Environmental Monitoring and Assessment (2006) **116:** 335–344
DOI: 10.1007/s10661-006-7571-z

COMPARISON AND VERIFICATION OF BACTERIAL WATER QUALITY INDICATOR MEASUREMENT METHODS USING AMBIENT COASTAL WATER SAMPLES

JOHN F. GRIFFITH[1,*], LARISSA A. AUMAND[2], IOANNICE M. LEE[3], CHARLES D. MCGEE[4], LAILA L. OTHMAN[5], KERRY J. RITTER[1], KATHY O. WALKER[6] and STEPHEN B. WEISBERG[1]

[1]*Southern California Coastal Water Research Project (SCCWRP), 7171 Fenwick Lane, Westminster, CA 92683, USA;* [2]*MEC Analytical Systems, Inc., Carlsbad, California, USA;* [3]*City of Los Angeles, Department of Sanitation, Los Angeles, California, USA;* [4]*Orange County Sanitation District, Fountain Valley, California, USA;* [5]*City of San Diego, Metropolitan Waste Water Department, San Diego, California, USA;* [6]*Los Angeles County Sanitation District, Whittier, California, USA*
*(*author for correspondence, e-mail: johng@sccwrp.org)*

(Received 9 November 2004; accepted 18 May 2005)

Abstract. More than 30 laboratories routinely monitor water along southern California's beaches for bacterial indicators of fecal contamination. Data from these efforts frequently are combined and compared even though three different methods (membrane filtration (MF), multiple tube fermentation (MTF), and chromogenic substrate (CS) methods) are used. To assess data comparability and quantify variability within method and across laboratories, 26 laboratories participated in an intercalibration exercise. Each laboratory processed three replicates from eight ambient water samples employing the method or methods they routinely use for water quality monitoring. Verification analyses also were conducted on a subset of wells from the CS analysis to confirm or exclude the presence of the target organism. Enterococci results were generally comparable across methods. Confirmation revealed a 9% false positive rate and a 4% false negative rate in the CS method for enterococci, though these errors were small in the context of within- and among-laboratory variability. Fecal coliforms also were comparable across all methods, though CS underestimated the other methods by about 10%, probably because it measures only *E. coli*, rather than the larger fecal coliform group measured by MF and MTF. CS overestimated total coliforms relative to the other methods by several fold and was found to have a 40% false positive rate in verification. Across-laboratory variability was small relative to within- and among-method variability, but only after data entry errors were corrected. One fifth of the laboratories committed data entry errors that were much larger than any method-related errors. These errors are particularly significant because these data were submitted in a test situation where laboratories were aware they would be under increased scrutiny. Under normal circumstances, it is unlikely that these errors would have been detected and managers would have been obliged to issue beach water quality warnings.

Keywords: fecal indicator bacteria, laboratory intercalibration, method comparison, water quality

1. Introduction

Southern California's beaches are monitored extensively to screen for fecal contamination from human activities, such as wastewater discharges, industrial input,

and surface runoff (Schiff *et al*., 2002). More than 30 groups are involved in this monitoring, including city and county health departments, treated wastewater dischargers, stormwater permittees, and non-profit environmental organizations. These groups all measure the same parameters enterococci (ENT), fecal coliforms (FC) and total coliforms (TC), but have the option of choosing from a number of different measurement methods. Wastewater dischargers primarily rely on membrane filtration (MF). Stormwater agencies and environmental groups primarily use the IDEXX® chromogenic substrate (CS) method. Health departments historically have relied on multiple-tube fermentation (MTF) and MF, but have begun to use CS more frequently in the last several years.

Data from these multiple providers are collated and used collectively in several ways. On a daily basis, they are used to assess beach water quality and as the basis for issuance of beach water quality warnings. On a long-term basis, they are integrated to identify chronically contaminated beaches for Section 303(d) listing under the federal Clean Water Act and for development of environmental report cards that compare water quality among locations and over time. Using these data interchangeably assumes that results from multiple laboratories using different measurement methods are comparable, even though the laboratories may have varying levels of proficiency and may employ detection methods that rely on widely different products of bacterial growth.

