4-2017

The Effectiveness of Sanitary Inspections as a Risk Assessment Tool for Thermotolerant Coliform Bacteria Contamination of Rural Drinking Water: A Review of Data from West Bengal, India

Christian Snoad
DelAgua Health

Corey L. Nagel
Oregon Health & Science University

Animesh Bhattacharya
West Bengal Public Health Engineering Department

Evan A. Thomas
Portland State University, evan.thomas@pdx.edu

Citation Details

Snoad, C., Nagel, C., Bhattacharya, A., & Thomas, E. (2017). The Effectiveness of Sanitary Inspections as a Risk Assessment Tool for Thermotolerant Coliform Bacteria Contamination of Rural Drinking Water: A Review of Data from West Bengal, India. The American Journal Of Tropical Medicine And Hygiene, doi:10.4269/ajtmh.16-0322

Am. J. Trop. Med. Hyg., 96(4), 2017, pp. 976–983
doi:10.4269/ajtmh.16-0322

The Effectiveness of Sanitary Inspections as a Risk Assessment Tool for Thermotolerant Coliform Bacteria Contamination of Rural Drinking Water: A Review of Data from West Bengal, India

Christian Snoad,[1] Corey Nagel,[2] Animesh Bhattacharya,[3] and Evan Thomas[1,4]*

[1]*DelAgua Health, The Old Dairy, Marlborough, Wiltshire, United Kingdom;* [2]*OHSU/PSU School of Public Health, Oregon Health and Science University, Portland, Oregon;* [3]*West Bengal Public Health Engineering Department, Kolkata, India;* [4]*Department of Mechanical and Materials Engineering, Portland State University, Portland, Oregon*

Abstract. The use of sanitary inspections combined with periodic water quality testing has been recommended in some cases as screening tools for fecal contamination. We conducted sanitary inspections and tested for thermotolerant coliforms (TTCs), a fecal indicator bacteria, among 7,317 unique water sources in West Bengal, India. Our results indicate that the sanitary inspection score has poor ability to identify TTC-contaminated sources. Among deep and shallow hand pumps, the area under curve (AUC) for prediction of TTC > 0 was 0.58 (95% confidence interval [CI] = 0.53–0.61) and 0.58 (95% CI = 0.54–0.62), respectively, indicating that the sanitary inspection score was only marginally better than chance in discriminating between contaminated and uncontaminated sources of this type. A slightly higher AUC value of 0.64 (95% CI=0.57–0.71) was observed when the sanitary inspection score was used for prediction of TTC > 0 among the gravity-fed piped sources. Among unprotected springs (AUC = 0.48, 95% CI = 0.38–0.55) and unprotected dug wells (AUC = 0.41, 95% CI = 0.20–0.66), the sanitary inspection score performed more poorly than chance in discriminating between sites with TTC < 1 and TTC > 0. Aggregating over all source types, the sensitivity (true positive rate) of a high/very high sanitary inspection score for TTC contamination (TTC > 1 CFU/100 mL) was 29.4% and the specificity (true negative rate) was 77.9%, resulting in substantial misclassification of the sites when using the established risk categories. These findings suggest that sanitary surveys are inappropriate screening tools for identifying TTC contamination at water points.

INTRODUCTION

Water-related diseases continue to constitute a significant health burden globally and in rural India.[1–3] A strong association between thermotolerant coliforms (TTCs), a fecal indicator bacteria (FIB), and diarrhea disease has been identified in the literature.[4,5] In India, national regulations stipulate that all rural drinking water sources should be tested twice per year for FIB, with sanitary inspections (SIs) conducted at the same time as sample collection.

First introduced in 1991 and published in the World Health Organization (WHO) monitoring guidelines in 1993, SIs have become a common component of global water quality surveillance programs.[6–8] They were developed to provide a rudimentary comparable method for quantifying risk factors that can contribute to microbiological contamination of water sources. SIs include a simple visual assessment of typically around 10 risk factor questions, specific to the source type, which are answered with yes or no responses. Each risk factor question is weighted equally.[9] The sum of all the questions answered "yes" is the sanitary inspection score (SIS). The higher the SIS value the higher the category of risk. The SIS and FIB results can be grouped into risk categories and combined on a risk prioritization matrix.[8,10] An example of risk prioritization matrix is shown in Supplemental Table 1. In this example, values that fall within red blocks may be prioritized at a higher level by remediation authorities. Sanitary Inspections may be adapted to local contexts[7] such as modifying the minimum distance to a latrine based on local lithological conditions.[11]

In some cases, sanitary surveys have been presented as useful in predicting the presence of FIB contamination, including in guidance provided by the WHO, stating, "It is possible to assess the likelihood of fecal contamination of water sources by a sanitary survey. This is often more valuable than bacteriological testing alone, because a sanitary survey makes it possible to see what needs to be done to protect the water source, and because fecal contamination may vary, so a water sample only represents the quality of the water at the time it was collected."[6]

However, existing literature have identified a poor statistical correlation between sanitary survey score and presence of FIB such as TTC.[9,12] Previous studies have analyzed the effectiveness of SIs using varying methods. Some studies have looked at the relationship between the overall SIS and FIB concentration.[12–15] Others have analyzed the relationship between individual SI risk questions and FIB concentration using logistic regression.[12,16] Further studies applied the multivariate analysis sanitary hazard index to prioritize remedial actions and assessed the resulting impact.[7,17] Of the studies that assessed the relationship between the overall SIS and FIB, only one, using a sample size of nine wells, found a statistically significant association.[14] Other studies found either no or weak associations using either linear regression, odds ratio (OR) or Spearman's rank.[12,13,15]

Notably, no study to date has specifically evaluated the discriminatory performance of the SI as a method to screen water sources for potential fecal contamination as measured by TTC. The distinction between statistical association and predictive performance is an important one, because the presence of a statistically significant relationship does not guarantee that the measure will prove useful in discriminating between cases with and without the target outcome. Therefore, in addition to examining the statistical association between the results of the SI and

---

*Address correspondence to Evan Thomas, Portland State University, 1930 SW 4th Ave., Suite 400, Portland, OR 97201. E-mail: evan.thomas@pdx.edu

microbiological testing, this study assesses the accuracy of the SI to identify TTC-contaminated water sources.

