

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

Published by Oxford University Press on behalf of the International Epidemiological Association.

International Journal of Epidemiology 2009;38:1651–1661

© The Author 2009; all rights reserved. Advance Access publication 2 July 2009

doi:10.1093/ije/dyp241

Evaluation of a pre-existing, 3-year household water treatment and handwashing intervention in rural Guatemala

Benjamin Arnold,^{1*} Byron Arana,² Daniel Mäusezahl,³ Alan Hubbard⁴ and John M Colford, Jr¹

Accepted 2 June 2009

Background The promotion of household water treatment and handwashing with soap has led to large reductions in child diarrhoea in randomized efficacy trials. Currently, we know little about the health effectiveness of behaviour-based water and hygiene interventions after the conclusion of intervention activities.

Methods We present an extension of previously published design (propensity score matching) and analysis (targeted maximum likelihood estimation) methods to evaluate the behavioural and health impacts of a pre-existing but non-randomized intervention (a 3-year, combined household water treatment and handwashing campaign in rural Guatemala). Six months after the intervention, we conducted a cross-sectional cohort study in 30 villages (15 intervention and 15 control) that included 600 households, and 929 children <5 years of age.

Results The study design created a sample of intervention and control villages that were comparable across more than 30 potentially confounding characteristics. The intervention led to modest gains in confirmed water treatment behaviour [risk difference = 0.05, 95% confidence interval (CI) 0.02–0.09]. We found, however, no difference between the intervention and control villages in self-reported handwashing behaviour, spot-check hygiene conditions, or the prevalence of child diarrhoea, clinical acute lower respiratory infections or child growth.

Conclusions To our knowledge this is the first post-intervention follow-up study of a combined household water treatment and handwashing behaviour change intervention, and the first post-intervention follow-up of either intervention type to include child health measurement. The lack of child health impacts is consistent with unsustainable behaviour adoption. Our findings highlight the difficulty of

¹ Division of Epidemiology, University of California, Berkeley, CA, USA.

² Center for Health Studies, Universidad del Valle de Guatemala, Guatemala City, Guatemala.

³ Swiss Tropical Institute, Department of Public Health and Epidemiology, University of Basel, Basel, Switzerland.

⁴ Division of Biostatistics, University of California, Berkeley, CA, USA.

* Corresponding author. Division of Epidemiology, University of California, 50 University Hall, Berkeley, CA 94720-7360, USA. E-mail: benarnold@berkeley.edu

implementing behaviour-based household water treatment and handwashing outside of intensive efficacy trials.

Keywords Household water treatment, handwashing, intervention, sustainability, propensity score matching, targeted maximum likelihood, Guatemala

Introduction

The prevalence of diarrhoea in developing countries has encouraged the development of low-cost, behaviour-based interventions to interrupt diarrhoea-causing pathogen transmission by improving water quality at the point-of-use and by washing hands using soap. Meta-analysis of efficacy studies indicate that household water treatment reduces diarrhoea in children <5 years of age by 30–40%,^{1–3} and handwashing with soap reduces diarrhoea and acute respiratory infections by 31 and 24%, respectively.^{4,5} The child health improvements documented in efficacy studies of the interventions reflect treatment effects in the short term (maximum study duration 12 months), typically under weekly or bi-weekly behavioural reinforcement.

In this article, we evaluate the effectiveness of a 3-year, combined water treatment and handwashing intervention in rural Guatemala through a novel extension of previously published design (propensity score matching) and statistical methods (targeted maximum likelihood estimation).^{6,7} Between October 2003 and September 2006, two non-governmental organizations, Caritas and Catholic Relief Services, implemented a large household water treatment and handwashing campaign in approximately 90 villages across three municipalities in rural eastern Guatemala. The implementing organizations had oversight from the SODIS Foundation (<http://www.fundacionsodis.org>). The promoted water treatment methods were boiling, solar disinfection (SODIS) and chlorination using diluted bleach; all having demonstrated health benefits.^{2,8} All villages received the same intervention package, and all activities were initiated at the same time. In each intervention village, Caritas technicians introduced the programme to village residents and recruited community-based health promoters for training. The trained health promoters later visited households with children ≤ 3 years of age or with pregnant mothers to promote water treatment and handwashing using soap. Promoters educated mothers about proper nutrition for their children, and at the end of each visit gave the family a small ration of rice, beans and oil. The visits were conducted monthly or bimonthly and lasted ~ 30 min each (see online appendix for additional details; available as supplementary data at *IJE* online). There exists no formal record of the proportion of eligible households that participated, but technicians on the ground suggest that the majority of

eligible households participated. At the conclusion of the intervention, the implementing organization conducted a survey of participating households and recorded water treatment behaviour based on self-report. The survey estimated that 70% of participating households regularly used some method of household water treatment (village level participation range: 29–100%). The SODIS Foundation provided these data at the start of our evaluation.