A number of studies have compared the response of MF and MTF, and a few studies have compared these methods to CS (Kinzelman *et al*., 2003; Francy and Darner, 2000; Abbott *et al*., 1998; Eckner, 1998; Budnick *et al*., 1996; Palmer *et al*., 1993; Bej *et al*., 1991; Edberg *et al*., 1990; Covert *et al*., 1989). Noble *et al*. (2003a) was the first to compare results among all three methods and place differences among methods into the context of variability among laboratories that use the same method. However, Noble *et al*. used fabricated samples created primarily from laboratory strains of bacteria seeded into clean matrices. Natural ambient water samples often contain contaminants, particularly suspended solids, which have the potential to interfere with these methods. Natural samples also contain native bacteria, such as *Aeromonas*, *Vibrio, Pseudomonas*, and *Flavobacteria* spp., which have been shown to produce positive reactions in substrates containing 4-methylumbelliferyl-β-glucuronide (MUG) and can lead to false positives in the CS test (Pisciotta *et al*., 2002; Landre *et al*., 1998; Davies *et al*., 1995; Hidalgo *et al*., 1977).

Here, we present an intercalibration study modeled after Noble *et al*. (2003a), but based on measurement of ambient water samples collected from sites known to have complex matrices. The study assessed comparability of results among 26 southern California laboratories that conduct routine bacterial monitoring using three bacterial indicator measurement methods and identified common causes of error in determining bacterial concentrations for water quality monitoring purposes. In addition, the study evaluated the reliability of CS methods through verification of target organisms. Reliability of the CS method is particularly important in southern

TABLE I
Median concentrations of fecal indicator bacteria per 100 ml reported in each sample across all methods

Location	Site description	Enterococci	Fecal coliforms (*E. coli*)	Total coliforms
Ballona Wetlands	Estuary	80	130	2,400
Cabrillo Beach	Embayment	285	500	820
Doheny Beach	Open Beach	22	20	98
MBW 6	Urban Creek	130	100	12,033
MBW 10	Urban Creek	199	41	2,000
Malibu Beach	Open Beach	10	488	1,194
Malibu Lagoon	Estuary	122	300	5,191
Mission Bay	Embayment	120	285	800

California because results from this method are increasingly the basis for decisions regarding the issuance of beach water quality warnings.

2. Methods

Ambient water samples were collected from eight sites throughout southern California, including open marine beaches, estuaries, and flowing creeks carrying dry-weather urban runoff (Table I). All samples were collected in sterile, 20 L carboys following Standard Methods 1060 protocol for aseptic sampling techniques (APHA, 1995). Samples were then transported on ice to the Orange County Sanitation District (OCSD) laboratory in Fountain Valley, California. Upon arrival, carboys were placed on magnetic stir-plates, a sterile stir bar was added, and samples were stirred continuously for a minimum of 20 minutes to ensure homogeneity. Water from each carboy was dispensed into 26 sets of sterile, pre-labeled 100 ml bottles, which were transported on ice to participating laboratories. Sample processing began simultaneously at all laboratories at a pre-arranged time to eliminate differences due to holding time.

Samples were analyzed for TC, FC, *Escherichia coli* (EC), and ENT, using the method or methods routinely performed by each laboratory. Three classes of methods were used: CS, MTF, and MF. Each laboratory analyzed multiple dilutions of each sample to minimize the number of samples occurring outside of a quantifiable range. All analyses were performed in triplicate.

Laboratories performing CS used IDEXX media and the Quanti-Tray/2000® system for all samples, following the manufacturer's instructions. Colilert®-18 media were used for enumeration of TC and EC, while Enterolert™ media were used for ENT.

Multiple-tube fermentation for TC recovery used APHA 9221B (LTB/BGB). FC recovery was by either APHA Method 9221E.1 (EC) or APHA Method 9221E.2

(A-1). The ENT group was enumerated using the APHA Method 9230B (azide dextrose/PSE; APHA, 1995).