MATERIALS AND METHODS

DelAgua Health, a UK-based social enterprise, was contracted by the West Bengal Public Health Engineering Department (WBPHED) to coordinate and manage the testing of 7,317 unique water sources as part of the State's annual routine testing program. The project involved testing for 14 parameters, conducting SIs and creating onsite community awareness. All the data used in this analysis were collected by this project.

Testing was carried out by 18 operators selected and provided by the WBPHED, 13 from PHED laboratories and 5 from laboratories run by nongovernmental organizations. In addition, DelAgua Health provided three mobile laboratory technicians for extra field support. All 21 operators were trained in conducting SIs and microbiological analysis. Of the operators, 20 hold relevant degree-level qualifications.

Target area. In support of a state-wide annual testing routine, water quality laboratories were operated in 18 districts in West Bengal, India. Each district tested at least 400 water sources within a 6-month period. The exact locations and water points to be tested were selected by the district PHED executive engineer, assistant engineer, and chemist. The area selected targeted villages that had not yet been tested within the annual testing program, thus forming a crude convenience sample. Figure 1 shows in blue the administrative blocks in West Bengal where testing was conducted. Darker shading approximates the block-level density of water points tested.

Sanitary surveys. The SI forms used for the program were provided by the Government of India Uniform Drinking Water Quality Monitoring Protocol (UDWQMP).[18] A different SI form, each with yes/no questions, was used for each of the water source types (Supplemental Table 2). Water source type was determined by the test operator, in line with the UDWQMP guidelines. As there is no formal consistent definition to determine whether a hand pump is deep or shallow, this was determined by the test operator based on their local knowledge and asking the local community.

The SIs were conducted on-site before collecting a sample for microbiological analysis. The forms were filled out directly into a smartphone form application designed by DelAgua through an enhanced commercial version of the Open Data Kit (ODK) (www.doforms.com). The form presented to the user loaded automatically depending on the type of water source selected.

Microbiological analysis. The TTCs were used as the FIB and are recognized by the Indian standard methods as an acceptable alternative to testing for *Escherichia coli* directly.[19] It has been reported that in nontropical climates at least 95% of TTC are *E. coli*.[8,20] However, in tropical environments, TTC may originate from non-fecal sources or multiply within certain tropical waters, thus overestimating the fecal contamination risk in some cases.[20–23]

The analysis of TTC bacteria was carried out using the Indian standard membrane filtration method with membrane lauryl sulfate broth growth media.[19] The number of colony forming units (CFU) was enumerated by counting the yellow colonies after an incubation at 44 ± 0.5°C for 16–18 hours. Incubation was performed in a DelAgua brand dual incubator portable test kit, with one incubator calibrated to 37 ± 0.5°C and the other to 44 ± 0.5°C. The samples were collected in line with Indian standards, but modified for a 2-minute initial purging for hand pumps. Each mobile laboratory system was equipped with cool bags and ice packs. Where freezing of ice packs was not possible, samples were kept in cool bags and processed as soon as possible. The samples were processed and incubation started between 2 and 7 hours from sample collection. Each laboratory conducted standard sterilization methods for sample bottles and reusable aluminum plates, using either an autoclave or pressure cooker with heating plate.

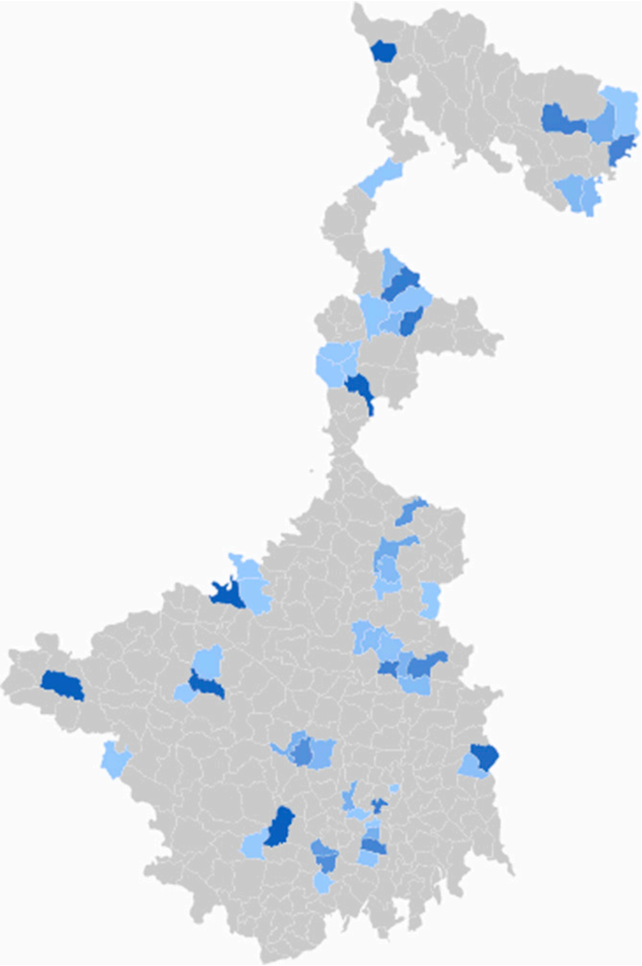

FIGURE 1. Water point locations tested in West Bengal administrative blocks.

A stack blank testing was conducted each day for each bottle of growth media used to verify that it had not become contaminated, as demonstrated by a color change on an incubated wetted media pad. A manifold blank testing, carried out with 100 mL of distilled water filtered through a filter paper, was targeted to be carried out as a negative control for 100% of samples before sample filtration to confirm successful sterilization of the filtration manifold. A duplicate split sample approach was used for 14% of samples to assess operator and process precision and provide a positive control. A duplicate set of plates were considered normal, if the second plates count was within a 95% confidence limit of the first, assuming a Poisson distribution of bacteria in the water.[1] Of the duplicate tests, 98% were considered normal. Further positive controls through *E. coli* testing and negative controls with *Enterobacter aerogenes* species were not conducted as part of this program.

Test results were recorded into the custom smartphone application and photos of the plates were taken. The data were transmitted to a secure server via the cellular data networks and processed and displayed on a project website dashboard. The incoming results were checked daily by the program manager and data analyst including verifying plate counts.

