Spot-check Observational Method for Assessing Hygiene Practices: Review of Experience and Implications for Programmes

Marie T. Ruel and Mary Arimond

International Food Policy Research Institute,
2033 K Street, NW, Washington, DC 20006, USA

ABSTRACT

Structured observations, considered to be the method of choice for assessing hygiene practices, are time-consuming and expensive and are subject to reactivity (households modifying their behaviours in the presence of an observer). Hygiene practices also vary from day to day. Spot-checks, an alternative approach for observation, have recently gained increased popularity. In this approach, a list of predetermined conditions is observed at one point in time during a home visit. This paper reviewed experience with spot-checks for measuring hygiene practices and discussed strengths and weaknesses of the approach and its potential usefulness to those undertaking programmes. Review of seven studies carried out in various contexts confirmed that spot-checks are a promising alternative to structured observations, because these are less intrusive, less time-consuming, more economical, and less reactive. These are, however, equally affected by day-to-day variations in practices and conditions, and repeated observations are, thus, recommended. Additional research is needed to assess the validity, reactivity, and repeatability of hygiene indicators derived from spot-checks.

Key words: Diarrhoea; Hygiene; Knowledge, attitudes, practice

INTRODUCTION

Several studies have focused on the measurement of hygiene practices (1-4), and various qualitative and quantitative methods have been used for assessing hygiene practices. Many studies have mixed and sequenced the following approaches: survey methods based on recall, in-depth and key-informant interviews, focus groups, other participatory rapid assessment tools, semi-structured and structured observations, and spot-checks. Microbiological methods, involving sampling bacteria from hands, have also been used as a proxy measure of effective hand-washing practices (5,6). The limitations of recall methods for measuring hygiene practices are well-recognized. Several studies have demonstrated that respondents consistently and significantly over-report ‘good’ practices (7-10). Because of this, many researchers have employed structured observations, involving systematic recording of practices by passive observers. While these can provide detailed information and quantitative estimates of the prevalence of various practices, structured observations have three disadvantages. First, these are time-consuming and labour-intensive and require careful standardization of skilled observers. Second, these are subject to ‘reactivity,’ which occurs when individuals modify their behaviour because of the presence of an observer. Finally, these are affected by day-to-day variability, which is common to many types of behaviours, including hygiene. These latter two concerns have led to questions regarding the...
validity (whether the measurement reflects the truth) and reliability (repeatability, i.e. whether repeated measurements provide the same answer) of structured observations. Reactivity compromises both validity and reliability: the observer cannot see the true, usual practice (so the measure is not valid); and the practice changes between visits (so the measure is not reliable).

A less-intrusive and possibly less-reactive approach for observation that has gained increasing popularity in recent years is the spot-check. In this approach, a list of predetermined conditions is observed at one point in time during a home visit. In contrast to structured observations, spot-checks can be performed rapidly and unobtrusively. This approach is intended to capture information on the product of hygiene behaviours rather than the behaviours themselves (11). For example, the spot-observation that hands and nails of mothers are dirty is presumed to reflect the fact that mothers do not wash their hands frequently (or carefully). Thus, spot-observations provide information on ‘proxies’ for behaviours and, by definition, do not require observation of the actual behaviours.

The present paper reviews recent experience with this relatively novel approach for the measurement of hygiene practices and discusses strengths and weaknesses of the approach and its potential usefulness to those undertaking programmes. We have also summarized our own experience with spot-checks to measure hygiene practices in the context of a representative survey in a major urban African centre in Accra, Ghana (12,13).

**MATERIALS AND METHODS**

The Medline and Agricola bibliographic databases were searched for literature on measurement of hygiene practices published during the 1990s, and we selected seven studies that used spot-check methods of observation. A brief summary of study designs, types of hygiene indicators and composite indices used, and key findings of analyses of validity, repeatability, and associations with childhood illnesses is presented in the table.

The main criterion for inclusion of studies in this review was that the spot-check methods of observation were employed to measure hygiene practices (or proxys for hygiene practices). The review did not select and rank studies based on their overall quality—study design, indicators used, control for confounding factors in statistical analysis—as would be required for a meta-analysis documenting the association between hygiene practices and childhood diarrhoea, for example. The main objective of the present review was, thus, purely methodological, i.e. to review the experience with spot-checks in various contexts and to document the types of indicators derived from this methodology to measure hygiene practices and their association with childhood illnesses.

