Scaling Up a Water, Sanitation, and Hygiene Program in Rural Bangladesh: The Role of Program Implementation

Jade Benjamin-Chung, PhD, MPH, Sonia Sultana, MBBS, MPH, Amal K. Halder, PhD, Mohammed Ali Ahsan, MA, Benjamin F. Arnold, PhD, MPH, Alan E. Hubbard, PhD, Leanne Unicomb, PhD, Stephen P. Luby, MD, and John M. Colford Jr, MD, PhD, MPH

Objectives. To evaluate whether the quality of implementation of a water, sanitation, and hygiene program called SHEWA-B and delivered by UNICEF to 20 million people in rural Bangladesh was associated with health behaviors and sanitation infrastructure access.

Methods. We surveyed 33 027 households targeted by SHEWA-B and 1110 SHEWA-B hygiene promoters in 2011 and 2012. We developed an implementation quality index and compared the probability of health behaviors and sanitation infrastructure access in counterfactual scenarios over the range of implementation quality.

Results. Forty-seven percent of households (n = 14 622) had met a SHEWA-B hygiene promoter, and 47% of hygiene promoters (n = 527) could recall all key program messages. The frequency of hygiene promoter visits was not associated with improved outcomes. Higher implementation quality was not associated with better health behaviors or infrastructure access. Outcomes differed by only 1% to 3% in scenarios in which all clusters received low versus high implementation quality.

Conclusions. SHEWA-B did not meet UNICEF's ideal implementation quality in any area. Improved implementation quality would have resulted in marginal changes in health behaviors or infrastructure access. This suggests that SHEWA-B's design was suboptimal for improving these outcomes. (*Am J Public Health.* 2017;107:694–701. doi:10.2105/AJPH.2017.303686)

Scientists and development stakeholders argue that health programs proven effective in randomized efficacy trials should be translated into large-scale programs to benefit public health.[1] Substantial evidence supports the scaling up of numerous health programs.[2–4] Since the establishment of the Millennium Development Goals, the funding and motivation for scaling up has grown.[5] However, translating efficacious, small-scale programs to a large scale can present implementation challenges.[6,7] A growing body of literature documents barriers and facilitators to scaling up, yet there is little empirical evidence about how best to scale up.[7–12] In the water, sanitation, and hygiene (WASH) sector in particular, several evaluations of large-scale programs have found limited health impacts and incomplete program uptake.[13–17] As stakeholders consider scaling up other WASH programs, there is a scientific imperative to evaluate program impacts on health and to document reasons for intervention success or failure.[6]

One of the largest WASH programs in history was the Sanitation Hygiene Education and Water Supply in Bangladesh (SHEWA-B) program, which was implemented by UNICEF and the Government of Bangladesh. SHEWA-B targeted approximately 20.4 million beneficiaries from 2007 to 2012. The program promoted hygiene practices and aimed to reduce diarrhea and WASH-related diseases among the poorest in rural Bangladesh. A 2009 interim assessment of SHEWA-B found little to no improvement in health behaviors (e.g., handwashing), access to hygiene or sanitation infrastructure, or prevalence of diarrhea and respiratory illness among children younger than 5 years.[18] These results could reflect a suboptimal program that needed to be better tailored to the target population or an appropriate program that needed to be better implemented. We conducted an observational study in SHEWA-B program areas to measure whether health behaviors and access to hygiene and sanitation infrastructure would have been better if SHEWA-B implementation had been of higher quality in all clusters. A positive association between implementation quality and health behaviors and access to hygiene and sanitation infrastructure would suggest that the SHEWA-B program had an appropriate design, even if it was imperfectly implemented, and a lack of association would suggest that the program needed to be better tailored to the population.

ABOUT THE AUTHORS

Jade Benjamin-Chung, Benjamin F. Arnold, and John M. Colford Jr are with the Division of Epidemiology, School of Public Health, University of California, Berkeley. Alan E. Hubbard is with the Division of Biostatistics, School of Public Health, University of California, Berkeley. Sonia Sultana, Amal K. Halder, Mohammed Ali Ahsan, and Leanne Unicomb are with the Environmental Interventions Unit, Infectious Disease Division, International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka. Stephen P. Luby is with the Department of Medicine, Stanford University, Stanford, CA.

Correspondence should be sent to Jade Benjamin-Chung, PhD, MPH, Division of Epidemiology, School of Public Health, 101 Haviland Hall, MC #7358, University of California, Berkeley, Berkeley, CA 94720-7358 (e-mail: jadebc@berkeley.edu).
Reprints can be ordered at http://www.ajph.org by clicking the "Reprints" link.