Data analysis method. Results of the SI conducted at each water source were scored to yield the SIS and sites were assigned a SIS risk category. Each "yes" answer scored a 1, and each "no" a 0. A total score indicated a SIS risk category. On a 10-question survey, categories are assigned as low risk (0–2), intermediate risk (3–5), high risk (6–8), very high risk (9–10), an established WHO scoring criteria.[10] The number and nature of survey questions differed between water source type (Supplemental Table 2). In the case of surveys with 8, 9, and 11 questions, we proportioned the scoring to match the 10-question structure. Raw TTC counts were collapsed into categories based on WHO guidelines: < 1 CFU/100 mL—in conformity with WHO guidelines (A category, Supplemental Table 1), 1–10 CFU/100 mL—low risk (B), 11–100 CFU/100 mL—intermediate risk (C), 101–1,000 CFU/100 mL—high risk (D), > 1,000 CFU/100 mL—very high risk (E). The results of the SI and microbiological testing were combined using the WHO risk prioritization matrix to calculate each site's prioritization for remedial action. An example of risk prioritization matrix is shown in Supplemental Table 1.[10]

Logistic regression was used to test the association between the results of the SI and the probability of site TTC contamination. We dichotomized the raw TTC count to create three binary indicators of TTC contamination: TTC < 1 CFU/100 mL versus TTC > 0 CFU/100 mL, TTC $\leq$ 10 CFU/100 mL versus TTC > 10 CFU/100 mL, and TTC $\leq$ 100 CFU/100 mL versus TTC > 100. We regressed each of these binary indicators against the raw SIS to examine whether any observed associations varied in relation to the magnitude of TTC contamination. Separate models were fitted for each water point type. We also fit bivariate logistic regression models to assess the association of each SI item with site contamination. The parameter estimates from fitted models were exponentiated to yield ORs. Significance tests and 95% confidence intervals (CIs) were calculated using robust standard errors to account for village-level clustering.

Next, we evaluated the discriminatory ability of the SI when used as a method of screening water points for TTC contamination. An effective screening method should have both high sensitivity and high specificity to predict the target condition.[24] In this study, sensitivity, also referred to as the true positive rate, quantifies the ability of the SI to correctly identify a contaminated water point. It was calculated as TP/(TP + FN), where TP are true positives (the number of contaminated water points with a SIS at or above a prespecified threshold) and FN are false negatives (the number of contaminated water points with a SIS below that same threshold). Specificity, or true negative rate, on the other hand, provides a metric of the ability of the SI to correctly identify an uncontaminated water point. It was calculated as TN/(TN + FP), where TN are true negatives (the number of uncontaminated water points with an SIS below a prespecified threshold) and FP are false positives (the number of uncontaminated water points with a SIS at or above that same threshold).

In a screening test with a range of possible scores, the sensitivity and specificity are dependent on the cut point chosen to distinguish cases with the condition from cases without the condition. The standard method of evaluating the overall discriminatory ability of a continuous screening test is the receiver operator characteristic (ROC) curve. An ROC curve plots the true positive rate (sensitivity) against the false positive rate (1 specificity) for each possible cut point (e.g., for each possible SIS) over the response range of a given screening test. The area under the ROC curve (AUC) provides a global metric of the measure's accuracy in predicting the outcome. An AUC of 1.0 indicates perfect prediction, whereas an AUC of 0.5 indicates a screening test that is no better than chance.[24] In addition, ROC curves are useful in identifying the specific threshold value of a screening test that maximizes sensitivity and specificity.

Using the raw SIS, we grouped water points by type and constructed ROC curves for each of the binary FIB cut points described earlier (TTC > 0, TTC > 10, and TTC > 100). We used a semi-parametric approach to calculate the AUC and estimated pointwise 95% CIs for the ROC curve using bootstrap resampling (5,000 replications).[25] We accounted for village-level clustering in the resampling procedure. In addition, we calculated the sensitivity and specificity, with corresponding bootstrap 95% CIs, of a "high-risk" SIS to identify sites with contamination levels exceeding each of the binary FIB cut points. Finally, we tested whether the discriminatory ability of the SI could be improved by removing items that were not significantly associated with site contamination in bivariate analyses and reverse coding items when the observed association was in the opposite direction from that specified in the original SI form. We recalculated the AUC for each comparison using this revised scoring, and recalculated the sensitivity (true positive rate) and specificity (true negative rate) of a high-risk score, defined as positive responses to at least 50% of inspection items (consistent with WHO categorization). To assess for potential bias in the point estimate of the AUC related to the choice of estimation method (semi-parametric versus nonparametric), we repeated all ROC analyses using nonparametric estimation methods and compared the AUC estimates.[25,26] In each case, the semi-parametric AUC estimate was within two percentage points of the corresponding nonparametric estimate, indicating that there was no bias in AUC results related to our choice of estimator. All statistical analyses were conducted using Stata 14 (Stata Corporation, College Station, TX).

RESULTS

A total of 7,317 water sources were tested by the project. Stack and/or manifold blank testing indicated that 48 sites had contaminated test samples and results from these sites were discarded. In addition, 27 of the water sources consisted of various uncommon source types and were not

TABLE 1

Sanitary inspection, microbiological testing, and WHO risk prioritization by source type

	Deep hand pump (*N* = 5,126)	Shallow hand pump (*N* = 1,347)	Unprotected spring (*N* = 381)	Gravity-fed piped (*N* = 356)	Unprotected dug well (*N* = 32)	Total (*N* = 7,242)
Sanitary inspection*						
Low risk, *N* (%)	2,148 (41.9)	190 (14.1)	175 (45.9)	326 (91.6)	21 (65.6)	2,860 (39.5)
Intermediate risk, *N* (%)	1,999 (39.0)	428 (31.8)	206 (54.1)	26 (7.3)	6 (18.8)	2,665 (36.8)
High risk, *N* (%)	899 (17.5)	612 (45.4)	0 (0)	4 (1.1)	4 (12.5)	1,519 (21.0)
Very high risk, *N* (%)	80 (1.6)	117 (8.7)	0 (0)	0 (0)	1 (3.1)	198 (2.73)
TTC						
< 1, *N* (%)	4,029 (78.6)	1,107 (82.2)	225 (59.1)	265 (74.4)	12 (37.5)	5,638 (77.9)
1–10, *N* (%)	646 (12.6)	37 (2.8)	31 (8.1)	34 (9.6)	1 (3.1)	749 (10.3)
11–100, N (%)	219 (4.3)	89 (6.6)	51 (13.4)	27 (7.6)	7 (21.9)	393 (5.4)
101–1,000, *N* (%)	171 (3.3)	43 (3.2)	56 (14.7)	23 (6.5)	6 (18.8)	299 (4.1)
> 1,000, *N* (%)	61 (1.2)	71 (5.3)	18 (4.7)	7 (2.0)	6 (18.8)	163 (2.3)
WHO risk prioritization						
No action (%)	11	6	0	39	16	11
Low action (%)	55	26	67	42	6	49
Higher action (%)	28	37	14	12	28	28
Urgent action (%)	6	31	19	7	50	12

TTC = thermotolerant coliforms; WHO = World Health Organization.

