

## Appendix

**Table A1.** Regression estimated for selection of matched control and treatment communities

Variable	Coefficient	
Total population (100s)	0.015***	(0.0036)
Scheduled caste (%)	-3.68*	(2.19)
Population under 6 yrs of age (%)	-44.4**	(22.5)
Female literacy (%)	2.49	(3.21)
Percentage of cultivators	-3.55	(3.00)
Percentage of agricultural laborers	3.12*	(1.84)
Percentage of permanent houses	-6.07**	(3.00)
Households with private taps (%)	1.35	(1.11)
Households without wastewater drainage (%)	-3.95**	(1.80)
Indicator for Guntur district	0.16	(0.74)
Indicator for West Godavari district	0.78	(0.75)
Constant	2.37	(5.15)
N	1411	
Pseudo-R <sup>2</sup>	0.191	

Notes: Data from 2001 Government of India Census and from 2003 National Habitation Survey. The model was estimated as a logit regression. Statistical significance is indicated as follows \*\*\*p<0.01; \*\* p<0.05; \* p<0.1.

**Table A2.** Determinants of sample attrition

Variable	Full model	Full model	Parsimonious model	Parsimonious model
Normalized wealth index	-0.024 (0.014)	-0.018 (0.013)	-0.025* (0.014)	-0.021* (0.012)
Scheduled caste	-0.029 (0.027)	-0.045* (0.024)	-0.035 (0.027)	-0.054** (0.024)
Household size	0.040*** (0.0080)	0.037*** (0.0089)	0.041*** (0.008)	0.040*** (0.009)
Number of children under 5 in household	-0.045** (0.019)	-0.052*** (0.019)	-0.051*** (0.019)	-0.058*** (0.019)
Household believes diarrhea is preventable	-0.26 (0.21)	-0.27 (0.20)	-0.34* (0.20)	-0.32* (0.19)
Household uses private tap for drinking	-0.066 (0.091)	0.0082 (0.071)	0.046 (0.043)	0.066 (0.046)
Household uses public tap for drinking	0.33*** (0.039)	0.35*** (0.037)	0.17*** (0.041)	0.16*** (0.041)
Household uses private well for drinking	0.29*** (0.054)	0.38*** (0.064)	-0.014 (0.048)	0.0021 (0.056)
Household uses surface water for drinking	0.25** (0.096)	0.33*** (0.066)	0.034 (0.075)	0.029 (0.075)
Distance to public tap (mins)	-0.025*** (0.005)	-0.021*** (0.005)	-0.034*** (0.005)	-0.032*** (0.0056)
Water consumption (liters/capita)	0.0005* (0.0003)	0.000 (0.0003)	0.0005** (0.0002)	0.0001 (0.0002)
Round trip time spent going to main drinking water source (min)	0.011*** (0.0040)	0.012*** (0.0043)	0.017*** (0.0046)	0.017*** (0.0049)
Household treats drinking water	0.034 (0.031)	0.048** (0.021)		
Expenses on drinking water treatment (Rs./month)	-0.000 (0.000)	-0.000** (0.000)		
Constant	0.32 (0.25)	0.43* (0.25)	0.31 (0.25)	0.37 (0.25)
Baseline source*perception interactions	YES	YES	NO	NO
Additional controls	YES	YES	NO	NO
Village FEs	NO	YES	NO	YES
N	2,634	2,634	2,640	2,640
R <sup>2</sup>	0.132	0.219	0.065	0.141

**Notes:** The dependent variable is an indicator variable that is equal to 1 if the household attrited from the sample after 2006. Regressors are survey measures at baseline. Standard errors, shown in parentheses, are clustered at the village level, and statistical significance is denoted by asterisks (1% level \*\*\*; 5% level \*\*, 10% level \*). We focus on variables that were statistically significant at the 10% level in at least 1 of the models. Other controls that are not shown (included in all models) are household income; normalized wealth; other backward caste status; average adult education; the number of water sources within 10 minutes round trip time; household reporting of hearing messages about treating drinking water; beliefs that infants with diarrhea should be breastfed, that diarrhea can be caused by bad water quality, that water pollution was a major problem in the village in 2006, that the government should pay for water supply improvements; participation in Gram Sabha meetings; distance to surface water sources; satisfaction with main drinking water source; open defecation; and total water purchases.

<sup>1</sup> Additional controls, which are not statistically significant, are a indicators for a belief that water or sanitation are major village problems, and diarrheal disease rate prevalence among children under five years old.