**RESULTS**

**Types of hygiene practices observed using spot-check approaches**

Various dimensions of hygiene were observed using spot-checks in the seven studies (Table). Indicators from the following five clusters of hygiene practices defined by Boot and Cairncross (1) were represented in these studies: (i) disposal of human faeces, (ii) use and protection of water sources, (iii) water and personal hygiene, (iv) food preparation and storage, and (v) domestic and environmental hygiene.

Indicators to assess personal hygiene (or more specifically proxies for these practices, such as cleanliness of mother and/or child) and domestic and environmental hygiene, e.g. cleanliness of floor and compound surfaces and garbage disposal, have been particularly popular in spot-checks and were included in all seven studies. Depending on the focus of each particular study, other dimensions of hygiene practices were emphasized. These included storage conditions of food and water in a longitudinal positive deviance study (14) and cleanliness of dishes and utensils in a focused longitudinal behaviour change intervention (6).

**Assessment of reactivity and repeatability of spot-check method**

Only one of the seven studies assessed the reactivity and repeatability of spot-check observations of hygiene practices. The study in Nicaragua compared indicators of hygiene measured by spot-check on two separate observation days and showed good agreement between the first and the second visits for most indicators, suggesting low reactivity (15). It is still possible that for some most unacceptable hygiene conditions, such as the presence of human faeces in the house or on the compound, reactivity could be hidden if the households were cleaned up before both the visits. The authors of the study did not specify whether families were informed of
the exact day and time of the field workers’ visits, but they indicated that the purpose (observation of hygiene practices) of the visit was not revealed to families. This might have helped reduce reactivity. Additionally, reactivity was not a major problem for other domestic hygiene indicators, since only about one-third of the households had clean kitchen or living-room floors.

Using a kappa statistic, the Nicaragua study also assessed the individual- and/or household-level repeatability of observations (15). The authors stated that scores from 0.40 to 0.75 were generally considered to indicate good repeatability, and scores over 0.75 indicate excellent repeatability. Seven of the 11 spot-check indicators in the Nicaragua study were in the range of good repeatability (0.40-0.75), and only four had a kappa-coefficient lower than 0.40, indicating poor repeatability. There was no apparent clustering of indicators by type of behaviours for those that had low repeatability compared to good repeatability; indicators from all five clusters of hygiene behaviours were distributed almost equally between low and good repeatability.

Little information on reactivity and repeatability of hygiene behaviours, assessed through spot-checks, is available. We reviewed the available information on these issues from structured observations. Several studies explored the validity and reliability of structured observations using repeated visits to households (8,15,16). These studies generally concluded that relevant hygiene practices had very low repeatability at the individual level. Practices vary significantly from one observation to the other—both due to reactivity and to natural day-to-day variation. At the population level, on the other hand, the repeatability of structured observations is relatively good on average, although it tends to vary between practices. A study of practices relating to child defaecation in Burkina Faso showed that some practices, such as disposal of stools in a latrine, washing the child’s bottom with water, and subsequent hand-washing, were not reactive and had good repeatability at the population level (16). Other related practices, such as frequency of child defaecation in the yard and hand-washing after visiting the latrine, were reactive and, thus, had poorer repeatability.

Examples of use of spot-check methods to derive indicators of hygiene and to document association between hygiene practices and childhood illnesses

Six of the seven studies examined associations between hygiene practices measured by spot-checks and diarrhoea. The study in India—a positive deviance study—looked at the association of hygiene practices with growth outcomes of children, but did not report information on childhood diarrhoea (14).

In Nicaragua, Gorter et al. found that many hygiene variables measured through spot checks were associated with diarrhoea in the expected direction, i.e. better practices were associated with less diarrhoea, although only four indicators reached statistical significance (15). These indicators were: clean living-room floor; clean kitchen floor; garbage organized in heaps or absent; and the use of diapers or underclothes by the child. Absence, or weakness, of the association between diarrhoea and various spot-check indicators at the household level might have been due to the low-to-moderate repeatability of some hygiene indicators. The number of families observed was smaller for some key practices (e.g. hand-washing), which might have resulted in lower statistical power for these practices. Additionally, as suggested by the authors in their discussion of structured observations, it is also likely that there might be some threshold number of good practices required to reduce the risk of diarrhoea. This would explain the lack of association with diarrhoea for many individual practices. For example, a spotless kitchen-floor and clean utensils will not protect a child who plays in an unhygienic compound. Some clusters of key practices may be necessary and may be better reflected by summary scales or indices. The authors did not report any multivariate analysis of the data.

