Is Structured Observation a Valid Technique to Measure Handwashing Behavior? Use of Acceleration Sensors Embedded in Soap to Assess Reactivity to Structured Observation

Pavani K. Ram,* Amal K. Halder, Stewart P. Granger, Therese Jones, Peter Hall, David Hitchcock, Richard Wright, Benjamin Nygren, M. Sirajul Islam, John W. Molyneaux, and Stephen P. Luby

University at Buffalo, Buffalo, New York; ICDDR,B: International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh; Unilever R&D, Bebington, Wirral, United Kingdom; 4-Front Research, Capenhurst, Cheshire, United Kingdom; Emory University, Atlanta, Georgia; Water and Sanitation Program, The World Bank, Washington, DC

Abstract. Structured observation is often used to evaluate handwashing behavior. We assessed reactivity to structured observation in rural Bangladesh by distributing soap containing acceleration sensors and performing structured observation 4 days later. Sensors recorded the number of times soap was moved. In 45 participating households, the median number of sensor soap movements during the 5-hour time block on pre-observation days was 3.7 (range 0.3–10.6). During the structured observation, the median number of sensor soap movements was 5.0 (range 0–18.0), a 35% increase, $P = 0.0004$. Compared with the same 5-hour time block on pre-observation days, the number of sensor soap movements increased during structured observation by $\geq 20\%$ in 62% of households, and by $\geq 100\%$ in 22% of households. The increase in sensor soap movements during structured observation, compared with pre-observation days, indicates substantial reactivity to the presence of the observer. These findings call into question the validity of structured observation for measurement of handwashing behavior.

INTRODUCTION

Handwashing with soap reduces the risk of diarrhea and pneumonia, the two leading causes of death among children < 5 years of age in low-income countries.^{1–4} Increasingly, small- and large-scale projects are being undertaken to improve rates of handwashing with soap among people living in high childhood mortality settings. Although proving the efficacy of handwashing with soap is typically achieved by examining the impact of the handwashing promotion on health (e.g., prevalence of diarrhea), most public health programs undertaking handwashing promotion do not have the resources to examine health impact as an outcome of their projects.⁵ Rather, they typically rely on measuring process indicators or the program's impact on handwashing behavior. Self-report is a commonly used method for recording information about behavior in program evaluations and research studies on handwashing promotion. However, studies have consistently showed that self-report overestimates handwashing behavior, when compared with structured observation.^{6–8} Structured observation consists of direct observation of the subject's behavior over a protracted period of time, with duration of observation typically lasting 3–7 hours.^{8–11} In contrast to proxy measures of handwashing, such as observation of the presence of soap in the home, structured observation yields rich behavioral detail, including when, how, and with what material hands are washed, and with contextual information about handwashing at critical times, such as after defecation. But structured observations place a stranger in the household for the duration of the observation period. Because this introduces the risk of a Hawthorne effect,¹² we are compelled to ask how the observer's presence influences handwashing behavior and, consequently, how the results of structured observation should be interpreted.

For this study, we used a novel technology to assess subjects' reactivity to structured observation. Acceleration sensors embedded in ordinary-appearing bars of soap capture

movement in the three dimensions, thereby tracking soap movement (Figure 1). Bars of soap containing acceleration sensors, hereafter referred to as “sensor soap” can be placed in homes for several days, enabling quantification of household soap movement over several days without having a human observer present. Previously unpublished data on validation of the sensor soaps are presented in the Appendix. As part of a larger study aimed at exploring currently used methods of measuring handwashing behavior, we used this technology in Bangladesh to examine whether and the extent to which subjects are reactive to the presence of a human observer during structured observation of handwashing behavior.

MATERIALS AND METHODS

We conducted this study in six rural villages, about 3 hours outside of Dhaka, Bangladesh's capital. There, households were typically located in clusters and shared a common courtyard. We enumerated all of the clusters in each of the six villages, using a school in the village as the central starting point. We went to each even-numbered cluster determined to have at least one child < 2 years of age. If there was only one child < 2 years of age in the cluster, we sought to enroll that child's primary caregiver. If there were multiple children < 2 years of age in the cluster, we selected one of the caregivers. First, the fieldworker identified all primary caregivers of at least one child < 2 years of age in the cluster; the fieldworker asked these primary caregivers to stand in front of their household. The fieldworker stood at the entrance of the cluster and, starting from her right and working in a circular fashion, she counted the primary caregivers from 1 to 5. The primary caregiver that was counted as “5” was requested to participate in the study. In each village, we visited as many clusters as needed to recruit 8–9 primary caregivers per village.

