International Journal of Environmental Science and Engineering Research (IJESER)

www.journal-ijeser.com
ISSN: 0976-3708 (print)
IJESER Vol 4(3):46-58, 2013

Towards establishing a human fecal contamination index in microbial source tracking

Yiping Cao[1], Charles Hagedorn[2], Orin C. Shanks[3], Dan Wang[4], Jared Ervin[5], John F. Griffith[1], Blythe A. Layton[1], Charles D. McGee[6], Timothy E. Riedel[7], Stephen B. Weisberg [1*]

[1] Southern California Coastal Water Research Project Authority, Costa Mesa, CA 92626 USA
[2] Department of Crop & Soil Environmental Sciences, Virginia Tech, Blacksburg, VA 24061 USA
[3] U.S. Environmental Protection Agency, Office of Research and Development, National Risk Management Research Laboratory, Cincinnati, OH 45268 USA
[4] Environmental and Water Studies, Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305 USA
[5] Earth Research Institute and Bren School of Environmental Science & Management, University of California, Santa Barbara, CA 93106 USA
[6] Orange County Sanitation District, Fountain Valley, CA 92708 USA
[7] Department of Civil and Environmental Engineering, University of California, Los Angeles, CA 90095 USA

ABSTRACT

The fecal indicator bacteria (FIB, such as Enterococcus) used to monitor recreational water quality do not differentiate human and animal fecal pollution, even though human fecal material represents a greater public health risk. Host-associated genetic markers that allow for source identification have been developed, but there is no agreed upon approach for integrating multiple samples exhibiting different marker signal strengths and varying levels of agreement among markers into an index that managers can use for prioritizing beaches with the greatest presence of human fecal contamination. As a first step towards developing such an index, we provided ten experts with a simulated dataset for 26 beaches where we systematically varied four factors: Enterococcus concentrations, frequency of detection for two human-associated markers, magnitude of the marker signal, and agreement between the markers. We then used the Delphi technique to establish consensus principles for how these factors should be used in ranking beaches with respect to human fecal contamination. The experts' initial ranking varied widely, but after three iterations of ranking and discussion, the experts converged on a consensus that: 1) frequency of samples that are positive for human-associated markers is of primary importance in ranking beaches; 2) magnitude of and consistency between the markers should be used to weigh marker frequency; and 3) general FIB data should receive the least weight. Using the expert's consensus, a conceptual mathematical algorithm is proposed to establish an index that consistently and transparently quantifies the relative probability of human fecal contamination at a beach.

Key words: *Microbial source identification, human fecal contamination, Delphi technique, pollution index, weight-of-evidence exercise*

*Corresponding author Email: stevew@sccwrp.org

1.0 INTRODUCTION

Recreational water quality is routinely monitored for fecal indicator bacteria (FIB), such as *Enterococcus* spp. and *E. coli*, as proxies for fecal contamination because they can be measured cheaper and faster than pathogens. Water bodies with FIB concentrations exceeding recreational water quality criteria (U.S. EPA 2012) are treated as a public health risk and management actions such as beach advisories and pollution remediation are typically implemented in response. However, FIB do not distinguish whether fecal contamination originates from human, animal or non-fecal sources. Human fecal material is considered a greater public health risk than non-human fecal material (Soller *et al.,* 2010) and it is desirable to prioritize sites for remediation based on the extent of human fecal contamination. The U.S. Environmental Protection Agency has even defined a quantitative microbial risk assessment (QMRA) process for developing alternative management strategies for beaches that have high FIB counts, but a low level of human fecal contamination (U.S. EPA 2012).

Many host-associated genetic markers that allow for fecal source identification have been developed over the last decade and recent method evaluation studies have demonstrated good sensitivity and specificity of these markers to their target hosts (Boehm *et al.,* 2013, Shanks *et al.,* 2010)). Studies have also illustrated how these marker assays can be used in combination with probabilistic approaches to detect a host-specific fecal contamination event in a particular water sample (Jenkins *et al.,* 2009, Kildare *et al.,* 2007, Lamendella *et al.,* 2009, Ryu *et al.,* 2012). Using host-associated marker and general fecal indicator measurement data allow estimation of contribution to total fecal pollution from different hosts (Stoeckel *et al.,* 2011, Wang *et al.,* 2010).

However, managers still lack an index that enables them to prioritize which beaches have the highest level of human fecal contamination. Establishing such a human fecal contamination index requires integration and prioritization of several factors, including frequency of human-associated MST marker detection, magnitude of the human-associated MST marker signal, consistency among MST markers when multiple markers are employed, and FIB (*Enterococcus*) concentration. These factors typically vary among the many samples collected to characterize conditions at a beach and mechanisms to facilitate appropriate incorporation of these factors are a prerequisite to such an index.

To begin the process of developing a human fecal contamination index for beaches, we used the Delphi technique (Linstone and Turoff 1975) to identify how experts in the field prioritize these factors. In this exercise, experts were provided a simulated dataset of 26 beaches where *Enterococcus* concentrations, frequency of detection for two human-associated MST markers, magnitude of the marker signal, and consistency between the two markers were systematically varied among beaches. The experts were not informed of the systematic variation, and the goal was to identify sources of variability among experts in weighting these factors, discuss these differences, and use this information for arriving at consensus principles that can form the basis for establishing a human fecal contamination index.

2.0 METHODS

2.1 The simulation design and ranking exercise

The simulated data set consisted of 22 scenarios (22 beaches each containing 20 samples) in which one of the four factors was varied while the other three were held constant (Table 1). *Enterococcus* concentrations were varied among the scenarios such that frequency and severity of violation of California's single-sample standard of 104 *Enterococcus* per 100ml (U.S. EPA 1986) decreases from *Enterococcus* concentration categories A, to B, and to C, with specific concentration values randomly generated from the corresponding concentration ranges (Table 1). Two human-associated MST markers ((Haugland *et al.,* 2010) and (Shanks *et al.,* 2009)) were included in the data set, and marker concentrations were placed into one of four ranges: not detected (ND), detected but below limit of quantification (BLOQ), barely above lower limit of quantification (Low), and two to four orders of magnitude above lower limit of quantification (High), with specific values randomly generated within these ranges. The frequency of marker detection among samples in different scenarios was varied among 10%, 30%, and 50%. In addition, four scenario replicates were included to assess each expert's internal consistency in ranking the beaches (producing a total of 26 beach scenarios in the simulated dataset). The scenario replicates had different randomly generated concentration values, but included the same number of samples in each of the concentration and frequency ranges described above. Ten water quality experts were asked to rank the 26 beaches from 1 (the most contaminated) to 26 (the least contaminated) with respect to the relative level of human fecal contamination.

The experts were chosen to represent research scientists and water quality managers from the federal government, a public research agency, academic institutions, and a wastewater treatment agency. The consensus-building exercise was conducted repeatedly until expert opinions converged. In the first iteration, the experts were asked to provide their rankings independently of each other. In the second iteration, the experts were allowed to confer, discuss differences in their initial rankings and work towards development of consensus principles before again providing their independent rankings. In the third iteration, the experts were again assembled to further identify the principles on which they agreed and on which they differed, to improve upon the degree of consensus attained in the second iteration.

Table 1: Scenario design of the simulated data set.