Membrane filtration for TC recovery was by APHA Method 9222B (M-Endo), both single-step and enrichment-step procedures. FC enumeration using MF was by APHA Method 9222D (M-FC). The 24-h EPA Method 1600 (M-EI) procedure (APHA, 1995) was used for ENT.

Five laboratories conducted confirmation testing on CS. When available, 10 positive wells were selected randomly for confirmation from trays with 80% or more positive wells. In certain instances, when a low number of positive wells was present, a smaller number of wells was selected from the tray. This resulted in confirmation testing for 71 ENT wells, 35 *E. coli* wells, and 153 TC wells. In addition, 55 non-fluorescing Enterolert® wells and 21 weakly fluorescing wells (scored as not containing ENT following manufacturer protocols) were subjected to verification analysis to test for false negatives.

Bacterial isolates for confirmation testing were obtained by wiping the back of the Quanti-Tray® with 70% isopropanol, puncturing the well with either a sterile syringe or sterile scalpel, and withdrawing the well contents. TC bacteria were confirmed by transferring well contents to either Tryptic Soy (TSB) or Brain Heart Infusion (BHI) broths, incubating at 35 °C, transferring again to Brilliant Green Bile broth (BGB), and subsequently plating to MacConkey agar. *E. coli* was confirmed by transferring well contents to either EC or A-1 broths, incubating at 44.5 °C and plating to either MacConkey or LES Endo agars. ENT was confirmed by inoculating either TSB or BHI broth with well contents, incubating at 35 °C, then filtering the broth, plating to mEI agar and incubating at 41 °C. Secondary confirmations were performed on each isolate that tested positive using a second EPA- or APHA-approved biochemical testing method (MF or MTF, as described previously) or by submitting isolates to the Vitek microbial identification system (bioMérieux, Hazelwood, MO).

Statistical differences in median concentrations among methods and within-lab variability among methods were assessed using ANOVA on ranks (Conover and Iman, 1981). Median values from replicate samples then were used to compute ranks across methods separately for each station. Where statistical differences among methods were detected ($p > 0.05$), individual stations were examined for possible station effects using Bonferroni-adjusted significance levels. Within-lab variability among methods was examined by taking the standard deviations of log counts across replicates for each laboratory sample and then ranking them within station. Bonferroni-adjusted significance levels were employed to control overall error rates at 0.05 when testing for station effects.

Estimates of variability for each method, based on the median standard deviation of log counts and the average median log ratios of these counts, were used to estimate the reported bacterial concentrations at which one would be 95% or 99% confident that a single sample analyzed would fall above or below California's standards at which public health warnings are issued.

To assess the importance of individual laboratory error, data were examined to determine how well individual laboratory results agreed with the results from the entire group of laboratories. This was done by identifying the number of samples produced by each laboratory that were more than a half log unit above or below the overall median. This criterion was chosen because it is roughly equal to the average within-lab method variability observed in previous studies (Noble *et al*., 2003a).

3. Results

Median concentrations of ENT exceeded California's single sample standard (104/100 ml) at five of the eight sample sites (Table I). State standards for FCs (400/100 ml) and TCs (10,000/100 ml) were exceeded at two stations and one station, respectively.

There was no statistical difference in median concentration between MF and MTF for any of the bacterial indicators, though there was a difference between CS and MF for all three indicators (Table II). For ENT, CS produced lower concentrations than MF, but most of the difference was attributable to a single station (Figure 1a). The median CS value at Doheny Beach was several-fold lower than that for either MF or MTF, but measured concentrations at that site were low for all methods. The median concentration was only 22/100 ml, with more than three-quarters of the CS values and almost half of the MF and MTF values reported as non-detectable values. When the Doheny Beach station data was removed from the analysis, there was no statistical difference among methods for ENT (Table II).

