Receiver Operating Characteristic Curve Analysis of Beach Water Quality Indicator Variables

Ann Michelle Morrison,[1] Kelly Coughlin,[1,2] James P. Shine,[1] Brent A. Coull,[3] and Andrea C. Rex[2]*

Environmental Science and Engineering Program, Department of Environmental Health,[1] and Department of Biostatistics,[3] Harvard School of Public Health, Boston, Massachusetts 02115, and Massachusetts Water Resources Authority, Charlestown Navy Yard, Boston, Massachusetts 02129[2]

Receiver operating characteristic (ROC) curve analysis is a simple and effective means to compare the accuracies of indicator variables of bacterial beach water quality. The indicator variables examined in this study were previous day's *Enterococcus* density and antecedent rainfall at 24, 48, and 96 h. Daily *Enterococcus* densities and 15-min rainfall values were collected during a 5-year (1996 to 2000) study of four Boston Harbor beaches. The indicator variables were assessed for their ability to correctly classify water as suitable or unsuitable for swimming at a maximum threshold *Enterococcus* density of 104 CFU/100 ml. Sensitivity and specificity values were determined for each unique previous day's *Enterococcus* density and antecedent rainfall volume and used to construct ROC curves. The area under the ROC curve was used to compare the accuracies of the indicator variables. Twenty-four-hour antecedent rainfall classified elevated *Enterococcus* densities more accurately than previous day's *Enterococcus* density ($P = 0.079$). An empirically derived threshold for 48-h antecedent rainfall, corresponding to a sensitivity of 0.75, was determined from the 1996 to 2000 data and evaluated to ascertain if the threshold would produce a 0.75 sensitivity with independent water quality data collected in 2001 from the same beaches.

Swimming-related illnesses can result from exposure to or ingestion of pathogens originating from human or animal feces. Many of these pathogens cannot be directly measured, and bodies of water often contain several different pathogens, making enumeration of each impractical (14). To protect public health from contaminated water, beach water quality criteria are based on indicator organisms, such as *Enterococcus* spp., that, if present in high densities, have been shown to be associated with an elevated risk of swimming-related illness (3–5, 9).

An effective indicator should be both sensitive and specific in predicting the presence and absence of all possible pathogens. Sensitivity is defined in this study as the ability of an indicator variable to correctly classify a beach as unsuitable for swimming (geometric mean *Enterococcus* density > 104 CFU/100 ml). Specificity is defined as the ability of an indicator variable to correctly classify a beach as suitable for swimming (geometric mean *Enterococcus* density ≤ 104 CFU/100 ml). An appropriate indicator variable should be easy to use and provide an accurate assessment of water quality in a timely fashion.

The U.S. Environmental Protection Agency's (EPA's) *Ambient Water Quality Criteria for Bacteria—1986* recommends that marine recreational waters not exceed a geometric mean density of *Enterococcus* of 35 CFU/100 ml and that no single sample exceed a maximum of 104 CFU/100 ml (13). An important limitation of using *Enterococcus* densities to manage beaches is that enumeration requires a minimum of 24 h.

Because microbial densities in the waters being tested can change significantly in this time (2, 6), results of samples collected 24 h previously may not provide an accurate assessment of water quality and exposure at the time of use. A recent study demonstrated that approximately 70% of single samples that exceeded a bacterial threshold standard at Huntington Beach, Calif., lasted less than 1 h and that approximately 40% of single samples that exceeded a bacterial threshold at this beach lasted less than 10 min (1).

Environmental variables can also be used as indicators of elevated pathogen concentrations. In many areas, rainfall is a major factor affecting beach water quality due to the impact of contaminated storm water and sewer overflows on the shoreline (10). Rainfall-based alert curves can be constructed to describe a statistical relationship between rainfall events and pathogen concentrations at a specific site (15). This analysis corresponds to a multiple regression model that includes the amount of rainfall, the storm duration time, the number of dry days between rainfall events, and the lag time between rainfall event and beach pathogen appearance. Rainfall-based alert curves can require input data on several variables that may or may not be easy and cost-effective to collect.

Receiver operating characteristic (ROC) curves were developed in the field of statistical decision theory and later used in the field of signal detection for analyzing radar images during World War II (7). ROC curves enabled radar operators to distinguish between an enemy target, a friendly ship, and noise. ROC curves assess the value of diagnostic tests by providing a standard measure of the ability of a test to correctly classify subjects. The biomedical field uses ROC curves extensively to assess the efficacy of diagnostic tests in discriminating between healthy and diseased individuals (12). A test with good dis-