Beach ID[a]	***Enterococcus***[b]	**Magnitude**[c]	**Frequency**[d]	**Consistency** [d] **(marker1 - marker2)**
1	A	High	10%	High - ND
2	A	High	10%	High - BLOQ
3	A	High	10%	High - High
4	A	High	30%	High - ND
5	A	Low	30%	Low - ND
6	A	Low	30%	Low - Low
7	A	BLOQ	50%	BLOQ - ND
8	A	BLOQ	50%	BLOQ - BLOQ
9	B	High	10%	High - ND
10	B	High	10%	High - High
11	B	High	30%	High - ND
12	B	Low	30%	Low - ND
13	B	BLOQ	50%	BLOQ - ND
14	B	BLOQ	50%	BLOQ - BLOQ
15	C	High	10%	High - ND
16	C	High	10%	High - High
17	C	High	30%	High - ND
18	C	Low	30%	Low - ND
19	C	BLOQ	50%	BLOQ - ND
20	C	BLOQ	50%	BLOQ - BLOQ
21	A	High	30%	High - High
22	C	BLOQ	10%	BLOQ - ND
23	A	High	10%	High - High
24	A	BLOQ	50%	BLOQ - BLOQ
25	B	High	10%	High - High
26	B	Low	30%	Low - ND

[a] Scenario replications: Beaches 23, 24, 25, and 26 were identical to beaches 3, 8, 10, and 12, respectively.
[b] *Enterococcus* concentrations: Frequency and severity of violation of the Enterococcus standard (104 cell/100ml) varied in a descending order from A: 30% (500-1000) + 70% (2-103), to B: 10% (500-1000) + 90% (2-103), to C: 50% (50-110) + 50% (2-10). For example, a beach with "A" *Enterococcus* would have 30% of the samples having *Enterococcus* concentrations between 500 and 1000 cell/100ml and 70% of the samples between 2 and 103 cell/100ml. Each data point was generated as a random number within the specified ranges.
[c] Marker concentrations: Each data point was generated as a random number within the specified ranges: High, Low, BLOQ (below limits of quantification), and ND (non-detectable) with the specified ranges being 10^5-10^7, 1500-2000, 400-500, and 0 copy/100ml, respectively.
[d] Frequency: Frequency of any detection (High, Low, BLOQ) of marker 1. Concentration and frequency of marker 2 was dictated by the specified "consistency" between markers 1 and 2.

2.2 Analysis of rankings

The internal consistency of the rankings by each expert was assessed by comparing their rankings for the paired beaches representing the same scenarios. If a pair of replicate beaches were assigned the same rank (for the few experts who assigned ties) or had ranks immediately below or above each other, the expert was considered to exhibit perfect internal consistency. Agreement among experts was evaluated by Spearman pair-wise correlation analysis of ranks. The 26 beaches were organized into groups such that within each group only one of the four factors (Table 1) was varied. Ranks for beaches within each group were then compared to reveal how variations in each factor influenced ranking of the beaches by the experts. All statistical analyses were conducted in R (R Core Development Team 2011).

3.0 RESULTS

3.1 Internal consistency of ranking among replicates

Internal consistency in ranking the beaches by the water quality experts was high (Table 2). Five experts (A, C, D, I, and J) and six experts (A, C, D, E, I, and J) exhibited perfect internal consistency for their rankings in iterations 1 and 2, respectively. Among the 80 possible pairings of duplicate beaches ranked by the experts in two iterations, 18 pairs (11 and 7 pairs for iterations 1 and 2, respectively) of duplicate beaches received identical ranking, while 43 pairs (20 and 23 pairs for iterations 1 and 2, respectively) were one rank apart (Table 2).

3.2 Overall agreement among experts

Overall agreement on beach ranking among the experts initially varied greatly, but increased from iteration 1 to iteration 2 (Fig. 1). The pair-wise correlation coefficients of the beach rankings among the experts ranged from -0.33 to 0.98 (average of 0.41) and from -0.14 to 0.98 (average of 0.47) for iterations 1 and 2, respectively.

3.3 *Enterococcus* CFU concentration

There was wide divergence among the experts regarding how *Enterococcus* information was used for beach ranking in iterations 1 and 2. Some experts completely disregarded *Enterococcus* concentration and used only the human-associated MST marker results for their rankings, while other experts used *Enterococcus* violations as the most important factor in their ranking. A few experts considered *Enterococcus* and human-associated marker information together, but usually assigned larger weights to human-associated marker data.

A comparison of the rankings for three beaches (beaches 3, 10, and 16) illustrates the different approaches in using *Enterococcus* data (Fig. 2). These three beaches had the highest concentrations for both human-associated MST markers in 10% of samples, but differed in their extent of *Enterococcus* violations, with beach 3 experiencing the most, beach 10 intermediate, and beach 16 the least enterococci pollution. In the first iteration (left panel, Fig. 2), experts B, C, H viewed *Enterococcus* as the most important factor, resulting in large rank differences between beaches 3 and 16 (the long vertical grey lines in Fig. 2), whereas in the second iteration experts C and H provided much closer rankings (shorter lines, right panel, Fig. 2). By contrast, rankings of expert G were unaffected by the *Enterococcus* in both iterations (short lines, both panels, Fig. 2).

Table 2: Expert ranking of the four pairs of duplicate beaches.

Pairs of duplicate beaches[a]	Iteration 1								Iteration 2							
	Pair#1		Pair#2		Pair#3		Pair#4		Pair#1		Pair#2		Pair#3		Pair#4	
Expert A	3	4	20	21	5	6	16	17	6	7	3	2	8	9	23	24
Expert B	3	4	15	17	5	6	13	12	2	3	12	19	5	4	15	17
Expert C	9	8	1	2	17	18	15	16	18	17	1	2	20	21	13	14
Expert D	5	5	19	19	5	5	16	16	3	3	6	6	8	8	15	15
Expert E	2	2	9	9	5	5	19	17	3	3	16	17	7	7	11	11
Expert F	15	12	2	1	16	13	9	10	18	19	4	1	21	22	13	15
Expert G	18	22	11	14	20	23	8	7	20	22	3	2	19	21	13	15
Expert H	9	11	3	4	8	10	17	18	14	16	2	1	13	15	10	12
Expert I	17	18	2	1	19	20	16	15	16	17	1	2	18	19	10	9
Expert J	3	3	8	8	5	5	20	20	14	15	2	1	17	16	11	12

[a] Pairs#1 to #4 refer to the four pairs of duplicate beaches (Beaches 3 and 23, 8 and 24, 10 and 25, 12 and 26) representing identical scenarios as described in Table 1.