This article was accepted January 20, 2017.

doi: 10.2105/AJPH.2017.303686

METHODS

UNICEF and the Bangladesh Department of Public Health Engineering (DPHE) delivered the SHEWA-B program to more than 20 million people in underserved areas without other large-scale WASH programs (Figure A, available as a supplement to the online version of this article at http://www.ajph.org). The program was implemented from 2007 until mid-2013. SHEWA-B consisted of rural, urban, and school-based programs; this assessment focuses on the rural program, which visited households and held community meetings to promote WASH behaviors to the mothers of young children. UNICEF and DPHE partnered with nongovernmental organizations that recruited, trained, and supervised hygiene promoters from the SHEWA-B communities. Hygiene promoters received 27 days of training between 2007 and 2012. They were responsible for delivering 6 key messages to SHEWA-B participants:

1. wash both hands with water and soap before eating or handling food,
2. wash both hands with water and soap or ash after defecation,
3. wash both hands with water and soap or ash after cleaning a baby's bottom,
4. safely collect and store drinking water,
5. dispose of children's feces in hygienic latrines, and
6. use hygienic latrine (all family members, including children).

Other messages included

1. clean and maintain latrines,
2. construct a new latrine if the existing one is full and fill latrine pits with soil or ash,
3. draw drinking water from an arsenic-safe water point,
4. wash raw fruits and vegetables with safe water before eating and cover food properly, and
5. manage menstruation period safely.

There was little migration in the target population during the program, and target beneficiaries were typically available to participate in program activities. At the time of this study, each hygiene promoter was instructed to visit 1200 to 1500 households every 3 months. Hygiene promoters were supposed to receive an incentive of 4200 Bangladeshi taka per month (approximately US $1.80 per day), which is roughly half the daily wage of an unskilled laborer in Bangladesh.

In early 2010, following the interim assessment of SHEWA-B in 2009,[18] UNICEF attempted to address delays in the disbursement of hygiene promoter funds, which may have hampered field implementation. UNICEF also partnered with the London School of Hygiene and Tropical Medicine to explore why interim improvements in WASH behaviors were only minimal. Following this, UNICEF developed a mass media campaign to relay more focused behavior change messages and attempted to strengthen the quality of communication between hygiene promoters and beneficiaries. We conducted this assessment during the period when these changes were implemented. Further details about the SHEWA-B program are provided elsewhere.[18]

Data Collection

To measure outcomes and implementation quality, we conducted a cross-sectional survey from June 2011 to April 2012 in a sample of clusters targeted by SHEWA-B. The survey included spot checks and demonstrations of health behaviors that UNICEF aimed to improve. We provide details about the sample size and sampling methods in the online Supplement. Field staff asked respondents about the nature and frequency of hygiene promoter home visits. Staff interviewed hygiene promoters in the sampled clusters about their job duties and assessed their recall of key messages.

Measurement of Implementation Quality

Because implementation quality was not defined a priori, we created an index with guidance from UNICEF about ideal program implementation. We constructed the index using variables from the survey of households targeted by SHEWA-B and the survey of hygiene promoters. To minimize reporting bias, we only included hygiene promoter survey items that were related to their knowledge of key SHEWA-B messages (e.g., we excluded items such as their reported frequency of visiting SHEWA-B households). We employed the Delphi method[19] to gather structured, qualitative feedback from 12 UNICEF staff that worked on SHEWA-B. Each participant independently assigned points to variables related to hygiene promoter performance measured in the cross-sectional and hygiene promoter surveys (1 = weak measure of implementation quality, 5 = strong measure of implementation quality). Following the first round, researchers calculated the mean points per item, reported the mean points per item to each participant, and asked participants if they wanted to change their initial point allocations. At each step, participants gave qualitative feedback about the items. Researchers solicited suggestions for additional items in both rounds, and participants assigned points to these items. Then, for each item considered relevant by UNICEF, we created weights equal to the average number of points per item in the second round. We scaled the index to a range of 0 to 100. A value of zero indicates that the respondent reported that they never met a SHEWA-B hygiene promoter and did not hear about or attend a community meeting, and that the hygiene promoter in their community could not recall any of the key messages. A value of 100 indicates that the hygiene promoter visited the household in the last month and the respondent heard of or attended a community event in the last month, knew the hygiene promoter's name, and reported that the hygiene promoter demonstrated key messages in the last year and that the hygiene promoter recalled all key messages.

Outcomes

The outcomes were as follows:

1. caregiver's correct demonstration of handwashing (she used soap, water, both hands);
2. presence of a dedicated handwashing location, with water and soap, within 10 feet of the place of defecation (or if soap was not present, the respondent could retrieve soap within 60 seconds);
3. clean child and caregiver hands (palms, finger pads, and fingernails were free of visible dirt);

4. availability of a private, improved latrine according to the Joint Monitoring Program (UNICEF–WHO) definition[20];
5. no feces on the latrine slab or floor;
6. a hygienic drinking water collection point (the platform at the water collection point was not broken or waterlogged, and there were no feces or garbage around it); and
7. a cover on drinking water containers.

All outcomes were measured through observation by field staff.