* Corresponding author. Mailing address: Massachusetts Water Resources Authority, Charlestown Navy Yard, 100 First Ave., Boston, MA 02129. Phone: (617) 788-4708. Fax: (617) 788-4889. E-mail: Andrea.Rex@mwra.state.ma.us.
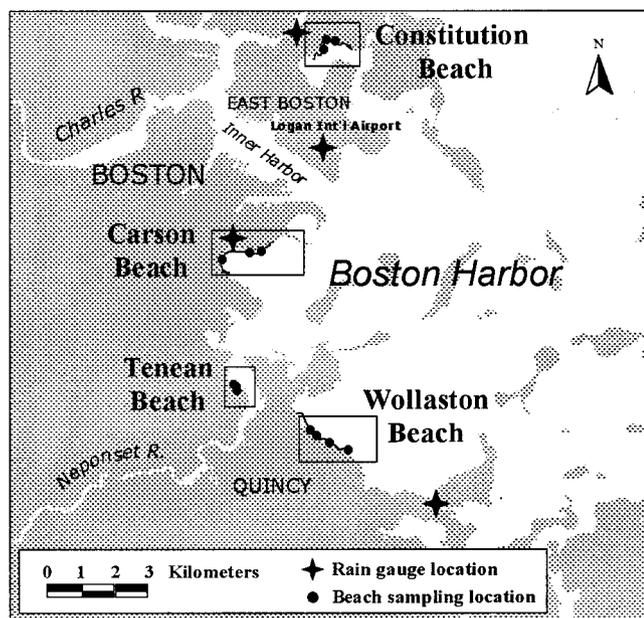
FIG. 1. Map of study beaches and rain gauge locations.

criminatory ability has both a high sensitivity and high specificity. ROC curves can (i) assess the overall discriminatory ability of different potential indicator variables by generating a common metric for comparison and (ii) aid in the selection of a specific value of an indicator variable to use as a threshold or limit that provides a desired trade-off in sensitivity and specificity. With respect to beach water quality indicator variables, ROC curves can quantify the overall effectiveness of different indicator variables to correctly or incorrectly classify a beach as suitable for swimming and generate a single metric by which the different indicator variables can be compared.

Our objective was to determine the ability of *Enterococcus* density to correctly classify beach water quality as suitable or unsuitable for swimming 24 h after sample collection and to compare this to the ability of antecedent-rainfall volumes to correctly classify beach water quality as suitable or unsuitable for swimming. Another goal was to determine a maximum value, or threshold, for each of the indicator variables that would provide an optimal trade-off in sensitivity and specificity. The use of 104 CFU/100 ml in this study as a delineator of water suitable or unsuitable for swimming was based solely on current EPA recommendations. This work is not intended to comment on the appropriateness of this number or *Enterococcus* spp. as an indicator organism.

MATERIALS AND METHODS

Beach water sample collection and analysis. Four marine beaches in Boston Harbor were studied: Constitution, Carson, Tenean, and Wollaston beaches (Fig. 1). Each of the study beaches is potentially impacted by different pollution sources, including treated and untreated combined sewer overflows (CSOs), storm drain discharges, and publicly owned treatment works (POTW) (Table 1). Beaches were sampled during the swimming season (mid-June to Labor Day weekend) from 1996 to 2001. Three to four multiple samples were collected at each beach on each sampling day. Samples were generally collected in the morning. Attempts were made to collect samples every day, but occasionally samples could not be collected due to an unusually low tide. Sample collectors waded into approximately 1 m of water, and water samples were collected directly into 250-ml sterile bottles at approximately 0.2 m depth up-current of the person collecting the sample. Samples were immediately placed in a cooler at <10°C and processed at the laboratory within 6 h. *Enterococcus* was enumerated by either of two membrane filtration methods. In 1996 to 1997, standard method 9230C (American Public Health Association, 1995) using m*Enterococcus* agar incubated at 35°C for 48 h was used, and, in 1998 to 2001, EPA method 1600 (EPA 1997) using mEI agar incubated at 35°C for 24 h was used. The two different culture methods do not significantly differ in their abilities to measure *Enterococcus* density (L. Wong and M. Gofshteyn, Massachusetts Water Resources Authority unpublished data) (11).

Rainfall data. Rainfall was measured at three locations with Sierra Misco tipping bucket rain gauges (see Fig. 1 for locations). The rain gauges used were already in place for other purposes when we began this study, and, because sudden, localized, summer rain showers have an impact directly at the beach, the gauge closest to each study beach was selected as the most appropriate. The gauges were calibrated to read and electronically record rainfall in 0.01-in increments. Rainfall records were stored electronically on-site, and 15-min-interval rainfall sums were obtained via telemetry with QuadraScan software (data storage and telemetry provided by ADS Equipment, Inc.; software was provided by QuadraScan 1995). The frequency of data collection allowed for precise calculation of rainfall sums prior to beach water sample collection. Rainfall collection gauges were located within 4.4 km of the sampling locations. To test whether more universally available rain data could effectively be used in this analysis, rainfall data for 2001 were obtained from the Logan Airport National Weather Service (NWS) rain gauge, which is within a 9.1-km radius of the four beaches.