The key respondent for the study was the primary caregiver of a child < 2 years of age. The informed consent document was read aloud to all potential respondents because literacy among women of childbearing age in Bangladesh remains low. During the informed consent process, we described the structured observation process, indicating that the fieldworker would be observing routine practices related to water,

*Address correspondence to Pavani K. Ram, University at Buffalo, Rm 270 Farber Hall, 3435 Main Street, Buffalo, NY 14221. E-mail: pkram@buffalo.edu

FIGURE 1. Bar of soap with acceleration sensor. This figure appears in color at www.ajtmh.org.

sanitation, and hygiene. In addition, we described the purpose and nature of the sensor soap. The sensor soap was of identical composition and color to commercially available Lifebuoy soap (Mumbai, India); individuals who had previously experienced skin reactions from the commercially available soap were cautioned to refrain from using the sensor soap. In addition, because all of the participating households had at least one child < 2 years of age and since the acceleration sensor contains a 3-volt battery similar to that found in many wristwatches and calculators, we advised that sensor soap should be kept away from children who might place such items in their mouths.

After obtaining written informed consent, we administered a questionnaire and recorded observations of handwashing stations, presence of soap in the home, and information relevant to socioeconomic status. We also asked respondents to demonstrate their usual handwashing behavior. Households were randomized to either a brief duration of sensor soap deployment (4 days) or a prolonged duration of sensor soap deployment (9 days). There were 25 households each in the brief deployment and the prolonged deployment groups. Following questionnaire and other data collection, the fieldworker collected the bar of soap in use at the time and provided sensor soap to the household. The fieldworker provided two sensor soap bars if the household indicated that two bars of soap were needed for the household's use during a 3 or 4 day period. If no bars were in use, she provided one sensor soap bar. The day that sensor soap bars were placed was counted as Day 0 (Figure 2). Recording of soap movement began at 12:01 AM on Day 1. In brief deployment households, the sensor soap bars recorded soap movement on Days 1–4; Day 4 was the structured observation day. In prolonged deployment households, the first set of sensor soap bars recorded soap movement on Days 1–4 and the second set of sensor soap bars recorded soap movement on Days 6–9; Day 9 was the structured observation day. Prolonged deployment households had a longer data collection time because of a secondary study objective not relevant to this work. Soap movement was recorded only until the

completion of the structured observation day on Day 4 in the brief deployment households and Day 9 in prolonged deployment households.

During the structured observation, the fieldworker seated herself so that she could keep the primary caregiver in her line of sight at all times. She recorded opportunities for handwashing, such as preparing food, and whether and how the caregiver washed her hands. Although her focus was on the primary caregiver of the child < 2 years of age, the fieldworker also recorded handwashing opportunities for other household members, as they occurred in her line of sight. Of the 50 households included in the study, 25 were randomly selected for hand microbiology testing during the course of structured observation. We collected a hand rinse upon arrival of the observer and then two more times when the observer identified a handwashing opportunity; results of hand rinse sampling are reported separately. On completion of the structured observation, the interviewer retrieved the sensor soap and terminated its recording. Here, we describe the methods of structured observation and hand rinse collection to inform the reader about the nature of the observer's interaction with study participants. However, structured observation and hand microbiology data were not used to address research questions of interest in this work.

All household-level data were collected by female fieldworkers after obtaining written informed consent from participants. This study was approved by the human subjects review committees of ICDDR,B: International Center for Diarrheal Disease Research, Bangladesh (Dhaka, Bangladesh) and the University at Buffalo (Buffalo, NY).

Data analysis. The acceleration sensors use accelerometer technology and measure movement in three dimensions: if the sensor soap is still or moving slowly, its orientation with respect to gravity can be inferred. Algorithms developed by Unilever decoded whether the sensor soap had changed orientation in any of the three dimensions. We defined a lack of movement as when the soap was still, when acceleration of the soap did not involve two or more dimensions, or when acceleration took place for 1 second or less. Movements were classified as genuine when acceleration occurred for more than 1 second in two or three dimensions. For those households that were given two sensor soap bars, the number of soap movements reflects the combined total recorded by both soaps.

The sensor soap replaced the soap that had been in use in the household. We did not ask the household to restrict the use of the sensor soap to just the primary caregiver of the child < 2 years of age, because that would dramatically alter the typical soap use behavior in the household. Therefore, the number of sensor soap movements recorded by the sensor soap reflects household-level soap use and not individual soap use.