**Table A3.** Sample descriptive statistics

Variable	2006		2008	
	N	Mean	N	Mean
<i>Demographics</i>				
Female Respondent	2751	1.00	2361	0.99
Respondent's Age	2751	23.0	2361	25.1
Average total monthly HH income (Rs.)	2750	2756	2349	4292
HH is below poverty line (bpl)	2724	0.91	2310	0.47
Average schooling of household adults (yrs)	2751	4.87	2361	4.12
Average HH size	2751	4.48	2361	4.60
Number of children aged 5 and under	2751	1.51	2361	1.43
<i>Community participation</i>				
HH participates in neighborhood cleaning	2750	0.13	2357	0.10
HH attended Gram Sabha, prior 6 months	2750	0.15	2357	0.22
<i>Water supply</i>				
HH uses a private connection for drinking	2751	0.29	2361	0.25
HH uses a public tap for drinking	2751	0.50	2361	0.47
HH uses a public well for drinking	2751	0.11	2361	0.08
HH uses a private well for drinking	2751	0.07	2361	0.11
HH uses surface water for drinking	2751	0.04	2361	0.02
Access to commercial treatment plants	2751	0.03	2361	0.45
HH uses water from CWS	2751	0.00	2361	0.08
Water consumption (lpcd)	2751	56.1	2361	27.7
Number of water sources used	2751	1.56	2361	1.67
Number of drinking water sources used	2751	1.08	2361	1.14
HH very satisfied with main drinking water source	2751	0.61	2361	0.70
<i>Water treatment, handling, sanitation, and hygiene</i>				
HH practices open defecation	2750	0.48	2357	0.40
Main drinking water vessel is fully covered	2750	0.97	2352	0.99
Main drinking water vessel has narrow mouth	2751	0.11	2361	0.12
Main drinking water vessel elevated > 3 feet	2683	0.08	2352	0.13
Wash drinking water storage containers daily	2751	0.84	2361	0.81
Water is removed from container with safe method	2749	0.13	2352	0.07
Treats or filters water before drinking	2750	0.48	2359	0.30
# of times adults wash hands w/soap at critical times	2751	1.83	2358	1.33
<i>Water quality and diarrheal disease</i>				
HH has positive <i>e. coli</i> test	1395	0.10	1161	0.20
HH had at least one diarrhea episode, children ≤ 3 yrs	3300	0.15	1910	0.06
HH had at least one diarrhea episode, children ≤ 5 yrs	4151	0.14	3370	0.05
HH had at least one diarrhea episode, adults	12310	0.07	10858	0.03
Cost of diarrheal illness (Rs./month)	2751	535	2361	142
<i>Coping costs</i>				
Time spent collecting water (minutes/month)	2751	664	2361	1027
Costs of treating water (Rs./month)	2751	296	2361	181
Water purchase cost – all sources (Rs./month)	2751	22.4	2361	17.6
Total coping costs (Rs./month)	2751	364	2361	479

**Notes:** Due to laboratory constraints, *E. coli* was only tested in a randomly selected half of the sample. Observation of drinking water containers was attempted, but was not possible in 102 households in 2006 and 534 households in 2008; outcomes for these households are self-reported or missing. Prevalence of diarrhea, measured based on having 3 or more loose stools within any 24 hour period during the past month, is among all sample individuals in the specified age groups. Coping costs are calculated as described in the text.

**Table A4.** Minimum detectable effect size for key outcomes, given data from follow-up comparison communities, and follow-up sample sizes

<b>Variable</b>	<b>Control mean (sd)</b>	<b>Intra-cluster correlation</b>	<b>MDE (<math>\alpha = 10\%</math>; <math>\beta = 0.8</math>)</b>	<b>MDE (<math>\alpha = 10\%</math>; <math>\beta = 0.9</math>)</b>
(1a) Access to commercial water treatment plants	0.22 (0.42)	0.774	0.30	0.36
(1b) Use of commercial water treatment plants	0.031 (0.17)	0.193	0.086	0.11
(2) Number of drinking water sources used	1.02 (0.33)	0.123	0.10	0.12
(3) % household samples with <i>e. coli</i>	0.17 (0.37)	0.104	0.15	0.19
(4) % of households treating water at home	0.30 (0.46)	0.094	0.12	0.14
(5) Diarrhea prevalence in children 3 years and under <sup>4</sup>	0.041 (0.20)	0.0144	0.035	0.042
(6) Diarrhea prevalence in children 5 years and under <sup>4</sup>	0.032 (0.18)	0.0178	0.024	0.029
(7) Diarrhea prevalence in adults <sup>4</sup>	0.020 (0.14)	0.0108	0.009	0.011

Notes: The intra-cluster correlation is for the whole sample, while the MDEs are calculated from the mean and variance in the control group alone. For *e. coli* and diarrhea disease outcomes, we consider the sample size of tested households, and of individuals in those age groups, at follow-up. Prevalence of diarrhea, measured based on having 3 or more loose stools within any 24 hour period in the prior month, is among all individuals within the households having children in the specified age groups.