Data preparation. In this study, multiple samples were collected from several sites along each beach. To eliminate pseudoreplication, we calculated the geometric mean *Enterococcus* density for each beach on each day and used this value in all subsequent analysis. As described earlier, we defined water suitable for swimming as water with a geometric mean *Enterococcus* density of 104 CFU/100 ml or less, which corresponds to the recommended EPA single-sample maximum. This definition of water suitable for swimming is not in conflict with the EPA recommendation that the geometric mean of *Enterococcus* counts not exceed 35 CFU/100 ml because that value is intended for samples collected over time. The replicate samples analyzed here can be considered a snapshot of a single point in time and analogous to a single sample.

Construction of ROC curves. An example of the information used to create an ROC curve is shown in Fig. 2. The hypothetical distributions of *Enterococcus* spp. in water suitable for swimming and water unsuitable for swimming on a given day are plotted with respect to previous day's *Enterococcus* density as the indicator variable. Due to the distribution of the indicator variable population, the two distributions overlap. For each unique value of the indicator variable there is an associated true-positive rate (TPR; sensitivity) and a false-positive rate (FPR; 1 − specificity), which are shown in Fig. 2. The ROC curve is a function relating the TPR to the FPR. The TPR for a given indicator variable is the true proportion of days having an *Enterococcus* density above 104 CFU/100 ml as identified

TABLE 1. Local pollution sources at the four study beaches, including the number of CSOs and storm drains

Beach	No. of:		Other possible pollution sources unrelated to rain
	CSOs	Storm drains	
Constitution	1 (treated)	5	Wildlife, adjacent yacht club (potential source of illegally discharged boat waste)
Carson	7 (untreated)	7	Wildlife, adjacent yacht club
Tenean	1 (treated)	3	Neponset River, adjacent marina
Wollaston	0	8	POTW[a] outfall discharging 0.8 miles offshore (closed July 1998), two yacht clubs on beach

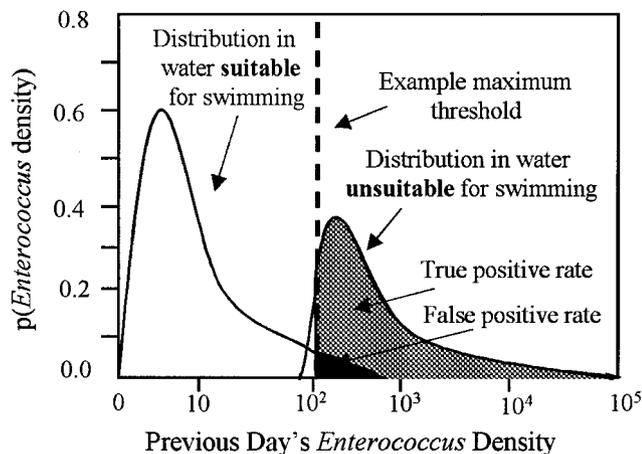[a] POTW, publicly owned treatment works.

FIG. 2. Hypothetical distributions of previous day's *Enterococcus* in water suitable for swimming and water unsuitable for swimming. Shown are a hypothetical TPR and FPR associated with a potential threshold value. An ROC curve was constructed by plotting the TPR (grey) and FPR (black) for each unique value of the indicator variable.
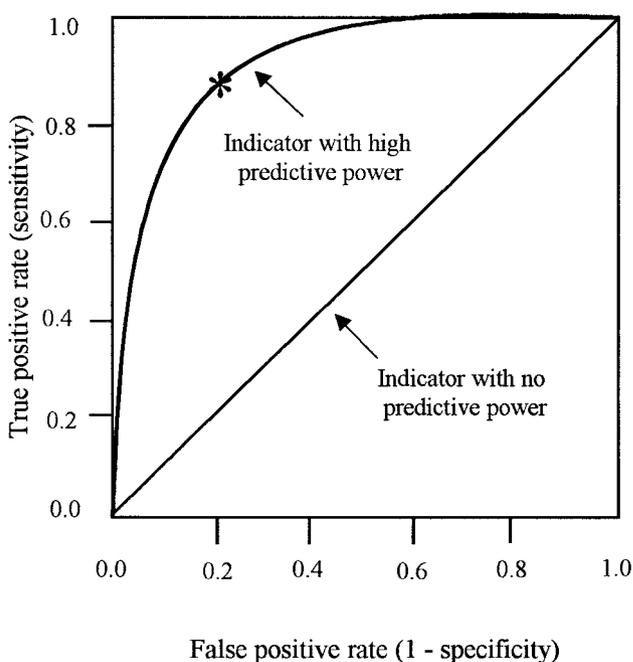

FIG. 3. ROC curve constructed by plotting the TPR and FPR associated with each unique value of the indicator variable. *, hypothetical TPR-FPR pair. An indicator variable with high discriminatory ability will have a curve with an AUC near 1, and an indicator variable with low discriminatory ability will have an AUC near 0.5.