The primary objective of this study was to revisit households to assess water treatment behaviour, basic hygiene knowledge and practices, and child health 6 months after the conclusion of the intervention. We measured child health using self-reported symptoms of acute diarrhoeal and respiratory illness. We used anthropometric measurements that have demonstrated utility as outcome measures for water and sanitation interventions.^{9–11}

Methods

Study design and population

This study was conducted in the Camotán municipality in the mountainous state of Chiquimula, Guatemala, near the eastern border with Honduras. We used a cross-sectional cohort design with a 7-day retrospective risk period.¹² All data collection followed protocols approved by institutional review boards at the University of California, Berkeley and the Universidad del Valle de Guatemala, and all participants provided informed consent.

Since the intervention was non-randomized and villages were purposely selected by the implementing organizations, intervention villages were likely different, on average, from other villages in the study region. Baseline differences between intervention and control villages could lead to differences in water and hygiene practices and child health, independent of the intervention. To help increase comparability between intervention and control groups, we used restriction and propensity score matching¹³ based on pre-intervention characteristics to select intervention and control villages. All study villages—intervention and control—were selected in 2007, after the intervention ended. We adapted this selection approach from prospective, non-randomized, community-level intervention studies.^{6,14,15}

We obtained village-level 2002 census data with detailed information about demographics, education,

housing conditions, water sources and sanitation for 88 villages (30 intervention, 58 control) in the study region.¹⁶ We restricted our sample to villages with at least 50 children <5 years of age to guarantee a sufficient sample in each village. After a rapid assessment, we excluded two additional potential control villages that were qualitatively wealthier and had a large fraction of residents living in the USA sending remittances. This newly generated wealth was not reflected in the 2002 census. Our final sample for the match included 23 intervention villages and 26 potential controls.

We modelled the probability of participation in the behaviour change intervention using a logit model: $\text{logit} [\text{Pr}(A=1|W)] = \alpha'W$, where A is an indicator variable equal to 1, if a village participated in the intervention, and 0 otherwise, and W is a vector of characteristics that included the percentages in each village of: males; children <5 years of age; literate females; individuals employed in agriculture; households with private water taps; households with private wells; households with private latrines; households with electricity; and households with soil floors. Additionally, we measured the number of households, people per household and distance to the municipal centre. Importantly, the covariates in W were selected after detailed discussions with program technicians in the implementing organizations, and include information that the organizations used to select intervention villages.

In both the propensity score model used in our design and the targeted maximum likelihood estimation used in our adjusted analyses, one must estimate regressions that are not of direct interest (nuisance parameters), but are necessary to estimate the parameter of interest. The consistency of our estimates is contingent on the consistency of these nuisance parameter estimates. To estimate the nuisance parameters, we used the Deletion/Substitution/Addition (D/S/A) algorithm, which is a flexible model-selection approach that fits polynomial terms and their tensor products using cross-validation.¹⁷ Our D/S/A-selected propensity score model included the covariates as main effects with no interactions after allowing for up to two-way interactions, a maximum quadratic order for each term, and up to 15 total terms.¹⁷ We excluded intervention and control villages outside the region of common support (overlap) on the propensity score.¹⁸ Finally, we matched each intervention village to a control village without replacement, using the linear predictor of the model with nearest neighbour matching. This step resulted in 19 matched pairs. Due to time-in-field constraints, we included the 15 pairs with the closest match in our study.

We selected households within each village, using a stratified systematic sample. Our team used village sketch maps from the municipal planning department to split each village into two geographic strata with roughly equal numbers of houses. Within each

stratum, the field supervisor chose a random start, and the interviewer teams visited every third house until 10 houses were sampled. The inclusion criteria for the study were: (i) at least one child <5 years of age living in the home and (ii) the family had lived in the village since 2003 or earlier (the time of intervention start). If a selected household met our inclusion criteria but the primary caretaker was away, the field team returned two additional times before choosing a replacement household.

Data collection and measurement

A team of four trained fieldworkers and a field supervisor conducted household interviews during the dry season between April and June of 2007. The survey instrument was pre-tested and validated over a 2-week period in nearby, non-study villages. We collected household water samples in a random sample of 48 households from eight study villages (four intervention and four control, distributed across major water catchments). Water samples were collected in 100-ml Whirl-Pack™ bags in a fashion that mimicked each household's water retrieval practices, and transported in a cooler to the laboratory at the Universidad del Valle de Guatemala for culturing within 20 h of collection. Samples were processed using the Colilert Quantitray 1000 kit (IDEXX Laboratories), and we used a most probable number (MPN) table to quantify *Escherichia coli*.

The primary health outcomes of our study were diarrhoea, clinical acute lower respiratory-tract infections (ALRI), and child growth measured by height, weight and mid-upper-arm circumference. We collected gastrointestinal and respiratory symptoms over the previous 7 days using a health calendar modelled after Goldman *et al.*¹⁹ Gastrointestinal and respiratory outcomes were measured using daily longitudinal prevalence^{20,21} with 2-day recall after we identified under-reporting of symptoms for recall periods >2 days.^{22,23}

Self-reported health outcomes

We defined diarrhoea as three or more loose or watery stools in 24 h, or a single stool with blood or mucus.²⁴ We recorded symptoms of highly credible gastrointestinal illness (HCGI), which includes any of the following four conditions: vomiting, watery diarrhoea, soft diarrhoea and abdominal cramps, or nausea and abdominal cramps.²⁵ We defined clinical ALRI according to the World Health Organization (WHO) clinical case definition: cough or difficulty breathing with a raised respiratory rate measured with a wristwatch (more than 60 breaths/min in children <60 days old, more than 50 breaths/min for children aged 60–364 days, more than 40 breaths/min for children aged 1–5 years).²⁶