*On a 10-question survey, categories are assigned as low risk (0–2), intermediate risk (3–5), high risk (6–8), very high risk (9–10), an established WHO scoring criteria.[10] The number and nature of survey questions differed between water source type (Supplementary Material B). In the case of surveys with 8, 9, and 11 questions, we proportioned the scoring to match the 10-question structure.

included in the analysis. This yielded a final sample of 7,242 water points, of which 5,126 were deep hand pumps, 1,347 were shallow hand pumps, 381 were unprotected springs, 356 were piped supplies, and 32 were unprotected dug wells.

SI of the water points resulted in 23.7% being categorized as high/very high risk of TTC contamination based on the established SIS threshold (Table 1). The results of the SIs varied considerably across water source types. Shallow hand pumps had the highest proportion of sites that were deemed high/very high risk (54.1%), followed by deep hand pumps (19.1%) and unprotected wells (15.6%). Notably, no unprotected springs had SISs indicative of high/very high risk of contamination. Gravity-fed piped sources had the highest proportion of sites (91.6%) with SISs in the low-risk category.

Microbiological testing (Table 1) revealed that 22.1% of sites had evidence of TTC contamination (TTC > 0 CFU/100 mL). About 11.8% of sites had TTC counts greater than 10 CFU/100 mL and 6.4% had TTC counts > 100 CFU/100 mL. The protected source types had the lower rates of TTC contamination compared with the unprotected source types. The lowest prevalence of TTC contamination (17.8%) was found among shallow hand pumps, notable given the high proportion of shallow hand pumps with SISs in the high/very high-risk category. Of the protected source types, the highest proportion of TTC-contaminated sites was among the gravity-fed piped sources (25.6%). The proportion of TTC-contaminated sites among unprotected springs was 40.9%, none of which had SIS at or exceeding the high-risk threshold. The highest prevalence of TTC contamination was found among unprotected dug wells (62.5%), also notable given the low number of sites of this type with high/very high-risk SISs (15.6%).

The results of logistic regression models of the association between the raw SIS and the presence of TTC are presented in Table 2. We observed significant associations between the SIS and the probability of a TTC count > 0 among deep hand pumps (OR = 1.16, 95% CI = 1.10–1.22), shallow hand pumps (OR = 1.11, 95% CI = 1.04–1.19), and gravity-fed piped supplies (OR = 1.46, 95% CI = 1.18–1.80). There was no significant relationship between the SIS and a TTC count of > 0 among unprotected springs (OR = 0.96, 95% CI = 0.67–1.39) or unprotected dug wells (OR = 0.92, 95% CI = 0.69–1.21). There were no significant associations between raw SIS and the probability of a TTC count > 10 or TTC > 100 for any of the water source types.

The association between each SI item and the probability of site contamination is presented in Supplemental Table 2. Across the water source types, the SI among deep hand pumps yielded the greatest number of significant items, with eight of the nine items significantly associated with a TTC count > 0. Of these items, five were consistently associated with site contamination across the specified TTC thresholds (TTC >10, TTC >100, see Supplemental Table 2, deep hand pumps HP1, HP2, HP4, HP5, and HP6). Unexpectedly, the two items regarding the proximity and location of latrines were negatively associated with the likelihood of site contamination. We observed reduced odds of

TABLE 2

Logistic regression of water source contamination on sanitary inspection score by water source type

	TTC > 0		TTC > 10		TTC > 100	
Source type	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Deep hand pump	1.16 (1.10–1.22)	< 0.001	1.09 (0.99–1.20)	0.087	0.99 (0.85–1.16)	0.931
Shallow hand pump	1.11 (1.04–1.19)	0.003	1.06 (0.99–1.14)	0.093	1.07 (0.98–1.18)	0.127
Unprotected spring	0.96 (0.67–1.39)	0.846	0.99 (0.64–1.54)	0.969	0.86 (0.66–1.13)	0.289
Gravity-fed piped supply	1.46 (1.18–1.80)	< 0.001	1.15 (0.92–1.43)	0.219	1.06 (0.78–1.46)	0.709
Unprotected dug well	0.92 (0.69–1.21)	0.534	0.87 (0.67–1.11)	0.261	0.823 (0.62–1.10)	0.182

CI = confidence interval; OR = odds ratio; TTC = thermotolerant coliforms.

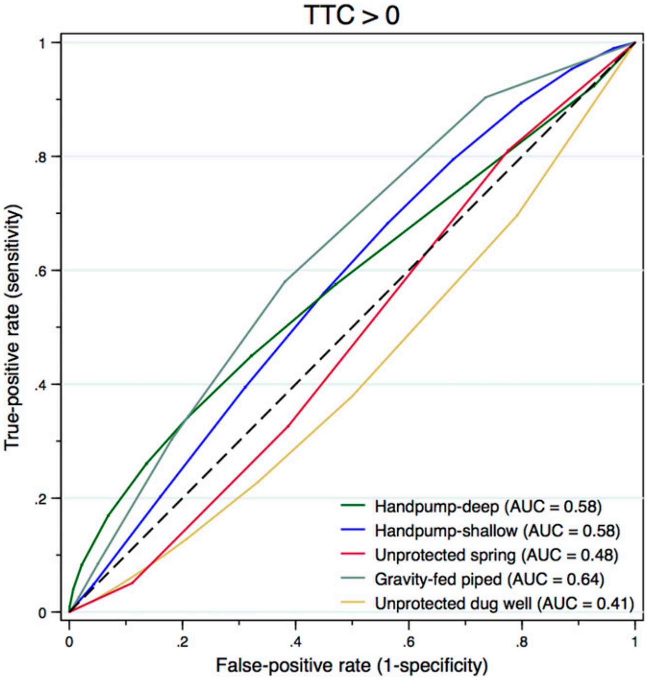

FIGURE 2. Receiver operator characteristic curves and area under the curve (AUC) values for prediction of thermotolerant coliforms (TTC) contamination using the sanitary inspection score for each water source type for TTC > 0. The dashed line (AUC = 0.5) represents a non-discriminating test; that is, one that performs no better than chance. Values above dashed line indicate better than chance predictive ability and below line indicate worse than chance.

contamination (TTC > 0) among sites with a latrine within 10 m (OR = 0.65, 95% CI = 0.045–0.093, *P* = 0.018) and with a latrine on higher ground (OR = 0.59, 95% CI = 0.41–0.84, *P* = 0.001). These associations were significant across contamination thresholds. Among shallow hand pumps, six of the nine hand pump items were associated with significantly greater odds of the site having a TTC count > 0. There was no observed association between the size of the cement floor around the well or the proximity or location of the nearest latrine and having a TTC count > 0. Only three of the survey items (cracks or defects on the cement floor, ponding on the cement floor, and priming required during the dry season) were significantly associated with a TTC count > 10 among shallow hand pumps, and no item was significantly associated with a TTC > 100.