We wondered whether the novelty of the sensor soap might itself result in a greater use of soap than would typically take place in the household. We hypothesized that, if there were a novelty effect, a greater number of sensor soap movements would be detected on Day 1 compared with subsequent days. To detect whether there was a novelty effect, we compared the number of sensor soap movements recorded during the 24 hours of Day 1 to the mean number of sensor soap movements recorded during subsequent days. Because sensor soap movement data were not normally distributed, we used the Wilcoxon signed-rank test, a non-parametric test for paired data, to assess the statistical significance of the difference

Day	0	1	2	3	4	5	6	7	8	9
Brief sensor soap deployment households (N=25)	Informed consent Questionnaire Household observations Provision of sensor soap	Full day*	Full day	Full day	Structured observation Retrieval of sensor soap	--	--	--	--	--
Prolonged sensor soap deployment households (N=25)	Informed consent Questionnaire Household observations Provision of sensor soap	Full day	Full day	Full day	Full day	Replacement of sensor soap No soap movement tracking	Full day	Full day	Full day	Structured observation Retrieval of sensor soap
* "Full day" implies that 24 hours of soap use data was collected by the sensor soap										

FIGURE 2. Timeline of data collection in study households, Bangladesh, 2007.

in sensor soap movements between Day 1 and subsequent days.

The 5-hour period during which the structured observation was conducted was labeled the *observation time block*. Because the sensor soap records the time of each movement, we were able to quantify the number of soap movements occurring during the observation time block on each of the pre-observation days: Days 1, 2, and 3 for brief deployment households and Days 6–8 for prolonged deployment households. To address our principal study question, to assess whether there was reactivity to the presence of the observer during the structured observation, we compared the number of sensor soap movements during the observation time block on the pre-observation days to the number of sensor soap movements recorded on the structured observation day. We used paired *t* tests to compare sensor soap movements during the observation time block on pre-observation days to sensor soap movements during the observation time block on the structured observation day if the data were normally distributed. If the data were non-normally distributed, we used the Wilcoxon signed-rank test.

We calculated the difference between the number of sensor soap movements recorded during the observation time block on the structured observation day and the mean number of sensor soap movements during the observation time block on the pre-observation days. We then stratified households into quartiles of difference. After reviewing the data, we classified those households in the highest quartile of difference as “reactors” and those in the lower three quartiles as non-reactors. We conducted bivariate analysis to examine the relationship between potential explanatory factors such as education level and ownership of household assets, and reactor status.

Budgetary and logistical constraints limited us to enrolling a total of 50 primary caregivers into the study.

RESULTS

All 50 respondents were the mothers of children < 2 years of age. The mean age of respondents was 26 years (range 16–39 years) and the mean age of the youngest child in the household was 0.8 years (range < 1 month to 1.9 years). The mean number of persons in the household was 5.6 (range 3–12) and the mean number of children < 5 years of age in the household was 1.5 (range 1–8). A total of 56% of respondents reported that at least one member of their household owned a mobile phone.

In total, 82 bars of soap containing acceleration sensors were distributed to brief deployment and prolonged deploy-

ment households (Table 1). All were retrieved at the end of the deployment period. For brief deployment households, soap movement data were recovered for 21 households; 4 sensors failed to yield movement data. For prolonged deployment households, 19 of 28 of the first set of sensors yielded movement data and 28 of 29 of the second set of sensors yielded data; sensors for prolonged deployment households were deployed by the field team without onsite technical expertise from Unilever.

We had data from 21 brief deployment households and 18 prolonged deployment households for our assessment of a novelty effect from the introduction of the sensor soap. Among brief deployment households, the median number of soap movements on Day 1 was 9.0 (SD 3.1, range 5–16) and on Days 2–3 was 7.5 (SD 3.6, range 2–18) ($P = 0.17$). Among prolonged deployment households, the median number of soap movements on Day 1 was 10 (SD 5.5, range 6–22) and on Days 2–4 and 6–8 was 8.5 (SD 5.7, range 5–23) ($P = 0.28$).

