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*

Follow this and additional works at: https://pdxscholar.library.pdx.edu/mengin_fac

Part of the Materials Science and Engineering Commons, and the Mechanical Engineering Commons

Let us know how access to this document benefits you.

Citation Details


This Article is brought to you for free and open access. It has been accepted for inclusion in Mechanical and Materials Engineering Faculty Publications and Presentations by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.
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 Thomas¹,⁴⁺

¹DelAgua Health, The Old Dairy, Marlborough, Wiltshire, United Kingdom; ²OHSU/PSU School of Public Health, Oregon Health and Science University, Portland, Oregon; ³West Bengal Public Health Engineering Department, Kolkata, India; ⁴Department 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.⁵,¹⁰ 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 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 faecal 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 faecal 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 faecal 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). 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. It has been reported that in nontropical climates at least 95% of TTC are E. coli. 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.

The analysis of TTC bacteria was carried out using the Indian standard membrane filtration method with membrane lauryl sulfate broth growth media. 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.

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.\textsuperscript{1} Of the duplicate tests, 98% were considered normal.
Further positive controls through \textit{E. coli} testing and nega-
tive controls with \textit{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.

\textbf{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.\textsuperscript{10} The number and nature of survey questions dif-
fered between water source type (Supplemental Table 2). In
the case of surveys with 8, 9, and 11 questions, we propor-
tioned 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 micro-
biological testing were combined using the WHO risk pri-
oritization matrix to calculate each site’s prioritization for
remedial action. An example of risk prioritization matrix is
shown in Supplemental Table 1.\textsuperscript{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 ≤ 10 CFU/100 mL
versus TTC > 10 CFU/100 mL, and TTC ≤ 100 CFU/100 mL
versus TTC > 100. We regressed each of these binary indica-
tors against the raw SIS to examine whether any observed
associations varied in relation to the magnitude of TTC con-
tamination. 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 contami-
nation. The parameter estimates from fitted models were
exponentiated to yield ORs. Significance tests and 95% confi-
dence 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 tar-
get condition.\textsuperscript{24} In this study, sensitivity, also referred to as
the true positive rate, quantifies the ability of the SI to cor-
rectly 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 pre-
specified 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 pre-
specified 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.\textsuperscript{24} In addition, ROC curves are
useful in identifying the specific threshold value of a screen-
ing 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).\textsuperscript{25} We accounted
for village-level clustering in the resampling procedure.
In addition, we calculated the sensitivity and specificity, with
AUC curve results, using 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).

\textbf{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 con-
stituted of various uncommon source types and were not
protected dug wells (62.5%), while unprotected springs showed a high proportion of sites (91.6%) with SISs in the low-risk category. The highest proportion of TTC-contaminated sites was found among shallow hand pumps, 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 contamination 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 = 1.19–1.80). There was no significant relationship between the SIS and a TTC count > 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. We observed significant associations between raw SIS and the probability of site contamination across the specified TTC thresholds (TTC >10, TTC >100, see 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 1

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

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

### Table 2

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

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

**Table 1**

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

<table>
<thead>
<tr>
<th>Source type</th>
<th>Deep hand pump (N = 5,126)</th>
<th>Shallow hand pump (N = 1,347)</th>
<th>Unprotected spring (N = 381)</th>
<th>Gravity-fed piped (N = 356)</th>
<th>Unprotected dug well (N = 32)</th>
<th>Total (N = 7,242)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low risk, N (%)</td>
<td>2,148 (41.9)</td>
<td>190 (14.1)</td>
<td>175 (45.9)</td>
<td>326 (91.6)</td>
<td>21 (65.6)</td>
<td>2,860 (39.5)</td>
</tr>
<tr>
<td>Intermediate risk, N (%)</td>
<td>1,999 (39.0)</td>
<td>428 (31.8)</td>
<td>206 (54.1)</td>
<td>26 (7.3)</td>
<td>6 (18.8)</td>
<td>2,665 (36.8)</td>
</tr>
<tr>
<td>High risk, N (%)</td>
<td>899 (17.5)</td>
<td>612 (45.4)</td>
<td>0 (0)</td>
<td>4 (1.1)</td>
<td>4 (12.5)</td>
<td>1,519 (21.0)</td>
</tr>
<tr>
<td>Very high risk, N (%)</td>
<td>80 (1.6)</td>
<td>117 (8.7)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (3.1)</td>
<td>198 (2.73)</td>
</tr>
<tr>
<td>WHO risk prioritization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No action (%)</td>
<td>11</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>Low action (%)</td>
<td>55</td>
<td>26</td>
<td>67</td>
<td>42</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>Higher action (%)</td>
<td>28</td>
<td>37</td>
<td>14</td>
<td>12</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>Urgent action (%)</td>
<td>6</td>
<td>31</td>
<td>19</td>
<td>7</td>
<td>50</td>
<td>12</td>
</tr>
</tbody>
</table>

WHO = World Health Organization.

**Table 2**

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

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

CI = confidence interval; OR = odds ratio; TTC = thermotolerant coliforms.
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 = 0.022) 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 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.