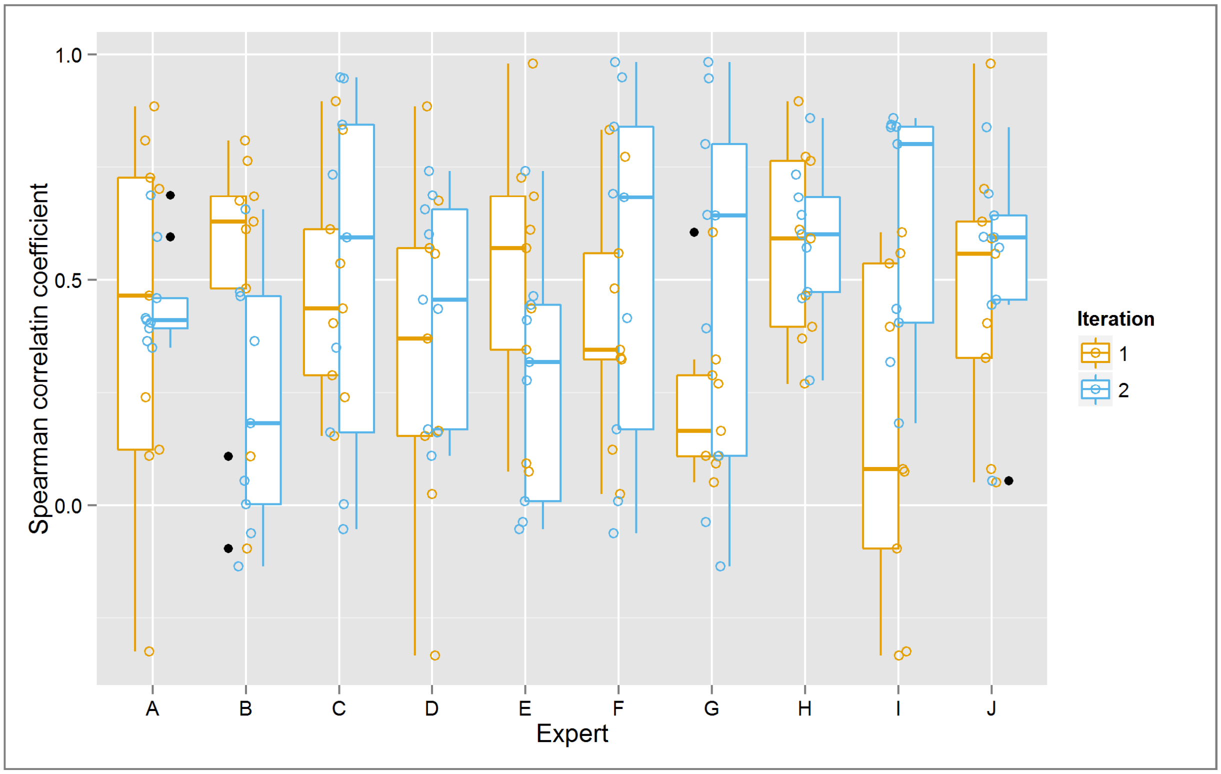

Figure 1: Correlation coefficients of ranks between each pair of experts (y-axis) vs. experts (Experts A to J; x-axis). Each colored circle (jittered to display potential overlapping points) represents one correlation coefficient in iterations 1 (yellow circles) or 2 (blue circles). Two side-by-side boxplots with corresponding colors show the summary statistics of the correlation coefficients for iterations 1 (yellow) and 2 (blue). On the boxplots, the central lines indicate median, ends of boxes (i.e., "hinges") indicate 1st and 3rd quartiles, extended lines indicate values within 1.5 times IQR (inter-quartile range) of the hinges, and black dots indicate extreme values. The relative long length of boxes indicates high variability in experts' rankings. The increase of medians from iterations 1 to 2 (i.e., blue central lines above yellow central lines) for most experts indicates improved general agreement among experts.

Subsequent discussion as part of iteration 3 led the experts to agree on two points with respect to *Enterococcus* standard violation. First, *Enterococcus* geometric mean concentrations for a site were more important to ranking than the concentration from individual samples. The second point of agreement was that the *Enterococcus* concentration should have much less effect on the rankings than human-associated marker data since *Enterococcus* is not specific to human fecal sources.

3.4 Human-associated MST markers: Frequency of detection, magnitude of signals, and consistency between MST methods

When other factors were held constant, higher frequency of marker detection resulted in a beach being ranked as more contaminated unanimously across all experts in both iterations 1 and 2. However initially, there was a big difference among experts in how strongly marker detection frequency influenced individual beach ranking. For beaches 3 (10% frequency) and 21 (30% frequency), the level of perceived increase in contamination depended on the expert (Fig. 3). For example, in iteration 1 (left panel, Fig. 3) expert G ranked the beach with 30% marker detection (beach 21) a total of 17 positions higher than the beach with only 10% frequency (beach 3), whereas expert E produced immediately adjacent ranks for these same two beaches. After discussion, 7 out of 10 experts gave beach 3 a less contaminated ranking in iteration 2 as compared to in iteration 1 (Fig. 3), and the average rank differences between these two beaches also increased from 6.1 to 7.9 ranks (while medians increased from 4.5 to 9.5 ranks), indicating frequency influenced ranking more strongly in iteration 2 than in iteration 1. Further consensus building (iteration 3) led to the expert's conclusion that higher frequency of detection was the key criterion for beach ranking.