by that indicator variable. A perfect TPR (1.0) means that all incidences of *Enterococcus* densities above 104 CFU/100 ml occur above the threshold value of the indicator variable. The FPR for a given indicator variable is the proportion of days having an *Enterococcus* density less than or equal to 104 CFU/100 ml incorrectly identified as being above 104 CFU/100 ml by that particular variable. This means that if a threshold value for an indicator variable has a zero FPR there will be zero incidences of an *Enterococcus* density below 104 CFU/100 ml above this threshold level of the indicator variable. The result of the calculation of the TPR and FPR is a unique pair of data points for each unique value of an indicator variable. For a given data set, the observed values of TPR and FPR are plotted for each unique value of the indicator variable to form a sample ROC curve (Fig. 3).

ROC curves provide three important ways to examine the efficacy of a test. First, ROC curves can evaluate the overall ability of an indicator variable to make correct classifications, as in this study, of whether beach water quality is suitable or unsuitable for swimming. The shape of the ROC curve shows the discriminatory ability of the indicator variable examined. An ideal indicator variable has a curve with an area under the curve (AUC) of 1, and an indicator variable with poor discriminatory ability has an AUC near 0.5 (Fig. 3). The second function of an ROC curve is to allow direct comparison of the abilities of different indicator variables to make correct classifications (e.g., beach water quality as suitable or unsuitable for swimming) through a common metric, the AUC. Finally, ROC curves facilitate the selection of a maximum threshold value of the indicator variable that best balances sensitivity and specificity. For each indicator variable, the curve will show the trade-off between sensitivity and specificity for any potential threshold.

Several different indicator variables for assessing beach water quality were examined, including 24-, 48-, and 96-h antecedent rainfall and previous day's *Enterococcus* density. ROC curves were constructed with Microsoft Excel software (Microsoft Corporation 1999). The resulting paired TPR-FPR points were plotted to form a sample ROC curve. Sample AUC values were calculated according to the trapezoid rule, and these values were verified by using the Mann-Whitney procedures described by DeLong et al. (8). Standard error values were calculated and chi-square tests of hypotheses involving ROC curves were performed by the methods of Delong et al. (8).

RESULTS

Water quality at the four beaches. The four beaches in this study are located in densely populated urban areas and are affected in various degrees by urban coastal pollution. Table 2 describes the water quality at the study beaches. All four study beaches complied with EPA criteria for a designated marine bathing beach. However, the beaches occasionally failed to meet EPA's limit of 104 CFU of *Enterococcus*/100 ml for individual samples. Carson and Constitution beaches had the best water quality; Tenean and Wollaston beaches had the poorest water quality. Water quality was generally poorer in wet weather due to the effects of urban storm water runoff and CSOs.

ROC curve analyses. Two of the ROC curves that were calculated for Constitution Beach are shown in Fig. 4 as examples of the curves that were constructed. Visual inspection of each curve provides a great deal of information regarding the utility of an indicator variable. The instantaneous slope of the curve demonstrates the rate at which the sensitivity changes with changes in specificity. An ideal indicator variable would have a curve with a steep incline from the origin that plateaus quickly to a high level of sensitivity. A steep slope means that the sensitivity increases more than the specificity decreases. In Fig. 4A, the slope of the ROC curve for previous day's *Enterococcus* density over the data is gradual, indicating that for almost every level of the variable there is an increase in sensitivity with a decrease in specificity of a similar magnitude. In contrast, the slope of the ROC curve for 48-h antecedent rainfall (Fig. 4B) is steep from the origin, indicating that the sensitivity increases more than the specificity decreases.

The AUC is a useful parameter for evaluating the overall value of an indicator variable and facilitates comparison of different indicator variables by providing a common metric. Table 3 presents the AUCs for the four indicator variables examined in this study. These values are the result of data averaged over all beaches. The AUC for each rain-based indicator is greater than the AUC for previous day's *Enterococcus*

TABLE 2. Descriptive statistical analysis on the *Enterococcus* data to characterize each beach[a]

Beach	5-yr geometric mean *Enterococcus* density (CFU/100 ml) (range)[b]	Total no. of sampling days	No. of sampling days exceeding:					
			35 CFU/100 ml			104 CFU/100 ml		
			Dry[c]	Wet[c]	Total	Dry	Wet	Total
Constitution	6 (0–18,300)	342	6	43	49	1	22	23
Carson	5 (0–2,340)	344	5	31	36	1	10	11
Tenean	11 (0–3,000)	300	16	64	80	4	23	27
Wollaston	13 (0–6,300)	362	20	60	80	5	21	26

[a] The values calculated were 5-year geometric means and daily geometric means for each beach. The numbers of days that each beach had a geometric mean *Enterococcus* density greater than 35 and 104 CFU/100 ml were determined, and these totals were classified by the weather conditions in which they occurred.
[b] Multiple samples were collected daily at each beach. *Enterococcus* counts from multiple samples were pooled by using a daily geometric mean. Ranges show individual values.
[c] Dry weather is defined as no rainfall for at least 48 h prior to sampling. Wet weather is any rainfall in the previous 48 h.

density (chi-square test, $P = 0.079$ for the comparison of differences in AUCs based on 24-h antecedent rainfall and previous day's *Enterococcus* density).