To ensure that there was no inherent reason why soap movement should have been different on the observation day compared with pre-observation days (e.g., religious observances),

TABLE 1
Placement, retrieval, and data yield of sensor soaps in households or primary caregivers of children < 2 years of age, Bangladesh, 2007

Placement and retrieval of sensor soaps	Brief sensor soap deployment households (N = 25)	Prolonged sensor soap deployment households (Week 1) (N = 25)	Prolonged sensor soap deployment households (Week 2) (N = 25)
One soap bar already in place in the home	22	19	–
Two soap bars already in place in the home	0	2	–
No soap bars in place in the home	3	4	–
One sensor soap bar provided	25	22	21
Two sensor soap bars provided	0	3	4
Total number of sensor soap bars placed	25	28	29
Number of sensor soap bars retrieved	25	28	29
Number of sensor soap bars that yielded data	21	19	28

we examined whether there was any difference in the number of soap movements during the 5 hours *preceding* structured observation. Thus, if the structured observation began at 9 AM, we compared the number of soap movements detected between 4 AM and 9 AM on pre-observation days to those detected on the observation day. Data for this analysis were available for 21 brief deployment households and 24 prolonged deployment households. During the pre-observation days, the median number of soap movements during the 5 hours preceding structured observation was 1.7 (range 0–8). On the day of the structured observation, the median number of soap movements was 2.0 (range 0–14) ($P = 0.79$).

To address our primary study question, we examined whether there was reactivity to the presence of the observer during the observation time block. Data for this analysis were available for 21 brief deployment households and 24 prolonged deployment households. The median number of soap movements during the 5-hour time block on pre-observation days was 3.7 (range 0.3–10.6). During the structured observation, the median number of soap movements was 5.0 (range 0–18.0), a 35% increase ($P = 0.0004$). When we restricted this analysis to those households that had one bar of soap before the study and one sensor soap provided to them, we had similar findings: the median number of soap movements on pre-observation days was 3.7 (0.3–10.6) and the median number of soap movements on the structured observation day was 5.0 (1–18), a 35% increase ($P < 0.0001$). Overall, compared with pre-observation days, 28 (62%) households increased the number of soap movements during the observation time block on the structured observation day by $\geq 20\%$. Ten (22%) households increased the number of soap movements by $\geq 100\%$ (i.e., at least doubling their movement of soap when the observer was present, compared with days when the observer was not present).

To describe households that were particularly reactive to the presence of the observer, we divided households with

soap movement data into quartiles of reactivity. There were 12 reactors (highest quartile of difference in soap movements between pre-observation and observation days) and 33 non-reactors (lowest 3 quartiles). The mean difference in the number of soap movements between the observation day and pre-observation days was 6.1 (range 3.0–10.3) for reactors and 0.3 (–3.7–3.3) for non-reactors ($P < 0.0001$).

Reactors reported a higher number of years of education for both the respondent and the head of household, compared with non-reactors (Table 2). Reactors were significantly more likely to report not having to share a toilet with other households, compared with non-reactors. Mobile phone ownership was reported by 11 (92%) reactors and 13 (42%) non-reactors ($P = 0.005$, odds ratio [OR] = 15.2, 95% confidence interval [CI] = 1.7–133.3). Household ownership of a watch was more frequent among reactors than among non-reactors. There was no significant difference between reactors and non-reactors with respect to social economic status (SES) score, as determined by principal components analysis ($P = 0.11$), although 83% of reactors were in the higher two SES quartiles compared with 48% of non-reactors. Visibility of the acceleration sensor at the time of retrieval was not associated with reactivity. Whereas 4 (33%) reactors participated in hand rinses for quantification of hand contamination, 17 (52%) of non-reactors did so ($P = 0.33$). Notably, reactors and non-reactors had similar median numbers of soap movements during the observation time block on pre-observation days (4.0 versus 3.7, $P = 0.53$).

Because of the small numbers in each subgroup, we did not conduct multivariate modeling to assess for confounding between the various factors associated with reactivity.

DISCUSSION

With the use of a novel measurement technique involving bars of soap with acceleration sensors, this study confirmed

TABLE 2

Factors associated with reactivity to structured observation among rural households observed after 3 days of placement of bars of soap with acceleration sensors, Bangladesh, 2007