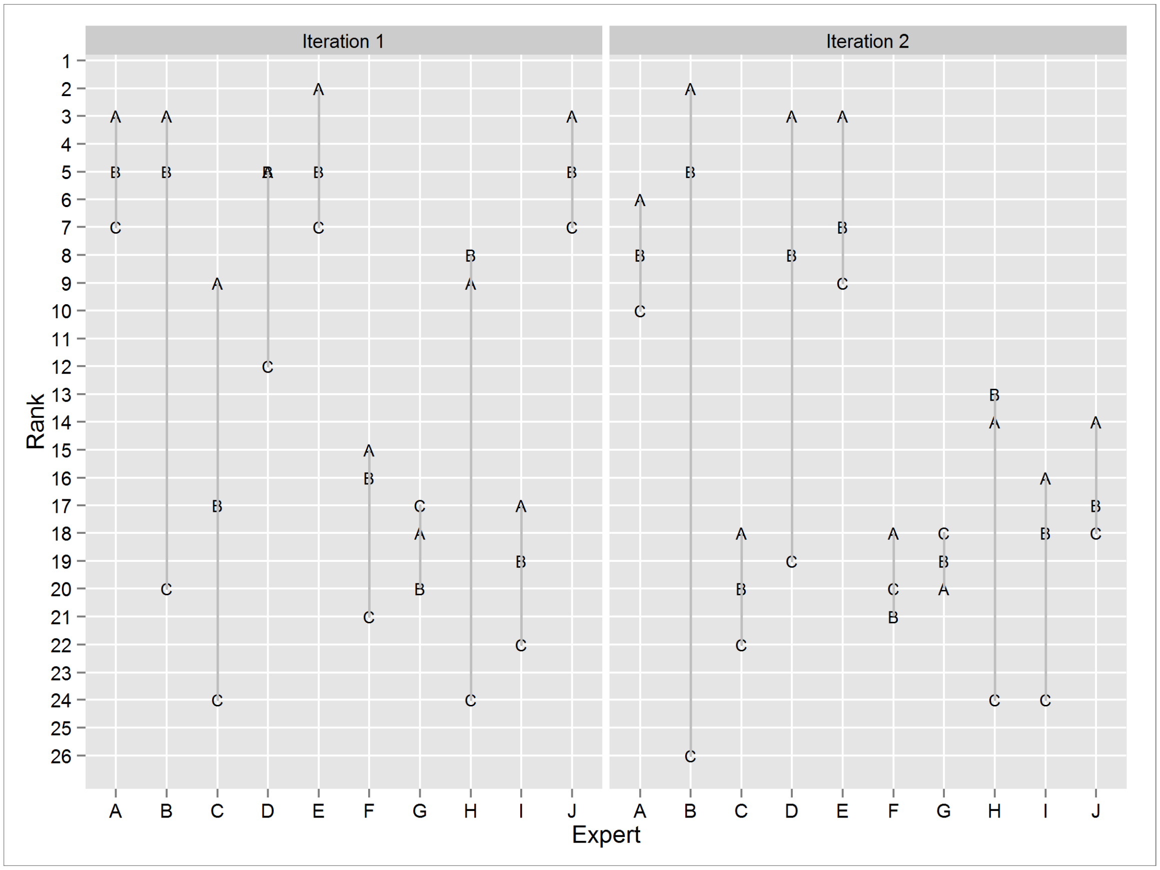

Figure 2: *Enterococcus* effect on beach ranking: Ranks (y-axis) provided by the experts (x-axis: experts A to J for iterations 1 and 2 on left and right panels, respectively) for beaches (3, 10, and 16) that only differed in extent of *Enterococcus* standard violations (denoted by "A", "B", "C" as defined in Table 1).

All three beaches had both human markers within the "High" magnitude range (10^5-10^7 copy/100 mL) in 10% of the samples. A higher rank (i.e. smaller number and higher position on the y-axis) indicates that the beach was regarded as more contaminated with human fecal sources. The difference in ranking the beaches is highlighted by the length of the grey line connecting the beaches for each expert: The further apart the ranks are, the more influence *Enterococcus* had on experts' ranking.

Similar to human-associated MST marker frequency, higher marker concentrations resulted in a beach being ranked as more contaminated unanimously across all experts in both iterations 1 and 2. However, there was a big difference among experts in how strongly magnitude affected beach ranking as well. For beaches 6 and 21, an increase of magnitude from Low to High always resulted in a more contaminated ranking by all experts, but the level of the ranking increase ranged from 1 to 13 positions, depending on the expert. For example, in iteration 1 (left panel, Fig. 4) expert D ranked the beach with High marker concentration 13 ranks higher than the beach with Low marker concentration, whereas majority of the experts (experts A, B, C, E, F, H, I, and J) produced immediately adjacent ranks for these two beaches. Consequently, median rank differences between these two beaches (differing by marker magnitude only) were only one rank in both iterations.

The experts disagreed on how to use consistency between human-associated MST markers for ranking. Some experts used the two human-associated MST markers together to assess the extent of contamination in each sample before integration across samples for ranking the beaches. Other experts treated the two human-associated MST markers for each sample as if they were two independent samples. However, the commonality between the two approaches and among the experts was that detection of both markers carried more weight than detection of just one marker. For example, beaches 1, 2, and 3 only differed in consistency between markers while all other factors were fixed, and an increased consistency between markers (*i.e*. High-ND [beach 1], to High-BLOQ [beach 2], to High-High [beach 3]) resulted in a more contaminated ranking by most experts in both iterations (Fig. 5).

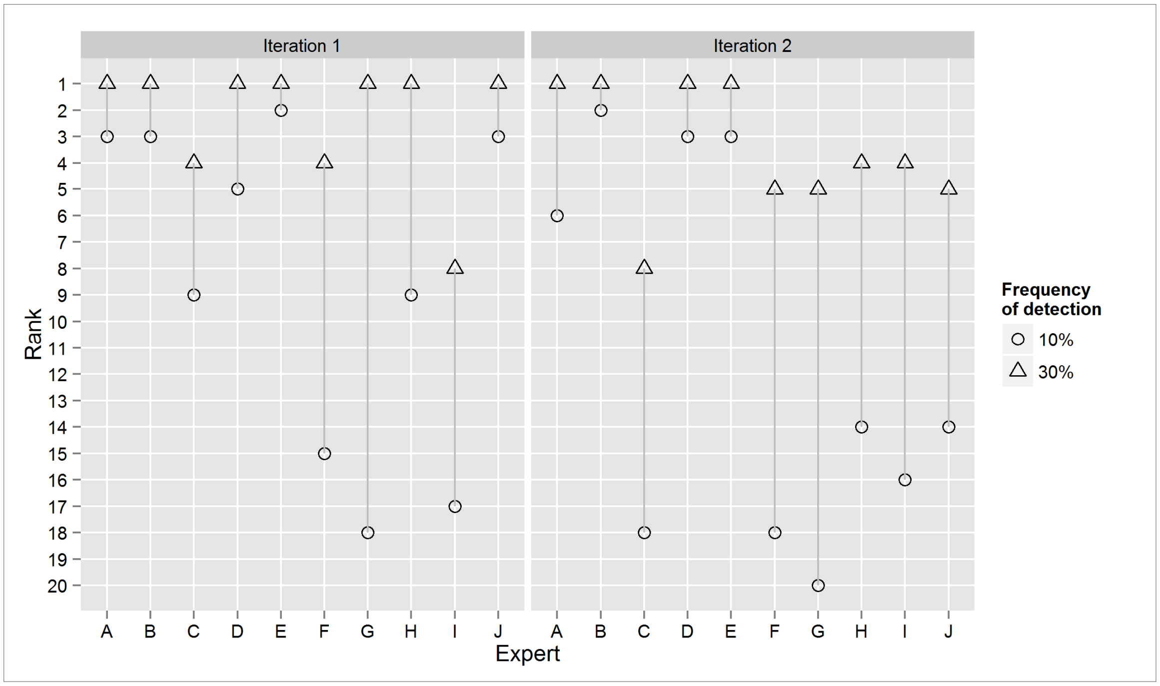

Figure 3: Frequency effect on beach ranking: Ranks (y-axis) provided by the experts (x-axis: experts A to J for iterations 1 and 2 on left and right panels, respectively) for beaches 3 (10%) and 21 (30%) that differed only in frequency of human marker detection (denoted by different symbols).

Both beaches had both marker concentrations within the "High" magnitude range (10^5-10^7 copy/100 mL) and *Enterococcus* concentrations within the range "A,"", but either 10% or 30% frequency of marker detection. A higher rank (i.e. smaller number and higher on the y-axis) indicates that the beach was regarded as more contaminated with human fecal sources. The difference in ranking the beaches is highlighted by the length of the grey line connecting the beaches for each expert: The further apart the ranks are, the more influence frequency had on experts' ranking.

