.10.059
- 851 87. Abbink J. Dam controversies: contested governance and developmental discourse on the  
852 Ethiopian Omo River dam: DAM CONTROVERSIES, ETHIOPIA. *Soc Anthropol.* 2012;20: 125–  
853 144. doi:10.1111/j.1469-8676.2012.00196.x
- 854 88. Adem TA. The Local Politics of Ethiopia’s Green Revolution in South Wollo. *Afr Stud Rev.*  
855 2012;55: 81–102. doi:10.1017/S0002020600007216
- 856 89. Maes K, Closser S, Vorel E, Tesfaye Y. A Women’s Development Army: Narratives of  
857 Community Health Worker Investment and Empowerment in Rural Ethiopia. *Stud Comp*  
858 *Int Dev.* 2015;50: 455–478. doi:10.1007/s12116-015-9197-z
- 859 90. Hayes RJ, Moulton LH. *Cluster Randomised Trials.* Taylor & Francis; 2009. Available:  
860 <https://books.google.co.uk/books?id=p-rPAAAACAAJ>

861 91. Haque SS, Freeman MC. The Applications of Implementation Science in Water, Sanitation,  
862 and Hygiene (WASH) Research and Practice. Environ Health Perspect. 2021;129: 065002.  
863 doi:10.1289/EHP7762

864

865



## Formative research and intervention design

## Intervention implementation and process evaluation

- **Formative research**, analysis of results, and intervention mapping
- **Intervention design** feedback from stakeholders and behavioral trials to test intervention approaches
- **Finalization of intervention** design and materials
- **Validation of survey prompts** and related answer choices used in the impact evaluation via feedback from formative research households on their meaning, comprehensiveness, and appropriateness
- **Intervention allocation** at the cluster-level using baseline data to secure balance across key variables

### Action planning and management activities:

District-level sensitization and action planning

Community-level 'Whole system in the room'

*Ongoing monitoring of implementation*

Adaptive management and cross fertilization visits

### Training and capacity building activities:

Training of trainers and facilitators for counseling visits with caregivers

*Ongoing supportive supervision*

Training of community conversations facilitators

Refresher trainings of trainers and household and group-level intervention activity facilitators

### Key behavior change activities:

**Household-level:** counseling visits with caregivers (as dictated by trained facilitators)

**Community-level:** mobilization and commitment events

**Group-level:** community conversations (as dictated by trained facilitators)



### Impact evaluation:

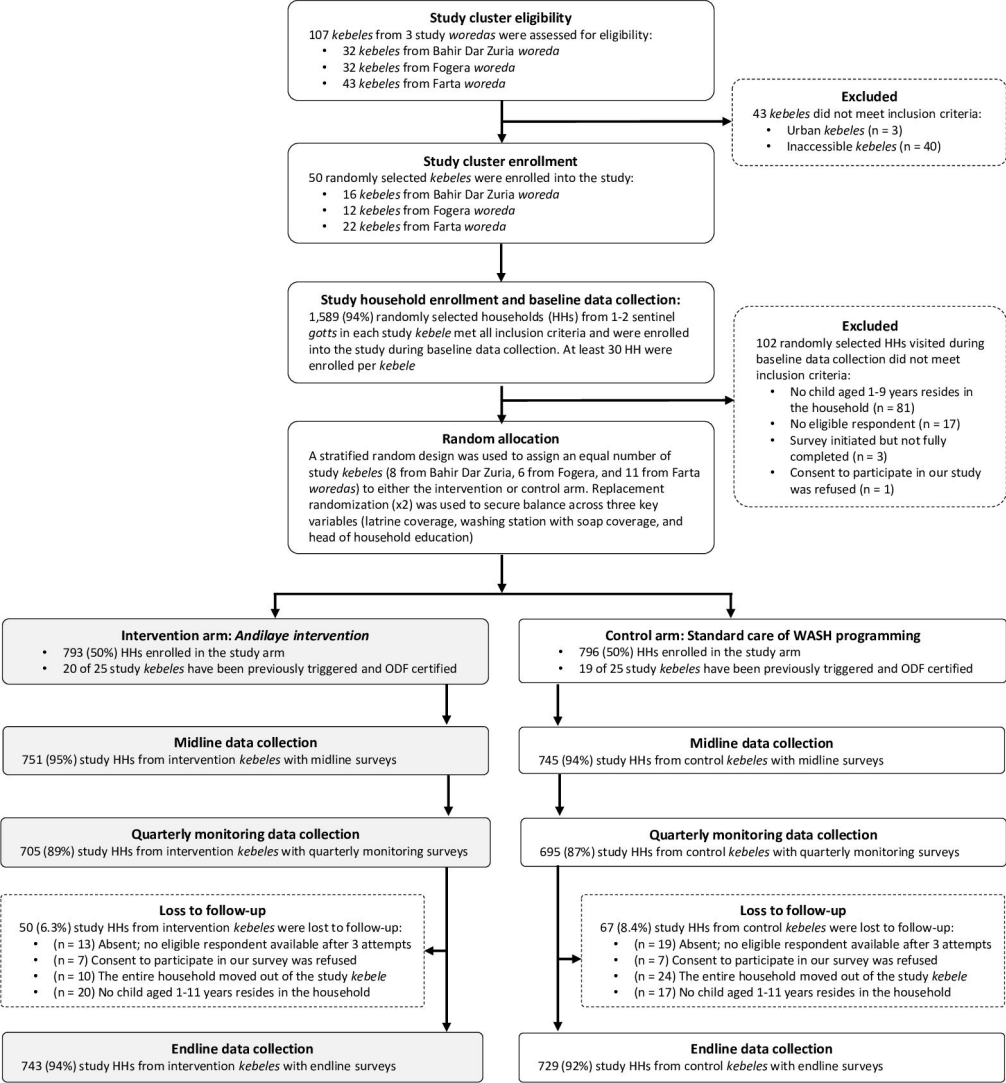
Enrollment and baseline data collection

Midline data collection

Quarterly monitoring visit

Quarterly monitoring visit

Endline data collection



○ Intervention kebeles (n=25)   ■ Bahir Dar Zuria   ◆ Farta   ▲ Fogera   – Overall