**Table A5.** Comparison of key outcomes for treatment and control households: OLS regression using community treatment status, omitting controls

Outcomes	Intent to Treat (ITT) <sup>1</sup>			Actual community-level treatment <sup>2</sup>		
	Assigned to treatment	R <sup>2</sup>	N	Assigned to treatment	R <sup>2</sup>	N
(1a) Access to commercial water treatment plants	0.43*** (0.11)	0.40	5107	0.84*** (0.044)	0.70	5107
(1b) Use of commercial water treatment plants	0.10*** (0.032)	0.076	5107	0.17*** (0.030)	0.14	5107
(2) Number of drinking water sources used	0.14*** (0.051)	0.026	5107	0.26*** (0.043)	0.11	5107
(3) % household samples with <i>E. coli</i> <sup>3</sup>	0.079 (0.051)	0.022	2555	0.13** (0.054)	0.042	2555
(4) % of households treating water at home	-0.073 (0.041)	0.036	5106	-0.033 (0.041)	0.088	5107
(5) Monthly diarrhea prevalence in children 3 years and under <sup>4</sup>	0.025 (0.019)	0.030	5198	0.030* (0.017)	0.039	5198
(5b) 2-week diarrhea prevalence in children 3 years and under <sup>4</sup>	0.019 (0.015)	0.019	5200	0.001 (0.014)	0.027	5200
(6) Monthly diarrhea prevalence in children 5 years and under <sup>4</sup>	0.025 (0.017)	0.038	7515	0.031* (0.016)	0.045	7515
(6b) 2-week diarrhea prevalence in children 5 years and under <sup>4</sup>	0.021 (0.014)	0.023	7518	0.004 (0.013)	0.029	7518
(7) Monthly diarrhea prevalence in all individuals <sup>4</sup>	0.015 (0.009)	0.021	23145	0.019** (0.008)	0.028	23145
(7b) 2-week diarrhea prevalence in all individuals <sup>4</sup>	0.011 (0.007)	0.013	23157	0.005 (0.007)	0.018	23157

**Notes:** Standard errors, shown in parentheses, are clustered at the community level. Statistical significance is adjusted across the set of outcomes using the multiple hypothesis testing correction for the false discovery rate (FDR) described in Anderson (2016), and is denoted by the asterisks (1% level \*\*\*; 5% level \*\*, 10% level \*).

<sup>1</sup> Coefficients were estimated using an OLS regression model without any controls  $Z$  except for age of the individual. We report the coefficient  $\kappa$  on  $T \cdot d$  shown in equation 2, where treatment is assigned at the community level based on planned CWS installation in 2006. Omitting the multiple hypothesis correction for the FDR across this set of outcomes, according to the procedure discussed in Anderson (2008), yields no change in significance, except for outcomes 4 and 7, which become significant at the 10% level.

<sup>2</sup> We report the same coefficient  $\kappa$  on  $T \cdot d$  shown in equation 2, except that treatment is assigned at the community level based on actual CWS installation in 2008. We also include the controls listed below Table 3. Omitting the multiple hypothesis correction across these outcomes yields no change in significance.

<sup>3</sup> Due to laboratory constraints, *E. coli* was only tested in a randomly selected half of the sample.

<sup>4</sup> Prevalence of diarrhea, measured based on having 3 or more loose stools within any 24 hour period, is for all sample individuals within the specified age groups.