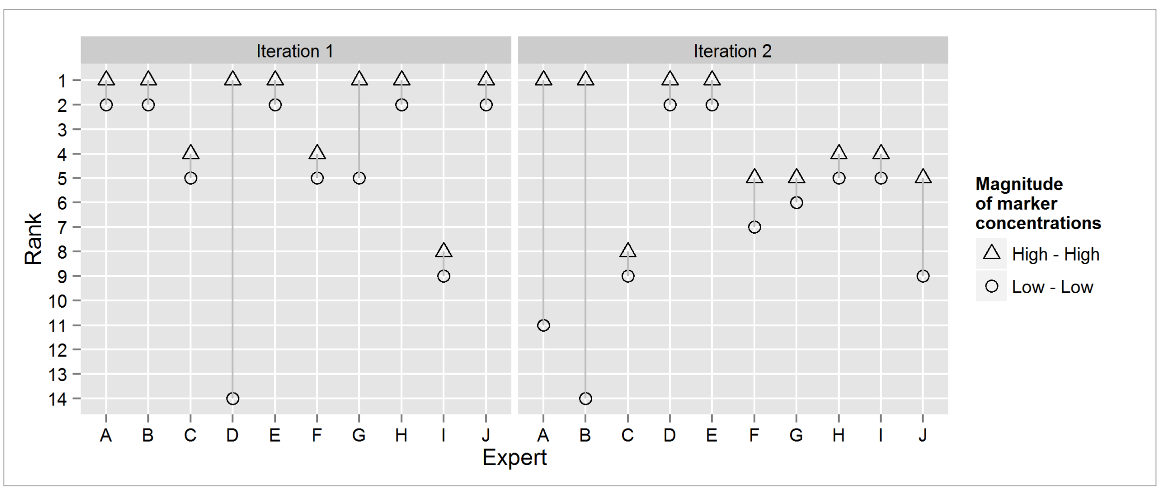

Figure 4: Magnitude effect on beach ranking: Ranks (y-axis) provided by the experts (x-axis: experts A to J for iterations 1 and 2 on left and right panels, respectively) for beaches 6 (Low-Low) and 21(High-High) that differed only in magnitude of human marker concentrations (denoted by different symbols).

Both beaches had both markers detected in 30% of samples and *Enterococcus* concentrations within the range "A,"", but both marker concentrations were within either the "LLOQ" or "High" magnitude range. A higher rank (i.e. smaller number and higher on the y-axis) indicates that the beach was regarded as more contaminated with human fecal sources. The difference in ranking the beaches is highlighted by the length of the grey line connecting the beaches for each expert: The further apart the ranks are, the more influence magnitude had on experts' ranking.

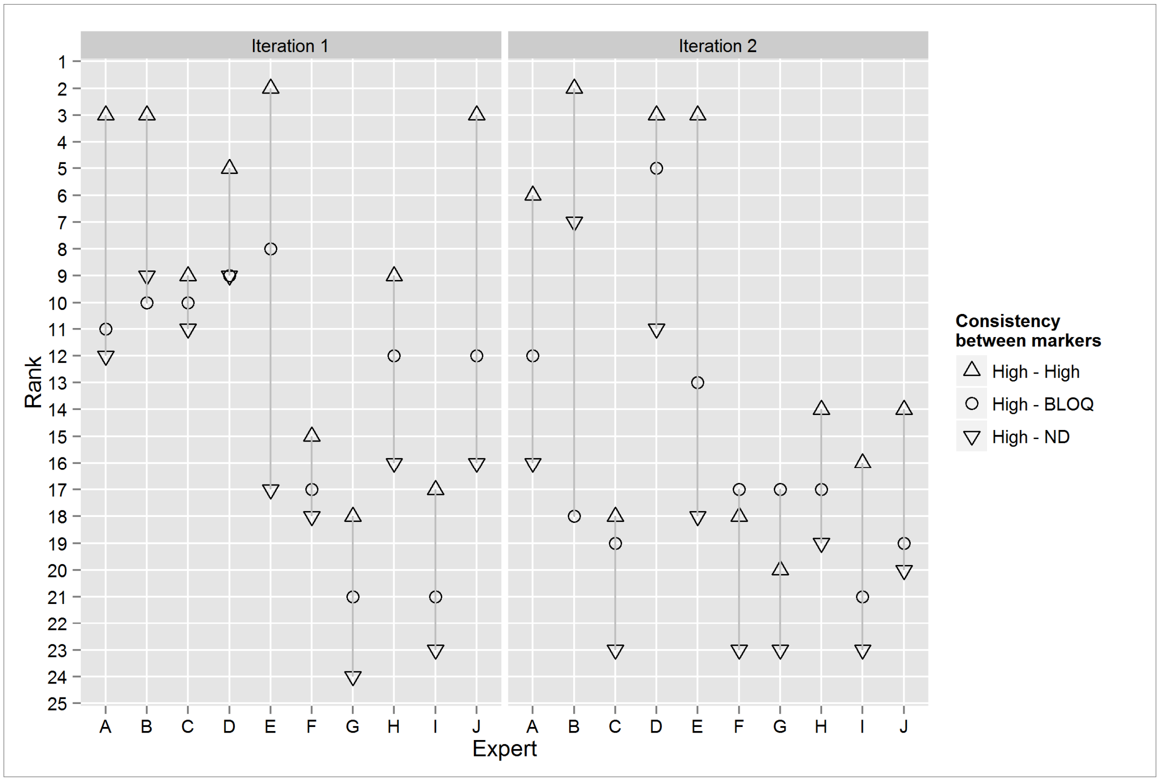

Figure 5: Effect of consistency between markers on beach ranking: Ranks (y-axis) provided by the experts (x-axis: experts A to J for iterations 1 and 2 on left and right panels, respectively) for beaches (1, 2, and 3) that differed in consistency between human marker concentrations (denoted by different symbols).
All three beaches had *Enterococcus* concentrations within range "A,"", 10% samples with "High" concentrations of the first marker but "ND" (beach 1), "BLOQ" (beach 2), or "High" (beach 3) ranges for concentrations of the second marker. A higher rank (i.e. smaller number and higher on the y-axis) indicates that the beach was regarded as more contaminated with human fecal sources. The difference in ranking the beaches is highlighted by the length of the grey line connecting the beaches for each expert: The further apart the ranks are, the more influence the factor "consistency between markers" had on experts' ranking.

3.5 Human-associated MST markers: Frequency of detection compared to magnitude of signals; Frequency of detection compared to consistency between markers

The highly variable, but unidirectional, effects that each of the three human marker factors exerted on beach ranking resulted from how differently experts weighed frequency, magnitude and consistency between markers. In iteration 1, more than half of the experts placed more weight on magnitude than frequency (left panel, Fig. 6): Six experts ranked beach 10 (10% samples with marker concentrations of High-High) as more contaminated than beach 14 (50% samples with marker concentrations of BLOQ-BLOQ). In iteration 2, the majority of experts (7 of 10) ranked beach 10 (higher magnitude) as less contaminated than beach 14 (higher frequency;) (right panel, Fig. 6). In subsequent discussions (iteration 3), the experts concluded that frequency of human-associated MST marker detection is more important than magnitude of, or consistency between, human markers.

Regarding frequency vs. consistency, frequency was generally of more importance to the experts. This can be seen in a comparison of beaches 2 (10% samples with marker concentrations of High-BLOQ) and 4 (30% samples with marker concentrations of High-ND), where 8 of 10 experts ranked beach 4 (higher frequency) as more contaminated in both iterations (Fig. 7). Discussion among the experts in iteration 3 confirmed their reliance on frequency of marker detection as more important than consistency between markers.