Anthropometrics

All fieldworkers were standardized on anthropometric measurement techniques over 2 days of training,²⁷

and they collected measurements in teams of two. Fieldworkers measured the weight and length of children <2 years of age in the lying position, and children aged 2–5 years standing using infant scales (Tanita 1380, 0.1 kg accuracy) and stadiometers (420 Measure All, 0.1 cm accuracy). Upper-arm circumference was measured for children aged ≥ 6 months at the mid-point of the upper right arm using an elastic tape (0.1 cm accuracy).

Knowledge and practices

We measured water-treatment practices using self-reported behaviour. Families that reported treating their water were classified as 'confirmed' if they met the three following criteria: (i) reported treating their water in the previous 7 days, (ii) had treated water in their home at the time of the interview and (iii) could produce the materials they used to treat water. Fieldworkers evaluated the presence of treated water based on self-reported information and a sample (not tested) provided by the family. Treatment materials included a designated pot and storage container for boiling water, plastic bottles for SODIS and liquid bleach or chlorine tablets and a designated storage container for chlorine treatment. Interviewers collected self-reported handwashing behaviour by asking an open question to mothers about when they washed their hands in the past 24 h and coding answers using five critical times: before cooking, eating, or feeding children and after defecation or changing the baby. Interviewers collected information about hygiene and water storage with discrete spot-check observations during the interview.

Statistical methods

Using daily symptoms reported at the time of the interview, we reconstructed the 48-h retrospective risk period for each child in the study.¹² The parameter of interest for all outcomes is the marginal treatment effect conditional on selection into the study based on restriction and propensity score matching. We estimate the parameter as:

$$E(Y | A = 1, W^*) - E(Y | A = 0, W^*)$$

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 ($W^* = W | A = 1$, more than 49 children <5 years of age). Thus, our inference is limited to the set of intervention villages for which there is a comparable control village based on the village selection method. For self-reported health outcomes, we calculated the difference in the daily longitudinal prevalence between the intervention and control groups. We converted the anthropometric measurements to age- and sex-specific Z-scores using a publicly available algorithm that references the 2006 WHO Growth Standards,²⁸ and calculated the difference in Z-score means.

For self-reported health and anthropometric outcomes, we attempted to improve the efficiency of the estimator and control for potential residual confounding using targeted maximum likelihood estimation (MLE).⁷ Non-technical details of our estimation strategy with complete notation and specifications are included in the appendix (available as supplementary data at *IJE* online). We calculated 95% confidence intervals (CIs) for unadjusted and adjusted estimates using a bootstrap with matched village pairs as the sampling unit to reflect the design and account for correlation between children within villages.

We estimated that a sample of 30 villages with 20 households per village would be sufficient to detect a difference of 5.5 percentage points in the longitudinal prevalence of diarrhoea, assuming a prevalence of 15% in the control group²⁹ and 80% power.

Results

Village selection and pre-intervention characteristics

The village selection process improved the comparability of intervention and control villages across a range of important, pre-intervention characteristics. Table 1 summarizes pre-intervention covariate means for control and intervention villages, their standardized difference (SD), which is equal to the difference in means in standard deviations and is a useful measure of balance.¹⁴ Before restriction and matching, intervention villages had more households on average (91 vs 55%; SD = 67) and a greater proportion of households with tap water (73 vs 53%; SD = 47), latrines (61 vs 53%; SD = 18) and electricity (61 vs 45%; SD = 32). After restriction and matching, balance improved for 7 of 12 covariates (Table 1).

Population characteristics

Interviewers visited 30 villages (660 households). Of these, 60 (9.0%) refused to participate in the study: 27/327 (8.3%) in intervention villages and 33/333 (9.9%) in control villages. The final sample included 600 households, 929 children <5 years of age and 1858 child-days of observation. Fieldworkers obtained complete anthropometric measurements for 872 (94%) children.

Intervention and control villages remained well balanced across a wide range of potentially confounding variables in 2007 (Table 2). Of 48 stored water samples, nearly all contained *Escherichia coli*: only two (4%) samples had MPN <1 per 100 ml and the mean (SD) \log_{10} *E. coli* concentration per 100 ml was 1.975 (0.870) in the control and 2.292 (1.033) in the intervention group. In intervention villages, 147 (49%) of study households reported participating in monthly visits from CRS/Caritas promoters at anytime since 2003 (only 5% of families could remember specific dates—month/year when asked).