Potential Confounders

We prespecified potential confounders as variables that could affect implementation quality and outcomes directly or through intermediates. Because the program was delivered by the same hygiene promoter in each cluster, we defined the following confounders at the cluster level rather than the individual level.[21]

Regional poverty level. Because nongovernmental organizations recruited and trained hygiene promoters, it is possible that in poorer areas, the education level of hygiene promoters was lower or the training they received was of poorer quality. Lower education of respondents in poorer areas may also have affected the extent of behavior change. We used data from the 2000 Bangladesh Household Income and Expenditure Survey to control for subdistrict poverty in our models.[22]

Season of data collection. In the rainy season, hygiene promoters may have had more trouble traveling to assigned villages. We defined cool season as September through February, hot season as March through May, and rainy season as June through August.

Geographic features. Hygiene promoters may have had trouble traveling to clusters in flood-prone areas, particularly during the cool season, when paths become muddy. Households in drought-prone areas may have had limited access to water mainly in the dry season, which may have affected their handwashing and drinking water storage behaviors. We divided study areas into 3 geographic types: regular, flood-prone, and drought-prone areas. We explored effect modification by each of these variables.

Statistical Analyses

Unadjusted analysis. We calculated the percentage of each outcome among household respondents and calculated 95% confidence intervals using robust standard errors to account for clustering at the cluster level. We explored crude associations between each outcome and the implementation quality index using scatter plots and locally weighted scatter plot smoothing (LOWESS) plots with normal-based point-wise 95% confidence bands.[23]

Counterfactual analysis. We conducted a counterfactual analysis to assess whether outcomes would have differed if all clusters had received a low, moderate, or high level of implementation quality. Although in practice it is unlikely that all clusters would receive the same level of implementation quality, using this approach allowed us to (1) compare outcomes over the range of implementation quality and (2) estimate the maximum effect of implementation quality if all clusters had received high- versus low-quality implementation. We conducted counterfactual analyses using the cluster-level mean of outcomes, implementation quality index, and potential confounders since SHEWA-B was effectively a cluster-level intervention: approximately 1 hygiene promoter was responsible for each cluster, and hygiene promoters not only visited individual households but also held communitywide meetings.

First, we used generalized linear models with a Gaussian family and identity link function to estimate associations between each outcome (Y) and implementation quality (A), controlling for potential confounders (W) ($\mathrm{E}[Y \mid A, W]$). We considered linear models to be appropriate given that our scatter plots indicated linear patterns between implementation quality and each outcome. To allow for nonlinear relationships, we also estimated $\mathrm{E}[Y \mid A, W]$ using a data-adaptive, ensemble machine learning algorithm that included the following learners: the simple mean, generalized linear models, Bayesian generalized linear models, lasso and elastic-net regularized generalized linear models, and generalized additive models.[24] Results were similar using both estimation methods, so we only present the results from the generalized linear models.

To compare outcomes over the range of implementation quality, we created counterfactual data sets in which all clusters were assigned an implementation quality level ($A = a$) ranging from the 10th to 90th percentile of the observed distribution in increments of 10. We then reestimated mean outcomes using model coefficients and averaging over values of confounders ($\mathrm{E}_W[Y \mid A = a, W]$). To estimate the maximum effect of implementation quality, we estimated the difference in the probability of each outcome at the 90th (index = 58) and 10th percentiles (index = 3) of implementation quality ($\mathrm{E}_W[Y \mid A = 58, W] - \mathrm{E}_W[Y \mid A = 3, W]$). We chose to use percentiles of the observed distribution rather than the theoretical minimum and maximum (index = 0 and index = 100) to avoid extrapolating beyond the information in the observed data. To detect possible residual confounding of the association between implementation quality and outcomes, we repeated the analysis using a negative control outcome[25,26]: the reported number of neonatal deaths in the respondent's household in the past 5 years. We used a nonparametric bootstrap with 1000 replicates to obtain percentile-based 95% confidence intervals for each scenario. We excluded households with missing data from analyses and thus assumed these data were missing completely at random (Table A, available as a supplement to the online version of this article at http://www.ajph.org).

RESULTS

To reach the planned sample of 32 480 households in 1160 clusters, field staff invited 33 134 households in 1182 clusters to participate; 33 027 households consented to participate (response rate = 99.7%; Table B, available as a supplement to the online version of this article at http://www.ajph.org). The majority (96%) of respondents were the mothers of the youngest child in the household. We reached 1110 of 1164 hygiene promoters (95.4%) in 1126 clusters. After merging cross-sectional survey data with hygiene promoter survey data, outcome data from 31 521 households in 1182 clusters and implementation quality data from 1126 clusters were available for analysis. In calculating the index, we excluded data from 56 clusters because of missing hygiene promoter survey data. Forty-three

percent of households sampled were in subdistricts in which 37% to 55% of residents were estimated to live below the poverty line (Table C, available as a supplement to the online version of this article at http://www.ajph.org).