Among piped supplies, two of the 10 SI items were associated with site TTC > 0. Sites with unchlorinated water in the reservoir (OR = 2.32, 95% CI = 1.13–4.76, *P* = 0.22) or with < 0.2 ppm free residual chlorine in the principal distribution pipes (OR = 2.41, 95% CI = 1.28–4.54, *P* = 0.006) had more than twice the odds of TTC > 0. Unchlorinated reservoir water was not significantly related with TTC > 10 or TTC > 100. Low levels of residual chlorine in the principal distribution pipes were more strongly associated with higher thresholds of site contamination. We were unable to model the association between the item querying leaks in the distribution system and TTC contamination, as there was only one positive response to this item, although we note that microbial testing at this site did indicate the presence of TTC contamination.

Two of the eight factors included in the SI of unprotected springs had no positive responses (latrine upstream of the spring and contaminant silt or animal excreta observed in the spring box). In addition, all the three sites observed to have unsanitary overflow pipes had TTC > 10, preventing calculation of ORs for contamination thresholds TTC > 0 and TTC > 10. Of the remaining five items with estimable odds ratios, none were associated with TTC >0 or TTC >10. An unsanitary overflow pipe was associated with significantly greater odds (OR = 8.5, 95% CI = 2.06–35.13, *P* = 0.003) of TTC > 100, as was an absent or nonfunctional surface water diversion ditch (OR = 2.57, 95% CI = 1.3–5.09, *P* = 0.007). Surprisingly, unfenced sites were significantly less likely than fenced sites to have TTC > 100 (OR = 0.58, 95% CI = 0.37–0.91, *P* = 0.019).

Among dug wells, there were no sites with positive responses for the items “Is the nearest latrine on higher ground?” or “Are there cracks/defects in the cement floor?” None of the items were significantly associated with TTC contamination at any of the defined thresholds.

The ROC curves of the SIS predicting TTC contamination are shown in Figures 2–4. Across source types and TTC contamination thresholds, the SIS was, at best, a poor predictor of site contamination. Among deep and shallow hand pumps, the AUC for prediction of TTC > 0 was 0.58 (95% CI = 0.53–0.61) and 0.58 (95% CI = 0.54–0.62), respectively, indicating that the SIS was only marginally better than chance in discriminating between contaminated and uncontaminated sources of this type. A slightly higher AUC value of 0.64 (95% CI = 0.57–0.71) was observed when the

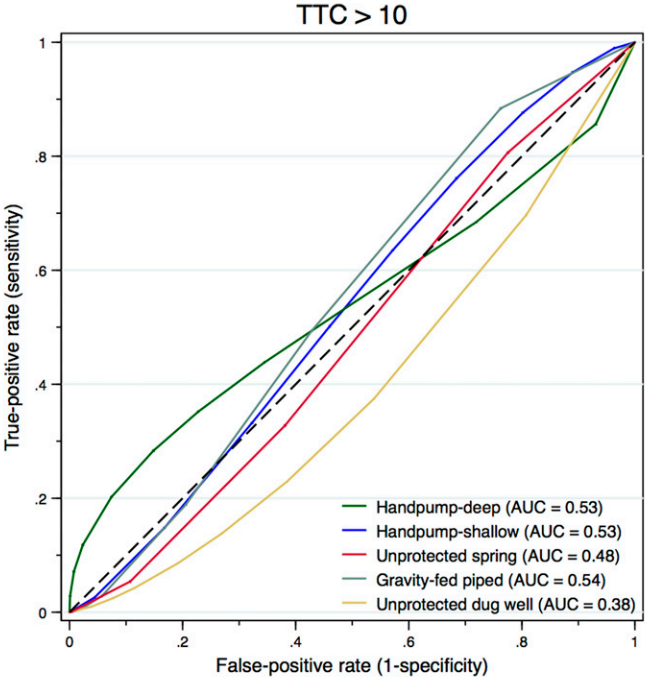

FIGURE 3. Receiver operator characteristic curves and area under the curve (AUC) values for prediction of thermotolerant coliforms (TTC) contamination using the sanitary inspection score for each water source type for TTC > 10. The dashed line (AUC = 0.5) represents a non-discriminating test; that is, one that performs no better than chance. Values above dashed line indicate better than chance predictive ability and below line indicate worse than chance.

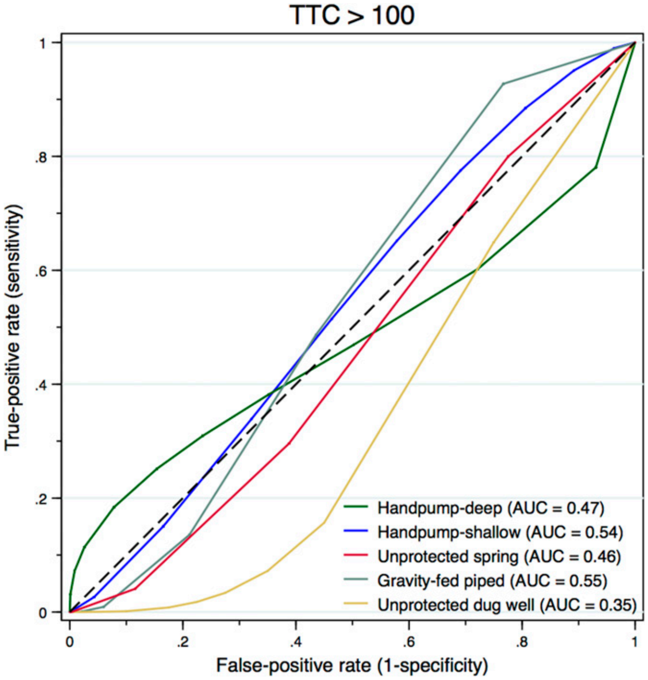

FIGURE 4. Receiver operator characteristic curves and area under the curve (AUC) values for prediction of thermotolerant coliforms (TTC) contamination using the sanitary inspection score for each water source type for TTC > 100. The dashed line (AUC = 0.5) represents a non-discriminating test; that is, one that performs no better than chance. Values above dashed line indicate better than chance predictive ability and below line indicate worse than chance.