4.0 DISCUSSION

Three consensus principles were reached through this exercise. First, the frequency of human-associated MST marker detection is the most important factor in ranking beaches for extent of human contamination because the management goal is to assess the typical condition at a beach, rather than the exceptional event. The elevation of marker frequency also stemmed from concern that magnitude is a less reliable line of evidence, as laboratory steps such as water filtration and DNA isolation lead to approximately half a log unit of variability in estimated marker concentrations (Ebentier *et al.,* 2013, Shanks *et al.,* 2012). Moreover, these markers may experience differential degradation and removal rates by predation/absorption compared to human pathogens (Walters *et al.,* 2009). As such, the experts placed higher confidence in presence/absence distinctions (i.e. frequency) than in precision of marker concentrations, consistent with Soule et al's (2006) recommendation to base conclusions on positive events, rather than magnitude of individual sample measurements.

The second principle is that magnitude and consistency between human-associated MST markers should also be considered, but used as secondary w eights to support the primary factor of frequency. While human markers are relatively sensitive and specific, there are examples of cross-reaction or inhibition that can affect performance (Layton *et al.,* In Press). Thus, confidence in counting presence in estimating frequency is enhanced when confirmed by a second marker or by large magnitude in marker detections.

The third principle is that *Enterococcus* concentration is of least importance in ranking beaches with regard to extent of human fecal contamination. The experts arrived at this principle because *Enterococcus* is not specific to human fecal contamination and is typically poorly correlated with presence of human pathogens (Harwood *et al.,* 2005). The experts felt that *Enterococcus* concentration should be used for identifying that a beach is of sufficient concern to be selected for collection of human-associated marker data, but beyond that it should be only a minor modifier in a ranking process.

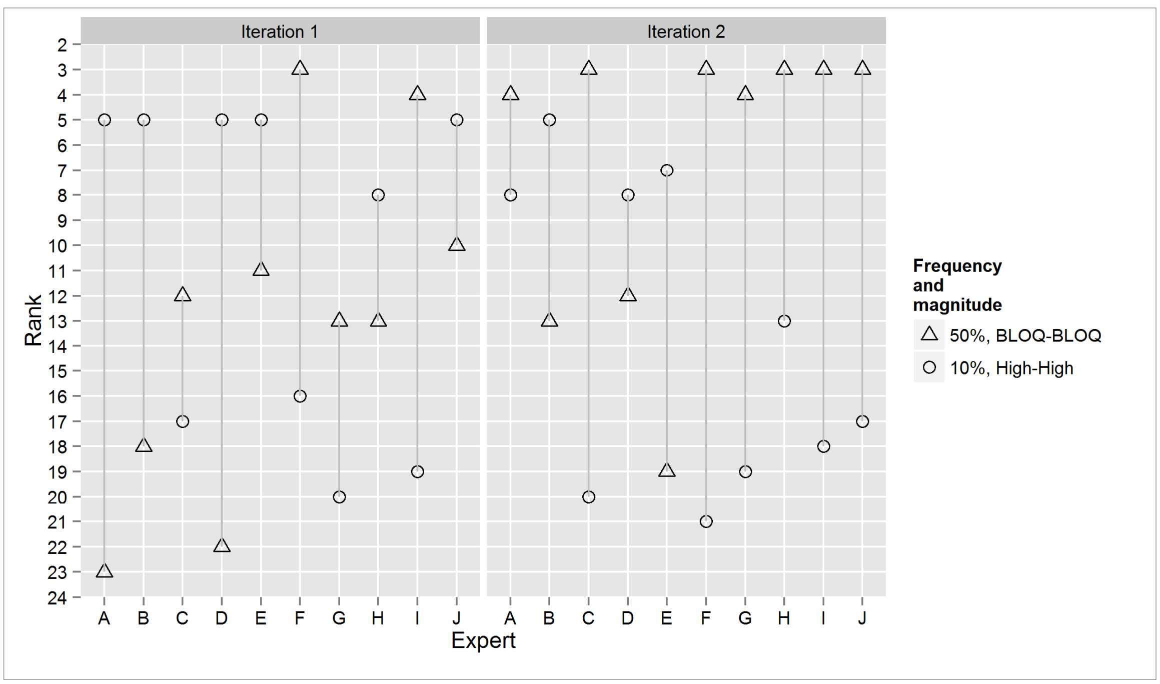

Figure 6: Frequency vs. magnitude effect on beach ranking: Ranks (y-axis) provided by the experts (x-axis: experts A to J for iterations 1 and 2 on left and right panels, respectively) for beaches 10 (High-High) and 14 (BLOQ-BLOQ) that had either lower frequency of high marker concentrations or higher frequency of lower marker concentrations.

Both beaches had *Enterococcus* concentrations within range "B."". A higher rank (i.e. smaller number and higher on the y-axis) indicates that the beach was regarded as more contaminated with human fecal sources. The difference in ranking the beaches is highlighted by the length of the grey line connecting the beaches for each expert: The direction and distance between the corresponding ranks indicated how the two factors (frequency vs. magnitude) were weighed against each other by the experts.

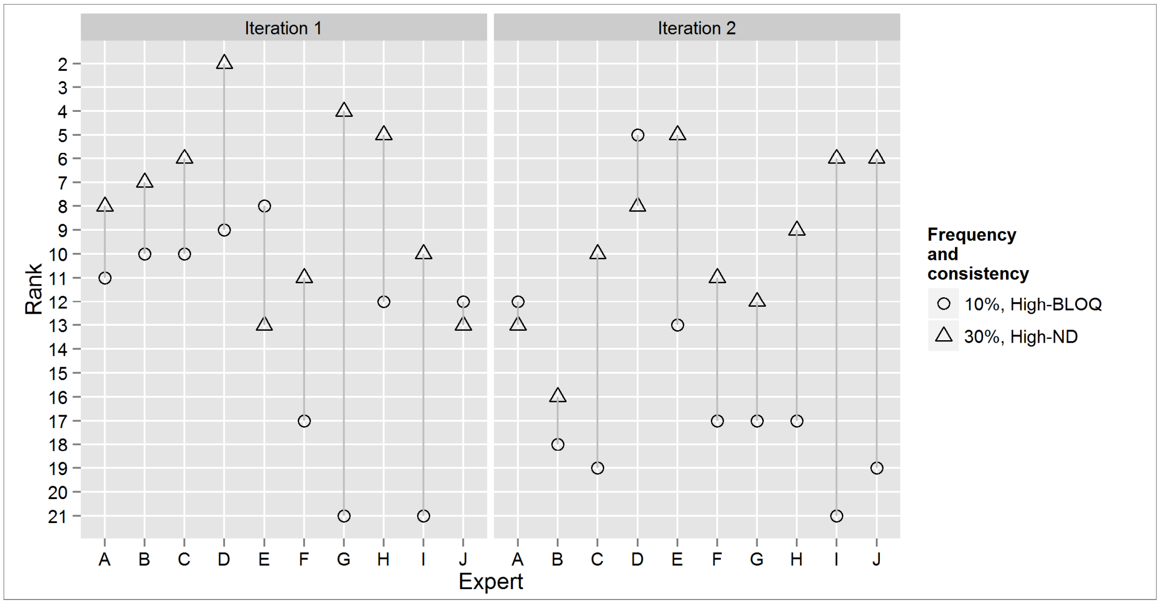

Figure 7: Frequency vs. consistency-between-markers effect on beach ranking: Ranks (y-axis) provided by the experts (x-axis: experts A to J for iterations 1 and 2 on left and right panels, respectively) for beaches 2 (High-BLOQ) and 4 (High-ND) that had either lower frequency of high consistency between marker concentrations or higher frequency of lower consistency between marker concentrations.