Table 1 Summary of pre-intervention characteristics before and after village-selection, Camotán, Guatemala, Census 2002

Mean	All villages			Study sample		
	Control	Intervention	SD ^a	Control	Intervention	SD ^a
Male (%)	51.2	50.7	-1.0	52.2	51.6	-1.3
Children <5 years of age (%)	18.8	19.5	1.7	19.5	19.4	-0.2
Age 7-14 that work (%)	11.1	11.1	0.1	12.5	11.3	-3.7
Female literacy (%)	42.4	41.8	-1.2	38.2	41.6	7.0
Work in agriculture (%)	85.8	89.6	12.5	88.1	90.5	8.0
Total households	54.5	91.0	66.6	85.8	97.9	28.3
Houses with tap water (%)	52.5	73.3	47.0	59.8	67.8	17.2
Houses with latrine (%)	52.6	61.3	17.8	45.4	63.3	37.2
Houses with electricity (%)	45.3	60.8	31.8	46.9	55.1	16.5
Houses with soil floors (%)	75.0	80.5	14.0	79.2	79.4	0.5
Houses with thatched roofs (%)	44.8	49.8	9.8	47.6	48.9	2.8
Distance to municipal centre (km)	18.9	16.2	-28.0	19.0	15.7	-32.7
Number of villages	58	30		15	15	
Number of households	3160	2731		1287	1469	

^aThe standardized difference is equal to the difference in standard deviation of the mean (I-C) multiplied by 100. It is calculated as $(\mu_I - \mu_C)/[(S_I^2 + S_C^2)/2]^{1/2} \times 100$. For example, a difference of 1 standard deviation is equal to 100. A value of zero indicates equality of the means.

Knowledge and practice outcomes

Overall, 85% of study households were satisfied with their drinking-water quality, but only 65% of respondents believed their drinking water was clean. The proportion of participating families in intervention villages reporting water treatment dropped from 70% at the end of the intervention to 37% 6 months later. Households in intervention villages were more likely to treat their water than control households based on self-reported activity [33.3 vs 21.0%; Risk Difference (RD)=0.12, 95% CI 0.01-0.24], and based on confirmed water treatment activity at the time of the visit (8.7 vs 3.3%; RD=0.05, 0.01-0.10) (Table 3). The primary reason families gave for not treating their water was that it was already clean (48%), followed by bad taste (14%), not interested (11%) and no time (7%).

We did not observe differences between intervention and control groups in self-reported handwashing behaviour, or spot-check observations of hygienic conditions (Table 4). Soap was present in most homes (90%), and its use was similar in intervention and control villages (RD=0.03, -0.05 to 0.11).

Child health

In children <5 years of age, the daily longitudinal prevalence of diarrhoea and HCGI during the measurement period was 11.9 and 12.6%, respectively. Intervention and control groups did not differ in diarrhoea [Longitudinal Prevalence Difference (LPD)=0.004, 95% CI -0.051 to 0.058] or HCGI (LPD=0.005, -0.054 to 0.065) (Table 5). Respiratory illness was common among children in the study: the

daily longitudinal prevalence of cough or difficulty breathing was 30.0% and clinical ALRI was 6.9%. We observed no differences between the intervention and control groups in the longitudinal prevalence of cough or difficulty breathing (LPD=0.012, -0.097 to 0.137) or ALRI (LPD=0.019, -0.028 to 0.078).

Study children were generally well nourished but, consistent with our acute self-reported health outcomes, we found no differences in anthropometric measures between children living in intervention and control villages (Table 6). Adjustment for a large set of potential confounding variables using targeted maximum likelihood did not change the unadjusted results (see appendix, available as supplementary data at *IJE* online).

Discussion

To our knowledge, this is the first post-intervention follow-up study of a combined household water treatment and handwashing behaviour change intervention, and the first to extend propensity score matching and targeted maximum likelihood estimation to the design and analysis of a pre-existing intervention. The absence of child health impacts is consistent with the modest improvement we observed in water treatment behaviour (Table 3), no detectable differences in handwashing behaviour, and highly contaminated living environments (Table 4). These findings are consistent with efficacy trials of household water treatment that have found that health impacts are contingent on compliance.^{2,3} The large difference between self-reported and confirmed water treatment

Table 2 Summary of post-intervention characteristics and measures of balance following a 3-year point-of-use water treatment and handwashing intervention, Camotán, Guatemala, 2007