Variable	Reactors (N = 12)	Non-reactors (N = 33)	P value*	Odds ratio (95% CI)
Mean difference in number of soap movements between structured observation day and mean of pre-observation days	6.1 (3.0–10.3, SD 2.1)	0.3 (–3.7–3.3, SD 2.0)	< 0.0001†	–
Mean number of soap movements during the structured observation time block on pre-observation days	4.7 (2.7–10.6, SD 2.5)	4.4 (0.3–10.6, SD 2.4)	0.53	–
At least one child < 5 years of age with cough during previous two weeks	3 (25%)	1 (3%)	0.06	10.0 (0.9–108.3)
Mean no. years education for the primary caregiver	7.9 (4.0–12.0, SD 2.5)	5.2 (0–14, SD 3.7)	0.04†	–
Mean no. years education for household head	7.5 (0–14, SD 4.1)	4.4 (0–12, SD 3.9)	0.06	–
Watch ownership	12 (100%)	20 (65%)	0.02†	14.0 (0.8–256.4)
Mobile phone ownership	11 (92%)	13 (42%)	0.005†	15.2 (1.7–133.3)
Defecation place not shared with other households	11 (92%)	11 (38%)	0.0002†	18.0 (2.0–159.3)
Socioeconomic status quartile				
Quartile 1 (poorest)	1 (8%)	8 (26%)	0.13	
Quartile 2	1 (8%)	8 (26%)		
Quartile 3	6 (50%)	6 (19%)		
Quartile 4 (wealthiest)	4 (33%)	9 (29%)		
Washed both hands upon request to demonstrate usual hand-washing practice	10 (83%)	15 (48%)	0.05†	5.3 (1.0–28.4)
Soap weight < 65 grams at the time of collection‡	4 (33%)	14 (42%)	0.74	0.7 (0.2–2.7)
Acceleration sensor visible at the time of collection	5 (42%)	10 (32%)	0.72	1.5 (0.4–5.9)
Participation in hand microbiology assessment	4 (33%)	17 (52%)	0.33	0.5 (0.1–1.9)

* P value calculated based on Mantel–Haenszel χ^2 , t test, or Wilcoxon rank sum, as appropriate.

† Statistically significant at the $P = 0.05$ level.

‡ Soap weight of 65 grams was chosen as the cutoff because of our previous experience indicating that the acceleration sensor is often visible when the overall weight of the sensor soap bar is 65 grams or less.

our hypothesis that households alter their handwashing behavior when observed. We found that, compared with previous days, sensor soap movements on the structured observation day increased by about 35%. Reactivity to observation was evident in a subset of the households characterized by greater social status, suggesting that structured observation captures information about different segments of the population differentially. Our findings call into question the validity of structured observation details because it appears that a majority of participating caregivers substantially altered their behavior in the presence of an observer.

A number of social psychology experiments have also confirmed the potential for reactivity to direct observation.¹³ In a university restroom in the United States, Pedersen and colleagues¹⁴ used visible and hidden observers to assess reactivity to observation. Only 3 of 19 subjects who were not aware of the observer's presence washed hands, compared with 18 of 20 subjects who were aware of the observer.¹⁴ In another study, users of a women's restroom more frequently washed their hands with soap when at least one other person was present at the sink area, compared with users who were alone at the sink.¹⁵ In their study also set in a college women's restroom and with findings similar to the two studies described above, Munger and Harris¹⁶ argue that theories of social influence may explain the increase in handwashing behavior when a subject is aware of an observer's presence. These restroom observation studies allowed detection of handwashing behavior at a critical juncture, specifically after possible fecal contact. The sensor soap does not allow us to detect such contextual reactivity. Compared with usual behavior, reactivity to structured observation may similarly affect handwashing behavior at all critical times or affect behavior at some critical times more than others.

Usual soap use behavior at critical times, and differential reactivity, may be mediated by awareness of social expectations regarding soap use. For example, in a society where hand contact with feces is considered particularly "dirty" and the social expectation is to wash hands with soap when such fecal contact occurs, caregivers under observation may be more likely to wash hands with soap than they would as usual practice.¹⁷ If no such expectation exists for washing hands with soap before feeding a child, a caregiver might not be prompted to do so any more while under observation than in her usual practice.

In studies of parent-child interaction, reactivity has been shown to be greater among mothers than among fathers.¹³ Because the sensor soap was used by all household members, and not only the caregiver, and since the caregiver was the primary target of the structured observation, we cannot know whether all household members, or caregivers primarily, altered soap use behavior in the presence of the observer.

Several correlates of reactivity were indicative of socioeconomic status, suggesting that those in higher socioeconomic strata are more aware of social expectations of handwashing than those in lower socioeconomic strata, and that structured observation may capture information on handwashing behavior differentially among subgroups of the population distinguished by socioeconomic status or education level. Those in lower socioeconomic strata in resource-poor settings may actually be less able to use more soap reactively, simply in an effort to impress or please an observer, because of the relative cost of soap use to their households.