Implementation Quality

Only 31% of households reported meeting a hygiene promoter in the last 4 months (n = 8328 households; Table 1). Fifty-three percent of households (n = 16 670 households) did not recall ever meeting a SHEWA-B hygiene promoter (Table 1). The range of the implementation quality index was 0 to 90, with a mean of 28 and SD of 21 (n = 1126 clusters; Figure B, available as a supplement to the online version of this article at http://www.ajph.org). Implementation quality was significantly higher in areas surveyed during the cool and hot season than in the rainy season (Table D, available as a supplement to the online version of this article at http://www.ajph.org). There was no statistically significant difference in mean implementation quality between areas prone to drought, flooding, or neither.

Hygiene Promoter Survey

The most commonly reported problems hygiene promoters reported were that beneficiaries did not have time to listen during community meetings (n = 495 hygiene promoters; 45%) and household visits (n = 376; 34%), and that beneficiaries did not have enough money to buy soap (n = 375; 34%). The majority reported that they met with their supervisor at least weekly (n = 814; 73%) and that supervision was sufficient (n = 908; 82%). Most were satisfied with the content (n = 934; 84%) and duration (n = 709; 64%) of their training, but reported that their stipend was insufficient (n = 961; 87%) and was not paid on time (n = 721; 65%). When asked to recall the 6 key messages of SHEWA-B listed in Methods, on average, hygiene promoters recalled 2.9 out of 5 messages about safe water storage, 4.5 out of 6 about handwashing, and 4.5 out of 9 about latrine usage. Fewer than half (47%) of the hygiene promoters surveyed could recall the main key messages of SHEWA-B (n = 527).

TABLE 1—Inputs Into the Implementation Quality Index: SHEWA-B Program, Bangladesh, 2011–2012

Input	No. of Households	% (95% CI)[a]
Hygiene promoter visited household at least once	14 622	47 (45, 48)
Hygiene promoter visited household in the last month	5 471	19 (17, 20)
Hygiene promoter visited household in the last 4 mo	8 328	31 (29, 32)
Respondent ever heard of or attended a community event	7 892	26 (24, 27)
Respondent heard of or attended a community event in the last month	3 341	11 (10, 12)
Respondent heard of or attended a community event in the last quarter	5 821	19 (18, 20)
Respondent knew a SHEWA-B hygiene promoter by name	8 663	28 (26, 29)
Respondent recalled that a hygiene promoter gave safe water messages in the last year	8 076	26 (24, 27)
Respondent recalled that a hygiene promoter gave handwashing messages in the last year	9 120	29 (28, 31)
Respondent recalled that a hygiene promoter gave sanitation messages in the last year	9 128	29 (28, 31)
Respondent recalled that a hygiene promoter gave at least 3 messages in the last year	7 584	24 (23, 26)
Hygiene promoter could recall all key SHEWA-B messages[b]	527	47 (44, 50)

Note. CI = confidence interval; SHEWA-B = Sanitation Hygiene Education and Water Supply in Bangladesh.

[a]Standard errors were adjusted for clustering at the cluster level.

[b]Data used to create this variable are from the hygiene promoter survey. If a cluster was covered by multiple hygiene promoters, the values were averaged so that the number is the total number of clusters. The number presented in the percentage column is the mean across all clusters.

Outcomes Compared With UNICEF Targets

Performance was close to UNICEF targets for (1) having no feces on the latrine slab or floor, (2) presence of a dedicated handwashing location, (3) no open defecation, and (4) covering drinking water containers (Table 2). Performance was substantially below target for access to a private, improved latrine (observed: 23%; target: 75%) and for having a hygienic drinking water point (observed: 28%; target: 82%).

Program Outcomes and Implementation Quality

Unadjusted analysis. The frequency of hygiene promoter visits was not associated with improved health behaviors or access to hygiene or sanitation infrastructure. The percentage of respondents performing health behaviors or who had access to hygiene or sanitation infrastructure was never more than 6% higher among those who saw a hygiene promoter in the last month than among those who had met a promoter only once (Tables E and F, available as a supplement to the online version of this article at http://www.ajph.org). In some cases, the percentage was lower among those who had a met promoter more recently. In scatter and LOWESS plots, the probability of each outcome had a monotonic pattern across the range of implementation quality, and the slope was relatively flat (Figure C, available as a supplement to the online version of this article at http://www.ajph.org).