Mean	Control	Intervention	Standardized difference ^a	Variance Ratio ^b
Child characteristics	n = 455	n = 474		
Male (%)	50.3	54.4	8.21	0.99
Age (months)	30.5	28.2	-13.56	0.98
Total months breastfed	16.4	15.1	-17.22	0.91
Mother's characteristics	n = 300	n = 300		
Age (years)	30.5	29.7	-9.64	0.83
Currently pregnant (%)	11.0	12.7	5.15	1.13
Works for money (%)	11.3	16.0	13.59	1.34
Literate (%)	25.0	28.7	8.27	1.09
Leaves village more than once per week (%)	38.5	28.1	-10.90	0.84
Household characteristics	n = 300	n = 300		
Total persons living in home	6.8	7.0	8.41	1.25
Number of children <15 years old	3.7	4.0	11.74	1.26
Number of children <5 years old	1.5	1.6	10.00	1.05
Electricity (%)	43.7	55.3	23.46	1.00
Dirt floor (%)	67.0	61.7	-11.13	1.07
Thatch roof (%)	25.3	25.0	-0.77	0.99
Home ownership (%)	95.3	92.7	-11.23	1.53
Land ownership (%)	51.3	49.7	-3.33	1.00
Use banking services (%)	8.7	7.3	-4.91	0.86
Have relatives in USA (%)	13.3	11.0	-7.13	0.85
Have relatives in the capital (%)	17.0	16.3	-1.79	0.97
Durable good ownership (%)				
Refrigerator	9.7	6.3	-12.29	0.68
Radio	81.0	83.7	6.98	0.89
Television	14.3	13.3	-2.89	0.94
Mobile phone	28.3	35.3	15.04	1.13
Bicycle	7.0	15.0	25.74	1.96
Automobile	3.7	4.0	1.73	1.09
Water supply				
Primary water source (%)				
Private tap	67.3	66.0	-2.82	1.02
Public tap	8.7	9.0	1.17	1.03
Public well	8.0	4.7	-13.70	0.60
Spring	12.3	10.7	-5.22	0.88
Surface water (river/lake)	3.0	8.3	23.18	2.63
Other	0.7	1.3	6.69	1.99
Minutes per day retrieving water	26.9	29.5	3.84	1.09
Satisfied with water quantity (%)	81.0	73.3	-18.31	1.27
Sanitation				
Latrine ownership (%)	55.3	64.0	17.71	0.93
Animals in living vicinity of house (%)				
Pigs	42.3	56.0	27.55	1.01
Chickens/ducks	94.7	95.0	1.50	0.94
Dogs/cats	92.7	93.3	2.61	0.92
Cows/horses/mules/donkeys	14.3	9.0	-16.64	0.67

^aThe standardised difference is equal to the difference in standard deviation of the mean (I-C) multiplied by 100. It is calculated as $(\mu_I - \mu_C)/[(S_I^2 + S_C^2)/2]^{1/2} \times 100$. For example, a difference of 1 standard deviation is equal to 100. A value of zero indicates equality of the means.

^bThe variance ratio is equal to S_I^2/S_C^2 . A value of 1 indicates equal variance.

Table 3 Water storage and treatment practices following a 3-year point-of-use water treatment and handwashing intervention, Camotán, Guatemala, 2007

Outcome	Control		Intervention		RD	
	<i>n</i>	(%)	<i>n</i>	(%)	(95% CI) ^a	
Water storage practices						
Stores drinking water in home	241	(80.3)	242	(80.7)	0.003	(−0.07 to 0.08)
Exclusively covered or narrow mouth	182	(60.7)	188	(62.7)	0.020	(−0.06 to 0.10)
Exclusively covered	174	(58.0)	187	(62.3)	0.043	(−0.04 to 0.12)
Self-reported water treatment						
Any method	63	(21.0)	100	(33.3)	0.123	(0.00 to 0.24)
Boiling	50	(16.7)	60	(20.0)	0.033	(−0.07 to 0.14)
SODIS	0	(0.0)	39	(13.0)	0.130	(0.07 to 0.19)
Chlorine	12	(4.0)	15	5.0	0.010	(−0.02 to 0.04)
Confirmed water treatment^b						
Any method	10	(3.3)	26	(8.7)	0.053	(0.02 to 0.09)
Boiling	8	(2.7)	15	(5.0)	0.023	(0.00 to 0.05)
SODIS	0	(0.0)	11	(3.7)	0.037	(0.01 to 0.06)
Chlorine	2	(0.7)	3	(1.0)	0.003	(−0.01 to 0.02)

^a95% CI calculated by bootstrap resampling matched village pairs with 1000 iterations.

^bWater treatment was confirmed if the family (i) self-reported treating water in the previous 7 days, (ii) had treated water at the time of the interview and (iii) could produce the materials they used to treat water.

n = 300 intervention and *n* = 300 control households.

Table 4 Handwashing and hygiene conditions following a 3-year point-of-use water treatment and handwashing intervention, Camotán, Guatemala, 2007

Outcome	Control		Intervention		RD	
	<i>n</i>	(%)	<i>n</i>	(%)	(95% CI) ^a	
Self-reported handwashing^b						
Before cooking	243	(81.0)	232	(77.3)	−0.037	(−0.11 to 0.04)
Before eating	100	(33.3)	101	(33.7)	0.003	(−0.09 to 0.09)
Before feeding children	61	(20.3)	49	(16.3)	−0.040	(−0.14 to 0.06)
After defaecation	157	(52.3)	152	(50.7)	−0.017	(−0.12 to 0.09)
After changing baby	38	(12.7)	38	(12.7)	0.000	(−0.10 to 0.10)
Spot-check observations						
Mother's hands are clean	271	(90.3)	267	(89.0)	−0.013	(−0.07 to 0.04)
Mother's nails are clean	220	(73.3)	216	(72.0)	−0.013	(−0.10 to 0.08)
Can produce a bar of soap	266	(88.7)	275	(91.7)	0.030	(−0.05 to 0.11)
Bar of soap is in plain view	170	(56.7)	177	(59.0)	0.023	(−0.07 to 0.12)
Food is covered	160	(53.3)	167	(55.7)	0.023	(−0.11 to 0.16)
Garbage present inside home	173	(57.7)	142	(47.3)	−0.103	(−0.23 to 0.02)
Feces observed in living area	211	(70.3)	231	(77.0)	0.067	(−0.05 to 0.18)

^a95% CI calculated by bootstrap resampling matched village pairs with 1000 iterations.