Why do we behave better in public, when being observed, than in private, when not observed? Leary¹⁸ suggests that modifying one's behavior in public is a highly adaptive response to living in society. We behave differently so that we may be inoffensive or attractive to those around us. We behave differently so that we may conform to social norms. We may also behave differently because we perceive greater self-worth when our public behavior is perceived as positive.¹⁹ Furthermore, we may behave differently because we wish to display knowledge or attitudes consistent with higher social standing. Similar themes of attractiveness, social status, and affiliation have been confirmed as important motivators of handwashing with soap, on the basis of formative research in 11 countries.²⁰ Thus, reactivity to structured observation emphasizes the importance of social expectations for improving individual handwashing behavior.

Structured observation is a widely used method in studies of health behavior and social psychology.²¹ One thorough review of studies using structured observations of parent-child interactions finds that the presence of the observer has not been shown to substantially alter such behaviors.²² Structured observations have been important to explain various child health-related behaviors in resource-poor settings, for example types of treatments given to young children with diarrhea.^{23,24} Our study findings do not suggest that structured observation should be discarded as a measurement technique for health behaviors, or even for handwashing specifically. Rather, given the degree of reactivity detected, we conclude that the results of structured observation should not be interpreted as a valid measure of the frequency of handwashing. Instead, it is one of several measures of hand washing that can provide insight, for example on the situational context that constrains hand washing, but it also has important limitations. We should attempt to assess the degree to which social desirability exists within the study population for the behavior under study, and account for socioeconomic status and education in our assessments of behavior.

There was a high rate of failures of the sensor soaps, particularly during the first week after the departure of the international consultant (PH) who had previously worked with the acceleration sensors in a number of other studies. When acceleration sensors failed to yield movement data, the failure was complete, indicating either faulty initialization of the acceleration sensor or mechanical failure; in our experience, mechanical failures are typically caused by crushing of components of the acceleration sensor during the soap preparation process. The high failure rate suggests a lack of robustness and the challenges of knowledge and skill transfer relevant to this technology. Notably, the failure rate declined dramatically from the first to the second week after the international consultant's departure, suggesting that initial hurdles were quickly overcome. Although the failure rate does mean that we have fewer data points than we had intended, it does not imply that the data from the devices that did work were somehow compromised or invalid. Acceleration sensor failures did not in any way impact our assessment of reactivity because the high failure rate occurred during the first of two deployments of sensor soaps in the prolonged deployment households (with structured observation taking place at the end of the second deployment). Using data from the second deployment, we were able to compare sensor soap movements during the structured observation to sensor soap movements during the

preceding three days and we found a statistically significant increase in soap movement during the structured observation.

We faced one principal limitation in this study: assessing whether reactivity was a function of the structured observation alone or resulted from the combination of delivery of the sensor soap (another observational tool) and structured observation. We found that the placement of the sensor soap had little impact on household behavior either during the pre-observation days or on the day of structured observation. However, the structured observation may have motivated respondents to use the soap that had been given to them by the study worker more than they would have had they been simply using the soap that had already been present in the home. With the data available to us, we examined households that had minimal change in the household soap availability (among households that had one bar of soap before the study who were given one bar of soap with a acceleration sensor) and found that reactivity was similar to that seen among the larger group. In either case, the fact that soap movement increases substantially during the structured observation compared with pre-observation days confirms that presence of the human observer can lead to substantial deviations from usual handwashing behavior.

Structured observation data were not used in any way to address our principal study question of whether there was reactivity to the presence of the observer during the structured observation. Only sensor soap movement data were used to address this principal study question. This study was not designed to compare soap use information provided by different measurement techniques. Even if we dismiss the idea that the sensor movements truly reflect soap use, the increase in sensor soap movement, as measured during the period of structured observation, compared with days preceding structured observation, indicates that household members were physically interacting with the soap more during the structured observation than on preceding days.