**Table A6.** Comparison of additional outcomes for treatment and control households

Outcomes	Intent to Treat (ITT) <sup>1</sup>		
	Assigned to treatment	R <sup>2</sup>	N
Time spent collecting water from main source (minutes/trip)	1.2** (0.58)	0.038	5107
Water consumption for all uses (lpcd)	-6.4* (3.5)	0.21	5107
% of households using drinking water storage vessels with a narrow mouth	0.021 (0.055)	0.023	5107
% of households covering drinking water storage vessel	-0.00 (0.016)	0.016	5107
% of households elevating drinking water storage vessel	0.016 (0.026)	0.031	5107
% of households cleaning drinking water storage vessel daily	-0.050 (0.061)	0.014	5107
% of households that use safe method to remove water from vessel (ladle/spigot)	-0.023 (0.039)	0.034	5107
# of critical occasions at which adult respondent washes hands with soap	-0.005 (0.15)	0.065	5107
Household members report practicing open defecation	0.012 (0.044)	0.081	5107
Water purchase cost – all sources (Rs./month)	12.1 (8.4)	0.022	5107
Monthly costs of coping with inadequate and unsafe water (Rs./month) <sup>2</sup>	-63.7 (56.6)	0.032	5107
Total time spent collecting water (minutes/month)	186.3 (130.0)	0.041	5107
Cost of treating water in-house (Rs./month)	-42.0 (46.4)	0.036	5107
Household cost of illness due to diarrhea in previous month (Rs./month) <sup>3</sup>	61.5 (96.5)	0.023	5107

**Notes:** Coefficients were estimated using OLS regression for the same specifications shown in Table 4. Omission of controls did not alter the results (not shown). Standard errors, shown in parentheses, are clustered at the community level. Statistical significance is denoted by the asterisks (1% level \*\*\*; 5% level \*\*; 10% level \*).

<sup>1</sup> We report the coefficient  $\kappa$  on T\*d shown in equation 2, where treatment is assigned at the community level based on planned CWS installation in 2006.

<sup>2</sup> Monthly coping costs were calculated as the sum of (1) the monetized value of time spent collecting water, (2) expenditures on water treatment (boiling, filtering, and using chemicals), and (3) expenditures on water storage. The value of the time spent collecting water was calculated as the time spent walking to the household's water source plus the time spent waiting at the main water source times the number of trips per day to the main water source times the average hourly wage of the person collecting water (as calculated as the village hourly wage over the rainy and dry seasons for men, women, and children, respectively)

<sup>3</sup> Cost of illness is measured as the sum of out-of-pocket expenditures and estimated lost work income.

**Table A7.** OLS model for selection into CWS use the 2008, 2010, and 2011 panel waves

Variable	Full model	Full model	Parsimonious model	Parsimonious model
Log household income	0.021* (0.011)	0.018** (0.008)	0.022** (0.010)	0.017** (0.008)
Normalized wealth index (in 2008)	0.014 (0.009)	0.027*** (0.007)	0.013 (0.009)	0.029*** (0.007)
Household size in 2008	-0.0001 (0.005)	-0.0004 (0.004)	0.001 (0.006)	-0.001 (0.004)
Number of children under 5 in household	0.003 (0.009)	-0.000 (0.008)	0.003 (0.009)	-0.001 (0.008)
Average education of household adults (yrs)	0.005** (0.002)	0.005** (0.002)	0.005** (0.002)	0.005** (0.002)
Count of water sources < 10 minutes away	0.003 (0.008)	0.009 (0.006)	0.001 (0.009)	0.010 (0.006)
Household reported hearing messages about treating drinking water (2006/2008)	0.029* (0.015)	0.034* (0.019)	0.031* (0.016)	0.034* (0.019)
Household believes diarrhea is preventable	-0.015 (0.021)	-0.014 (0.016)	-0.017 (0.023)	-0.013 (0.016)
Household thinks government should pay for water supply improvements	-0.012 (0.016)	-0.002 (0.010)	-0.008 (0.015)	-0.003 (0.011)
Household participates in village cleaning activities	-0.0041 (0.024)	0.000 (0.019)	0.006 (0.025)	0.007 (0.019)
Household participates in Gram Sabha meetings	-0.0028 (0.021)	0.002 (0.021)	0.003 (0.026)	-0.000 (0.022)
Household uses private tap for drinking	-0.073 (0.049)	-0.065 (0.045)	0.084** (0.035)	-0.016 (0.021)
Household uses public tap for drinking	0.056 (0.034)	-0.003 (0.020)	0.003 (0.030)	-0.028 (0.019)
Household uses private well for drinking	-0.023 (0.037)	-0.015 (0.029)	-0.065** (0.032)	-0.024 (0.032)
Household uses surface water for drinking	0.16* (0.090)	0.054 (0.062)	0.073 (0.069)	0.013 (0.043)
Distance to public tap (mins)	-0.002 (0.003)	0.001 (0.002)	-0.004 (0.003)	-0.001 (0.001)
Distance to surface water source (mins)	0.003 (0.005)	0.002 (0.004)	-0.002 (0.004)	0.002 (0.004)
Respondent was very satisfied with main drinking water source in 2006	-0.010 (0.025)	0.005 (0.009)	-0.024 (0.027)	0.002 (0.010)
Household practices open defecation	0.024 (0.016)	-0.015 (0.013)	0.029 (0.019)	-0.016 (0.013)
Round trip time spent going to main drinking water source	0.003 (0.002)	0.001 (0.002)	0.0048* (0.0026)	0.002 (0.002)
Household thinks water supply is the biggest village environmental problem	0.057* (0.030)	0.025 (0.021)		
Household treats drinking water	0.044*** (0.016)	0.014 (0.014)		
Reported monthly household expense for treating water (Rs)	-0.000** (0.000)	-0.000 (0.000)		
Year: 2010	0.15*** (0.037)	0.15*** (0.038)	0.15*** (0.038)	0.15*** (0.038)

