Supporting Information

Arnold et al. 10.1073/pnas.1008944107

SI Materials and Methods

Setting. The study area is in the Tiruchirappalli (Trichy) district in Tamil Nadu, India. Intervention villages are located in the subdistricts of Thottiyam, Thuraiyur, and Thathaiyangarpet. Control villages are located in adjacent subdistricts of Manachanallur and Uppiliyapuram. Villages are between 17 and 55 km from the nearest city, Tiruchirappalli, and are accessed mainly by paved roads (median walking distance to an all-weather road is 3 min). The climate is tropical, hot, and subject to heavy rains during the monsoon season (August–December). During the study period the maximum temperature ranged between 23.0 and 40.7 °C, and there were 17 d with >25 mm of rain. The primary occupation is rice agriculture and cultivation (66% of the working adults in our sample). Other major occupations include self-employed businesses (8.5%), truck drivers (6.0%), factory workers (2.7%), and skilled artisans (2.4%). In 2004 and 2005, 22.8% of rural households in Tamil Nadu were below the poverty line; in the Trichy district, life expectancy at birth was 72.8 y and the district was ranked 13th in the state by the United Nations human development index (HDI = 0.732) (1).

Description of the Intervention. Between 2003 and 2007, two NGOs, Water.org, and their local partner Gramalaya, implemented the program in 12 rural villages. The pilot intervention was implemented in eight projects (five time periods) between 2004 and 2007 (Table S1). Gramalaya selected the 12 villages from a list of prospective villages that had requested their assistance. Gramalaya conducted a participatory rural appraisal (PRA) of the sanitation, water, and hygiene conditions of the prospective villages. Using information collected in the PRAs, Gramalaya chose ~1 village for every 5 that requested assistance, selecting the villages with the worst sanitation, water, and hygiene conditions. (Note that we could not use the "queue" of prospective villages as potential controls because the pilot intervention had been scaled up to the entire subdistricts by the time of our evaluation.) In the 12 selected villages, Gramalaya developed a combined intervention that was informed by each village's PRA. Although there was some variation in implementation due to differences in local conditions (Table S1), the main components of the intervention were the following:

Sanitation: (*i*) Community mobilization campaigns to build private toilets, which include walks of shame, defecation mapping, and identification of community representatives to lead the effort (2); (*ii*) assistance with the formation of village water and sanitation committees; (*iii*) construction or renovation of primary school toilets (sanitary blocks); (*iv*) formation of self-help groups (SHGs) to promote toilet use and construction; (*v*) technical support and local training for toilet construction (typically pour–flush single- or double-pit latrines with a water seal); (*vi*) capital cost assistance for families with subsidized, revolving loans provided through the local SHG; and (*vii*) certification of villages as "open defecation free" (ODF).

Water: (*i*) Renovation of community tubewell hand pumps; (*ii*) construction or renovation of primary school water taps; (*iii*) promotion of kitchen gardens and soak pits for waste water; and (*iv*) promotion of private household tap connections through subsidized loans (see the microcredit component, below).

Hygiene: (*i*) Village-wide hygiene education campaigns; (*ii*) formation of SHGs for women and children to promote good hygienic practices; and (*iii*) formation of school health clubs to promote good hygiene.

Microcredit: In 8 of the 12 villages (village nos. 5–12 in Table S1) Gramalaya promoted the construction of private toilets and private water tap connections through revolving loan funds. Gramalaya provided the loans directly to local SHGs and the SHG members distributed the loans to individual borrowers in their village. Each SHG was responsible for repaying the entire loan in full, thus harnessing the communal responsibility for a single loan. By December 2007, Gramalaya had disbursed $98,883 USD in loans in the intervention villages for 496 water-related loans and 1,177 sanitation-related loans (average loan size: $59 USD). Arney et al. provide additional details of the microcredit program in the intervention villages (3).

Details of Control Community Selection. Fig. 2 in the main text summarizes this process. We obtained panchayat-level data (panchayats are groups of one to five villages) from the 2001 Indian national census and supplemented it with 2003 Tamil Nadu Water Supply and Drainage Board survey data that included details about population, water supply, and cattle ownership at the village level. Our sampling frame for control villages included all villages in two subdistricts that neighbor intervention village subdistricts. We sampled from neighboring subdistricts because of heightened NGO water and sanitation activity in intervention subdistricts; the geographic separation between intervention and control villages also helped to prevent spillover effects. In our enrollment survey just 1.5% of study households reported receiving assistance from programs other than the intervention of interest (including agriculture, housing, credit, and prenatal care), and no households in control villages reported receiving sanitation-, water-, or hygiene-related interventions in the previous year.

There were 240 potential control villages in the original frame. We excluded 45 villages on the basis of scheduled caste membership, village size, and biofuel use to remove potential control villages that had obvious differences from the intervention villages. In the remaining 195 potential control and 12 intervention villages, we modeled the probability of receiving the intervention (A) conditional on baseline covariates (W), using a logistic regression model: $P(A=1|W)=[1+\exp(-\beta W)]^{-1}$. We used an iterative approach to selecting covariates in W by reestimating the model for different specifications and selecting the specification that selected a control group most closely balanced with the intervention group on the basis of standardized differences in means and nominal P values of baseline characteristics (4). The final model included the following main effects after ruling out other covariates and higher interactions: number of households in the village; per-capita cattle ownership in the village; the panchayat-level income; and the proportions of scheduled caste population, households with access to in-home or public tap water, literate female population, and households that use banking services (Table 2 of the main text includes the full set of covariates considered). After fitting the model, we matched two control villages to each intervention village without replacement, using a nearest-neighbor approach and the linear predictor from the model (5).