SIS was used for prediction of TTC > 0 among the gravity-fed piped sources. Among unprotected springs (AUC = 0.48, 95% CI = 0.38–0.55) and unprotected dug wells (AUC = 0.41, 95% CI = 0.20–0.66), the SIS performed more poorly than chance in discriminating between sites with TTC < 1 and TTC > 0. Across source types, the performance of the SIS decreased when attempting to identify sites with greater levels of contamination (Table 3).

Using the established SIS categorization for each source type survey, the performance of a high/very high SIS to screen for TTC contamination is presented in Table 4. Aggregating over all source types, the sensitivity of a high/very high risk SIS indicating a true positive rate for sites with TTC > 0 was 29.4% (95% CI = 23.4–34.5), whereas the specificity was 77.9% (95% CI = 74.5–80.1). The performance varied considerably by source type. For all source types but shallow hand pumps, the sensitivity was less than 30%, indicating that more than two-thirds of contaminated sites had SISs indicating low/intermediate risk. No unprotected spring received a high/very high SIS, although 40.9% of unprotected springs had TTC > 0. Similarly, only three of the 91 gravity-fed piped sources with TTC > 0 had high/very high SISs, resulting in a sensitivity of 3.3 (95% CI = 0.9–8.3). Among shallow hand pumps, the sensitivity was 66.3% (58.5–73.6), but this was accompanied by a specificity of 48.5% (42.5–54.6). Consistent with the results of the ROC analysis, there were only modest differences in sensitivity (true positive rate) and specificity (true negative rate) at higher levels of TTC contamination.

We achieved little improvement in discriminatory ability of the SI among hand pumps and unprotected springs when the scoring was limited to the subset of items observed to have statistically significant associations with source quality (Supplemental Table 3). The notable exception was among gravity-fed piped supplies. The single SI item that was significantly associated with TTC > 10 and TTC > 100, the presence of < 0.2 ppm free residual chlorine in the principal distribution pipes, displayed better ability than the full SI to identify sites with moderate (AUC = 0.66 versus 0.54) and high (AUC = 0.78 versus 0.55) contamination.

DISCUSSION

The use of SIs has been recommended by the WHO as a simple and cost-effective method of identifying microbiological risks to water quality, although the evidence to support this recommendation has been mixed. This study of 7,242 water points in West Bengal found that the SIS showed poor ability to identify TTC-contaminated sources. This finding was consistent across the five source types represented in the study population and largely invariant to increasing the threshold for defining sites as contaminated. For hand pumps and gravity-fed piped supplies, ROC analysis revealed that the SIS was marginally better than chance in predicting whether water points were TTC contaminated. Among unprotected springs and unprotected dug wells, the SIS performed worse than chance at predicting TTC contamination whether it was defined as a TTC count of > 0, > 10, or > 100 CFU/100 mL.

Aggregating over all source types, the sensitivity (true positive rate) of a high/very high SIS for TTC contamination (TTC > 0 CFU/100 mL) was 29.4% and the specificity (true negative rate) was 77.9%. This resulted in substantial misclassification of the sites when using the established risk categories for the SIS. For every water point with confirmed TTC contamination that was correctly labeled as high risk based on the SIS, 2.6 contaminated sites had a score below the high-risk threshold. Similarly, for every

TABLE 3

Area under the curve by water source type and TTC count threshold

	TTC > 0	TTC > 10	TTC > 100
Source type	AUC (95% CI)	AUC (95% CI)	AUC (95% CI)
Deep hand pump	0.58 (0.53–0.61)	0.53 (0.46 –0.59)	0.47 (0.37–0.58)
Shallow hand pump	0.58 (0.54–0.63)	0.53 (0.48–0.57)	0.54 (0.47–0.59)
Unprotected spring	0.48 (0.38–0.55)	0.48 (0.37–0.56)	0.46 (0.38–0.53)
Gravity-fed piped supplies	0.64 (0.57–0.71)	0.54 (0.48–0.64)	0.55 (0.44–0.69)
Unprotected dug well	0.41 (0.20–0.66)	0.38 (0.22–0.60)	0.35 (0.17–0.50)

AUC = area under curve; CI = confidence interval; TTC = thermotolerant coliforms.

TABLE 4
Sensitivity (true positive rate) of a high/very high sanitary inspection score