Year: 2011	0.067** (0.026)	0.066** (0.026)	0.066** (0.026)	0.065** (0.026)
Constant	-0.15 (0.10)	-0.28*** (0.077)	-0.18 (0.11)	-0.28*** (0.076)
Baseline source*perception interactions	YES	YES	NO	NO
Additional controls <sup>1</sup>	YES	YES	NO	NO
Village FEs	NO	YES	NO	YES
N	4,161	4,161	4,161	4,161
R <sup>2</sup>	0.097	0.200	0.071	0.193

Notes: Dependent variable is use of the CWS at the time of the survey in 2008, and regressors are survey measures at baseline, in 2006, unless otherwise specified. Standard errors, shown in parentheses, are clustered at the village level, and statistical significance is denoted by the asterisks (1% level \*\*\*; 5% level \*\*; 10% level \*). These models (without village fixed effects) are used to generate the propensity scores used for obtaining the matched samples that are considered in Table 6. Other controls that are not shown due to lack of explanatory power are caste status (scheduled or other backward caste); beliefs that infants with diarrhea should be breastfed and that diarrhea can be caused by bad water quality; per capita water consumption; and total water purchases.

<sup>1</sup> Additional controls, all of which had limited statistical significance are a perception that sanitation was a major village problem in 2006, and diarrheal disease rate prevalence among children under five years old in 2006.



**Table A8.** Comparison of additional outcomes for treatment and control households: Propensity Score Matching

Outcomes	ATT: Radius matching Outcomes in 2008			
	Full model		Basic model	
<b>Users of the CWS in 2008 (treated communities)</b>				
Time spent collecting water from main source (minutes/trip)	7.8***	(0.81)	7.5***	(0.82)
Water consumption for all uses (lpcd)	-3.62 **	(1.79)	-0.70	(1.48)
% of households using narrow-mouthed drinking water storage vessels	0.18***	(0.046)	0.18***	(0.043)
% of households covering drinking water storage vessel	-0.006	(0.008)	-0.005	(0.008)
% of households elevating drinking water storage vessel	0.003	(0.040)	0.002	(0.037)
% of households cleaning drinking water storage vessel daily	-0.11**	(0.047)	-0.14***	(0.044)
% of households that use safe method to remove water from vessel	0.13***	(0.044)	0.14***	(0.041)
# of occasions at which adult respondent washes hands w/ soap	0.19*	(0.14)	0.15	(0.13)
Uses private tap	0.080*	(0.047)	0.15***	(0.044)
Public tap is a major drinking water source	-0.19***	(0.039)	-0.16***	(0.039)
Public well is a major drinking water source	-0.065***	(0.010)	-0.056***	(0.008)
Private well is a major drinking water source	-0.024**	(0.011)	-0.15***	(0.018)
Household members practice open defecation – self-reported	0.14***	(0.047)	0.13***	(0.045)
Water purchase cost – all sources (Rs./month)	33.2***	(7.28)	36.5***	(6.8)
Cost of storage and maintenance of storage (Rs./month)	-3.5***	(0.93)	-3.3***	(0.94)
Time spent collecting water (min/month)	259***	(95.4)	206***	(84)
Cost of treating water in-house	-69.2*	(45.7)	-73.5*	(45.2)
Household cost of illness due to diarrhea in previous month <sup>2</sup>	69.0	(59.5)	35.4	(63.0)
<b>Non-users of the CWS in 2008 (treated communities)</b>				
Time spent collecting water from main source (minutes/trip)	0.90***	(0.25)	0.91***	(0.24)
Water consumption for all uses (lpcd)	-3.03***	(0.75)	-3.06***	(0.78)
% of households using drinking water storage vessels with a narrow mouth	-0.009	(0.015)	-0.008	(0.014)
% of households covering drinking water storage vessel	0.002	(0.003)	0.003	(0.003)
% of households elevating drinking water storage vessel	-0.007	(0.016)	-0.004	(0.017)
% of households cleaning drinking water storage vessel daily	-0.019	(0.018)	-0.019	(0.019)
% of households that use safe method to remove water from vessel	-0.028**	(0.011)	-0.028**	(0.010)
# of occasions at which adult respondent washes hands w/ soap	-0.028	(0.061)	-0.019	(0.063)
Uses private tap	0.059**	(0.023)	0.063**	(0.022)
Public tap is a major drinking water source	0.061**	(0.023)	0.063**	(0.025)
Public well is a major drinking water source	0.025*	(0.015)	0.022	(0.015)
Private well is a major drinking water source	-0.13***	(0.016)	-0.13***	(0.016)
Household members practice open defecation – self-reported	0.068**	(0.025)	0.060**	(0.025)
Water purchase cost – all sources (Rs./month)	9.8***	(2.6)	9.6***	(2.8)
Monthly costs of coping with inadequate and unsafe water <sup>1</sup>	-0.56	(0.42)	-0.56	(0.41)
Time spent collecting water (min/month)	267***	(77)	273***	(76.9)
Cost of treating water in-house	-30.7	(23.5)	-24.8	(23.1)
Household cost of illness due to diarrhea in previous month <sup>2</sup>	40.6	(40.8)	37.4	(45.5)