We also examined the ability of each indicator variable at each beach to correctly classify beach water quality as suitable or unsuitable for swimming. This enabled us to ascertain the utilities of different indicator variables at beaches affected by different environmental conditions. Table 4 lists the AUCs for each beach and each of the four indicator variables. The AUCs from previous day's *Enterococcus* density were more variable than the AUCs from the rain indicator variables; however, within each indicator variable there was no statistically significant difference in AUCs among beaches (chi-square test, $P > 0.05$).

Low interannual variability of AUCs indicates that the ability of the indicator variable to classify beach water quality is consistent over time. AUCs calculated by year from data for all beaches combined are shown in Table 5. Forty-eight-hour antecedent rainfall showed the least variability among years. All of the rain variables were more consistent among years than previous day's *Enterococcus* density.

A useful aspect of ROC analysis is the determination of the sensitivity and specificity associated with particular values of an indicator variable. We evaluated the sensitivity and specificity associated with previous day's *Enterococcus* density at three threshold values: 104, 35, and >0 CFU/100 ml (*Enterococcus* density above the level of detection) (Table 6). Table 6 shows that none of the previous day's *Enterococcus* density threshold values provide both a high sensitivity and high specificity at the four Boston Harbor beaches studied.

To evaluate the specificity of the indicator variables at a common sensitivity, we compared them at a uniform and moderately high sensitivity of 0.75 (Table 7). Ideally, the specificity associated with these thresholds would also be high. As an example we chose a sensitivity of 0.75 because it is analogous to the 75% confidence limit recommended by the EPA to calculate a beach-specific single-sample maximum for *Enterococcus* density (13). Forty-eight-hour antecedent rainfall had high specificity associated with a 0.75 sensitivity and had a practical threshold value of 0.21 in. of rain, which is typical of a light-to-moderate rainstorm. In contrast, 0.75 sensitivity was associated with a previous day's *Enterococcus* density of only 7 CFU/100 ml, which is only slightly higher than the limit of detection, and a very low specificity of 0.57.

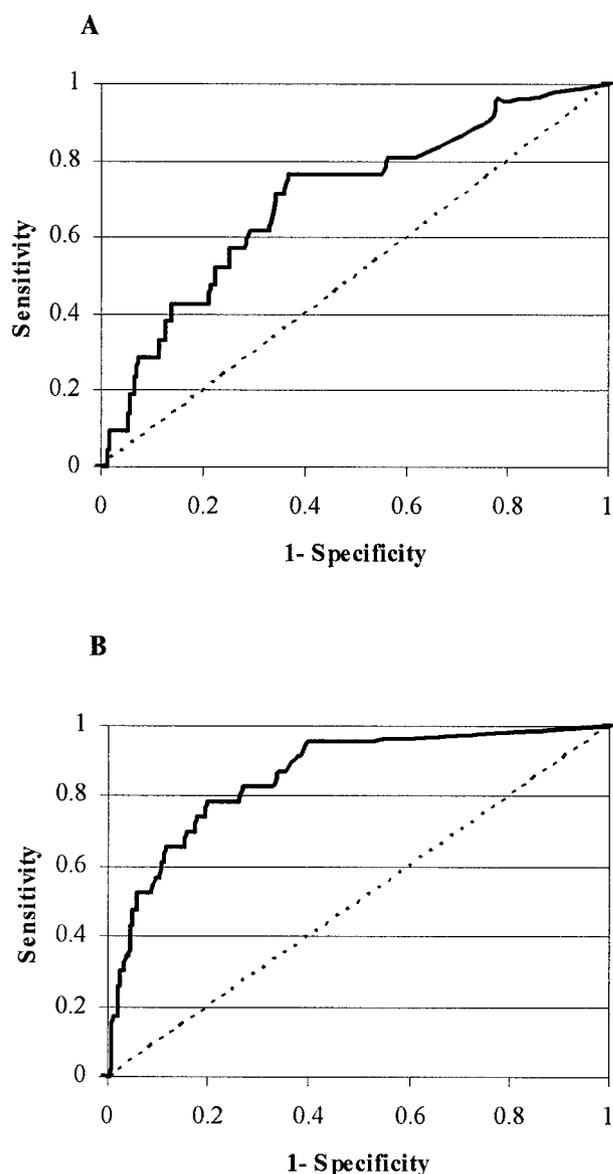We used water quality monitoring data from 2001 to deter-

FIG. 4. ROC curves of previous day's *Enterococcus* density (A) and 48-h antecedent rainfall (B) for Constitution Beach. These curves demonstrate the difference in the slopes of the ROC curves of previous day's *Enterococcus* density and 48-h antecedent rainfall.