We then conducted a rapid assessment of the 12 intervention and 24 potential control villages to measure information about interventions since 2003, the number of active self-help groups, school and administrative facilities, primary livelihoods, car and tractor ownership, and basic water infrastructure. The goal of the exercise was to reduce the control sample to one matched village per intervention village on the basis of more detailed information not available in the matching data. Our team found that 2 of the 24 villages were a single village. We excluded extremely small and large villages because they were qualitatively different from the intervention sample (e.g., the very large villages have their own hospitals). Finally, we excluded villages with fewer than two self-help groups (the minimum in intervention villages) on the grounds that they may be less socially organized than the intervention group. After these exclusions 1 village retained both matched control villages.

We listed and mapped all households in the 25 selected villages (12 intervention, 13 control) that had at least one child <5 y old. We then selected a random sample of 50 households per village. If a village had <50 households with children <5 y old, then the sample included all of its households.

Demographic, Socioeconomic, and Environmental Exposure Surveys. To ensure that the village selection procedure led to comparable intervention and control groups, interviewers collected demographic, socioeconomic, and environmental exposure information in the enrollment interview. The survey recorded self-reported defecation practices that included frequency of toilet use, reasons for use, location, and use patterns by different demographic groups (men, women, and children <5 y old). If a household owned a private toilet, interviewers asked household members to estimate its age. Interviewers also discretely collected information from female respondents in the household about their perceived safety and privacy while defecating. The survey recorded the use of eight different water sources and reasons for a household's use or nonuse of each source. If a household reported using a water source, then field staff collected details about the source including distance; number of trips per day; use of the water from each source; and perceived safety, reliability, and quality. Respondents were also asked to identify their primary water source and when they began to use their primary water source (an estimate of its age).

In both the enrollment and all follow-up interviews, interviewers assessed key environmental exposure information using discrete spot-check observations of hygienic conditions. Hygiene observations included details about whether a household had a dedicated handwashing station and whether it was stocked with water and soap. Interviewers also collected observations of animals and their feces in the home living area during the interviews and general cleanliness measures such as the presence of garbage in the home. Private toilets inspections (if owned) included information about evidence of use, cleanliness, and the availability of soap and water for handwashing after defecation.

A team of 10 locally hired fieldworkers (5 women, 5 men) and two supervisors conducted household interviews. The survey instruments were pretested and validated during a 3-wk period in nearby, nonstudy villages. All instruments were translated and independently back translated to ensure accuracy. Data were entered into Access (Microsoft), and we conducted standard quality control procedures including logic checks and double entry of 11% of the questionnaires. Of the 1,200 questionnaires double-entered, there were an average of 0.56 entry errors per questionnaire and 0.0008 errors per entry field.

Water Quality Sample Collection and Analysis. Beginning in the third survey round field staff collected water samples from all village sources and household drinking water. The 25 villages in our study have between one and seven village sources, and all village sources were tested in 10 survey rounds. Field staff collected 125 mL of water from village sources in a fashion that mimicked villager water retrieval practices. Field staff also collected drinking water samples from participant households during follow-up survey rounds. Participant households were randomly allocated into four groups. Two of the groups were measured in rounds 3 and 5, and the other two groups were measured in rounds 4 and 6. In survey rounds 7–12, households in one of the four groups were tested. Each household's drinking water was tested between one and four times over the study period.

Water samples were collected in 125-mL sterilized plastic bottles in a fashion that mimicked each household's water retrieval practices: by dipping a household cup into the vessel to transfer the water, by pouring water from the storage container into the sample container, or, if a household did not store drinking water, by retrieving water directly from the tap. Along with the water samples, field staff recorded basic characteristics of the water conditions at the time of collection (such as storage container type).

The field team transported all the water samples in a cooler to a laboratory at a nearby university for culturing within 24 h. In the first round of sampling, sample water was diluted at a ratio of 1:10 before filtering and analysis. After the first round of sampling, lab protocol was changed to a dilution ratio of 1:100 to more accurately quantify total coliform concentrations. After dilution, sample water was passed through a 0.45-μm membrane filter and incubated on HiCrome M-Tec Agar (HiMedia M1571) at 37 °C for 24 h. The number of purple colonies was counted and recorded as coliform bacteria. The number of blue-green colonies was counted and recorded as *Escherichia coli*. Samples without detectable levels of indicator organisms were set to 0.1 before analysis of quantitative counts on the $\log_{10}$ scale. We analyzed *E. coli* data only from the first round of samples that were analyzed with a 1:10 dilution ratio. For these samples, we compared groups using differences in *E. coli* on the $\log_{10}$ scale and using differences in the proportion of samples with ≥10 *E. coli* colony forming units (cfu) per 100 mL.

Each sample was also analyzed for hydrogen sulfide (H_2S)-producing bacteria, using the HiH2S test kit (HiMedia K020). Samples were left to incubate at room temperature for 24 h and, if room temperature fell below 30 °C, for an additional 12 h. Samples were recorded as positive for H_2S if they turned black.

Child Health Measurement. In each visit interviewers collected child illness symptoms over the previous 14 d from each child's caregiver, using a health calendar modeled after Goldman et al. (6). The calendar records each day that the child has individual symptoms. We defined diarrhea as three or more loose or watery stools in 24 h or a single stool with blood or mucus (7).

During the first and last interviews (rounds 1 and 12), fieldworkers collected anthropometric measurements in teams of two following standard protocols from the Demographic and Health Survey (8). Fieldworkers weighed children in the standing position when possible. They weighed children that were too young to stand in their caregiver's arms and reweighed the caregiver separately (the values were later subtracted during the analysis). Fieldworkers used scales accurate to 0.1 kg (Tanita 1631), and the scales were tested for accuracy each morning with a standardized 10-kg weight. Fieldworkers measured the length of children under age 2 y in the reclining position and children aged 2–5 y in the standing position, using portable stadiometers accurate to 0.1 cm (Seca 214).