**Notes:** The first stage propensity scores were obtained using logit model specifications with the same variables as the OLS models shown in Columns 1 and 3 of Table 5 (excluding village fixed effects). Here we report only the radius matching results (kernel matching yields no substantive change in outcomes); to maintain the same samples as in Tables 6 and 7, only households present during the baseline and follow-up survey are included. Total eligible 2008 user households is 131, except for *E.coli* (only 70 users tested) and diarrhea among children under 3 and 5 (85 and 130 users in 2008, respectively). These were matched to an eligible sample of 930 households from control communities (473 of whom had e-coli results) using radius and kernel matching methods. All 131 user households fell within the zone of common support; and 709 of 746 non-user households did. Standard errors, reported in parentheses, are bootstrapped with 500 replications. Statistical significance is adjusted across the set of outcomes using the multiple hypothesis testing correction for the false discovery rate (FDR)

described in Anderson (2016), and is denoted by the asterisks (1% level \*\*\*; 5% level \*\*; 10% level \*). Omitting the multiple hypothesis correction across the full set of these outcomes leads to some gain in significance.

<sup>1</sup> Monthly coping costs were calculated as the sum of (1) the monetized value of time spent collecting water, (2) expenditures on water treatment (boiling, filtering, and using chemicals), and (3) expenditures on water storage. The value of the time spent collecting water was calculated as the time spent walking to the household's water source plus the time spent waiting at the main water source times the number of trips per day to the main water source times the average hourly wage of the person collecting water (as calculated as the village hourly wage over the rainy and dry seasons for men, women, and children, respectively)

<sup>2</sup> Cost of illness is measured as the sum of out-of-pocket expenditures and lost income.

**Table A9.** Naïve “triple differences” OLS comparison of key outcomes for CWS user and non-user households:

Outcomes	Average Treatment Effect <sup>1</sup>			
	ATT Users	ATT Nonusers	R <sup>2</sup>	N
(2) Number of drinking water sources used	1.19*** (0.10)	-0.053 (0.048)	0.35	4163
(3) % household samples with <i>E. coli</i> <sup>2</sup>	0.14* (0.093)	0.084 (0.047)	0.0	2092
(4) % of households treating water at home	-0.26*** (0.063)	-0.041 (0.050)	0.050	4163
(5) Diarrhea prevalence in children 3 years and under <sup>3</sup>	0.048* (0.028)	0.012 (0.018)	0.022	4092
(6) Diarrhea prevalence in children 5 years and under <sup>3</sup>	0.042* (0.022)	0.015 (0.015)	0.025	6117
(7) Diarrhea prevalence in all individuals <sup>3</sup>	0.019* (0.011)	0.011 (0.008)	0.015	18773

**Notes:** Coefficients were estimated using OLS regression, including the controls listed under Table 4. Omission of controls did not alter the results (not shown). Standard errors, shown in parentheses, are clustered at the community level. Statistical significance is adjusted across the set of outcomes using the multiple hypothesis testing correction for the false discovery rate (FDR) described in Anderson (2016), and is denoted by the asterisks (1% level \*\*\*; 5% level \*\*; 10% level \*).