Counterfactual analysis. In the counterfactual scenarios comparing outcomes over the observed range of implementation quality, we found that the probability of each outcome was not substantially higher at higher levels of implementation quality

TABLE 2—Outcomes and UNICEF End-of-Program Targets: SHEWA-B Program, Bangladesh, 2011–2012

Outcome	No. of Households Observed (Total No.[a])	% (95% CI)[b]	UNICEF End-of-ProgramTarget, %
Private, improved latrine available[c]	6 725 (29 586)	23 (22, 24)	75
No observed feces on latrine slab or floor	14 284 (28 360)	50 (49, 51)	59
Observed dedicated handwashing location[d]	17 988 (31 431)	57 (56, 59)	55
No open defecation reported by respondent	29 523 (31 449)	94 (93, 94)	97
Correct caregiver handwashing demonstration[e]	18 096 (29 894)	61 (59, 62)	
Caregiver hands observed to be clean[f]	13 587 (31 291)	43 (42, 44)	
Child hands observed to be clean[f,g]	8 455 (29 340)	29 (28, 30)	
Observed hygienic drinking water collection point[h]	8 801 (31 312)	28 (27, 29)	82
Drinking water container observed to be covered	5 305 (12 253)	43 (42, 45)	45

Note. CI = confidence interval; SHEWA-B = Sanitation Hygiene Education and Water Supply in Bangladesh.

[a]Although the index is calculated at the cluster level, outcomes are calculated at the household level, so the total number is shown at the household level. A total of 1126 clusters had nonmissing implementation quality index values.

[b]Standard errors were adjusted for clustering at the cluster level.

[c]Improved latrine was defined according to the Joint Monitoring Program (UNICEF–WHO) definition. Field staff observed whether the latrine was improved or not, and private access to the latrine was determined through respondent self-report.

[d]Handwashing location with water and soap within 10 feet of the place of defecation, or if soap was not present, the respondent could retrieve soap within 60 seconds.

[e]Demonstration was considered correct if the respondent used soap, water, and both hands.

[f]No visible presence of dirt on nails, palms, or finger pads.

[g]The UNICEF log frame corresponding to these items was somewhat general, so these variables may estimate something slightly different than what UNICEF intended.

[h]Environmental sanitation is considered maintained if the water point's platform is not broken and not waterlogged and has no garbage, dirt, or feces around it. UNICEF's log frame indicator was defined as demonstrating skills to operate and maintain water points.

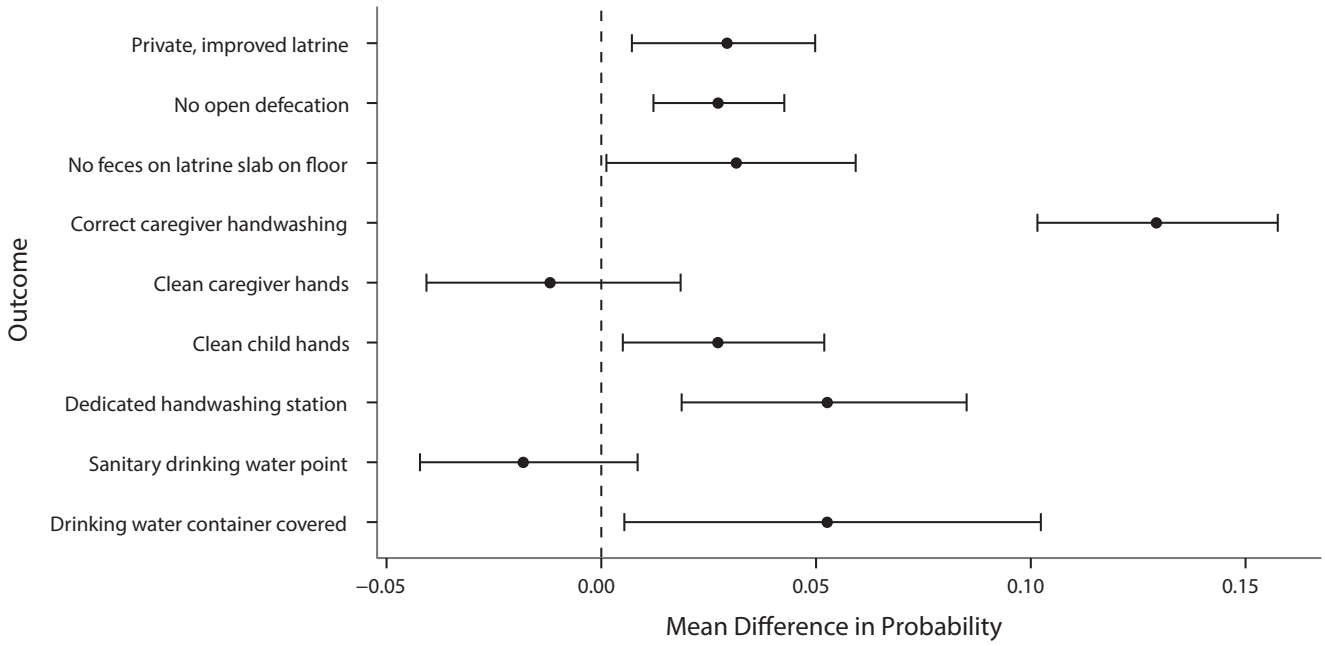
(Figure D, available as a supplement to the online version of this article at http://www.ajph.org). There was evidence of effect modification by subdistrict poverty level, season of data collection, and geographic features for certain outcomes (Figures E–G, available as a supplement to the online version of this article at http://www.ajph.org). For example, the association between having an improved latrine and implementation quality was stronger in areas with lower poverty levels. The association between implementation quality and having a dedicated handwashing station was stronger in areas that were not drought-prone.