TABLE 3. AUC associated with each indicator variable[b]

Indicator variable	AUC (SE)	n[a]
Previous day's *Enterococcus* density	0.73 (0.029)	561
24-h antecedent rainfall	0.81 (0.026)	186
48-h antecedent rainfall	0.81 (0.027)	177
96-h antecedent rainfall	0.79 (0.026)	225

[a] *n*, number of unique data points that went into the calculation of AUC.
[b] Results are from data averaged over all beaches.

mine if the 0.75 sensitivity threshold value for 48-h antecedent rainfall shown in Table 7, developed with data from 1996 to 2000, could classify bacterial water quality on an independent data set with a similar sensitivity and specificity (Table 8). The previous day's *Enterococcus* density threshold of 104 CFU/100 ml in the 2001 data analysis shows sensitivity and specificity values roughly similar to those determined from the 1996 to 2000 data, which are shown in Table 6. The threshold of 104 CFU/100 ml is important because it is the recommended threshold used by beach managers to post swimming advisories. To test the utility of a rain-based indicator variable, we used a somewhat cruder but more practical measure of antecedent rainfall, rainfall reports from a NWS rain gauge in the Boston Harbor area. The threshold value for 48-h antecedent rainfall used with the NWS data was 0.21 in. of rainfall during the preceding two calendar days, excluding any rainfall from the day of sample collection. Table 8 shows that the sensitivity of 48-h antecedent rainfall was near 0.75 at Wollaston and Tenean beaches but that the sensitivity was much less than 0.75 at Carson and Constitution beaches. The sensitivity of a previous day's *Enterococcus* density of 104 CFU/100 ml was similar to that from the 1996 to 2000 data and less accurate than 48-h antecedent rainfall at every beach except Carson. There were very few incidences of *Enterococcus* densities above 104 CFU/100 ml at any of the beaches in 2001, and particularly at Carson and Constitution beaches.

TABLE 4. AUCs for the four indicator variables at each beach

Indicator variable	Beach	AUC (SE)	n[a]
Previous day's *Enterococcus* density	Carson	0.81 (0.067)	127
	Constitution	0.70 (0.061)	139
	Tenean	0.75 (0.050)	162
	Wollaston	0.68 (0.061)	231
24-h antecedent rainfall	Carson	0.81 (0.029)	62
	Constitution	0.84 (0.046)	62
	Tenean	0.80 (0.049)	59
	Wollaston	0.81 (0.052)	70
48-h antecedent rainfall	Carson	0.86 (0.065)	82
	Constitution	0.86 (0.039)	82
	Tenean	0.78 (0.056)	78
	Wollaston	0.79 (0.056)	92
96-h antecedent rainfall	Carson	0.86 (0.039)	106
	Constitution	0.80 (0.053)	103
	Tenean	0.77 (0.053)	99
	Wollaston	0.77 (0.046)	113

[a] *n*, number of unique data points that went into the calculation of AUC.

TABLE 5. Interannual variability of AUCs for the four indicator variables[b]

Indicator variable	Yr	AUC (SE)	n[a]
Previous day's *Enterococcus* density	1996	0.55 (0.088)	109
	1997	0.78 (0.067)	167
	1998	0.78 (0.051)	165
	1999	0.60 (0.136)	114
	2000	0.75 (0.045)	184
24-h antecedent rainfall	1996	0.79 (0.068)	41
	1997	0.79 (0.081)	44
	1998	0.82 (0.048)	61
	1999	0.90 (0.069)	45
	2000	0.78 (0.048)	56
48-h antecedent rainfall	1996	0.81 (0.061)	55
	1997	0.79 (0.081)	53
	1998	0.82 (0.051)	77
	1999	0.86 (0.077)	55
	2000	0.78 (0.053)	67
96-h antecedent rainfall	1996	0.75 (0.044)	65
	1997	0.82 (0.081)	61
	1998	0.83 (0.049)	98
	1999	0.81 (0.089)	72
	2000	0.71 (0.049)	95

[a] *n*, number of unique data points that went into the calculation of AUC.
[b] Results are from data averaged over all beaches.

DISCUSSION

Beach managers post swimming advisories at the beaches in this study based on results from the previous day's *Enterococcus* density. If the previous day's *Enterococcus* density at a beach exceeds 104 CFU/100 ml of water, it is assumed that the current day's *Enterococcus* density exceeds 104 CFU/100 ml and a swimming advisory is posted. To our knowledge, this study is the first attempt to quantitatively assess the accuracy of this method of beach management and to compare the accuracy to that of an alternative indicator variable, namely, antecedent rainfall.