Statistical Analysis. We estimated that 12 villages per group would provide 80% power to detect differences in prevalence of 2.5 percentage points, assuming 10% diarrhea prevalence in control villages (9), an average of 50 children per village, 10% dropout, and a design effect of 3.5 for repeated measures (10).

The parameter of interest for all outcomes (both unadjusted and adjusted) is the marginal treatment effect conditional on selection into the study based on restriction and propensity score matching. This parameter of interest is equal to the average treatment effect among the treated (see the main text for additional details). We estimate the parameter as

$$\theta = E_{W^*}\{E(Y\,|A=1, W^*) - E(Y\,|A=0, W^*)\}, \qquad \textbf{[S1]}$$

where Y is the outcome of interest, A is an indicator equal to 1 if a child lives in an intervention village and 0 otherwise, and W^* is the set of characteristics among intervention villages in the study sample. For child diarrhea and other binary sanitation, water, and hygiene outcomes, we calculated the difference in prevalence (risk difference) of each outcome between the intervention and control groups.

In addition to calculating mean differences in 2008, for private toilet and tap construction we calculated the difference between intervention and control villages in newly constructed toilets and taps during the 5-y intervention period. For these two outcomes, this difference in the change in private amenities is a difference-in-difference (DID) parameter that removes residual time-invariant confounding between groups (11).

We explored whether the intervention's impact on private toilet construction varied by household wealth and scheduled caste status by stratifying the population by quintiles of a wealth index and by caste status. To create the wealth index we used the first component (eigenvector) from a principal components analysis of 20 household characteristics, which has been used as a wealth index score in low-income country studies (12, 13). The first component's eigenvalue was 3.74 and it explained 18.4% of the variability in household materials and assets. The wealth index was unimodal and approximately Gaussian and so we categorized households into quintiles.

Child anthropometry. We converted the anthropometric measurements to age- and sex-specific Z-scores using a publicly available Stata algorithm that references the 2006 WHO Growth Standards and calculated the difference in Z-score means (14). The algorithm identifies outliers for each Z-score, which we subsequently excluded from our analyses [ranges include (−6, 6) height-for-age, (−6, 5) weight-for-age, (−5, 5) weight-for-height, and (−5, 5) upper-arm-circumference-for-age]. We classified children as stunted, underweight, or malnourished if their Z-scores fell below −2 for height, weight, and weight-for-height/upper-arm-circumference Z-scores, respectively (14).

We calculated adjusted estimates of the intervention on child growth using linear models with household- and individual-level covariates and calculated marginal effects averaged over the covariate distribution in our population. This approach is often referred to as a g-computation estimator and is useful for estimating population intervention parameters (10, 15, 16). Although we measured child anthropometry at two points for most children in the study, we ignore the longitudinal data structure and analyze the data as a posttreatment comparison of means. Let Y be a child's Z-score and let A be an indicator variable equal to 1 if a child lives in an intervention village and 0 otherwise. Finally, let X be a set of covariates that could potentially confound or modify the relationship between A and Y, despite matching in the design. We include in X only covariates that could not reasonably be influenced by the intervention (Table S6). We assume that potential outcomes of Y are independent of treatment A conditional on X (strong ignorability).

We estimate the adjusted marginal difference between intervention and control groups using the following steps:

(i) Estimate the conditional expectation of Y given A and X using a generalized linear model with maximum likelihood $E[Y|A,X] = m(A,X)$ for some function of the covariates $m(\cdot)$. We used machine learning to specify $m(\cdot)$ (more below).

(ii) Evaluate the regression at $A = 1$ and $A = 0$ to get two predicted outcomes, $\widehat{Y}^a = \widehat{m}(A=a, X)$ for $a \in (0,1)$, for each individual.

(iii) Calculate the difference of the imputed values for each individual, and average the difference over individuals (j) and villages (i) to obtain an average, marginal effect of the intervention:

$$\widehat{\theta} = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{n_i}\sum_{j=1}^{n_i}\widehat{Y}_{ij}^{1} - \widehat{Y}_{ij}^{0}. \qquad \textbf{[S2]}$$

Child diarrhea. We quantified diarrhea using weekly longitudinal prevalence (total number of weeks with illness divided by the total weeks of observation), a disease measure that has been shown to be more strongly correlated with child mortality than incidence (17). We limited the longitudinal prevalence data to a 7-d recall window after identifying underreporting of symptoms for recall periods >7 d, using a standard data-adaptive approach (18, 19).

For our adjusted diarrhea analysis, we used an approach very similar to the anthropometry analysis, but allowed for time-varying covariates. The data are longitudinal with at most 12 monthly measurements. The outcome of interest $Y(t)$ is an individual-level indicator of a new episode of diarrhea in visit t (for $t = 0, \ldots, 11$). As with anthropometric outcomes, we model child-level diarrhea even though village is the treatment unit because there are household- and individual-level covariates that may be highly predictive of diarrhea. As before, let $X(t)$ be a set of covariates that may still confound or modify the relationship between A and Y and could not reasonably be influenced by the intervention (Table S6). The time-varying covariates in $X(t)$ include the child's age and month of follow-up (to control for seasonality in the outcome). We estimate the adjusted marginal difference between intervention and control groups using the following steps:

(i) Estimate probability of $Y(t)$ given A and $X(t)$ with maximum likelihood using a logistic regression model over all individuals and all time periods observed,

$$P(Y(t)|A,\, X(t)) = [1 + \exp - m(A,\, X(t))]^{-1} \qquad \textbf{[S3]}$$

for some function $m(\cdot)$ of the covariates A and $X(t)$.