^bResponses to an open-ended question about handwashing in the 24 h before the interview.

n = 300 intervention and *n* = 300 control households.

(Table 3) suggests that self-reported water treatment behaviour overestimates actual practice. Schmidt and Cairncross recently outlined the problems of self-reported health outcomes in non-blinded studies of household water treatment.³⁰ Our self-reported health outcomes likely suffer from less reporting bias

because we do not have frequent, repeated visits, we used a health calendar to collect symptoms, and we minimized recall to 48 h. Our objective anthropometric outcomes are an important complement to the self-reported outcomes, and the null treatments effect is consistent across all outcomes.

Table 5 Days of observation, days of illness and difference in longitudinal prevalence in children <5 years of age following a 3-year point-of-use water treatment and handwashing intervention, Camotán, Guatemala, 2007

Outcome	Control		Intervention		LPD (95% CI) ^a
	Days of observations	Days of illness	Days of observations	Days of illness	
Diarrhoea	910	107	948	115	0.004 (−0.051 to 0.058)
HCGI	910	113	948	122	0.005 (−0.054 to 0.065)
Cough or difficulty breathing	910	268	948	291	0.012 (−0.097 to 0.137)
Congestion or eoryza	910	144	948	173	0.024 (−0.026 to 0.071)
ALRI	910	54	948	71	0.019 (−0.028 to 0.078)

^a95% Confidence Intervals calculated by bootstrap resampling matched village pairs with 1000 iterations.

Table 6 Anthropometric measurements in children <5 years of age following a 3-year point-of-use water treatment and handwashing intervention, Camotán, Guatemala, 2007

Z-score ^a	Control			Intervention			Difference	
	<i>n</i>	Mean	(SD)	<i>n</i>	Mean	(SD)	(95% CI) ^b	
Weight	423	−1.312	(1.325)	453	−1.365	(1.219)	−0.053	(−0.331 to 0.206)
Height	424	−2.177	(1.880)	453	−2.136	(1.596)	0.041	(−0.305 to 0.326)
Weight-for-height	421	−0.122	(1.728)	451	−0.187	(1.421)	−0.066	(−0.248 to 0.124)
Mid-upper-arm circumference	401	0.348	(0.884)	426	0.335	(0.825)	−0.014	(−0.166 to 0.145)

^aZ-scores were calculated using a standard WHO Stata algorithm and 2006 world reference data.

^b95% CIs calculated by bootstrap resampling matched village pairs with 1000 iterations.

Our confirmed water treatment adoption in intervention households (9%) is lower than water treatment adoption reported after a CARE/Madagascar Safe Water System (SWS) campaign, which promoted chlorine treatment with safe storage. Ram *et al.* found that 54% (29/54) of households had detectable free chlorine in their stored water 18 months after the campaign.³¹ Parker *et al.* also report higher sustained adoption after a clinic-based SWS and handwashing intervention: 71% (36/51) of households had detectable free chlorine 1 year after the intervention.³² Our water treatment behaviour results are consistent with Luby *et al.*, who found 5% (22/462) of households regularly treating their water 6 months after the completion of a year-long household flocculent-disinfectant intervention trial in Guatemala.³³ Our handwashing and hygiene findings suggest that the presence of soap is common even in the absence of heightened promotion, but that self-reported handwashing remains infrequent around all key activities except cooking (Table 4). This finding contrasts with two earlier studies that report sustained handwashing behaviour change many years after short-duration interventions, though neither study included an adequate control group.^{34,35}

Our results demonstrate that with available pre-intervention secondary data, the careful selection of a study population in the design stage can greatly improve the baseline comparability of intervention and control groups in the evaluation of a pre-existing intervention. Prospective, randomized designs have implemented pair matching on one or two variables

such as baseline illness or community size to help improve the comparability of treatment arms.³⁶ The limitation of one- or two-variable matching in non-randomized designs is that implementing organizations usually rely on many (or ill-defined) characteristics to choose intervention recipients, and matching on one or two covariates is unlikely to balance a large set of potential confounders. Propensity score matching simplifies multivariate matching by accommodating continuous covariates and reducing a large set of matching characteristics to a single scalar. Restriction and matching limit inference to the population ultimately included in the study, but when interventions are targeted to a subset of the population, making inference to segments of the population that do not share characteristics with those treated must rely on extrapolation.

There are limitations to our study. Our design does not include baseline outcome measurement. It is possible that intervention villages were in worse health condition than controls before the intervention, and that their health improved to control levels by 2007. We think this scenario is unlikely given the limited behaviour change we observed and the comparability of intervention and control villages across a broad range of demographic, socio-economic and environmental characteristics in both 2002 and 2007. Secondly, we only measured outcomes at one point in time, and it is possible that we misclassified families with respect to behaviour and illness, since these characteristics likely vary over time. We attempted to reduce misclassification by using measures of water

treatment and hygiene that did not change rapidly over time, and by supplementing self-reported health outcomes with anthropometric measurements. Thirdly, only 49% of intervention households reported participating in the intervention. This modest participation rate may have diluted the treatment effect sufficiently to lead to a null finding with respect to effectiveness but is itself an important finding with respect to future implementation. Comparing the subgroup of participating intervention households to non-participants in unadjusted and adjusted analyses did not change our conclusions (see appendix available as supplementary data at *IJE* online).