This study confirms our hypothesis that individuals substantially alter handwashing behavior in the presence of an observer, and that this effect may be most pronounced among those most likely to be aware of social expectations of hand hygiene. Social marketing and other behavior change strategies attempt to influence handwashing behavior by modifying social norms regarding handwashing behavior, or by appealing to aspirations for higher socioeconomic status. Given our findings, reactivity to structured observation may be expected to increase, especially in response to such behavior change approaches. As handwashing promotion is increasingly scaled up globally for prevention of infectious diseases, there is a pressing need to identify valid and practical methods to measure handwashing behavior to monitor and evaluate such programs. Further investigation into strategies to mitigate the reactivity of individuals to structured observation is needed, because this technique is also used to measure a variety of human behaviors relevant to child survival, including child nurturing and treatment of watery diarrhea.¹³

Received December 15, 2009. Accepted for publication June 15, 2010.

Acknowledgments: We sincerely thank the participants for allowing us to spend many hours in their homes, occasionally during challenging circumstances. We also thank our excellent team of fieldworkers, who were diligent, good-humored, and thorough, even during acute and intense flooding. Bars of soap with acceleration sensors were provided by Unilever.

Note: Some information in this manuscript may have been posted in a report on the World Bank website.

Financial support: This study was commissioned by the Water and Sanitation Program with funding from the Bill and Melinda Gates Foundation to help the Water and Sanitation Program learn how to scale-up hand washing behavior change interventions. ICDDR, B acknowledges with gratitude the commitment of these institutions to the Centre's research efforts.

Authors' addresses: Pavani K. Ram, University at Buffalo, Buffalo, NY, E-mail: pkram@buffalo.edu. Amal K. Halder, M. Sirajul Islam, and Stephen P. Luby, ICDDR,B, Shaheed Tajuddin Ahmed Sharani, Mohakhali, Dhaka, E-mails: Amal_Halder@wvi.org, sislam@icddr.org, and sluby@icddr.org. Stewart P. Granger, Therese Jones, David Hitchcock, and Richard Wright, Unilever R&D, Port Sunlight, Bebington, Wirral, CH63 3JW, UK, E-mails: stewart.granger@southcentral.nhs.uk, therese.jones@unilever.com, davidh182@aim.com, and Richard.L.wright@unilever.com. Peter Hall, Unit 6, Capenhurst Technology Park, Capenhurst, Cheshire, CH16EH, United Kingdom, E-mail: Peter.Hall@4FrontResearch.co.uk. Benjamin Nygren, Rollins School of Public Health, Emory University, Atlanta, GA, E-mail: ghz8@cdc.gov. John W. Molyneaux, Water and Sanitation Program, The World Bank Group, MSN MC 6-622, Washington, DC, E-mail: jack_molyneaux@usa.net.

REFERENCES

1. Aiello AE, Coulborn RM, Perez V, Larson EL, 2008. Effect of hand hygiene on infectious disease risk in the community setting: a meta-analysis. *Am J Public Health* 98: 1372–1381.
2. Curtis V, Cairncross S, 2003. Effect of washing hands with soap on diarrhoea risk in the community: a systematic review. *Lancet Infect Dis* 3: 275–281.
3. Ejemot RI, Ehiri JE, Meremikwu MM, Critchley JA, 2008. Hand washing for preventing diarrhoea. *Cochrane Database Syst Rev* CD004265.
4. Luby SP, Agboatwalla M, Feikin DR, Painter J, Billhimer W, Altat A, Hoekstra RM, 2005. Effect of handwashing on child health: a randomised controlled trial. *Lancet* 366: 225–233.
5. Curtis V, Kanki B, Cousens S, Diallo I, Kpozehouen A, Sangare M, Nikiema M, 2001. Evidence of behavior change following a hygiene promotion programme in Burkina Faso. *Bull World Health Organ* 79: 518–527.
6. Manun'Ebo M, Cousens S, Haggerty P, Kalengaie M, Ashworth A, Kirkwood B, 1997. Measuring hygiene practices: a comparison of questionnaires with direct observations in rural Zaire. *Trop Med Int Health* 2: 1015–1021.
7. Stanton BF, Clemens JD, Aziz KM, Rahman M, 1987. Twenty-four-hour recall, knowledge-attitude-practice questionnaires, and direct observations of sanitary practices: a comparative study. *Bull World Health Organ* 65: 217–222.
8. Biran A, Rabie T, Schmidt W, Juvekar S, Hirve S, Curtis V, 2008. Comparing the performance of indicators of hand-washing practices in rural Indian households. *Trop Med Int Health* 13: 278–285.
9. Bentley M, Boot M, Gittelsohn J, Stallings R, 1994. *The Use of Structured Observations in the Study of Health Behaviour*. The Hague, The Netherlands: IRC International Water and Sanitation Centre.
10. Cousens S, Kanki B, Toure S, Diallo I, Curtis V, 1996. Reactivity and repeatability of hygiene behaviour: structured observations from Burkina Faso. *Soc Sci Med* 43: 1299–1308.
11. Curtis V, Cousens S, Mertens T, Trarore E, Kanki B, Diallo I, 1993. Structured observations of hygiene behaviors in Burkina Faso: validity, variability, and utility. *Bull World Health Organ* 71: 23–32.
12. Last JM, 2001. *A Dictionary of Epidemiology*. New York: Oxford University Press.
13. Russell A, Russell G, Midwinter D, 1992. Observer influences on mothers and fathers: self-reported influence during a home observation. *Merrill-Palmer Q* 38: 263–283.
14. Pedersen DM, Keithly S, Brady K, 1986. Effects of an observer on conformity to handwashing norm. *Percept Mot Skills* 62: 169–170.