Bartlett et al. incorporated 26 spot-check indicators in their longitudinal study of persistent diarrhoea among children aged 0-30 month(s) in Guatemala (11). Diarrhoea was measured by active surveillance of households through weekly visits by field staff throughout the year and every two days during episodes of diarrhoea. The 26 hygiene spot-check indicators were

Field workers were trained in spot-check methods during a one-week workshop, followed by field training in the presence of supervisors. Field workers were blinded to the diarrhoea status of children in the households observed, and the purpose of the visit was not described to family members. Diarrhoea was measured using daily diarrhoea calendars designed for illiterate mothers. The calendars were collected weekly by the field workers.
based on extensive preliminary observations in the community and included cleanliness of the mother, child, sibling, and household environment (Table). For each indicator, an index was calculated by dividing the number of times the condition was unhygienic by the number of times the indicator was observed for that household. These indices were then classified as high or low, based on their respective distributions. Finally, the dichotomous variable giving a score of high or low on each condition was compared between households where the index child had no episodes of persistent diarrhoea during the year and households where the child had one or more episode(s).

In bivariate analyses, 11 of the 26 variables were associated with persistent diarrhoea. In logistic regression analyses, 9 of these variables remained significantly associated, after controlling for the child’s age and a number of other potential confounders. The nine variables were: toys or baby bottles on the ground; dirty hands of mother; trash on floor of house; animals loose inside house; faeces on ground; or faecally-soiled diapers on child or on ground, either inside or outside living area.

The authors made three other important points in this report. First, they rightly noted that indicators relating to the presence of faecally-soiled diapers can be effects rather than (or as well as) causes of diarrhoea and that the possibility of reverse causality should receive attention when these indicators are used. Second, and consistently with others, they noted that behaviours clustered. Of the 55 possible paired correlations among the 11 conditions associated with diarrhoea, they reported that 38 (69%) were significant. Finally, the authors analyzed the cumulative effect of hygiene practices on the risk of diarrhoea. They created a composite index by summing six of the nine variables that were significant, after controlling for the child’s age and a number of other potential confounders. The scales were condensed to ‘1’ for analysis if completely clean and to ‘0’ otherwise. An overall cleanliness scale was created by averaging the cleanliness scores for the ground, child’s face, child’s hands, and mother’s sari. Reliability of the scales was assessed using Cronbach’s alpha and deemed satisfactory. The cleanliness score was strongly and negatively correlated with the prevalence of diarrhoea, measured weekly by day-to-day recall, in multivariate models that controlled for education of mothers and socioeconomic status.

Pinfold and Horan also reported the use of a spot-check indicator as part of evaluation of a focused behaviour change intervention in Thailand (6). This intervention used spot-check observations for assessing the cleanliness of dishes, along with microbiological methods for assessing faecal streptococci on fingertips taken as an indicator of the effectiveness of hand-washing. The spot-check method was used for assessing behaviour change relating to a very specific project message. The authors reported ‘significant improvements’ in intervention villages, although the magnitude of observed changes is not reported. In control villages, the authors reported a 45% difference in the incidence of diarrhoea between homes where no dirty dishes were seen on three separate occasions compared to homes where dirty dishes were seen on all three occasions. No multivariate analyses were reported.

In Bangladesh, Ahmed et al. used spot-check indicators of cleanliness, in addition to those from structured observations and other methods, in a pre-post evaluation of an educational intervention (17). Cleanliness of the child’s hands, face, and clothes, of the mother’s sari, and of two areas on the ground where children most often played was assessed and rated on a three-point scale. The scales were condensed to ‘1’ for analysis if completely clean and to ‘0’ otherwise. An overall cleanliness scale was created by averaging the cleanliness scores for the ground, child’s face, child’s hands, and mother’s sari. Reliability of the scales was assessed using Cronbach’s alpha and deemed satisfactory. The cleanliness score was strongly and negatively correlated with the prevalence of diarrhoea, measured weekly by day-to-day recall, in multivariate models that controlled for education of mothers and socioeconomic status.

First, they rightly noted that indicators relating to the presence of faecally-soiled diapers can be effects rather than (or as well as) causes of diarrhoea and that the possibility of reverse causality should receive attention when these indicators are used. Second, and consistently with others, they noted that behaviours clustered. Of the 55 possible paired correlations among the 11 conditions associated with diarrhoea, they reported that 38 (69%) were significant. Finally, the authors analyzed the cumulative effect of hygiene practices on the risk of diarrhoea. They created a composite index by summing six of the nine variables that were significant, after controlling for the child’s age and a number of other potential confounders. The scales were condensed to ‘1’ for analysis if completely clean and to ‘0’ otherwise. An overall cleanliness scale was created by averaging the cleanliness scores for the ground, child’s face, child’s hands, and mother’s sari. Reliability of the scales was assessed using Cronbach’s alpha and deemed satisfactory. The cleanliness score was strongly and negatively correlated with the prevalence of diarrhoea, measured weekly by day-to-day recall, in multivariate models that controlled for education of mothers and socioeconomic status.