Both beaches had *Enterococcus* concentrations within range "A."". A higher rank (i.e. smaller number and higher on the y-axis) indicates that the beach was regarded as more contaminated with human fecal sources. The difference in ranking the beaches is highlighted by the length of the grey line connecting the beaches for each expert: The direction and distance between the corresponding ranks indicated how the two factors (frequency vs. consistency) were weighed against each other by the experts.

While the experts agree on these general principles, the exercise revealed considerable variability in the experts' application of these principles, suggesting the need for standardization of MST data interpretation. Delphi-based exercises (Hsu and Sandford, 2007) frequently find that the experts' professional backgrounds affect their values regarding scientific evidence, leading to high variability in data interpretation (Cormier *et al.,* 2008). For example, experts with beach management or water quality regulatory background placed heavier emphasis on *Enterococcus* standard violations because those are the data they work with most often, whereas experts from research institutes mainly utilized the human-associated MST marker data for beach ranking. This is consistent with the recognition that more quantitative approaches are needed to better define certainty elements in an open framework process (Chapman *et al.,* 2002). This also corresponds to efforts by U.S.EPA and other federal agencies to develop a highly quantifiable, transparent, and repeatable approach to decision-making frameworks in risk analysis and ecological risk assessment (Chapman 2007, Linkov *et al.,* 2009, Suter II and Cormier 2011).

We propose that an algorithm based on the Bayesian approach, which has been previously used to determine if a particular detection represents a true positive "event" (Jenkins *et al.,* 2009, Kildare *et al.,* 2007, Lamendella *et al.,* 2009, Ryu *et al.,* 2012), can provide that consistency and transparency. Such an algorithm includes three basic steps:

1. Calculate a sample score, i.e. a weighted "event", using a Bayesian probabilistic model based on human-associated marker data and the markers' performance metrics (in the form of conditional probabilities). This would generate a sample score that represents the probability of human fecal presence in each sample.
2. Calculate a site score that reflects an average condition of a site and has a unified range (e.g. 0-100), from all sample scores. This site score serves as a human fecal contamination index for the extent of human fecal contamination at the site.
3. Use the index for water quality management applications such as beach ranking. A beach with a higher site score will be ranked as more contaminated with human fecal material than one with a lower site score.

A mathematical algorithm such as this would provide many advantages over an expert-decision approach. First, the "weights" for the weighted-frequency consensus approach are mathematically defined as conditional probabilities that can be scientifically obtained via MST method evaluation studies. Second, such an algorithm provides standardization of MST data interpretation, which allows consistent data interpretation across sites and time, aids in reproducibility across laboratories, and provides a benchmark for the systematic comparison of source identification results. Third, a mathematically defined model system will enable formal statistical analysis to assist in management decision-making by providing a comparative index. For example, one decision that managers often face relates to resource allocation: to assess the extent of human contamination given limited resources and a particular management goal, should more samples be taken and analyzed for one MST marker or should more markers be run on fewer samples? This question may be answered quantitatively by calculating which scenario (more samples compared to more markers) affords the more precise estimation of site scores (i.e. human fecal contamination index) by the algorithm (Fig. 8).

Such an algorithm would also be applicable to other management decisions, such as determining if a beach has low enough human fecal contamination to be eligible for QMRA studies (Soller *et al.,* 2010). The algorithm enables construction of statistically based decision rules with a predetermined confidence level (Fig. 9). Substituting other markers such as those for gulls or dogs, instead of human-associated MST markers, would also enable the algorithm to produce site scores based on the marker(s) of choice (i.e. gull or dog fecal contamination indices).

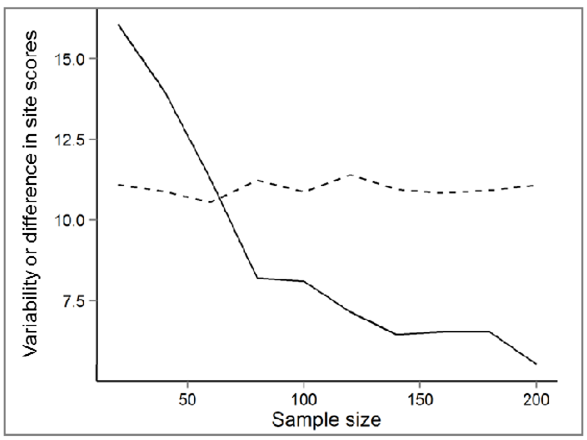

Figure 8: An example for using the algorithm to assist management decisions regarding resource allocation: more samples with each sample analyzed for one human-associated marker or fewer samples with each sample analyzed for two human-associated markers. The solid line represents how variability in site score decreases as sample size increases. The dashed line represents how the site score difference between when one or two markers are analyzed changes as sample size increases. This mock graph illustrates that resources are better spent to analyze more samples when the sample size is small.

Figure 9: An example of using site score in a distribution-based power analysis for determining if a beach has sufficiently low human contamination to be eligible for QMRA. For the purpose of this mock graph: a Binomial distribution for the site score was assumed appropriate for the power analysis, a true site score < 10 was assumed to indicate sufficiently low human contamination, and simulation was run for observed site scores ranging from 0 to 5. Utilizing the power analysis, a decision curve can be constructed such that the area above the decision curve represents sufficient evidence for low human contamination. Multiple decision curves can be constructed to represent different degree of confidence (i.e. certainty in the graph) in the decision.

5.0 CONCLUSIONS

- The Delphi exercise using a simulated data set of 26 beaches revealed large variability in how water quality experts interpreted data from an MST study, indicating the need for standardization.
- Three iterative components of the exercise identified consensus principles toward standardized MST data interpretation needed for developing a human fecal contamination index: 1) frequency of samples that are positive for human-associated markers is of primary importance; 2) magnitude of and consistency between human-associated markers should be used to weight MST marker frequency; and 3) general FIB data (e.g. *Enterococcus*) should receive the least weight.
- A mathematical algorithm based on a Bayesian approach could be used to quantitatively realize the consensus principles and produce an index that could be used to rank beaches with respect to the degree of human fecal contamination. While the algorithm holds great promise conceptually, additional work is needed to further develop, validate, and demonstrate the algorithm for use by the water quality community.

ACKNOWLEDGEMENTS

Funding for this project has been provided in part through an agreement with the California State Water Resources Control Board. The contents of this document do not necessarily reflect the views and policies of the California State Water Resources Control Board. Information has been subjected to the U.S. Environmental Protection Agency's peer and administrative review and has been approved for external publication. Any opinions expressed in this paper are those of the author (s) and do not necessarily reflect the official positions and policies of the agency. Any mention of trade names or commercial products does not constitute endorsement or recommendation for use.

REFERENCES

[1] Boehm, A.B., Van De Werfhorst, L.C., Griffith, J.F., Holden, P.A., Jay, J.A., Shanks, O.C., Wang, D. and Weisberg, S.B., 2013 Performance of forty-one microbial source tracking methods: A twenty-seven lab evaluation study. *Water Research*. In press.