In our counterfactual analysis of the maximum possible effect of implementation quality (Figure 1), the difference in the probability of each outcome was no more than 1% to 3%, with the exception of 3 outcomes: in the high- versus low-quality implementation scenarios, the probability (1) that caregivers correctly demonstrated handwashing was 13% higher (95% confidence interval [CI] = 11%, 16%), (2) that they had a dedicated handwashing location was 5% higher (95% CI = 2%, 9%), and (3) that they stored their drinking water in a covered container was 5% higher (95% CI = 1%, 10%). In our negative control analysis using the reported number of neonatal deaths in the respondent's household in the past 5 years as the outcome, we found no association (estimate = –0.003; 95% CI = –0.056, 0.044).

DISCUSSION

Our assessment of SHEWA-B is among the largest assessments (n = 33 027 households) ever conducted of a scaled-up public health program. The implementation of SHEWA-B did not meet UNICEF's ideal in any area. Approximately half of respondents did not recall ever meeting a SHEWA-B hygiene promoter, and thus they effectively were not offered the program. Health behaviors and access to hygiene and sanitation infrastructure were similar whether or not participants had met a SHEWA-B promoter. Our counterfactual analysis suggested that the difference in outcomes was very small among those who received the highest versus the lowest implementation quality. Evidence of an association between program implementation and health behaviors or access to hygiene or sanitation infrastructure would have justified estimation of other more realistic parameters, such as population intervention effects.[27–29]

Our results suggest that a redesign of some elements of the SHEWA-B program were needed to achieve desired targets of the remaining program outcomes at the time of this study. Several outcomes, such as the open defecation rate, nearly met UNICEF's target levels; however, considering the lack of association between implementation quality and health behaviors and access to hygiene or sanitation infrastructure, these outcomes likely resulted from factors other than SHEWA-B, such as secular trends. Our findings complement those of a separate evaluation of SHEWA-B, which found



Note. SHEWA-B = Sanitation Hygiene Education and Water Supply in Bangladesh. Whiskers indicate 95% confidence intervals (CIs). Each point estimate is the mean difference in the probability of the outcome if all individuals received implementation quality at the 90th vs 10th percentile of implementation quality. We used generalized linear models with a Gaussian link function to estimate associations between each outcome and implementation quality, controlling for potential confounders. For all outcomes except improved latrine access, we adjusted for subdistrict poverty level, geographic features, and season of data collection. For improved latrine access, we adjusted for subdistrict poverty level. Estimates average over values of confounders. We obtained 95% CIs using a nonparametric bootstrap with 1000 replicates.

FIGURE 1—Probability of Each Outcome if All Clusters Received Implementation Quality at the 90th vs 10th Percentile of the Observed Distribution: SHEWA-B Program, Bangladesh, 2011–2012

no differences in health behaviors or infrastructure access between intervention and control groups.[30] Although they found small decreases in child diarrhea and respiratory illness, because of the lack of impact of the program on outcomes of interest, evaluators concluded that differences in illness between the 2 groups at the end of the program were likely attributable to selection bias or unmeasured confounding rather than the SHEWA-B program.

Factors Affecting Poor Implementation of SHEWA-B

There were a number of potential factors that may have affected SHEWA-B implementation. First, although hygiene promoters reported satisfaction with their training, it is possible that the training was insufficient, as is suggested by our finding that promoters' recall of key messages was low. Second, assessments of hygiene promoter supervision and performance may have been insufficient. Improved supervision and audits with feedback may improve hygiene promoter performance and increase their job satisfaction and motivation.[11] Although UNICEF and DPHE conducted performance assessments of hygiene promoters, our results suggest that assessments did not result in high-level hygiene promoter performance in most areas. Third, most hygiene promoters reported that their stipend was insufficient and not paid on time. Given promoters' modest remuneration, the number of households SHEWA-B hygiene promoters were responsible for (1200–1500 per promoter) may have been unreasonable. Others have reported that community health worker attrition in Bangladesh was associated with workers' dissatisfaction with remuneration.[31,32] The health promoter literature suggests that sufficient remuneration contributes to promoter success.[11] Finally, the program may have been more successful if SHEWA-B, instead of only promoting behaviors, had also provided hardware such as handwashing stations or improved latrines, similar to what other local organizations have done.[33]