A final limitation is that our cross-sectional measurement does not ultimately resolve whether the intervention was sustainable. Two scenarios are consistent with our results: (i) the intervention successfully increased water treatment behaviour among participating families, but the new behaviours were not sustained after intervention completion, or (ii) the intervention never led to behaviour change and there was nothing to sustain. The only available reference point to evaluate these scenarios was an end-of-intervention survey conducted by the implementing organization in which 70% of participating households in our study villages reported consistent household water treatment. Whereas this estimate is likely biased upward, in our survey 6 months after the intervention, 33% of intervention village households self-reported that they treat their water, a measurement prone to similar upward bias (Table 3). Taken together, these measurements suggest that water treatment likely tapered off after activities ceased. Future studies could address sustainability more rigorously by collecting measurements at the end of the intervention period followed by identical measures later to capture changes over time.

Conclusion

Six months after a 3-year intervention in rural Guatemala we observed minimal sustained water

treatment and handwashing behaviour, which consequently led to no impacts on acute gastrointestinal, respiratory or anthropometric measures. Our findings highlight the difficulty of achieving sustained new behaviour adoption in the context of non-research intervention campaigns. Future research in this sector should focus on identifying techniques to improve and sustain behaviour adoption that implementing organizations can use in development programs. Our study design provides a useful template for effectiveness evaluations of pre-existing intervention campaigns initiated outside of formal research activities.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

Institute for Public Health and Water Research (<http://www.ipwr.org>). Funding to pay the Open Access publication charges for this article was provided by The Berkeley Research Impact Initiative (BRII).

Acknowledgements

The authors gratefully acknowledge Nazario Lopez and the interviewer team for long hours in the field, and Maricruz Alvarez for all laboratory analyses. Rodrigo Gramajo Rodriguez and Andri Christen provided valuable input into the questionnaire design and its translation. Fundación SODIS personnel Matthias Saladin and Alvaro Solano, and former Caritas personnel Marina de Lantán, Edna Mendoza and Carlos Miguel Loyo provided important information about the intervention and input into the questionnaire. Jan Hattendorf provided valuable comments on a draft of this manuscript.

Conflict of interest: None declared.

KEY MESSAGES

- Six months after a 3-year intervention that promoted household water treatment and handwashing with soap in rural Guatemala, we measured behavioural and child health outcomes in 600 households from 15 intervention and 15 matched control villages.
- We found that 9% of intervention households treat their water compared with 3% in control villages.
- Self-reported handwashing behaviour and observed hygienic conditions were similarly poor in both intervention and control villages.
- Consistent with the low sustained behaviour adoption, we found no difference between intervention and control villages in child diarrhoea, respiratory infections or growth.
- Our findings highlight the difficulty of sustaining behaviour-based interventions outside of controlled efficacy trials.