The median CS values for FCs were similar to those produced by MTF, but about 30% lower than those produced using MF. Though CS methods are known to underestimate FC levels because they detect only *E. coli*, the majority of the difference was attributable to low bacterial concentrations at two stations, Doheny Beach and MBW10 (Figure 1b). Median concentrations at the Doheny Beach and MBW10 stations were only 20/100 mL and 41/100 ml, respectively. When these stations were removed from the calculation, the difference between results from CS and MF was only 9%.

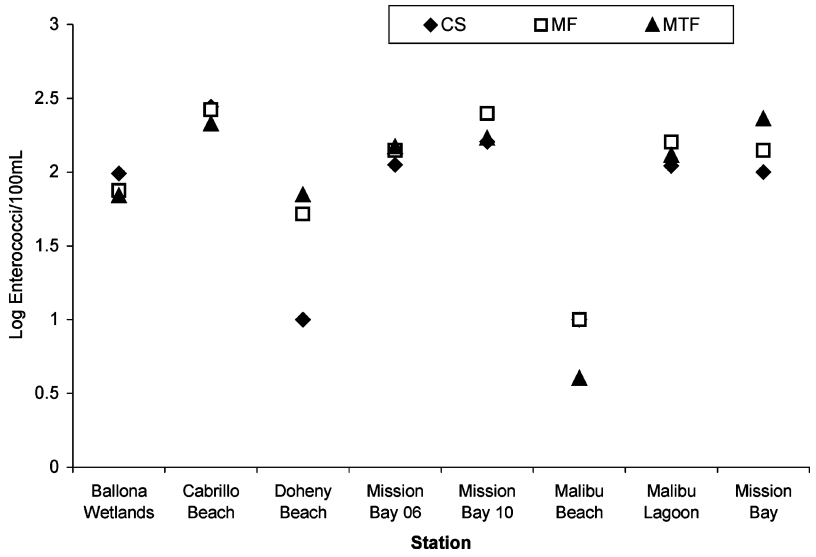
TABLE II
Estimated median ratios of log counts between methods, for each indicator

	Enterococci		Fecal coliform/*E. coli*		Total coliform
Comparison	All sites	Without Doheny	All sites	Without Doheny or MBW10	All sites
CS/MF	0.71*	0.86	0.69*	0.91	2.54*
CS/MTF	0.80	1.03	1.03	0.95	3.80*
MF/MTF	1.13	1.20	1.46	1.04	1.49

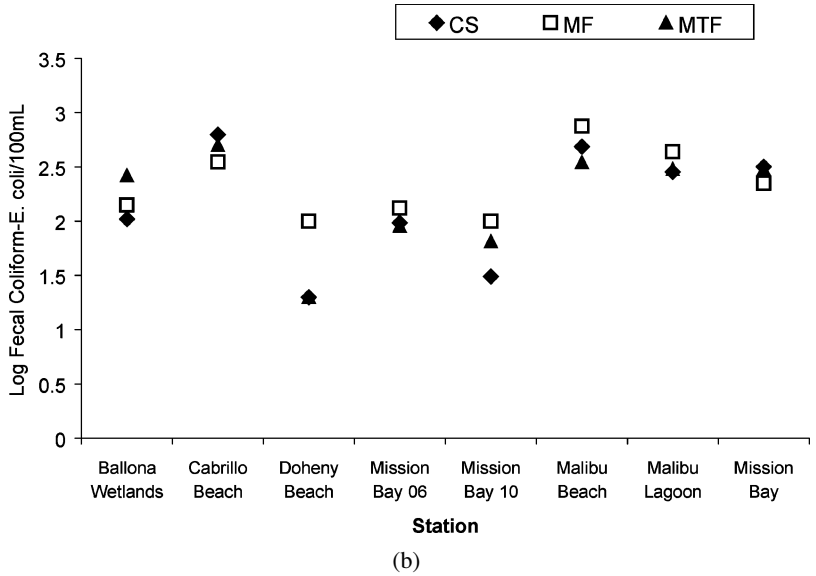
Asterisk indicates statistically different than 1.