Field workers were standardized in observational definitions using photographs, in which field workers individually scored photographs of typical situations, which were then discussed among the group. This was followed by field observations, where the field workers were paired to visit households. Standardization was achieved once full concordance between pairs (and with supervisors) was achieved.

No information on training and standardization of field staff in spot-check methods was provided.

Since it is often impossible for researchers to perform repeated measurements at two points in time, various statistical techniques have been developed for assessing reliability. Cronbach’s alpha is a widely-accepted statistic for assessing reliability and internal consistency of scales developed from a single set of measurements (one point in time) (24). Ahmed et al. (17) did not report alpha individually for each scale but stated that all had an alpha of at least 0.57, and that this was satisfactory.

Diarrhoea was measured in this study through active surveillance of children aged less than five years using calendars collected and verified by field workers on a weekly basis.
<table>
<thead>
<tr>
<th>Author/ reference/country</th>
<th>Type of study</th>
<th>Type of hygiene indicators</th>
<th>Index creation</th>
<th>Analyses and findings</th>
<th>Association with childhood diarrhoea (and growth)</th>
</tr>
</thead>
</table>
| Gorter et al. 1998 (15) (Nicaragua) | Prospective follow-up study (172 households with children aged less than 24 months) | -Food, utensils, baby-bottle covered  
-Cleanliness of the environment, floor, absence of faeces  
-Animals in compound  
-Type of garbage disposal  
-Water availability, storage, quantity  
-Sanitation facilities: availability, cleanliness  
-Child uses diapers, underclothes | Index: created for each of the 26 indicators: number of times unhygienic behaviour observed/number of times indicators observed; created two categories: high/low, based on individual distribution of each variable  
Composite index: summed six variables significant in multivariate analysis | Reactivity: compared results of two observations. Good agreement between two measures (low reactivity)  
Repeatability: used kappa statistic to assess individual- and household-level repeatability  
Results: low repeatability for 10 indicators (kappa < 40), good for 12 indicators (kappa 40-75) | Individual practices  
Four indicators were associated with diarrhoea (kitchen floor clean, living room floor clean, garbage organized or absent, child uses diaper or underclothes) |
| Bartlett et al. 1992 (11) (Guatemala) | Prospective year-long study of persistent diarrhoea (280 children aged 0-30 month(s)) | -Cleanliness of mother, child, siblings  
-Water storage type  
-Children's toys, bottles on floor  
-Faeces on the ground  
-Faecally-soiled diapers  
-Animals in compound  
-Garbage in living area | Index: created for each of the 26 indicators: number of times unhygienic behaviour observed/number of times indicators observed; created two categories: high/low, based on individual distribution of each variable  
Composite index: summed six variables significant in multivariate analysis | Individual indices: Compared children with no persistent diarrhoea (during 1 year) with those who had ≥1 episode. Eleven of the 26 indices associated with diarrhoea; 9 remained significant in multivariate analyses (logistic regression)  
Composite index: Ten percent of children with 0-1 unhygienic condition had an episode of persistent diarrhoea in previous year compared to >50% among those with all six unhygienic conditions (dose-response relationship was found) | |
| Ahmed et al. 1993 (17) (Bangladesh) | Longitudinal study of impact of behavioural change intervention [185 households with children aged 0-18 month(s)] | -Cleanliness of child's hands, face, clothes  
-Cleanliness of mothers' sari  
-Cleanliness of two areas on ground where children most often played | Scale: average cleanliness score for ground, child's face, hands, and mother's sari | Reliability:  
Chronbach's alpha >0.57 | Composite index  
Strong negative association between hygiene practices scale and prevalence of diarrhoea (controlling for maternal education and socioeconomic status)  
Contd.. |
<table>
<thead>
<tr>
<th>Author/Reference/country</th>
<th>Type of study</th>
<th>Type of hygiene indicators</th>
<th>Index creation</th>
<th>Analyses and findings</th>
<th>Association with childhood diarrhoea (and growth)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pinfold and Horan 1996 (6) (Thailand)</td>