The beaches in this study are relatively clean and conform to the EPA's 30-day geometric mean criterion of an *Enterococcus*

TABLE 6. Sensitivity and specificity associated with previous day's *Enterococcus* density threshold values at each beach in this study[a]

Beach	*Enterococcus* threshold (CFU/100 ml)	Sensitivity	Specificity
Carson	104	0.33	0.99
	35	0.44	0.91
	>0	1.0	0.19
Constitution	104	0.14	0.95
	35	0.29	0.90
	>0	0.95	0.20
Tenean	104	0.29	0.93
	35	0.57	0.76
	>0	1.0	0.13
Wollaston	104	0.28	0.94
	35	0.48	0.80
	>0	1.0	0.05

[a] An ideal threshold would have both a sensitivity and specificity close to 1.

TABLE 7. Comparison of the specificities associated with 0.75 sensitivity for each of the indicator variables[a]

Indicator variable	75% sensitivity threshold value	Specificity
Previous day's *Enterococcus* density	7 CFU/100 ml	0.57
24-h antecedent rainfall	0.04 in.	0.69
48-h antecedent rainfall	0.21 in.	0.80
96-h antecedent rainfall	0.44 in.	0.73

[a] The threshold values necessary for 0.75 sensitivity are also given. Results are from data averaged over all beaches.

density less than 35 CFU/100 ml. Statistical analyses of the 5-year data set showed that increased *Enterococcus* densities at these beaches were associated with wet weather. Therefore, we chose to compare rainfall indicator variables of beach water quality to the currently used previous day's *Enterococcus* density.

Using sample ROC curves we were able to compare the abilities of previous day's *Enterococcus* density and antecedent rainfall variables to correctly classify beach water quality as suitable or unsuitable for swimming by the common metric of the AUC. This analysis suggests that antecedent rainfall was both a more sensitive and more specific indicator of poor bacterial water quality than *Enterococcus* densities collected 24 h previously at the four beaches in the study. Each of the rainfall variables examined consistently had larger AUCs and less variability among beaches and among years than previous day's *Enterococcus* density.

The sensitivity and specificity associated with potential threshold values of previous day's *Enterococcus* density varied widely among beaches. This variability compromises the use of a uniform *Enterococcus* threshold as the sole indicator of water quality. The sensitivities associated with a threshold of previous day's *Enterococcus* density of 104 CFU/100 ml ranged from 0.14 to 0.33. Using previous day's *Enterococcus* density, beach managers posted the swimming advisory accurately only one-third of the time or less. At Boston Harbor beaches, the threshold value of previous day's *Enterococcus* density greater than 104 CFU/100 ml has a very high false-negative rate and is,

therefore, a poor indicator variable to protect public health at these beaches if used as the only management criterion.

The desired level of sensitivity can be determined a priori; however, the results of this a priori selection may prove impractical. A sensitivity of 75% implies that 25% of the incidences of *Enterococcus* densities above 104 CFU/100 ml will not be correctly discriminated by the threshold value. Increasing the desired sensitivity with a lower threshold value will decrease the probability of failing to predict an *Enterococcus* density above 104 CFU/100 ml, but specificity will decrease as sensitivity increases, meaning that the beach will be closed more often. With respect to public health, sensitivity is a more important parameter than specificity.

Beach management balances two competing priorities: (i) maintaining the beach as an accessible recreational resource by minimizing unnecessary swimming advisories and (ii) minimizing public health risk by appropriately issuing swimming advisories. A constructive approach is to evaluate the practical consequences of using different indicator variables, in addition to directly comparing AUC values. Thresholds of rainfall variables providing a desired sensitivity offer a more reasonable trade-off between beach accessibility and public health risk.

The sensitivities and specificities associated with 0.21 in. of 48-h antecedent rainfall determined with the 2001 validation data at Tenean and Wollaston beaches were very close to the expected values based on ROC curve analysis of the 1996 to 2000 data. However, at beaches with few incidences of *Enterococcus* densities above 104 CFU/100 ml, such as Carson and Constitution, an accurate indicator variable is difficult to identify because the sources of contamination may not be associated with rain and may be transient. This underscores the necessity of gathering enough monitoring data to adequately characterize a beach.