(ii) Evaluate the regression fit at $A = 1$ and $A = 0$ to get two predicted probabilities for each individual at each time, where $\widehat{Y}^a(t) = [1 + \exp - \widehat{m}(A=a, X(t))]^{-1}$ for $a \in (0,\ 1)$.

(iii) Calculate the mean difference of imputed illness probabilities over all villages (i), individuals (j), and times (t) to obtain an average, marginal difference in longitudinal prevalence due to the intervention:

$$\widehat{\theta} = \frac{1}{12}\frac{1}{m}\sum_{t=0}^{11}\sum_{i=1}^{m}\frac{1}{n_i}\sum_{j=1}^{n_i}\widehat{Y}^1(t)_{ij} - \widehat{Y}^0(t)_{ij}. \qquad \textbf{[S4]}$$

Model selection approach and inference. In all adjusted analyses, the estimators require that we specify models for the mean outcomes $m(A, X)$ and $m(A, X(t))$. The functional forms of these models are unknown and could be a complex combination of child-, household-, and village-level covariates. However, the consistency of the adjusted estimators relies on the correct specification of these models.

To reduce potential bias from model misspecification we used a flexible machine algorithm called Super Learner that calculates predicted outcomes given a large set of covariates (20). Super Learner is implemented in R in the SuperLearner package. Super Learner is a metalearning algorithm that uses V-fold cross-validation to combine individual candidate learners into a single prediction using optimal weights. Each individual candidate learner is fit using V-fold cross-validation. We included the following candidate learners in the Super Learner: generalized linear models with main effects, elastic net regression (a hybrid of lasso and ridge regression) (21), and generalized additive models (22).

We estimated SEs and confidence intervals for all unadjusted and adjusted estimates using a stratified bootstrap at the village level, resampling households within each village with replacement and 1,000 iterations (23). This approach reflects the design and treats the 25 study villages as the target population. All model selection algorithms were applied within each bootstrap iteration, so the SEs and corresponding confidence intervals include variability from both household sampling and model selection. All analyses were conducted in R (version 2.9.1, www.R-project.org).

Table S6 includes covariates included in model selection for all models. For each specific outcome we subset these covariates to those that had a univariate positive association with the outcome. We defined a positive association as a univariate association with $P \leq 0.20$, or a difference of 0.2 SDs (anthropometry outcomes), or an odds ratio ≤ 0.83, ≥ 1.2 (diarrhea).

Subgroup Analysis of Toilet Construction. Stratified analyses based on quintiles of the principal components analysis-derived wealth index and social class show that the intervention expanded private toilets to the most marginalized population subgroups. In the poorest wealth index quintile, 1% of control households built toilets, whereas 28% of the poorest intervention households built toilets [risk difference (RD) = 0.27; 95% CI = 0.20, 0.35] (Fig. S1). In control villages, 11% of scheduled caste households built toilets vs. 74% in intervention villages (RD = 0.63; 95% CI = 0.50, 0.75). (Scheduled castes include historically disadvantaged, low-rank Indian castes, which are currently under government protection.)

Applications to Sustainability Research. This study design can quickly gather information about intervention sustainability, but it requires nuanced interpretation if the field study takes place during the postintervention period only. Without a measurement at the end of the intervention activities it can be difficult to interpret a null finding (Fig. S4). For example, this study suggests that the hardware improvements (toilets and water taps) have been highly sustainable over a 4-y period (scenario 1, Fig. S4), but necessary hygiene and sanitation behavior change either improved and then got worse by the time of the evaluation or never improved at all (scenarios 2 and 3, Fig. S4).

A limitation of the design for sustainability research is that it does not guarantee a "pure" control group in highly dynamic populations or in studies with long periods between baseline and follow-up surveys. In this evaluation, control villages had similar improvements in water sources independent of the NGO intervention, so the primary difference between groups is in sanitation conditions (Fig. 3 and Table S4), and health impacts must be interpreted in that context. The problem of maintaining a pure control group applies to any sustainability evaluation, even those that are randomized with prospective outcome measurement. The central issue is the practical and ethical dilemma created by measuring intervention sustainability using outcomes that require a control, where it becomes increasingly difficult to expect the control group to remain intervention-free over long periods of time. A prospective approach could establish intervention impacts during a shorter randomized evaluation and then measure compliance to the intervention (as a proxy for impacts) in the postintervention period.