For TC, median concentrations from CS were several-fold higher than from either MF or MTF. CS produced the highest median concentration among the three methods at five of the eight stations (Figure 1c).

Verification analyses of CS results revealed a large percentage of false positive wells for TC. Only 93 of 153 positive wells (61%) from IDEXX Quanti-Trays



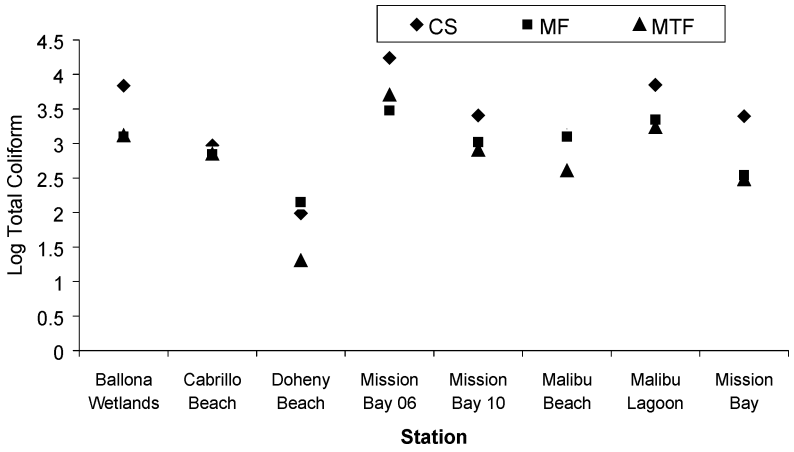
(a)



(b)

Figure 1. Median log counts vs. station for (a) ENT, (b) FC/*E. coli*, and (c) TC.

(*Continued on next page*)



(c)

Figure 1. (*Continued*)

were confirmed to contain bacteria from the TC group, a finding consistent with the higher median concentrations measured using CS. In contrast, *E. coli* was successfully isolated from 100% of the positive wells tested.

For Enterolert, 67 of 71 (94.3%) fluorescing wells and 5 of 55 (9.1%) empty wells were confirmed to contain ENT. Only 3 of the 21 tested wells exhibiting weak fluorescence contained ENT.

MTF exhibited the highest within-method variability of the three methods across all three classes of bacterial indicators (Table III). In contrast, MF exhibited the lowest variability among the three methods. When method variability was expressed in terms of a confidence interval surrounding existing California water quality standards, values from MTF measurements needed to be three times the numerical single sample standard to be 95% confident that the true number of indicator bacteria in the sample had exceeded the standard. For MF, a single measured value only

TABLE III

Confidence intervals for each method applied to concentrations at California's single sample standards values

	Enterococci (Std = 104)				Fecal coliform/*E. coli* (Std = 400)				Total coliform (Std = 10,000)			
	Above		Below		Above		Below		Above		Below	
Method	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%	95%	99%
CS	166	217	65	50	624	809	256	197	14,955	18,880	5,297	6,681
MF	146	178	73	61	767	1,120	208	143	14,696	18,365	5,445	6,800
MTF	209	314	52	34	856	1,331	186	120	21,409	33,266	3,000	4,670

needed to be 50% above the standard to be 95% confident that the true value exceeded the standard.

There appeared to be little effect associated with individual laboratories. Only two laboratories had more than one ENT sample for which the result differed by greater than a half log unit from the group median. One of those laboratories later was found to have a defective incubator that did not hold the proper temperature. For FC, two laboratories performing MF and one lab performing MTF reported values above the target range, but this reflected the comparatively large number of CS observations, which measure only *E. coli* and reduced the grand median. When comparisons were limited to within-method median, no results differed by more than half a log unit from the interlaboratory median. Comparisons were limited to within-method for TC because of CS bias for this indicator. In that comparison, no lab results differed by more than a half log unit from the group median.