**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.
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 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 SIS 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.

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

**Table 3**

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

**AUC** = area under curve; **CI** = confidence interval; **TTC** = thermotolerant coliforms.
have described the relationship between 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 \(E.\ \text{coli}\). Previous studies have suggested that \(E.\ \text{coli}\) is a more reliable indicator of pathogenic contamination of water sources than TTC.\(^{30}\) However, TTC is widely used in many settings as a more reliable 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

<table>
<thead>
<tr>
<th>Source type</th>
<th>True positive, (N)</th>
<th>True negative, (N)</th>
<th>False positive, (N)</th>
<th>False negative, (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTC &gt; 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deep hand pump</td>
<td>306 (6.0)</td>
<td>3,356 (65.5)</td>
<td>673 (13.1)</td>
<td>791 (15.4)</td>
</tr>
<tr>
<td>Shallow hand pump</td>
<td>159 (11.8)</td>
<td>537 (39.9)</td>
<td>570 (42.3)</td>
<td>81 (6.0)</td>
</tr>
<tr>
<td>Unprotected Spring</td>
<td>0 (0)</td>
<td>225 (59.1)</td>
<td>0 (0)</td>
<td>156 (40.9)</td>
</tr>
<tr>
<td>Gravity-fed piped supplies</td>
<td>3 (0.8)</td>
<td>264 (74.2)</td>
<td>1 (0.3)</td>
<td>88 (24.7)</td>
</tr>
<tr>
<td>Unprotected dug well</td>
<td>3 (9.4)</td>
<td>10 (31.3)</td>
<td>2 (6.3)</td>
<td>17 (53.1)</td>
</tr>
<tr>
<td>All source types</td>
<td>471 (6.5)</td>
<td>4,392 (60.7)</td>
<td>1,133 (15.7)</td>
<td>1,246 (17.2)</td>
</tr>
<tr>
<td>TTC &gt; 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deep hand pump</td>
<td>128 (2.5)</td>
<td>3,824 (74.6)</td>
<td>851 (16.6)</td>
<td>323 (6.3)</td>
</tr>
<tr>
<td>Shallow hand pump</td>
<td>129 (9.6)</td>
<td>544 (40.4)</td>
<td>600 (44.5)</td>
<td>74 (5.5)</td>
</tr>
<tr>
<td>Unprotected spring</td>
<td>0 (0)</td>
<td>256 (67.2)</td>
<td>0 (0)</td>
<td>125 (32.8)</td>
</tr>
<tr>
<td>Gravity-fed piped supplies</td>
<td>2 (0.6)</td>
<td>297 (83.4)</td>
<td>2 (0.6)</td>
<td>55 (15.5)</td>
</tr>
<tr>
<td>Unprotected dug well</td>
<td>2 (6.3)</td>
<td>10 (31.3)</td>
<td>3 (9.4)</td>
<td>17 (53.1)</td>
</tr>
<tr>
<td>All source types</td>
<td>261 (3.6)</td>
<td>4,931 (68.1)</td>
<td>1,456 (20.1)</td>
<td>594 (8.2)</td>
</tr>
<tr>
<td>TTC &gt; 100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deep hand pump</td>
<td>56 (1.1)</td>
<td>3,971 (77.5)</td>
<td>923 (18.0)</td>
<td>176 (3.4)</td>
</tr>
<tr>
<td>Shallow hand pump</td>
<td>76 (6.6)</td>
<td>580 (43.1)</td>
<td>653 (48.5)</td>
<td>38 (2.8)</td>
</tr>
<tr>
<td>Unprotected spring</td>
<td>0</td>
<td>307 (80.6)</td>
<td>0</td>
<td>74 (19.4)</td>
</tr>
<tr>
<td>Gravity-fed piped supplies</td>
<td>1 (0.3)</td>
<td>323 (80.7)</td>
<td>3 (0.8)</td>
<td>29 (8.2)</td>
</tr>
<tr>
<td>Unprotected dug well</td>
<td>0 (0.0)</td>
<td>15 (46.9)</td>
<td>5 (15.6)</td>
<td>12 (37.5)</td>
</tr>
<tr>
<td>All source types</td>
<td>133 (1.8)</td>
<td>5,196 (71.8)</td>
<td>1,584 (21.9)</td>
<td>329 (4.6)</td>
</tr>
</tbody>
</table>

\(95\% \text{ CI} = \) confidence interval; TTC = thermotolerant coliforms.

\(Sensitivity \% = 32) \) 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 \(E.\ \text{coli}\). Previous studies have suggested that \(E.\ \text{coli}\) is a more reliable indicator of pathogenic contamination of water sources than TTC.\(^{30}\) However, TTC is widely used in many settings as an acceptable indicator of fecal contamination.\(^{19}\)
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.


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: animesh.bhattacharya@gmail.com. Evan Thomas, Department of Mechanical and Materials Engineering, Portland State University, Portland, OR, E-mail: evan.thomas@pdx.edu.

REFERENCES