<td>Longitudinal study of focused behaviour change intervention: 25 intervention and 12 control villages (4,874 households with children aged less than 60 months)</td>
<td>-Bacteria on fingertips (faecal streptococci) -Cleanliness of dishes</td>
<td>Each indicator was rated as poor, average, good</td>
<td>Difference of 45% in incidence of diarrhoea between homes where no dirty dishes were seen on three visits compared to those with dirty dishes on three visits</td>
<td></td>
</tr>
<tr>
<td>Merchant and Udipi 1997 (14) (India)</td>
<td>Longitudinal positive and negative deviance in growth (150 children aged 6-36 months). Cross-sectional comparison of positive and negative deviants</td>
<td>-Cleanliness of floor, storage vessels -Storage conditions of food, water -Cleanliness of child’s nails, skin, face, clothes</td>
<td></td>
<td>No analysis of association with diarrhoea, only with growth: linear relationship between growth and cleanliness and storage of water and food; good practices were observed in 72% of positive deviants, 56% of median growers, and 39% of negative deviants (no control for confounding factors reported)</td>
<td></td>
</tr>
<tr>
<td>Kaltenthaler and Drasar 1996 (18) (Botswana)</td>
<td>Cross-sectional, mixed method study to guide design of hygiene intervention (116 families with children aged less than 6 years)</td>
<td>-Water storage -Presence of animals -Faeces in compound -Presence of leftover food, infant-bottles -Compound conditions (+point scale) -Presence of unwashed dishes, washing water -Presence of toilet</td>
<td>Index: included compound condition, plate, cloth, bacterial counts on fingertip of caregiver, distance to water source, toilet, animals in kitchen, faeces in compounds</td>
<td>Individual practices: None of the individual practices was significantly associated with diarrhoea. Composite index: Significantly more diarrhoea among the highest compared to the lowest quartile of hygiene index (no information on magnitude of effect)</td>
<td></td>
</tr>
<tr>
<td>Armar-Klemasu et al. 2000 (12) (Ghana)</td>
<td>Cross-sectional survey of 556 households with children aged less than 36 months</td>
<td>-Mother clean -Child clean -Diaper/bottom clean -Compound swept -No poultry faeces -No stagnant water -No human faeces -No unwashed utensils -Drinking water covered -House swept -No garbage container in house</td>
<td>Index was created by summing scores for first six indicators. Each practice was rated 1 (=good) or 0 (=poor). Scores were divided into three categories (based on distribution): poor hygiene (0-3 good practice(s)), average (4-5), and good hygiene (6 good practices)</td>
<td>Reliability: Cronbach’s alpha 0.69 (good internal consistency)</td>
<td>Individual practices: Nine of 11 practices associated with lower prevalence of diarrhoea (only three statistically significant: child clean, compound swept, no stagnant water) Composite index: Diarrhoea prevalence: 40% for those with poor hygiene, 32% for those with average hygiene, and 25% for those with good hygiene (statistically significant) Multidrive analysis: Stronger association among poorer and wealthier households than those with average socioeconomic status</td>
</tr>
</tbody>
</table>
In Bombay, Merchant and Udipi included spot-check indicators of hygiene in their positive deviance study (14). Repeated measurements of child growth were used for defining positive deviants, median growers, and negative deviants. Subsequently, households with children in each category were surveyed and observed in a cross-sectional study. The authors explored the association between hygiene indicators and growth, but did not study the association with diarrhoea. Spot-observations included cleanliness of floor; storage vessels; storage conditions of food and water; and cleanliness of the child’s nails, skin, face, and clothes (Table). Each indicator was rated as poor, average, or good7. Results indicated linear relationships between growth status and cleanliness and storage of water and food. These were rated as ‘good’ in 72% of homes of positive deviant children, in 56% of median growers, and in 39% of negative deviants. Conversely, these conditions were rated as ‘poor’ in 41% of homes of negative deviants, 20% of median growers, and 12% of positive deviants. Negative deviants were more likely than either median growers or positive deviants to have dirty fingernails and boils, or rashes. These findings should be interpreted with caution however, because control for socioeconomic status or for age of the child was not reported either in study design (by matching subjects) or in statistical analysis (using multivariate analyses).