In conclusion, this study has shown that, for Boston Harbor beaches, previous day's *Enterococcus* density was frequently a poor indicator of elevated *Enterococcus* densities at the time of use, had a high false-negative rate, and may not adequately protect bathers from increased pathogen concentrations. Antecedent rainfall was both a more sensitive and specific indicator of *Enterococcus* densities above 104 CFU/100 ml. Ante-

TABLE 8. Utility of 48-h antecedent rainfall for classifying bacterial water quality[a]

Indicator variable	Beach	No. of incidences of >104[b]	No. of correct predictions	Sensitivity (%)	Specificity (%)
48-h rainfall > 0.2 in. during preceding 2 calendar days	All beaches	31	19	61	74
	Carson	5	1	20	69
	Constitution	5	2	40	70
	Tenean	12	9	75	80
	Wollaston	9	7	78	77
Previous day's *Enterococcus* density > 104 CFU/100 ml	All beaches	31	13	42	94
	Carson	5	2	40	96
	Constitution	5	0	0	93
	Tenean	12	5	42	90
	Wollaston	9	6	67	96

[a] The 0.75 sensitivity threshold for 48-h antecedent rainfall that was calculated from the 1996 to 2000 data (Table 7) was used with the 2001 daily water quality data to determine if a sensitivity of 0.75 is achieved when this threshold is applied to independent data from the same beaches. Previous day's *Enterococcus* density with a threshold of 104 CFU/100 ml was also examined to determine if the sensitivity of 0.14 to 0.33 calculated in the 1996 to 2000 data is reproducible in an independent data set.
[b] Number of incidences of an *Enterococcus* density of >104 CFU/100 ml.

cedent-rainfall threshold values did not result in unacceptably high posting rates and provided more spatial and temporal consistency. Furthermore, antecedent rainfall is easily available at the time a beach manager must make a decision about issuing a swimming advisory.

This study has also demonstrated that ROC analysis is a simple and practical tool for quantifying the ability of indicator variables to assess beach water quality and that ROC curves facilitate the selection of a beach-specific threshold for an indicator variable that yields a desirable sensitivity and specificity. ROC analysis can effectively evaluate the relationships between a risk-related variable and candidate indicator variables used to actually manage the beach.

REFERENCES

1. **Boehm, A. B., S. B. Grant, J. H. Kim, S. L. Mowbray, C. D. McGee, C. D. Clark, D. M. Foley, and D. E. Wellman.** 2002. Decadal and shorter period variability of surf zone water quality at Huntington Beach, California. Environ. Sci. Technol. **36:**3885–3892.

2. **Brenniman, G. R., S. H. Rosenberg, and R. L. Northrop.** 1981. Microbial sampling variables and recreational water quality standards. Am. J. Public Health **71:**283–289.

3. **Cabelli, V. J., A. P. Dufour, M. A. Levin, L. J. McCabe, and P. W. Haberman.** 1979. Relationship of microbial indicators to health effects at marine bathing beaches. Am. J. Public Health **69:**690–696.

4. **Cabelli, V. J., A. P. Dufour, L. J. McCabe, and M. A. Levin.** 1983. A marine recreational water quality criterion consistent with indicator concepts and risk analysis. J. Water Pollut. Control Fed. **55:**1306–1314.

5. **Cabelli, V. J., A. P. Dufour, L. J. McCabe, and M. A. Levin.** 1982. Swimming-associated gastroenteritis and water quality. Am. J. Epidemiol. **115:**606–616.

6. **Cheung, W. H. S., K. C. K. Chang, and R. P. S. Hung.** 1991. Variations in microbial densities in beach waters and health-related assessment of bathing water quality. Epidemiol Infect. **106:**329–344.

7. **Collison, P.** 1998. Of bombers, radiologists, and cardiologists: time to ROC. Heart **80:**215–217.

8. **DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson.** 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics **44:**837–845.

9. **Dufour, A. P.** 1984. Bacterial indicators of recreational water quality. Can. J. Public Health **75:**49–56.

10. **Haile, R. W., J. S. Witte, M. Gold, R. Cressey, C. McGee, R. C. Milikan, A. Glasser, N. Harawa, C. Ervin, P. Harmon, J. Harper, J. Dermand, J. Alamillo, K. Barrett, M. Nides, and G.-Y. Wang.** 1999. The health effects of swimming in ocean water contaminated by storm drain runoff. Epidemiology **10:**355–363.

11. **Messer, J. W., and A. P. Dufour.** 1998. A rapid, specific membrane filtration procedure for enumeration of enterococci in recreational waters. Appl. Environ. Microbiol. **64:**678–680.

12. **Metz, C.** 1978. Basic principles of ROC analysis. Semin. Nucl. Med. **8:**283–298.

13. **U.S. Environmental Protection Agency.** 1986. Ambient water quality criteria for bacteria. EPA440/5–84–002. U.S. EPA Office of Water, Washington, D.C.

14. **U.S. Environmental Protection Agency.** 2000. Implementation guidance for ambient water quality criteria for bacteria—1986. Draft EPA-823-D-00–001. U.S. EPA Office of Water, Washington, D.C.

15. **U.S. Environmental Protection Agency.** 1999. Review of potential modeling tools and approaches to support the BEACH program. 823-R-99–002. U.S. EPA Office of Science and Technology, Washington, D.C.