[2] Chapman, P.M., McDonald, B.G. and Lawrence, G.S., 2002 Weight-of-evidence issues and frameworks for sediment quality (and other) assessments. *Human and Ecological Risk Assessment: An International Journal*. 8(7), 1489-1515.

[3] Chapman, P.M., 2007 Determining when contamination is pollution — Weight of evidence determinations for sediments and effluents. *Environment International*. 33(4), 492-501.

[4] Ebentier, D.L., Hanley, K.T., Cao, Y., Badgley, B., Boehm, A., Ervin, J., Goodwin, K.D., Gourmelon, M., Griffith, J., Holden, P., Kelty, C.A., Lozach, S., McGee, C., Peed, L., Raith, M., Sadowsky, M.J., Scott, E., Santodomingo, J., Sinigalliano, C., Shanks, O.C., Werfhorst, L.C.V.D., Wang, D., Wuertz, S. and Jay, J., 2013 Evaluation of the repeatability and reproducibility of a suite of PCR-based microbial source tracking methods. *Water Research*. In press.

[5] Harwood, V.J., Levine, A.D., Scott, T.M., Chivukula, V., Lukasik, J., Farrah, S.R. and Rose, J.B., 2005 Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. *Applied and Environmental Microbiology*. 71(6), 3163-3170.

[6] Haugland, R.A., Varma, M., Sivaganesan, M., Kelty, C., Peed, L. and Shanks, O.C., 2010 Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection of selected Bacteroidales species and human fecal waste by qPCR. *Systematic and Applied Microbiology*. 33(6), 348-357.

[7] Jenkins, M.W., Tiwari, S., Lorente, M., Gichaba, C.M. and Wuertz, S., 2009 Identifying human and livestock sources of fecal contamination in Kenya with host-specific Bacteroidales assays. *Water Research*. 43(19), 4956-4966.

[8] Kildare, B.J., Leutenegger, C.M., McSwain, B.S., Bambic, D.G., Rajal, V.B. and Wuertz, S., 2007 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales: A Bayesian approach. *Water Research*. 41(16), 3701-3715.

[9] Lamendella, R., Santo Domingo, J.W., Yannarell, A.C., Ghosh, S., Di Giovanni, G., Mackie, R.I. and Oerther, D.B., 2009 Evaluation of Swine-Specific PCR Assays Used for Fecal Source Tracking and Analysis of Molecular Diversity of Swine-Specific "Bacteroidales" Populations. *Applied and Environmental Microbiology*. 75(18), 5787-5796.

[10] Layton, B.A., Cao, Y., Ebentier, D.L., Ballesté, E., Brandão, J., Byappanahalli, M.N., Converse, R.R., Farnleitner, A.H., Gentry-Shields, J., Gidley, M.L., Gourmelon, M., Lee, C.S., Lee, J., Lozach, S., Madi, T., Meijer, W.G., Noble, R.T., Peed, L.A., Reischer, G.H., Rodrigues, R., Rose, J.B., Schriewer, A., Sinigalliano, C.D., Srinivasan, S., Stewart, J.R., Van De Werfhorst, L.C., Wang, D., Whitman, R.L., Wuertz, S., Jay, J.A., Holden, P.A., Boehm, A.B., Shanks, O.C. and Griffith, J.F., In Press Performance of Human Fecal Anaerobe-Associated PCR-Based Assays in a Multi-Laboratory Method Evaluation Study. *Water Research*.

[11] Linkov, I., Loney, D., Cormier, S., Satterstrom, F.K. and Bridges, T., 2009 Weight-of-evidence evaluation in environmental assessment: Review of qualitative and quantitative approaches. *Science of the Total Environment*. 407, 5199-5205.

[12] Linstone, H.A. and Turoff, M., 1975 *The Delphi Method: Techniques and Applications*, Reading, Mass.: Addison-Wesley, ISBN 978-0-201-04294-8.

[13] R Core Development Team, 2011 *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.*

[14] Ryu, H., Lu, J., Vogel, J., Elk, M., Chávez-Ramírez, F., Ashbolt, N. and Santo Domingo, J., 2012 Development and Evaluation of a Quantitative PCR Assay Targeting Sandhill Crane (Grus canadensis) Fecal Pollution. *Applied and Environmental Microbiology*. 78(12), 4338-4345.

[15] Shanks, O.C., Kelty, C.A., Sivaganesan, M., Varma, M. and Haugland, R.A., 2009 Quantitative PCR for Genetic Markers of Human Fecal Pollution. *Applied and Environmental Microbiology*. 75(17), 5507-5513.

[16] Shanks, O.C., White, K., Kelty, C.A., Sivaganesan, M., Blannon, J., Meckes, M., Varma, M. and Haugland, R.A., 2010 Performance of PCR-Based Assays Targeting Bacteroidales Genetic Markers of Human Fecal Pollution in Sewage and Fecal Samples. *Environmental Science & Technology*. 44(16), 6281-6288.

[17] Shanks, O.C., Sivaganesan, M., Peed, L., Kelty, C.A., Blackwood, A.D., Greene, M.R., Noble, R.T., Bushon, R.N., Stelzer, E.A., Kinzelman, J., Anan'eva, T., Sinigalliano, C., Wanless, D., Griffith, J., Cao, Y., Weisberg, S., Harwood, V.J., Staley, C., Oshima, K.H., Varma, M. and Haugland, R.A., 2012 Inter-laboratory general fecal indicator quantitative real-time PCR

methods comparison study. *Environmental Science & Technology*. 46(2), 945–953.

[18] Soller, J.A., Schoen, M.E., Bartrand, T., Ravenscroft, J.E. and Ashbolt, N.J., 2010 Estimated human health risks from exposure to recreational waters impacted by human and non-human sources of faecal contamination. *Water Research*. 44(16), 4674-4691.

[19] Soule, M., Kuhn, E., Loge, F., Gay, J. and Call, D.R., 2006 Using DNA microarrays to identify library-independent markers for bacterial source tracking. *Applied and Environmental Microbiology*. 72(3), 1843-1851.

[20] Stoeckel, D.M., Stelzer, E.A., Stogner, R.W. and Mau, D.P., 2011 Semi-quantitative evaluation of fecal contamination potential by human and ruminant sources using multiple lines of evidence. *Water Research*. 45(10), 3225-3244.

[21] Suter II, G.W. and Cormier, S.M., 2011 Why and how to combine evidence in environmental assessments: Weighing evidence and building cases. *Science of the Total Environment*. 409(8), 1406-1417.

[22] U.S. EPA, 1986 Ambient Water Quality Criteria for Bacteria –1986. EPA440/5-84-002, Washington, DC.

[23] U.S. EPA, 2012 Recreational Water Quality Criteria, Office of Water 820-F-12-058, Washington, DC.

[24] Walters, S.P., Yamahara, K.M. and Boehm, A.B., 2009 Persistence of nucleic acid markers of health-relevant organisms in seawater microcosms: Implications for their use in assessing risk in recreational waters. *Water Research*. 43(19), 4929-4939.

[25] Wang, D., Silkie, S.S., Nelson, K.L. and Wuertz, S., 2010 Estimating true human and animal host source contribution in quantitative microbial source tracking using the Monte Carlo method. *Water Research*. 44(16), 4760-4775.