The study in Botswana used mixed methods for examining the association between hygiene behaviour and diarrhoea (18). Qualitative methods were used first (focus-group discussions, unstructured observations, and in-depth key-informant interviews), followed by a quantitative survey which included data on socioeconomic status, diarrhoeal morbidity, and child anthropometry, in addition to spot-checks for hygiene and bacteriological sample collection from water, dishes, cleaning-rags, infant-feeding bottles, and hands8. The final phase included additional focus-group discussions and in-depth interviews. Spot-check indicators used in this study included various aspects relating to the use and storage of water, general hygiene conditions of the house and compound, and the availability of services9 (Table). Individually, neither spot-check indicators nor bacteriological indicators were associated with diarrhoea. Since hygiene practices clustered, an index was developed, combining the conditions of the compound, bacterial counts from plates, rages, and caretakers’ fingertips, distance to water source, ownership of toilet, presence of animals in the kitchen, and of faeces in the compound. The index was scored from 8 to 20 and divided into quartiles. Comparison of the lowest to the highest quartile showed significantly ‘more diarrhoea’ among the households in the lowest quartile, although the magnitude of the difference was not reported.

Finally, in our study in Accra, 11 spot-check indicators were used (12,13). The indicators included: maternal and child cleanliness10, child’s diaper/bottom cleanliness, whether house and compound were swept, presence or absence of poultry or human faeces, presence of garbage in house, stagnant water in compound, presence of unwashed utensils or plates and whether drinking water was covered11. Each variable was scored ‘1’ when the practice was good and ‘0’ otherwise. A simple hygiene index was created by summing up scores for all items that were available for at least 90% of the households; this meant exclusion of four indicators. In addition, the ‘human faeces on the compound’ indicator was eliminated because of lack of variability; faeces were observed in only 2% of the households. The final index included six of the 11 hygiene indicators observed (Table). A categorical variable was created based on the distribution of index scores. Households with 0-3 good practice(s) were classified as having ‘poor’ hygiene, those with 4-5 good practices had ‘average’ hygiene, and those with the maximum score of 6 were classified as having ‘good’ hygiene practices. Cronbach’s alpha was 0.69, indicating a good level of internal consistency.

Most variables were associated in the expected direction with the prevalence of childhood diarrhoea,

---

7 No information was provided about training and standardization of field workers in spot-check methods.
8 To rate cleanliness of children and mothers, field workers were instructed to focus on appearance of their hands, nails, clothes, and hair. For children, the cleanliness of their face was also assessed.
9 Field workers were carefully trained and standardized in assessments of each indicator, until perfect concordance was achieved between pairs of fieldworkers and between each field worker and supervisor.
measured by two-week recall: poor practices were associated with a higher prevalence of diarrhoea. Few relationships, however, reached statistical significance. As discussed, there are several possible reasons for this, including potential unreliability in spot-check indicators and the possibility that a cluster of good practices—rather than any single practice—is necessary to decrease risk. In this instance, the limited nature and potential unreliability of information available on diarrhoea (a single two-week recall) was also a constraint.

To verify the notion that a minimum cluster of hygiene practices is necessary to reduce risk, we looked at the association between our composite index and various childhood illnesses. The composite index was significantly associated with diarrhoea, and the difference followed a strong dose-response relationship. Forty percent of children from households with poor hygiene (scores of 0-3) had diarrhoea in the previous two weeks compared to 32% among those with average hygiene (scores of 4-5) and 25% among those with good hygiene (score of 6).

Since previous analyses of the data showed a strong association of the hygiene index with several socioeconomic variables (12), we used a logistic regression model for controlling for socioeconomic and environmental factors when testing the association between hygiene practices and diarrhoea. In this model, the hygiene index was not significantly associated with childhood diarrhoea, but a two-way interaction was found between the quality of housing and asset index (a proxy for household wealth) and the hygiene index. The significant interaction, illustrated in the figure, revealed that benefits of good hygiene practices in reducing the prevalence of diarrhoea were larger among households from the two extreme wealth terciles. Children from the lowest socioeconomic group were approximately half as likely to have had diarrhoea in the previous two weeks if their household was in the good hygiene group compared to the poor hygiene group (23% and 41% respectively). Although children from the highest socioeconomic group were generally less likely to have had diarrhoea than children from the lowest socioeconomic group, good hygiene practices were also associated with a much lower prevalence of diarrhoea among the higher socioeconomic group (16% and 30% among households with good and poor hygiene respectively). We cannot explain why good practices were not beneficial to households from the middle socioeconomic group; further investigation is needed.