<sup>1</sup> We report the coefficient  $\beta^u$  from an OLS model for the interaction term between and indicator for treatment status (as assigned in 2006) and an indicator for use of the CWS in 2008, and a similar  $\beta^{nu}$  for the interaction with not using CWS water in 2008. For all but the diarrheal disease outcomes, there are 163 users and 1,011 non-users, with the balance being control households. Omitting the multiple hypothesis correction for the false discovery rate across the set of 6 outcomes for users, according to the procedure discussed in Anderson (2008), yields no change in significance, except that outcome 3 becomes insignificant ( $p = 0.14$ ). Omitting the correction to the estimates for nonusers leads to outcome 3 becoming significant at the 10% level.

<sup>2</sup> Due to laboratory constraints, *E. coli* was only tested in a randomly selected half of the sample.

<sup>3</sup> Prevalence of diarrhea, measured based on having 3 or more loose stools within any 24 hour period, is for all sample individuals within the specified age groups.

**Table A10.** Comparison of key outcomes for treatment and control households, controlling for community user density

Users of the CWS in 2008 (treated communities)	Users (Radius Matching) <sup>1</sup>			Non-Users (Radius Matching) <sup>1</sup>		
	User	Village use density	R <sup>2</sup>	Non-User	Village use density	R <sup>2</sup>
Time spent collecting water from main source (minutes/trip)	7.09*** (1.71)	2.40 (4.38)	0.24	1.51 (0.72)	-5.1* (2.6)	0.019
Water consumption for all uses (lpcd)	-7.78 (4.94)	13.2 (13.2)	0.053	-4.82 (2.62)	15.3* (8.4)	0.056
% of households using narrow mouth drinking water storage vessel	0.19* (0.10)	-0.062 (0.31)	0.046	0.004 (0.037)	-0.10 (0.18)	0.003
% of households covering drinking water storage vessel	-0.019 (0.021)	0.047 (0.045)	0.009	0.001 (0.004)	0.012 (0.009)	0.002
% of households elevating drinking water storage vessel	0.057 (0.069)	-0.20 (0.15)	0.015	0.001 (0.028)	-0.068 (0.10)	0.008
% of households cleaning drinking water storage vessel daily	-0.091 (0.12)	-0.056 (0.36)	0.016	-0.053 (0.057)	0.28* (0.15)	0.009
% of households that use safe method to remove water from vessel	0.26** (0.086)	-0.45 (0.17)	0.059	-0.026 (0.017)	-0.016 (0.062)	0.004
# of occasions at which adult respondent washes hands w/ soap	0.19 (0.30)	-0.022 (1.08)	0.006	0.027 (0.11)	-0.46 (0.63)	0.004
Uses private tap	-0.057 (0.13)	0.42 (0.48)	0.10	0.016 (0.093)	0.36 (0.34)	0.024
Public tap is a major drinking water source	-0.050 (0.11)	0.47 (0.21)	0.032	0.046 (0.099)	0.12 (0.38)	0.006
Public well is a major drinking water source	-0.066** (0.022)	0.009 (0.011)	0.039	0.057 (0.076)	-0.27 (0.26)	0.010
Private well is a major drinking water source	-0.030 (0.022)	-0.20 (0.054)	0.006	-0.11 (0.058)	-0.17 (0.10)	0.053
Household members practice open defecation – self-reported	-0.16 (0.10)	1.09** (0.30)	0.14	-0.026 (0.075)	0.79** (0.23)	0.030
Water purchase cost – all sources (Rs./month)	44.8* (20.4)	-45.0 (45.5)	0.099	17.1 (17.9)	-60.4 (67.5)	0.027
Cost of storage and maintenance of storage (Rs./month)	-5.57*** (1.33)	6.67 (3.53)	0.045	-0.48 (1.11)	-0.67 (3.41)	0.001
Time spent collecting water (min/month)	468** (230)	-684 (486)	0.029	356 (172)	-748 (511)	0.010
Cost of treating water in-house	56.5 (122)	-424 (229)	0.022	-17.4 (43.3)	-108 (160)	0.010
Household cost of illness due to diarrhea in previous month <sup>2</sup>	67.7 (107.1)	8.51 (212.3)	0.003	65.9 (63.4)	-208 (239)	0.005