4. Discussion

All three methods produced similar results for ENT. Verifications confirmed both false positives and false negatives using CS, but both rates were small relative to the within-laboratory measurement variability. This is consistent with Fleischer (1990) and Noble *et al.* (2003a), but differs from Kinzelman *et al.* (2003), who found nearly 50% false positives for ENT verifications with CS. However, Kinzelman *et al.* suggested that their false positives occurred primarily for weakly fluorescing wells, which is consistent with our results that 86% of weakly fluorescing wells did not contain ENT. The CS manufacturer presently recommends that poorly fluorescing wells not be counted, which was the protocol used by laboratories in this study.

One shortcoming of our study was that we did not have the resources to perform confirmation testing on all three methods. The very nature of this study, in which laboratories donated their time and supplies, necessarily constrained confirmation tests to CS. Each of the methods has characteristics that may affect the reliability of results depending on the type of water sample analyzed and the number and type of cells in the water. For example, a recent study in southern California using ambient marine water samples found false-positive rates for ENT using MF similar to and in some cases greater than those observed for CS in our study (Ferguson *et al.*, in press).

The only large difference among methods observed was the severe overestimation of TC density using CS relative to the other two methods. The high rate of false positives likely results from interference by non-coliform organisms, such as *Aeromonas*, *Vibrio, Pseudomonas*, and *Flavobacteria* spp., which are known to metabolize MUG (Pisciotta *et al.*, 2002; Landre *et al.*, 1998; Davies *et al.*, 1995; Hidalgo *et al.*, 1977). In practical terms, though, this overestimation seems to have little effect on beach warning systems in California, as the TC standard is so high

that the standard is almost never exceeded without FCs or ENT also exceeding the standard (Noble *et al*., 2003b).

While such a systematic methodological error is of concern, we found that the largest source of error was attributable to data processing. These data entry errors occurred for samples that the laboratories were aware were part of an intercalibration exercise, in which their results would come under greater scrutiny. Prior to the data analysis presented in this article, preliminary screening indicated that results from four of the labs differed by an order of magnitude from those of other labs. Upon inspection of original laboratory data sheets, we discovered that these labs failed to correct for dilution before data submission (which we subsequently corrected before conducting the analysis in this article). We also found that a fifth lab misaligned the sample numbers on the bottle with their internal tracking numbers, leading to values being submitted with the wrong sample number (again confirmed by examination of the original laboratory data sheets and corrected before our data analysis). These labs produced comparable data to all other labs after correcting for data submission errors. However, in typical applications, data from other labs is not available for comparison. These errors would have gone undetected and errant results would have been reported to managers for use in regulatory or public health decisions. These data management errors were far larger and more prevalent than any variability introduced by method or commingling of data across laboratories.

Acknowledgments

The authors thank the many individuals and agencies that provided the time and resources to make this study possible.

References

Abbott, S., Caughley, B. and Scott, G.: 1998, 'Evaluation of Enterolert® for enumeration of enterococci in the marine environment', *New Zealand J. Mar. Freshwater Res.* **32**, 505–515.

APHA (American Public Health Association): 1995, *Standard Methods for the Examination of Water and Wastewater*, 18th Edition, A. D. Eaton, L. S. Clesceri and A. E. Greenberg (eds), Washington, DC.

Bej, A. K., McCarty, S. C. and Atlas, R. M.: 1991, 'Detection of coliform bacteria and *Escherichia coli* by multiplex polymerase chain reaction: Comparison with defined substrate and plating methods for water quality monitoring', *Appl. Environ. Microbiol.* **57**, 2429–2432.

Budnick, G. E., Howard, R. T. and Mayo, D. R.: 1996, 'Evaluation of Enterolert® for enumeration of enterococci in recreational waters', *Appl. Environ. Microbiol.* **62**, 3881–3884.

Conover, W. J. and Iman, R. L.: 1981, 'Rank transform as a bridge between parametric and nonparametric statistics', *Am. Statistician* **35**, 125–141.