**DISCUSSION**

The two essential characteristics of good indicators are their validity and reliability. For hygiene indicators measured through either structured or spot-checks, Gorter et al. suggest the following additional criteria: (i) non-reactive (not affected by presence of observers); (ii) related to the outcome of interest; (iii) representative of the range of practices in the particular context where they are used; (iv) easily observed in most households (not hidden, and frequently performed, thus yielding limited missing data); (v) easy to standardize within and between observers (unambiguous, yielding limited intra- and inter-observer differences); and (vi) variable between individuals or households in a particular context, i.e. the practice should not be uniformly performed in one single way, so that it can distinguish between individuals or households.

Validity, reliability, and reactivity

Validity and reliability are two characteristics universally recognized as essential for indicators to be useful in the
Spot-check observational method for assessing hygiene practices

contexts of research, programme, or policy analysis. The results of our review showed that, although many indicators were used for measuring hygiene behaviours by spot-checks, few were actually tested for their validity and reliability.

Limited experience with spot-check methods suggests that these are less reactive than structured observations, probably because these can be carried out quickly and discretely. Like structured observations, however, spot-checks are susceptible to problems of day-to-day variability. The only study that examined reactivity and repeatability in spot-check observations did, indeed, show little evidence of reactivity but showed significant day-to-day variability (15). The magnitude of day-to-day changes also varied markedly between practices.

Recommendations for programmes to address problems of reactivity and day-to-day variations in behaviour

Reactivity in spot-checks can be reduced by carrying out observations discretely and unobtrusively and by not informing families of the exact day and time observers are planning to visit their households. This may prevent the problem of families making ‘special preparations’ for field workers’ visits and, thus, help reduce reactivity. Normal, non-reactive day-to-day variability in practices, however, is inevitable and presents the same challenge to spot-checks as to structured observations. The main consequence of day-to-day variability is that it makes indicators more ‘noisy’, or less precise because of larger variability. Depending on the purpose of data collection, the consequences of this increased variability will be more or less critical. In contexts where only population-level estimates of prevalence are needed—for example, in many pre- and post-intervention surveys—random day-to-day variability may not be a major concern. ‘Noisy’ indicators require larger sample sizes to increase precision of estimates, but they do not introduce a systematic bias. In the case of risk assessment of an individual household or studies relating hygiene behaviours to outcomes, such as diarrhoea, day-to-day variability results in misclassification of exposure status and, thus, can mask underlying associations.

The approach most commonly recommended to address the problem of day-to-day variability for both spot-check and structured observations is to repeat the observation on different days, so that information from multiple days can be averaged. This is particularly useful for investigations that aim at assessing individual- or household-level causal associations between hygiene practices and health outcomes as described above.
Repeating observations can also address the problem of reactivity, which, as indicated above, may be more of a concern for structured observations. In a longitudinal structured observational study of feeding interactions and behaviours, Gittelsohn et al. showed that reactivity declined rapidly between the first and the second visits and remained somewhat stable in subsequent visits (19). To address the problem of reactivity, the authors recommend either discarding data from the first day, or standardizing the first day’s data based on subsequent observations. Although reactivity may be less problematic for spot-check, repeating observations when possible, should help address both day-to-day variability and potential reactivity problems.

**Context-specificity**

Evidence from the literature on structured observations suggests that the validity, reliability, and reactivity of indicators are context-specific. Studies in Burkina Faso (16) and in Nicaragua (15), for example, provide evidence of differences in reactivity in the practice of disposal of children’s stools between the two contexts. Similar evidence on context-specific differences in reactivity and repeatability of spot-check indicators is not yet available. However, given the experience with structured observations, there is reason to believe that such differences could be observed.

In addition to being non-reactive and repeatable, indicators should vary between households and should be easily observable. In the studies reviewed, these characteristics varied by cultural and/or geographic context. For example, the presence of human faeces on the compound was not useful in the Accra study, because it was observed too infrequently (12). In Nigeria, Omotade et al. reported observing faeces near the house in 10% of peri-urban compounds and 25% of rural compounds, making it a potentially discriminating variable (20). Similarly, cleanliness of the floor inside the house in the Accra study was unobservable by spot-checks in one-half of households, because field workers were not invited inside the house. In Nicaragua, on the other hand, the condition of floor surfaces was an easily-observed and discriminating indicator (15).

Validity, reliability, reactivity, and other characteristics of indicators measured through spot-checks should be tested in a wider range of contexts to identify commonalities and differences between contexts.