1. Tamil Nadu State Planning Commission (2008) *Tamil Nadu 11th Five Year (2007–2012) Plan Document*, Technical Report (Tamil Nadu State Planning Commission, Tamil Nadu, India).
2. Kar K (2003) *Subsidy or Self-Respect? Participatory Total Community Sanitation in Bangladesh*, Technical Report, Working Paper 184 (Institute of Development Studies, Brighton, UK).
3. Arney H, Damodaran S, Meckel M, Barenberg A, White G (2008) *Creating Access to Credit for Water and Sanitation: Women's Self-Help Groups in India*, IRC Symposium: Sanitation for the Urban Poor. Partnerships and Governance (International Water and Sanitation Centre, Delft, The Netherlands), pp 1–10.
4. Rosenbaum PR, Rubin DB (1984) Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc* 79:516–524.
5. Rosenbaum PR, Rubin DB (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat* 39:33–38.
6. Goldman N, Vaughan B, Pebley AR (1998) The use of calendars to measure child illness in health interview surveys. *Int J Epidemiol* 27:505–512.
7. Baqui AH, et al. (1991) Methodological issues in diarrhoeal diseases epidemiology: Definition of diarrhoeal episodes. *Int J Epidemiol* 20:1057–1063.
8. ORC Macro (2006) *Demographic and Health Survey Interviewer's Manual*. MEASURE DHS Basic Documentation No. 2, Technical Report (ORC Macro, Atlanta).
9. Rahmathullah L, Underwood BA, Thulasiraj RD, Milton RC (1991) Diarrhea, respiratory infections, and growth are not affected by a weekly low-dose vitamin A supplement: A masked, controlled field trial in children in southern India. *Am J Clin Nutr* 54: 568–577.
10. Diggle PJ, Heagerty P, Liang KY, Zeger SL (2002) *Analysis of Longitudinal Data* (Oxford Univ Press, Oxford), 2nd Ed.
11. Meyer BD (1995) Natural and quasi-experiments in economics. *J Bus Econ Stat* 13: 151–161.
12. Filmer D, Pritchett LH (2001) Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of India. *Demography* 38: 115–132.
13. Vyas S, Kumaranayake L (2006) Constructing socio-economic status indices: How to use principal components analysis. *Health Policy Plan* 21:459–468.
14. World Health Organization (2006) *WHO Child Growth Standards: Length/Height-for-Age, Weight-for-Age, Weight-for-Length, Weight-for-Height and Body Mass Index-for-Age: Methods and Developments*, Technical Report (World Health Organization, Geneva).
15. Robins J (1987) A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *J Chronic Dis* 40 (Suppl 2):139S–161S.
16. Ahern J, Hubbard A, Galea S (2009) Estimating the effects of potential public health interventions on population disease burden: A step-by-step illustration of causal inference methods. *Am J Epidemiol* 169:1140–1147.
17. Morris SS, Cousens SN, Kirkwood BR, Arthur P, Ross DA (1996) Is prevalence of diarrhea a better predictor of subsequent mortality and weight gain than diarrhea incidence? *Am J Epidemiol* 144:582–588.
18. Alam N, Henry FJ, Rahaman MM (1989) Reporting errors in one-week diarrhoea recall surveys: Experience from a prospective study in rural Bangladesh. *Int J Epidemiol* 18: 697–700.
19. Feikin DR, et al. (2010) Evaluation of the optimal recall period for disease symptoms in home-based morbidity surveillance in rural and urban Kenya. *Int J Epidemiol* 39: 450–458.
20. van der Laan MJ, Polley EC, Hubbard AE (2007) Super learner. *Stat Appl Genet Mol Biol* 6:25.
21. Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33:1–22.
22. Hastie T, Tibshirani R (1990) *Generalized Additive Models* (Chapman & Hall, London).
23. Freedman DA (2005) *Statistical Models: Theory and Application* (Cambridge Univ Press, New York).

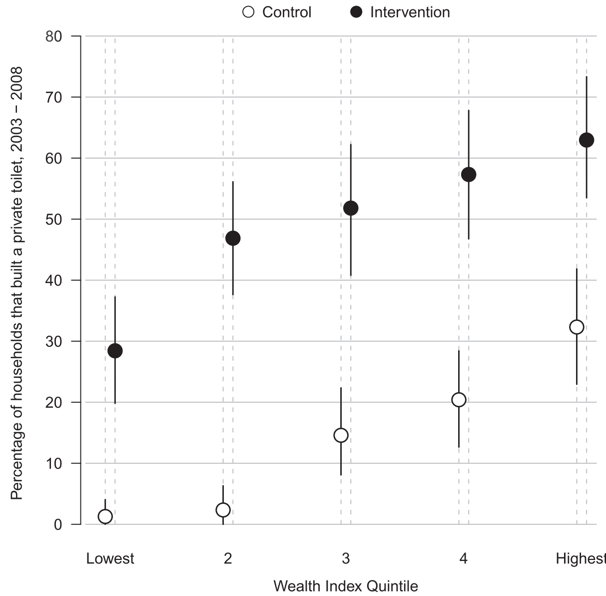

Fig. S1. Private toilet construction 2003–2008 by wealth index quintile. Vertical lines mark bootstrapped 95% confidence intervals. The wealth index was calculated with principal components analysis using housing materials and durable goods ownership (see *SI Materials and Methods* for details).

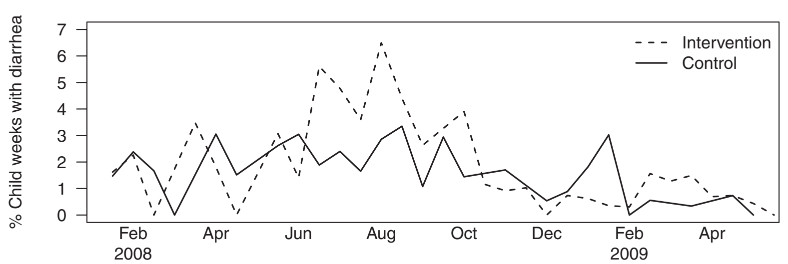

Fig. S2. Diarrhea prevalence in children <5 y old during the study period. Data are aggregated into 2-wk periods, with at least 50 children per group in each period. n = 120 cases in 7,183 child weeks (control) and n = 139 cases in 7,076 child weeks (intervention). Of the 1,284 children, 83% (1,070) had no diarrhea in the 12 wk of monitoring, and children had an average of 0.944 episodes per year.