Source type	True positive, *N* (%)	True negative, *N* (%)	False positive, *N* (%)	False negative, *N* (%)	Sensitivity % (true positive rate) (95% CI)	Specificity % (true negative rate) (95% CI)
TTC > 0						
Deep hand pump	306 (6.0)	3,356 (65.5)	673 (13.1)	791 (15.4)	27.9 (21.9–34.0)	83.3 (79.5–86.6)
Shallow hand pump	159 (11.8)	537 (39.9)	570 (42.3)	81 (6.0)	66.3 (58.5–73.6)	48.5 (42.5–54.6)
Unprotected Spring	0 (0)	225 (59.1)	0 (0)	156 (40.9)	0 (n/a)	100 (n/a)
Gravity-fed piped supplies	3 (0.8)	264 (74.2)	1 (0.3)	88 (24.7)	3.3 (0.9–8.3)	99.6 (99.2–99.7)
Unprotected dug well	3 (9.4)	10 (31.3)	2 (6.3)	17 (53.1)	15.0 (4.8–30.0)	83.3 (63.6–92.9)
All source types	471 (6.5)	4,392 (60.7)	1,133 (15.7)	1,246 (17.2)	29.4 (23.4–34.5)	77.9 (74.5–80.1)
TTC > 10						
Deep hand pump	128 (2.5)	3,824 (74.6)	851 (16.6)	323 (6.3)	28.4 (20.7–36.3)	81.8 (77.4–85.3)
Shallow hand pump	129 (9.6)	544 (40.4)	600 (44.5)	74 (5.5)	63.6 (55.6–71.1)	47.6 (41.3–53.7)
Unprotected spring	0 (0)	256 (67.2)	0 (0)	125 (32.8)	0 (n/a)	100 (n/a)
Gravity-fed piped supplies	2 (0.6)	297 (83.4)	2 (0.6)	55 (15.5)	3.5 (1.2–9.8)	99.3 (98.2–99.7)
Unprotected dug well	2 (6.3)	10 (31.3)	3 (9.4)	17 (53.1)	10.53 (4.6–23.8)	76.9 (46.2–92.9)
All source types	261 (3.6)	4,931 (68.1)	1,456 (20.1)	594 (8.2)	30.5 (24.2–37.0)	77.2 (73.5–80.4)
TTC > 100						
Deep hand pump	56 (1.1)	3,971 (77.5)	923 (18.0)	176 (3.4)	24.1 (15.2–33.5)	81.1 (76.7–84.7)
Shallow hand pump	76 (5.6)	580 (43.1)	653 (48.5)	38 (2.8)	66.7 (55.2–77.2)	47.0 (41.3–53.1)
Unprotected spring	0	307 (80.6)	0	74 (19.4)	0 (n/a)	100 (n/a)
Gravity-fed piped supplies	1 (0.3)	323 (90.7)	3 (0.8)	29 (8.2)	3.3 (1.5–7.4)	99.1 (97.7–99.7)
Unprotected dug well	0 (0.00)	15 (46.9)	5 (15.6)	12 (37.5)	0 (n/a)	75.0 (57.9–90.9)
All source types	133 (1.8)	5,196 (71.8)	1,584 (21.9)	329 (4.6)	28.8 (21.2–37.1)	76.6 (73.0–80.0)

CI = confidence interval; TTC = thermotolerant coliforms.

contaminated site with a high-risk SIS, there were 2.4 uncontaminated sites that were also labeled as high risk.

Although there were only minimal differences in the AUC values between the five water source types represented in the study sample, the accuracy of a high/very high SIS differed considerably across source types. For example, the highest sensitivity (true positive rate, 66.3%) of a high/very high score for any TTC contamination (TTC > 0 CFU/100 mL) was observed among shallow hand pumps, although this was accompanied by a false positive rate of 51.5%. Among deep hand pumps, the sensitivity of a high/very high score was reduced to 27.9%, albeit with a reduction in the rate of false positives to 16.7%. Of the 91 gravity-fed piped sources that had confirmed TTC contamination, only three had high/very high SISs (sensitivity = 3.3%) and none of the unprotected springs were classified as high/very high risk, although 41% had confirmed TTC contamination. Notably, inspection of the ROC curves for all source types failed to identify a SIS cut point that could be used to classify contaminated sites with both high sensitivity and high specificity, reflecting the poor separation in the distribution of SISs between sites with and without TTC contamination. These results may suggest that the sanitary survey questions are not sufficient for identifying contamination risk at these source types.

It should be noted that we observed statistically significant associations between the SIS and the presence of TTC contamination (TTC > 0 CFU/100 mL) among three of the five source types included in the study sample (deep and shallow hand pumps and gravity-fed supplies). Given the poor performance of the SI to accurately discriminate between contaminated and uncontaminated sites, this finding may appear counterintuitive. However, it simply reflects the often unappreciated limitation of statistical measures of association to assess predictive performance.[27] Pepe and others[28] have described the relationship between odds ratios and discriminatory accuracy and have shown that screening tools with adequate discrimination must have observed odds ratios far larger than those seen in this study (or indeed in most epidemiological studies). This underscores the need to use appropriate analytic methods to evaluate the accuracy of measures intended for screening purposes.

This study has important limitations that warrant mention. First, microbiological testing was only conducted at a single time point for each water point, concurrent with the SI. Thus, this study was unable to capture temporal variation in water quality or examine whether high SIS resulted in increased risk of contamination over time. In addition, a single TTC sample at a water point may not identify contamination in soil, sand, or biofilms, nor does it account for microbial deactivation from chlorine or other disinfection that may not address all TTC contamination. We also acknowledge that concentrations of TTC are known to vary by orders of magnitude between samples at the same water points,[29] including between and during weather events and changes wherein runoff can dramatically impact TTC concentrations.[10,14] Future studies may consider collecting several water samples at separate times from each water point. However, the results of this study are consistent with the majority of previous studies that found no association between the SIS and likelihood of microbiological contamination. Second, the small sample size for unprotected dug wells (*N* = 32) limits the generalizability of the findings for this source type. Finally, we were not able to test for the presence of specific organisms, such as *E. coli*. Previous studies have suggested that *E. coli* is a more reliable indicator of pathogenic contamination of water sources than TTC.[30] However, TTC is widely used in many settings and is recognized by the Indian standard methods as an acceptable indicator of fecal contamination.[19]

Although common sense suggests that visually identifying contamination risks is an important component of efforts to ensure water quality and reduce waterborne disease, we found that the SIS had poor predictive performance for identifying water points with microbiological contamination. These results suggest that the SI, as it is most commonly implemented, is an ineffective strategy of screening water points for targeted microbiological testing or remediation.

Received April 25, 2016. Accepted for publication January 5, 2015.

Note: Supplemental tables appear at www.ajtmh.org.

Acknowledgments: We gratefully acknowledge the contributions of professional chemists, engineers, and water quality technicians that conducted the data collection and laboratory analysis, including those employed by the WBPHED and DelAgua Health Limited.

Financial support: The data collection and analysis carried out for this article was funded by the West Bengal Public Health and Engineering Department and DelAgua Health Limited.

Disclosures: Christian Snoad and Evan Thomas were compensated consultants to DelAgua Health, commissioned to carry out the testing program through a contract from the West Bengal Public Health Engineering Department headed by Animesh Bhattacharya.

Authors' addresses: Christian Snoad, DelAgua Health, The Old Dairy, Marlborough, Wiltshire, United Kingdom, E-mail: christian.snoad@delagua.org. Corey Nagel, OHSU/PSU School of Public Health, Oregon Health and Science University, Portland, OR, E-mail: nagelc@ohsu.edu. Animesh Bhattacharya, West Bengal Public Health Engineering Department, Kolkata, India, E-mail: animeshbhat@gmail.com. Evan Thomas, Department of Mechanical and Materials Engineering, Portland State University, Portland, OR, E-mail: evan.thomas@pdx.edu.