Other Evaluations of Large-Scale WASH Programs

The results of this large assessment demonstrate the difficulty of maintaining program quality while scaling up, as has been reported by others.[7] On the whole, our findings echo those of the few, existing assessments of large-scale WASH programs. Evaluations of large-scale handwashing campaigns reported increased exposure to handwashing messages (e.g., 9% –16% increase in Peru[13] and 10% increase in Vietnam[14]). These studies found increases in certain handwashing behaviors but no impact on health. Two randomized trials of large-scale sanitation programs in India found moderate increases in sanitation coverage from baseline to follow-up (in Madhya Pradesh, change from 18% to 44% in intervention, 21% to 24% in control[16]; in Odisha, change from 9% to 63% in

intervention, 8% to 12% in control[17]). Neither study found reductions in diarrhea or enteric parasite infections. A trial of a large-scale sanitation program in Indonesia found that at follow-up, household toilet construction occurred in 16% of treatment households and 13% of control households; they reported 3.3% diarrhea prevalence in treated communities and 4.6% prevalence in control communities at follow-up ($P < .001$).[15]

Limitations

This study was subject to several limitations. First, our implementation quality index was developed through a systematic process with UNICEF staff that designed and implemented SHEWA-B, but it remains possible that it was poorly defined. Future studies of large-scale programs would benefit from a priori definition of program fidelity measures to allow for rigorous assessment.[34] Second, it is possible that recall bias, respondent bias, and measurement error occurred. We used rapid observations of hygiene practices and conditions because they are efficient and have been shown to be valid, reliable indicators for many hygiene outcomes.[35] We found no association between the index and a negative control outcome (neonatal deaths), suggesting that the association we report between implementation quality and outcomes was not likely to be a result of residual confounding. It is also possible that respondents did not recall hygiene promoter visits, which would have caused us to underestimate SHEWA-B implementation quality; however, even if respondents failed to recall visits, poor recall of the visits suggests that they were likely not to have been impactful in changing the behaviors targeted by SHEWA-B. Fourth, information about promotion activities provided in interviews with hygiene promoters was not consistent with information reported by SHEWA-B beneficiaries, which suggests considerable response bias from hygiene promoters. In addition, the 5% of hygiene promoters who did not participate in our survey may have been systematically different from those who participated, which may have influenced our findings. Finally, we were unable to explore the importance of factors such as management and financing of SHEWA-B because of the complexity of implementation. Large-scale programs have been more successful when there is strong leadership and management and realistic arrangements for financing.[7]

Public Health Implications

We found that the implementation of SHEWA-B, one of the largest public health WASH programs in history, was not optimal in any area. Even if implementation quality had been higher, our findings suggest that outcomes would not have been substantially better. Modifications to SHEWA-B's design would likely have been needed for the program to improve health behaviors and access to hygiene and sanitation infrastructure and to reduce disease. AJPH

CONTRIBUTORS

J. Benjamin-Chung, S. Sultana, A. K. Halder, M. A. Ahsan, L. Unicomb, and S. P. Luby designed the study. J. Benjamin-Chung, S. Sultana, A. K. Halder, and M. A. Ahsan led data collection. J. Benjamin-Chung wrote the first draft of the article. J. Benjamin-Chung, A. K. Halder, B. F. Arnold, A. E. Hubbard, and J. M. Colford Jr developed the statistical analysis plan. J. Benjamin-Chung and A. K. Halder conducted statistical analyses. All authors revised the article.

ACKNOWLEDGMENTS

We thank the study participants and acknowledge the contributions of Abul Kasham Shoab and Mohammed Abdur Raquib to the field coordination of this study.

HUMAN PARTICIPANT PROTECTION

Study participants provided written consent. The study was approved by the Ethical Review Committee at the International Centre for Diarrhoeal Disease Research, Bangladesh, and the institutional review board at the Centers for Disease Control and Prevention.

REFERENCES

1. *World Development Report 1993: Investing in Health*. New York, NY: Oxford University Press; 1993.

2. Bryce J, Black RE, Walker N, Bhutta ZA, Lawn JE, Steketee RW. Can the world afford to save the lives of 6 million children each year? *Lancet*. 2005;365(9478):2193–2200.

3. Jha P, Mills A, Hanson K, et al. Improving the health of the global poor. *Science*. 2002;295(5562):2036–2039.

4. Victora CG, Hanson K, Bryce J, Vaughan JP. Achieving universal coverage with health interventions. *Lancet*. 2004;364(9444):1541–1548.

5. Sachs JD. Health in the developing world: achieving the Millennium Development Goals. *Bull World Health Organ*. 2004;82(12):947–949.

6. Madon T, Hofman KJ, Kupfer L, Glass RI. Implementation science. *Science*. 2007;318(5857):1728–1729.

7. Mangham LJ, Hanson K. Scaling up in international health: what are the key issues? *Health Policy Plan*. 2010;25(2):85–96.

8. Durlak JA. Why program implementation is important. *J Prev Interv Community*. 1998;17(2):5–18.

9. Durlak J, DuPre E. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *Am J Community Psychol*. 2008;41(3–4):327–350.