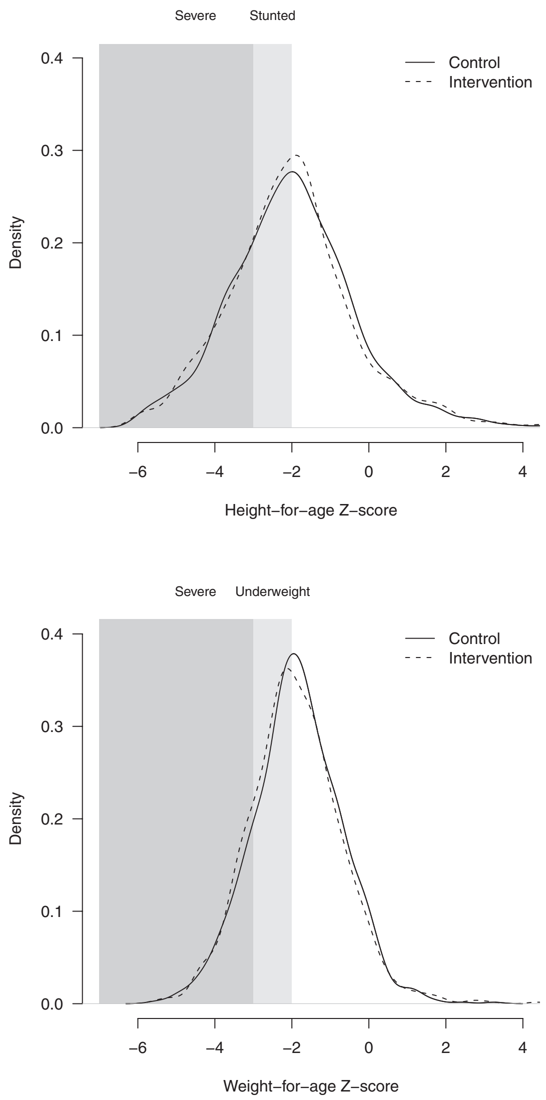

Fig. S3. Height-for-age and weight-for-age *Z*-score distributions. Smoothed kernel density distributions for *Z*-score values by treatment group for children <5 y old are shown. Shaded regions indicate the segments of the population that are stunted or underweight (*Z*-scores < −2) or are severely stunted or underweight (*Z*-scores < −3) compared with international reference values. Adjusted mean differences (intervention minus control) for height-for-age $Z = 0.01$ (95% CI = −0.15, 0.19), $n = 1{,}968$, and weight-for-age $Z = 0.03$ (95% CI = −0.11, 0.17), $n = 1{,}989$.

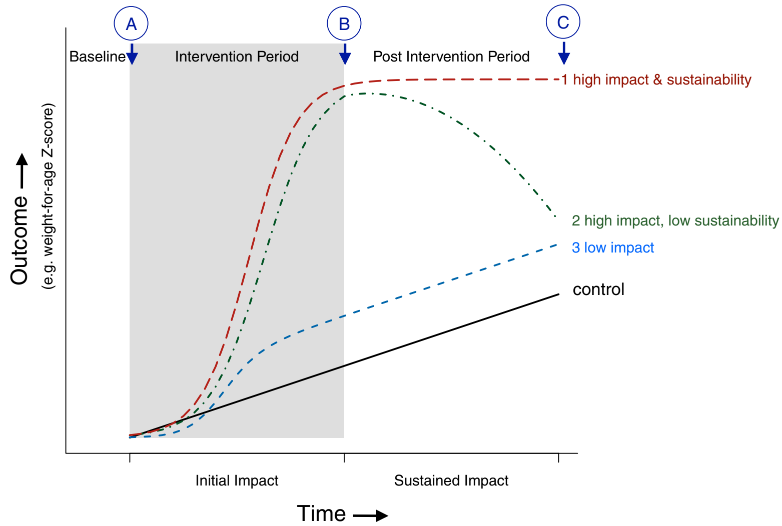

Fig. S4. Measuring sustainability. Three measurements (A, B, and C) are required to conclusively differentiate between the three hypothetical impact and sustainability scenarios.

Table S1. Summary of major intervention components in the 12 study villages

N	Village name	*N* households, census 2001	Project dates, mo/y	Age, mo	Brief intervention description
1	Keelakarthigaipatti	194	04/2003–03/2004	47	Water: community tube wells capped by hand pumps
2	Sakkampatti	140			Sanitation: School sanitary block, mobilization campaign (220 HH toilets)
					Hygiene: Village-wide hygiene education campaign, school health clubs
3	Mettupatti	70	04/2003–04/2004	46	Water: Community hand pumps
4	Periyanachipatti	80			Sanitation: School sanitary block, mobilization campaign (>63 HH toilets)
					Hygiene: Village-wide hygiene education campaign, school health clubs
5	Ponnusangampatti	290	01/2004–12/2004	38	Water: ≈279 HH taps
6	Melakothampatti	90			Sanitation: Mobilization campaign (273 HH toilets)
7	Theverappampatti	125			Hygiene: Village-wide hygiene education campaign
8	Ayinapatti	114	01/2005–03/2006	23	Water: ≈45 HH taps, new school water taps
					Sanitation: Renovated school toilets, mobilization campaign (97 HH toilets)
					Hygiene: Village-wide hygiene education campaign
9	Melakarthikaipatti	289	01/2005–03/2006	23	Water: ≈21 HH taps, 1 hand pump renovated, new school tap
					Sanitation: Renovated school toilets, mobilization campaign (370 HH toilets)
					Hygiene: Village-wide hygiene education campaign, school health clubs
10	Melanaduvalur	160	10/2005–10/2006	17	Water: ≈50 HH taps, 1 hand pump renovated, 14 public stand posts renovated
					Sanitation: Renovated school toilets, mobilization campaign (118 HH toilets)
					Hygiene: Village-wide hygiene education campaign
11	Kanganipatti	160	10/2005–10/2006	17	Water: 50 HH taps, 2 hand pumps renovated, new school tap
					Sanitation: Renovated school toilets, mobilization campaign (115 HH toilets)
					Hygiene: Village-wide hygiene education campaign
12	Kollapatti	220	10/2006–09/2007	5	Water: 100 HH taps, restored/repaired 4 existing hand pumps and school water facilities
					Sanitation: Renovated school toilets, mobilization campaign (118 HH toilets)
					Hygiene: Village-wide hygiene education campaign

Information was provided by the implementing organizations Water.org and Gramalaya. Villages 5–12 had access to microcredit loans for private household water and sanitation improvements. The age in months is the time elapsed from the intervention completion to the middle of the first round of data collection (February 2008). HH, household.