**Mixed methods and triangulation of results**

The purpose of this review was to highlight the potential usefulness of spot-checks for measuring hygiene practices. No single method is perfect. The potential usefulness of mixed methods is a theme consistently emphasized in the literature on hygiene measurement. Various qualitative and quantitative methods may be employed, including an array of rapid and/or participatory techniques (e.g. focus-group discussions, structured and unstructured key-informant interviews, community meetings), structured and unstructured observations, spot-checks, and various types of surveys to measure hygiene practices. Although each approach has its strengths and weaknesses, mixed methods produce complementary information and allow triangulation of findings, validation of new tools, and development of simplified data-collection tools.

**Use of composite indices for programme monitoring and evaluation**

This review highlighted the potential usefulness of creating composite indices of behaviours for specific purposes. Indices may be particularly helpful in instances where practices cluster, or where a minimum number of good practices is necessary to achieve a significant impact on a particular outcome. This seems to be the case for hygiene practices.

Using indices, it is also possible to combine multiple dimensions of a concept. Reciprocally, it is a main disadvantage that summary indicators may conceal associations between individual practices and outcomes of interest. While this does not constitute a problem for some applications, it does limit interpretation for others.

Composite indices are not likely to be useful for programme monitoring. The main purpose of monitoring is to collect and review information on project implementation (activities, coverage, and use), which can then be used for re-orienting the programme and strengthening implementation and the quality of service-delivery in an ongoing fashion (21-22). For this purpose, clear and simple process indicators that can be easily measured and interpreted are required. Composite indices do not meet these criteria. However, when interim measures of programme impact are needed, indices may be useful, as described below in the context of evaluations.

Composite indices may be useful for evaluating interventions, especially those that target specific
behaviours for modification. Selection of variables to include in the index should be based on behaviours targeted by the intervention. With this approach, the global impact of the intervention on practices can be assessed. In addition, creating the index does not prevent evaluators from examining the impact of the intervention on individual practices as well, if desirable. The main advantage of using an index is that it allows a summary of the impact of the programme using a single indicator. This can be particularly useful to illustrate the findings graphically and for advocacy purposes. The index also may be used in bivariate or multivariate analyses for examining the association between practices and outcomes, or alternatively, it may be used for studying the determinants of changes in behaviours.

As emphasized earlier, selection of variables for an index should be context-specific. This was clearly reflected in creation of the hygiene index in our Accra study. Four variables had to be eliminated from the index: three because they could not be observed in many households and, thus, yielded too many missing values and one because the practice was uniformly positive (it lacked variability). Reliability of the index should also be tested in the context in which it is used and it should be validated or triangulated against other indicators whenever possible.

This review confirms that spot-check observations are a promising alternative to structured observations, because these are less time-consuming, less costly, and less reactive. Since—in common with structured observations—these are affected by day-to-day variations in practices and conditions, repeated observations are optimal, particularly when the need for precision is high.

As with all observational approaches, spot-checks require intensive training and standardization of field workers, because such methods usually involve making subjective judgements. Once training and standardization are completed, spot-check methods become much less time-intensive than structured observations and, therefore, constitute a time-saving and economic alternative approach to structured observations.

Additional research is needed to assess the validity, reactivity, and repeatability of hygiene indicators and composite indices derived from spot-checks in various cultures. In some cases, it is possible that the existing data from longitudinal studies could be re-analyzed to address reactivity and repeatability. Examples of useful indicators and composite indices that have been used in spot-check observations of hygiene practices are available from the seven studies reviewed. Experience with these indicators should be used for developing and validating similar observation tools in other contexts.

Programmes to improve hygiene practices should be preceded by qualitative research, even if it is relatively rapid and simple. At a minimum, basic information should be collected to illuminate local practices of interest and cultural norms surrounding these practices and to assess whether the practices of interest are amenable to change. Information should also be gathered on practical aspects, such as which key practices are observable, whether these vary within the population, and whether these are likely to be more or less reactive in a particular culture. This information will be valuable for selecting the practices to be targeted, for designing the implementation strategy, and for selecting suitable indicators and methods for monitoring and evaluation in different contexts. Mixed methods are also recommended in the monitoring and evaluation stages to allow for triangulation of findings and validation of indicators.

Measurement of hygiene practices is far more advanced than measurement of other areas of childcare, such as psychosocial care and mother-child interactions (23). Spot-check methodologies, however, still require more validation of research and experience from a sufficient variety of contexts. The richness of experience from structured observations should be used for accelerating progress in understanding possibilities and limitations of the spot-check method. This experience should also be used for developing methodologies and indicators for measuring other key dimensions of caregiving to children.

ACKNOWLEDGEMENTS

Financial support for this study was provided by the Food and Nutrition Technical Assistance (FANTA) Project and by IFPRI.

REFERENCES