10. Hanson K, Ranson MK, Oliveira-Cruz V, Mills A. Expanding access to priority health interventions: a framework for understanding the constraints to scaling-up. *J Int Dev*. 2003;15(1):1–14.

11. Rowe AK, de Savigny D, Lanata CF, Victora CG. How can we achieve and maintain high-quality performance of health workers in low-resource settings? *Lancet*. 2005;366(9490):1026–1035.

12. Yamey G. Scaling up global health interventions: a proposed framework for success. *PLoS Med*. 2011;8(6):e1001049.

13. Galiani S, Gertler P, Orsola-Vidal A. Promoting handwashing behavior in Peru: the effect of large-scale mass-media and community level interventions. Social Science Research Network, Rochester, NY, 2012. Available at: http://papers.ssrn.com/abstract=2170640. Accessed April 29, 2013.

14. Chase C, Do QT. Handwashing behavior change at scale: evidence from a randomized evaluation in Vietnam. 2012. Available at: http://papers.ssrn.com/abstract=2151769. Accessed April 29, 2013.

15. Cameron LA, Shah M, Olivia S. Impact evaluation of a large-scale rural sanitation project in Indonesia. The World Bank, 2013. Available at: http://econpapers.repec.org/paper/wbkwbrwps/6360.htm. Accessed April 29, 2013.

16. Patil SR, Arnold BF, Salvatore AL, et al. The effect of India's total sanitation campaign on defecation behaviors and child health in rural Madhya Pradesh: a cluster randomized controlled trial. *PLoS Med*. 2014;11(8):e1001709.

17. Clasen T, Boisson S, Routray P, et al. Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: a cluster-randomised trial. *Lancet Glob Health*. 2014;2(11):e645–e653.

18. Huda TMN, Unicomb L, Johnston RB, Halder AK, Yushuf Sharker MA, Luby SP. Interim evaluation of a large scale sanitation, hygiene and water improvement programme on childhood diarrhea and respiratory disease in rural Bangladesh. *Soc Sci Med*. 2012;75(4):604–611.

19. Linstone HA, Turoff M. *Delphi Method: Techniques and Applications*. Boston, MA: Addison-Wesley Publishing; 1975.

20. World Health Organization. Core questions on drinking-water and sanitation for household surveys. Published 2006. Available at: http://www.who.int/water_sanitation_health/monitoring/household_surveys/en. Accessed December 23, 2013.

21. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37–48.

22. Bangladesh Bureau of Statistics. Household income & expenditure survey. Published 2000. Available at: http://catalog.ihsn.org/index.php/catalog/2257. Accessed February 13, 2017.

23. Hastie T, Tibshirani R, Friedman JH. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York, NY: Springer; 2009.

24. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol*. 2007;6(1):1–21.

25. Lipsitch M, Tchetgen ET, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21(3):383–388.

26. Arnold B, Ercumen A, Benjamin-Chung J, Colford J. Negative controls to detect selection bias and measurement bias in epidemiologic studies. *Epidemiology*. 2016; 27(5):637–641.

27. Hubbard AE, van der Laan MJ. Population intervention models in causal inference. *Biometrika*. 2008; 95(1):35–47.

28. Westreich D. From exposures to population interventions: pregnancy and response to HIV therapy. *Am J Epidemiol*. 2014;179(7):797–806.

29. Ahern J, Colson KE, Margerson-Zilko C, Hubbard A, Galea S. Predicting the population health impacts of community interventions: the case of alcohol outlets and binge drinking. *Am J Public Health*. 2016;106(11): 1938–1943.

30. UNICEF. SHEWA-B Programme health impact study report. 2014. Available at: http://www.unicef.org/evaldatabase/files/WASH_SHEWA-Bangladesh_health_impact_eval.pdf. Accessed February 13, 2017.

31. Khan SH, Chowdhury AM, Karim F, Barua MK. Training and retaining Shasthyo Shebika: reasons for turnover of community health workers in Bangladesh. *Health Care Superv*. 1998;17(1):37–47.

32. Rahman SM, Ali NA, Jennings L, et al. Factors affecting recruitment and retention of community health workers in a newborn care intervention in Bangladesh. *Hum Resour Health*. 2010;8(1):12.

33. Dreibelbis R, Winch PJ, Leontsini E, et al. The Integrated Behavioural Model for Water, Sanitation, and Hygiene: a systematic review of behavioural models and a framework for designing and evaluating behaviour change interventions in infrastructure-restricted settings. *BMC Public Health*. 2013;13(1):1015.

34. Campbell M. Framework for design and evaluation of complex interventions to improve health. *BMJ*. 2000; 321(7262):694–696.

35. Ram P. Practical guidance for measuring handwashing behavior: 2013 update. World Bank Water and Sanitation Program, 2013. Available at: https://www.wsp.org/sites/wsp.org/files/publications/WSP-Practical-Guidance-Measuring-Handwashing-Behavior-2013-Update.pdf. Accessed February 13, 2017.