Table S2. Summary of postintervention characteristics at the beginning of data collection in 2008

Characteristics	Control mean (SE)	Intervention mean (SE)	Difference
Children <5 y old			
Female	0.52 (0.018)	0.47 (0.019)	−0.042
Age in mo	30.40 (0.816)	31.72 (0.774)	1.318
Currently breastfeeding	0.28 (0.016)	0.24 (0.018)	−0.034
Adults			
Work in agriculture	0.35 (0.040)	0.46 (0.024)	0.110**
Female literacy	0.70 (0.014)	0.65 (0.021)	−0.049*
Female education			
No education	0.21 (0.018)	0.24 (0.022)	0.031
Primary school	0.26 (0.022)	0.25 (0.011)	−0.016
Middle school	0.21 (0.018)	0.22 (0.017)	0.016
High school	0.20 (0.015)	0.20 (0.010)	−0.000
Higher secondary or more	0.12 (0.016)	0.08 (0.012)	−0.034*
Mother's age, y	26.98 (0.235)	26.76 (0.228)	−0.217
Households			
Scheduled caste	0.14 (0.037)	0.12 (0.068)	−0.021
Community group member	0.48 (0.037)	0.45 (0.057)	−0.032
Women members in SHG	0.35 (0.041)	0.34 (0.040)	−0.013
House has soil floor	0.28 (0.034)	0.35 (0.022)	0.073*
House has thatched roof	0.21 (0.030)	0.28 (0.031)	0.073*
Total persons in house	4.76 (0.090)	4.78 (0.050)	0.021
Sleeping rooms in house	1.79 (0.069)	1.76 (0.064)	−0.033
House has electricity	0.92 (0.018)	0.88 (0.032)	−0.038
Own their house	0.89 (0.023)	0.97 (0.009)	0.078***
Have a bank account	0.22 (0.017)	0.21 (0.028)	−0.012
Own radio	0.59 (0.028)	0.52 (0.032)	−0.067
Own television	0.73 (0.058)	0.58 (0.059)	−0.151*
Own mobile phone	0.32 (0.032)	0.33 (0.032)	0.009
Own motorcycle/scooter	0.27 (0.028)	0.24 (0.021)	−0.033
Own bicycle	0.74 (0.037)	0.79 (0.031)	0.054
Own mosquito net	0.12 (0.019)	0.14 (0.015)	0.019

Robust SEs account for clustering at the village level. Control sample sizes: n = 596 children <5 y old, 1,453 adults ≥15 y old, and 456 households. Intervention sample sizes: n = 577 children <5 y old, 1,465 adults ≥15 y old, 444 households. SHG: microcredit or finance self-help group.
*Different from 0 at the 10% level.
**Different from 0 at the 5% level.
***Different from 0 at the 1% level.

Table S3. Summary of water quality measures for village source and household drinking water samples

	Control		Intervention		
Water quality measure	*N*	Mean	*N*	Mean	Difference (95% CI)
Village source samples					
Log_{10} total coliforms	365	2.76	329	2.70	0.06 (−0.25, 0.20)
Log_{10} *E. coli*	32	0.96	30	0.76	−0.20 (−0.43, −0.06)
E. coli ≥10 cfu, %	32	21.88	30	10.00	−11.88 (−31.36, −2.49)
Positive for H_2S, %	366	62.84	329	53.80	−9.04 (−16.35, 0.67)
Household water samples					
Log_{10} total coliforms	1,540	3.35	1,475	3.25	−0.09 (−0.17, −0.01)
Log_{10} *E. coli*	227	0.97	214	0.99	0.02 (−0.08, 0.12)
E. coli ≥10 cfu, %	227	26.87	214	27.57	0.70 (−7.61, 8.10)
Positive for H_2S, %	1,543	85.81	1,483	82.74	−3.07 (−5.91, 0.21)

Samples were collected January 2008 through April 2009. Log_{10} total coliform and *E. coli* concentrations are in colony forming units (cfu) per 100 mL.

Table S4. Summary of toilet ownership, defecation practices, and perceived safety for women and girls

Outcome	Control, %	Intervention, %	Risk difference (95% CI)
Private toilets			
Have toilet in 2008	26	57	0.31 (0.26, 0.37)
New toilet since 2003	15	48	0.33 (0.28, 0.39)
Open defecation			
Any OD	88	77	−0.11 (−0.16, −0.06)
Adult men OD	84	68	−0.16 (−0.21, −0.10)
Adult women OD	81	61	−0.19 (−0.25, −0.14)
Children <5 y old OD	88	76	−0.12 (−0.17, −0.07)
Perception of women and girls during defecation			
Have privacy	59	72	0.13 (0.08, 0.19)
Safe, daytime	59	72	0.13 (0.07, 0.18)
Safe, nighttime	59	71	0.13 (0.07, 0.19)
Never harassed	59	74	0.15 (0.10, 0.21)

Open defecation (OD) practices were self-reported by household members. n = 456 control and n = 444 intervention households.