**Notes:** The first stage propensity scores were obtained using logit model specifications with the same variables as the OLS models shown in Columns 1 and 3 of Table 5 (excluding village fixed effects). We compute the user density as the sum of households sourcing from the CWS in the village divided by the total number of sample households, excluding the household (or members) being considered in the analysis. Coefficients shown above were then estimated using an OLS regression model controlling for the propensity score and this community user density. Standard errors, shown in parentheses, are adjusted to account for sample weighting from the radius or kernel PSM procedures, respectively. Statistical significance is also adjusted across the set of outcomes using the multiple hypothesis testing correction for the false discovery rate (FDR) described in Anderson (2016), and is denoted by the asterisks (1% level \*\*\*; 5% level \*\*; 10% level \*). This leads to modest loss in significance for several coefficients in these models. For all but the diarrheal disease and *E.coli* outcomes, there are 131 users, 709, and 930 control households in the common support region.

<sup>1</sup> We only report results for the full specification and radius matching, with a caliper of 0.1. Results with kernel matching or the parsimonious specification of the first stage are not substantively different.

<sup>2</sup> Cost of illness is measured as the sum of out-of-pocket expenditures and lost income.

**Table A11.** Outcomes for treatment households over the longer term

Outcomes	A. Intent-to-treat <sup>1</sup> (OLS) 2006-2011			B. Users only: Originally treated vs. comparison villages <sup>2</sup>		C. Nonusers only: Originally treated vs. comparison villages <sup>2</sup>	
	1. 2006-2008	2. 2008-2010	3. 2010-2011	1. 2010	2. 2011	1. 2010	2. 2011
(1a) Access to commercial water treatment plants	0.43*** (0.11)	-0.28* (0.11)	-0.068 (0.059)	n.a.	n.a.	0.17 (0.11)	0.11 (0.11)
(1b) Use of commercial water treatment plants	0.10** (0.032)	-0.040 (0.050)	-0.043 (0.032)	n.a.	n.a.	n.a.	n.a.
(4) % of households treating water at home	-0.069 (0.040)	0.062 (0.064)	-0.011 (0.062)	0.000 (0.088)	0.025 (0.053)	0.065 (0.059)	-0.051 (0.062)
(5) Diarrhea prevalence in children under 3 years of age <sup>3</sup>	0.017 (0.015)	n.a.	-0.025 (0.017)	n.a.	0.066 (0.058)	n.a.	-0.015 (0.019)
(6) Diarrhea prevalence in children under 5 years of age <sup>3</sup>	0.021 (0.014)	n.a.	-0.006 (0.013)	n.a.	0.018 (0.028)	n.a.	0.011 (0.014)
(7) Diarrhea prevalence in all individuals <sup>3</sup>	0.011 (0.007)	n.a.	0.001 (0.004)	n.a.	0.014 (0.0098)	n.a.	0.006 (0.004)
Total sample size (n)	8903			192	159	1727	1701
Sample size from original treatment communities	4484			128	93	851	857
Sample size from original comparison communities	4419			64	66	876	844

**Notes:**

<sup>1</sup> In column A1, we report the coefficient  $\kappa$  on T\*d shown in equation 2, where treatment is assigned at the community level based on planned CWS installation in 2006 (this is equivalent to the change reported in Table 4). Columns 2 and 3 are the coefficients on additional interactions between the indicator d (which denotes assignment to treatment in 2006) and indicators for years 2010 and 2011, respectively. We also control for time indicators. Standard errors, shown in parentheses, are clustered at the community level. Omitting the multiple hypothesis correction for the false discovery rate across each set of 2010 and 2011 ITT outcomes, according to the procedure discussed in Anderson (2008), yields no change in significance, except for outcome 1 in the 2008-2010 period ( $p = 0.017$ ). Significance is denoted by the asterisks (1% level \*\*\*; 5% level \*\*; 10% level \*).

<sup>2</sup> In columns B1 and B2, we compare outcomes for users of CWS water in original treatment villages in 2008 and their matched comparisons at that time, in the years 2010 and 2011. In columns C1 and C2, we compare outcomes for nonusers of CWS water in original treatment villages in 2008, and their matched comparison at that time, in the years 2010 and 2011. Omitting the multiple hypothesis correction for the false discovery rate across the set of user only, and non-user only, outcomes yields no change in significance.

<sup>3</sup> Prevalence of diarrhea, measured based on having 3 or more loose stools within any 24 hour period within the prior month, is for all sample individuals within the specified age groups.