Covert, T. C., Shadix, L. C., Rice, E. W., Haines, J. R. and Freyberg, R. W.: 1989, 'Evaluation of the autoanalysis Colilert test for detection and enumeration of total coliforms', *Appl. Environ. Microbiol.* **55**, 2443–2447.

Davies, C. M., Apte, S. C., Peterson, S. M. and Stauber, J. L.: 1995, 'Possible interference of lactose-fermenting marine vibrios in coliform *B*-D-galactosidase assays', *J. Appl. Bacteriol.* **78**, 287–393.

Eckner, K. F.: 1998, 'Comparison of membrane filtration and multiple-tube fermentation by Colilert® and Enterolert® methods for detection of waterborne coliform bacteria, *Escherichia coli*, and enterococci used in drinking and bathing water quality monitoring in southern Sweden', *Appl. Environ. Microbiol.* **64**, 3079–3083.

Edberg, S. C., Allen, M. J., Smith, D. B. and Kriz, N. J.: 1990, 'Enumeration of total coliforms and *Escherichia coli* from source water by defined substrate technology', *Appl. Environ. Microbiol.* **8**, 366–369.

Ferguson, D. M., Moore, D. F., Getrich, M. A. and M. H. Zhowandai: in press, 'Enumeration and speciation of enterococci found in marine and intertidal sediments and coastal water in southern California', *Journal of Applied Microbiology*.

Fleisher, J. M.: 1990, 'The effects of measurement error on previously reported mathematical relationships between indicator organism density and swimming associated illness: A quantitative estimate of bias', *Int. J. Epidemiol.* **19**, 1100–1106.

Francy, D. S. and Darner, R. A.: 2000, 'Comparison of methods for determining *Escherichia coli* concentrations in recreational waters', *Water Res.* **24**, 2770–2778.

Hidalgo, C., Reyes, J. and Goldschmidt, R.: 1977, 'Induction and properties of B-galactosidase and B-galactosidase permease in *Pseudomonas* BAL-31', *J. Bacteriol.* **129**, 821–829.

Kinzelman, J., Ng, C., Jackson, E., Gradus, S. and Bagley, R.: 2003, 'Enterococci as indicators of Lake Michigan recreational water quality: Comparison of two methodologies and their impacts on public health regulatory events', *Appl. Environ. Microbiol.* **69**, 92–96.

Landre, J. P., Gavriel, A. A. and Lamb, A. J.: 1998, 'False-positive coliform reaction mediated by *Aeromonas* in the Colilert defined substrate technology system', *Lett. Appl. Microbiol.* **26**, 352–354.

Noble, R. T., Weisberg, S. B., Leecaster, M. K., McGee, C. D., Ritter, K., Walker, K. O. and Vainik, P. M.: 2003a, 'Comparison of beach bacterial water quality indicator measurement methods', *Environ. Monit. Assess.* **81**, 301–312.

Noble, R. T., Moore, D. F., Leecaster, M., McGee, C. D. and Weisberg, S. B.: 2003b, 'Comparison of total coliforms, fecal coliforms, and enterococci bacterial indicator response for ocean recreational water quality testing', *Water Res.* **37**, 1637–1643.

Palmer, C. J., Tsai, Y., Lang, A. L. and Sangermano, L. R.: 1993, 'Evaluation of Colilert®-marine water for detection of total coliforms and Escherichia coli in the marine environment', *Appl. Environ. Microbiol.* **59**, 786–790.

Pisciotta, J. M., Rath, D. F., Stanek, P. A., Flanery, D. M. and Harwood, V. J.: 2002, 'Marine bacteria cause false-positive results in the Colilert-18 rapid identification test for *Escherichia coli* in Florida waters', *Appl. Environ. Microbiol.* **68**, 539–544.

Schiff, K. C., Weisberg, S. B. and Raco-Rands, V. E.: 2002, 'Inventory of ocean monitoring in the Southern California Bight', *Environ. Manage.* **29**, 871–876.