Table S5. Summary of hygiene and handwashing indicators

	Control		Intervention		
Hygiene indicator	*N*	(%)	*N*	(%)	Risk difference (95% CI)
Handwashing station spot check					
Station with water	5,297	(72)	5,130	(70)	−0.02 (−0.06, 0.02)
Station with water and soap/detergent	5,297	(55)	5,130	(52)	−0.03 (−0.07, 0.01)
Latrine spot check					
Hole is covered	1,048	(4)	2,291	(4)	−0.00 (−0.02, 0.01)
Water available for handwashing	1,048	(85)	2,291	(85)	0.00 (−0.06, 0.06)
Soap available for handwashing	1,048	(59)	2,290	(44)	−0.14 (−0.21, −0.08)
Feces on ground (not in hole)	951	(1)	2,069	(1)	−0.00 (−0.01, 0.00)
Household spot check					
Feces observed in living area	5,293	(26)	5,120	(34)	0.08 (0.04, 0.11)
Staff could smell feces during interview	5,297	(11)	5,130	(15)	0.04 (0.02, 0.07)
Garbage present inside home*	454	(8)	436	(6)	−0.02 (−0.06, 0.01)
Can produce a bar of soap*	456	(86)	444	(85)	−0.02 (−0.06, 0.03)
Soap is in plain view*	454	(26)	436	(19)	−0.07 (−0.12, −0.01)
Caregiver self-reported handwashing with soap†					
After changing baby/handling baby's feces	1,349	(15)	1,308	(17)	0.02 (−0.02, 0.06)
After defecation	1,349	(24)	1,308	(25)	0.01 (−0.03, 0.05)

Except where noted, data were collected during 12 monthly visits in 456 control households and 444 intervention households. Total *N* varies slightly by indicator.

*Measured in round 1 only.

†Measured in rounds 1–3 only: coded responses to an open-ended question.

Table S6. Covariates used in model selection for adjusted analyses

Category	Covariate
Children	Sex
	Age, mo
	Currently breast feeding
	Food types consumed in the previous 24 h: meat, eggs, leafy green vegetables, vitamin A-rich foods, milk, nuts/beans, cheese
Adults	Primary caregiver's education level (factor)
	Participate in a community group
	At least one parent works in agriculture
	Scheduled caste
	Use banking services
	Mother works
	Mother's age
Household	Soil floor (vs. concrete or tile)
	Thatch roof (vs. improved materials)
	Household has electricity
	Family owns its home
	Family owns its land
	Total persons living in the home
Durable goods ownership	Television
	Mobile phone
	Motorcycle or scooter
	Bicycle
	Mosquito net

Table S7. Child and household characteristics at enrollment for groups with complete and incomplete follow-up over the 12-mo study

	Complete follow-up [mean (SE)]		Incomplete follow-up [mean (SE)]		Difference
Child characteristics					
Female	0.49	(0.013)	0.56	(0.065)	−0.07
Age, mo	28.59	(0.564)	25.53	(2.238)	3.06
Currently breastfeeding	0.32	(0.013)	0.40	(0.052)	−0.09
Diarrhea in last 7 d	0.01	(0.004)	0.02	(0.018)	0.00
Cough in last 7 d	0.10	(0.007)	0.04	(0.025)	0.06**
Congestion in last 7 d	0.21	(0.014)	0.18	(0.054)	0.03
Height-for-age Z-score	−2.14	(0.109)	−2.18	(0.275)	0.03
Weight-for-age Z-score	−1.94	(0.063)	−2.14	(0.168)	0.20
Household characteristics					
Mother's age, y	26.91	(0.173)	25.23	(0.823)	1.69
Caregiver education					
No education	0.15	(0.016)	0.13	(0.053)	0.02
Primary school	0.13	(0.013)	0.08	(0.041)	0.06
Middle school	0.27	(0.018)	0.33	(0.085)	−0.05
High school	0.31	(0.014)	0.30	(0.078)	0.01
Higher secondary or more	0.14	(0.014)	0.18	(0.048)	−0.04
Household works in agriculture	0.65	(0.035)	0.40	(0.089)	0.25***
Scheduled caste	0.14	(0.041)	0.00	(0.000)	0.14***
Community group member	0.47	(0.034)	0.35	(0.067)	0.12
Women members in SHG	0.35	(0.029)	0.30	(0.060)	0.05
House has soil floor	0.31	(0.021)	0.43	(0.082)	−0.12
House has thatched roof	0.24	(0.024)	0.30	(0.071)	−0.06
Total persons in house	4.78	(0.054)	4.55	(0.149)	0.23
House has electricity	0.90	(0.018)	0.85	(0.053)	0.05
Own their house	0.93	(0.016)	0.88	(0.067)	0.06
Have a bank account	0.21	(0.016)	0.30	(0.081)	−0.09
Own television	0.66	(0.044)	0.60	(0.095)	0.06
Own mobile phone	0.32	(0.022)	0.40	(0.059)	−0.08
Own motorcycle/scooter	0.26	(0.019)	0.18	(0.046)	0.08*
Own bicycle	0.77	(0.024)	0.70	(0.074)	0.07
Own mosquito net	0.13	(0.012)	0.18	(0.080)	−0.04

Robust SE s account for clustering at the village level. SHG, microcredit or finance self-help group. N = 1,243 children with complete follow-up; *N* = 57 children with incomplete follow-up. *N* = 860 households with complete follow-up; *N* = 40 households with incomplete follow-up.

*Different from 0 at the 10% level.

**Different from 0 at the 5% level.

***Different from 0 at the 1% level.