References

- ¹ Fewtrell L, Kaufmann RB, Kay D, Enanoria W, Haller L, Colford JM Jr. Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *Lancet Infect Dis* 2005;**5**:42–52.
- ² Clasen T, Schmidt WP, Rabie T, Roberts I, Cairncross S. Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. *Br Med J* 2007;**334**:782.
- ³ Arnold BF, Colford JM Jr. Treating water with chlorine at point-of-use to improve water quality and reduce child diarrhea in developing countries: a systematic review and meta-analysis. *Am J Trop Med Hyg* 2007;**76**:354–64.
- ⁴ Ejemot RI, Ehiri JE, Meremikwu MM, Critchley JA. Hand washing for preventing diarrhoea. *Cochrane Database Syst Rev* Issue 1. Art. No.: CD004265. DOI: 10.1002/14651858.CD004265.pub2. (online at: <http://www.cochrane.org/reviews/en/ab004265.html>).
- ⁵ Rabie T, Curtis V. Handwashing and risk of respiratory infections: a quantitative systematic review. *Trop Med Int Health* 2006;**11**:258–67.
- ⁶ Preisser JS, Young ML, Zaccaro DJ, Wolfson M. An integrated population-averaged approach to the design, analysis and sample size determination of cluster-unit trials. *Stat Med* 2003;**22**:1235–54.
- ⁷ van der Laan M, Rubin D. Targeted Maximum Likelihood Learning. *Int J Biostatistics* 2006;**2**:1–38.
- ⁸ Vanderslice J, Briscoe J. All coliforms are not created equal – a comparison of the effects of water source and in-house water contamination on infantile diarrheal disease. *Water Resour Res* 1993;**29**:1983–95.
- ⁹ Esrey SA, Habicht JP, Casella G. The complementary effect of latrines and increased water usage on the growth of infants in rural Lesotho. *Am J Epidemiol* 1992;**135**:659–66.
- ¹⁰ Esrey SA. Water, waste, and well-being: a multicountry study. *Am J Epidemiol* 1996;**143**:608–23.
- ¹¹ Checkley W, Gilman RH, Black RE *et al.* Effect of water and sanitation on childhood health in a poor Peruvian peri-urban community. *Lancet* 2004;**363**:112–18.
- ¹² Hudson JI, Pope HG Jr, Glynn RJ. The cross-sectional cohort study: an underutilized design. *Epidemiology* 2005;**16**:355–59.
- ¹³ Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;**70**:41–55.
- ¹⁴ Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Stat Med* 2007;**26**:20–36.
- ¹⁵ Pattanayak SK, Poulos C, Yang JC, Patil SR, Wendland KJ. Of taps and toilets: quasi-experimental protocol for evaluating community-demand-driven projects. *J Water Health* 2009;**7**:434–51.
- ¹⁶ INE. Censos Nacionales XI de Poblacion y VI de Habitacion, Guatemala 2002. Instituto Nacional de Estadística, 2002.
- ¹⁷ Sinisi SE, van der Laan MJ. Deletion/substitution/addition algorithm in learning with applications in genomics. *Stat Appl Genet Mol Biol* 2004;**3**:Article18.
- ¹⁸ Heckman J, Ichimura H, Smith J, Todd P. Characterizing selection bias using experimental data. *Econometrica* 1998;**66**:1017–98.
- ¹⁹ Goldman N, Vaughan B, Pebley AR. The use of calendars to measure child illness in health interview surveys. *Int J Epidemiol* 1998;**27**:505–12.
- ²⁰ Morris SS, Cousens SN, Kirkwood BR, Arthur P, Ross DA. Is prevalence of diarrhea a better predictor of subsequent mortality and weight gain than diarrhea incidence? *Am J Epidemiol* 1996;**144**:582–88.
- ²¹ Schmidt WP, Luby SP, Genser B, Barreto ML, Clasen T. Estimating the longitudinal prevalence of diarrhea and other episodic diseases: continuous versus intermittent surveillance. *Epidemiology* 2007;**18**:537–43.
- ²² Alam N, Henry FJ, Rahaman MM. Reporting errors in one-week diarrhoea recall surveys: experience from a prospective study in rural Bangladesh. *Int J Epidemiol* 1989;**18**:697–700.
- ²³ Heuveline P, Goldman N. A description of child illness and treatment behavior in Guatemala. *Soc Sci Med* 2000;**50**:345–64.
- ²⁴ Baqui AH, Black RE, Yunus M, Hoque AR, Chowdhury HR, Sack RB. Methodological issues in diarrhoeal diseases epidemiology: definition of diarrhoeal episodes. *Int J Epidemiol* 1991;**20**:1057–63.
- ²⁵ Colford JM Jr, Wade TJ, Sandhu SK *et al.* A randomized, controlled trial of in-home drinking water intervention to reduce gastrointestinal illness. *Am J Epidemiol* 2005;**161**:472–82.
- ²⁶ Gove S. Integrated management of childhood illness by outpatient health workers: technical basis and overview. The WHO Working Group on Guidelines for Integrated Management of the Sick Child. *Bull World Health Organ* 1997;**75 (Suppl 1)**:7–24.
- ²⁷ ORC Macro. 2006. *Demographic and Health Survey Interviewers Manual*. MEASURE DHS Basic Documentation No. 2. Calverton, Maryland, USA.: ORC Macro.
- ²⁸ WHO. *WHO Anthro Software*. 2.02 edn. Geneva: WHO, 2008.
- ²⁹ Reller ME, Mendoza CE, Lopez MB *et al.* A randomized controlled trial of household-based flocculant-disinfectant drinking water treatment for diarrhea prevention in rural Guatemala. *Am J Trop Med Hyg* 2003;**69**:411–19.
- ³⁰ Schmidt W-P, Cairncross S. Household Water Treatment in Poor Populations: Is There Enough Evidence for Scaling up Now? *Environ Sci Technol* 2009;**43**:986–92.
- ³¹ Ram PK, Kelsey E, Rasoatiana *et al.* Bringing safe water to remote populations: an evaluation of a portable point-of-use intervention in rural Madagascar. *Am J Public Health* 2007;**97**:398–400.
- ³² Parker AA, Stephenson R, Riley PL *et al.* Sustained high levels of stored drinking water treatment and retention of hand-washing knowledge in rural Kenyan households following a clinic-based intervention. *Epidemiol Infect* 2006;**134**:1029–36.
- ³³ Luby SP, Mendoza C, Keswick BH, Chiller TM, Hoekstra RM. Difficulties in bringing point-of-use water treatment to scale in rural Guatemala. *Am J Trop Med Hyg* 2008;**78**:382–87.

- ³⁴ Cairncross S, Shordt K, Zacharia S, Govindan BK. What causes sustainable changes in hygiene behaviour? A cross-sectional study from Kerala, India. *Soc Sci Med* 2005;**61**:2212–20.
- ³⁵ Wilson JM, Chandler GN. Sustained improvements in hygiene behaviour amongst village women in Lombok, Indonesia. *Trans R Soc Trop Med Hyg* 1993;**87**: 615–16.
- ³⁶ Campbell MJ, Donner A, Klar N. Developments in cluster randomized trials and Statistics in Medicine. *Stat Med* 2007;**26**:2–19.