REFERENCES

1. World Health Organization, 2015. *Drinking Water, Fact sheet no.391*. Available at: http://www.who.int/mediacentre/factsheets/fs391/en/. Accessed September 5, 2015.
2. Central Bureau of Health Intelligence, Directorate General of Health Services, Ministry of Health and Family Welfare, 2013. *National Health Profile 2013*. New Delhi, India: Government of India.
3. UNICEF/WHO, 2009. *Diarrhoea: Why Children Are Still Dying and What Can Be Done*. Geneva, Switzerland: WHO.
4. Hodge J, Chang HH, Boisson S, Collin SM, Peletz R, Clasen T, 2016. Assessing the association between thermotolerant coliforms in drinking water and diarrhea: an analysis of individual level data from multiple studies. *Environ Health Perspect.*, doi: 10.1289/EHP156.
5. Fewetrell L, Kauffman R, Kay D, Enanoria W, Haller L, Colford J, 2005. Water, sanitation, and hygiene interventions to reduce diarrhea in less developed countries: a systematic review and meta-analysis. *Lancet 5:* 42–52.
6. Wisner B, Adams J, 2003. *Environmental Health in Emergencies and Disasters: A Practical Guide*. Geneva, Switzerland: World Health Organization.
7. Lloyd BJ, Helmer R, 1991. *Surveillance of Drinking Water Quality in Rural Areas*. New York, NY: Logman Scientific and Technical, Co-published in the United States with John Wiley and Sons, Inc.
8. WHO, 1993. *Guidelines for Drinking-Water Quality, Vol. 1: Recommendation*, 2nd edition. Geneva, Switzerland: World Health Organisation.
9. Lloyd BJ, Bartram JK, 1991. Surveillance solutions to microbiological problems in water quality control in developing countries. 1991. *J. Water Science Technol 24:* 61–75.
10. WHO, 1997. *Guidelines for Drinking-Water Quality, Vol. 3: Surveillance and Control of Community Supplies*, 2nd edition. Geneva, Switzerland: World Health Organization.
11. ARGOSS, 2001. *Guidelines for Assessing the Risk to Groundwater from On-Site Sanitation*. Swindon, United Kingdom: British Geological Survey Commissioned Report, CR/01/142, 97. Available at: http://www.susana.org/_resources/documents/default/2-1926-argoss-manual.pdf.
12. Luby S, Gupta S, Sheikh M, Johnston R, Ram R, Islam S, 2008. Tubewell water quality and predictors of contamination in three flood-prone areas in Bangladesh. *J Appl Microbiol 105:* 1002–1008.
13. Barthiban S, Lloyd B, Maier M, 2012. Sanitary hazards and microbial quality of open dug wells in the Maldives Islands. *J Water Resource Prot 4:* 474–486.
14. Mushi D, Byamukama D, Kirscher A, Mach R, Brunner K, Farnleitner A, 2012. Sanitary inspection of wells using risk-of-contamination scoring indicates a high predictive ability for bacterial faecal pollution in the peri-urban tropical lowlands of Dar es Salaam, Tanzania. *J Water Health 10:* 236–243.
15. Parker A, Youlten R, Dillon D, Nussbaumer T, Carter R, Tyrell S, Webster J, 2010. An assessment of microbiological water quality of six water source categories in north-east Uganda. *J Water Health 8:* 550–560.
16. Howard G, Pedley S, Barrett M, Nalubega M, Johal K, 2003. Risk factors contributing to microbiological contamination of shallow ground water in Kampala, Uganda. *Water Res 37:* 3421–3429.
17. Lloyd BJ, Boonyakarnkul T, 1991. *Combined Assessment of Sanitary Hazards and Faecal Coliform Intensity for Rural Groundwater Supply Improvement in Thailand*. National conference on Geologic Resources of Thailand: Potential for Future Development. November17–24, 1992. Department of Mineral Resources, Bangkok, Thailand.
18. Ministry of Drinking Water and Sanitation, Government of India, 2013. *Uniform Drinking Water Quality Monitoring Protocol*. New Delhi, India: s.n. Ministry Of Drinking Water And Sanitation, Government Of India.
19. Bureau of Indian Standards, 1981. *Methods of Sampling and Microbiological Examination of Water (First Revision)*. IS 1622. New Delhi, India: s.n. Bureau of Indian Standards.
20. Byamukama D, Mach RL, Kansiime F, Manafi M, Farnleitner AH, 2005. Discrimination efficacy of faecal pollution detection in different aquatic habitats of a high altitude tropical country using presumptive coliform, *Escherichia coli* and *Clostridium perfringes* spores. *Appl Environ Microbiol 71:* 65–71.
21. Solo-Gabriele HM, Wolfert MA, Desmarais TR, Palmer CJ, 2000. Sources of *Escherichia coli* in a coastal subtropical environment. *Appl Environ Microbiol 66:* 230–237.
22. Ishii S, Sadowsky MJ, 2008. *Escherichia coli* in the environment: implications for water quality. *Microbes Environ 23:* 101–108.
23. WHO, 1996. *Guidelines for Drinking-Water Quality, Vol. 2: Health and Supporting Criteria*, 2nd edition. Geneva, Switzerland: World Health Organization.
24. Zou KH, O'Malley AJ, Mauri L, 2007. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation 115:* 654–657.
25. Alonzo TA, Pepe MS, 2002. Distribution-free ROC analysis using binary regression techniques. *Biostatistics 3:* 421–432.
26. Colak E, Mutlu F, Bal C, Oner S, Ozdamar K, Gok B, Cavusoglu Y, 2012. Comparison of semiparametric, parametric, and nonparametric ROC analysis for continuous diagnostic tests using a simulation study and acute coronary syndrome data. *Comput Math Methods Med 2012:* 2012.
27. Ware JH, 2006. The limitations of risk factors as prognostic tools. *N Engl J Med 355:* 2615–2617.
28. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P, 2004. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol 159:* 882–890.
29. Bain R, Cronk R, Wright J, Yang H, Skaymaker T, Bartram J, 2014. Fecal contamination of drinking-water in low- and middle-income countries: a systematic review and meta-analysis. *PLoS Med 11:* e1001644.
30. Leclerc HDAA, Mossel DAA, Edberg SC, Struijk CB, 2001. Advances in the bacteriology of the coliform group: their suitability as markers of microbial water safety. *Annu Rev Microbiol 55:* 201–234.