15. Drankiewicz D, Dundes L, 2003. Handwashing among female college students. *Am J Infect Control* 31: 67–71.
16. Munger K, Harris SJ, 1989. Effects of an observer on handwashing in a public restroom. *Percept Mot Skills* 69: 733–734.
17. Zeitlyn S, Islam F, 1991. The use of soap and water in two Bangladeshi communities: implications for the transmission of diarrhea. *Rev Infect Dis* 13 (Suppl 4): S259–S264.
18. Leary MR, 1996. *Self-Presentation: Impression Management and Interpersonal Behavior*. Boulder, CO: Westview Press.
19. Tice DM, 1999. The looking glass self is also a magnifying glass. Baumeister RF, ed. *The Self in Social Psychology*. Ann Arbor, MI: Taylor & Francis, 195–218.
20. Curtis VA, Danquah LO, Aunger RV, 2009. Planned, motivated and habitual hygiene behaviour: an eleven country review. *Health Educ Res* 24: 655–673.
21. Gubbels JS, Kremers SP, Stafleu A, Dagnelie PC, de Vries NK, Thijs C, 2010. Child-care environment and dietary intake of 2- and 3-year-old children. *J Hum Nutr Diet* 23: 97–101.
22. Gardner F, 2000. Methodological issues in the direct observation of parent-child interaction: do observational findings reflect the natural behavior of participants? *Clin Child Fam Psychol Rev* 3: 185–198.
23. Bentley ME, 1988. The household management of childhood diarrhea in rural north India. *Soc Sci Med* 27: 75–85.
24. Halvorson SJ, 2004. Women's management of the household health environment: responding to childhood diarrhea in the Northern Areas, Pakistan. *Health Place* 10: 43–58.

APPENDIX: VALIDATION OF ACCELERATION SENSOR SOAPS FOR MEASUREMENT OF HANDWASHING BEHAVIOR

Briefly, members of our research team (TJ, RW, SG) explored the relationship between soap movements detected through accelerometers and handwashing behavior in a pilot study in

Dokur, a small rural village in Andhra Pradesh, India. A convenience sample of 26 households, each with at least 2 children < 13 years of age and one child < 5 years of age, was selected. Households were excluded if they contained individuals with adverse skin conditions. Informed consent was obtained and households were paid 70 rupees (~US\$ 1.57) for participation. Sensor soaps were set up so, when moved, they created 15-second-long records, which were time- and date- stamped. Data were available for analysis from 23 households; data from three households were not included because of sensor soap failure or illness in the household, which might have resulted in unusual soap movement patterns. Excluding placement and retrieval days, the mean length of deployment was 6 full days. Several indicators of validity were observed. More sensor records were created during the peak washing hours from 4 AM–10 AM (48% of all records) than at other times of day (28% and 24% between 10 AM–4 PM and 4 PM–10 PM, respectively). The fewest records were created at night (< 1% between 10 PM and 4 AM). There was a significant correlation between household size and number of records ($r = 0.55, P < 0.01$). During the peak body washing period (8 AM–10 AM) there was a tendency for records to cluster together compared with a time period associated with defecation and handwashing (5 AM–7 AM). This was established by creating “record clusters” by grouping two or more records together where there was less than a 5 minute gap between contiguous records. The mean cluster size between 8 AM and 10 AM was 4.33 records compared with 2.23 records between 5 AM and 7 AM. The overall number of records predicted the weight of soap on retrieval ($r = -0.57, P < 0.01$). This pilot study suggested that sensor soaps validly detect the level of household soap use.