Use of predictive models and rapid methods to nowcast bacteria levels at coastal beaches

By Donna S. Francy*

U.S. Geological Survey, Ohio Water Science Center
6480 Doubletree Avenue, Columbus, OH 43229
**Corresponding author: dsfrancy@usgs.gov*

The need for rapid assessments of recreational water quality to better protect public health is well accepted throughout the research and regulatory communities. Rapid analytical methods, such as quantitative polymerase chain reaction (qPCR) and immunomagnetic separation/adenosine triphosphate (ATP) analysis, are being tested but are not yet ready for widespread use.

Another solution is the use of predictive models, wherein variable(s) that are easily and quickly measured are surrogates for concentrations of fecal-indicator bacteria. Rainfall-based alerts, the simplest type of model, have been used by several communities for a number of years. Deterministic models use mathematical representations of the processes that affect bacteria concentrations; this type of model is being used for beach-closure decisions at one location in the USA. Multivariable statistical models are being developed and tested in many areas of the USA; however, they are only used in three areas of the Great Lakes to aid in notifications of beach advisories or closings. These "operational" statistical models can result in more accurate assessments of recreational water quality than use of the previous day's Escherichia coli (E. coli) *concentration as determined by traditional culture methods. The Ohio Nowcast, at Huntington Beach, Bay Village, Ohio, is described in this paper as an example of an operational statistical model. Because predictive modeling is a dynamic process, water-resource managers continue to collect additional data to improve the predictive ability of the nowcast and expand the nowcast to other Ohio beaches and a recreational river. Although predictive models have been shown to work well at some beaches and are becoming more widely accepted, implementation in many areas is limited by funding, lack of coordinated technical leadership, and lack of supporting epidemiological data.*

Keywords: bacteria models, *E. coli*, recreational water, beach water quality, statistical models, multiple linear regression

Introduction

The need for rapid assessments of recreational water quality to better protect public health is well accepted throughout the research and regulatory communities. In the USA, concentrations of fecal-indicator bacteria–*Escherichia coli* (*E. coli*) or enterococci for fresh waters and enterococci for marine waters—are used to issue beach advisories or closings. The widely acknowledged shortcoming of using this approach is that standard culture methods for these bacteria take at least 18−24 hrs before results are available. The beach is posted with an advisory or closing or is determined to be acceptable for swimming on the basis of the previous day's concentration of *E. coli* or enterococci. Sanitary conditions

Aquatic Ecosystem Health & Management, 12(2):177–182, 2009.
DOI: 10.1080/14634980902905767

may change overnight and even throughout the day (Boehm et al., 2002), so beach managers may issue water-quality advisories based on outdated information of current public-health risk. Because of this time-lag issue, water-resource managers are seeking solutions that provide near-real-time estimates of recreational water quality.

Rapid analytical methods

One obvious solution is the use of a rapid analytical method. For fecal-indicator bacteria, three types of rapid methods, described in detail by Noble and Weisberg (2005), are undergoing development and testing: (1) surface and whole-cell recognition methods, (2) molecular methods, and (3) enzyme/substrate methods. The first two types are being tested by the U.S. Geological Survey (USGS) and others—immunomagnetic separation/adenosine triphosphate (IMS/ATP) and real-time quantitative polymerase-chain reaction (QPCR).

Lee and Deininger (2004) originally developed the IMS/ATP method, a surface recognition method. IMS/ATP involves mixing the sample with antibody-coated magnetic beads specific to cell-surface antigens of the bacteria species of interest, separating the tagged bacteria from the remainder of the sample in a magnetic field, and using a bioluminescent assay to obtain a concentration of bacteria based on the amount of ATP. (ATP is the energy molecule present in all cells and is a measure of cell viability.) The results are reported in relative light units (RLUs). In a study in a recreational river, investigators found significant relations ($r = 0.62$ and $r = 0.77$) between RLUs and the standard culture method for *E. coli* ($n = 228$) and enterococci ($n = 35$), respectively (Bushon et al., 2009).

Quantitative-PCR is a molecular method that targets DNA specific to an organism, such as *E. coli* and enterococci. In QCPR, preprocessing steps involve lysing of cells, extraction of DNA, and removal of inhibitory substances. Results may be reported in cell equivalents (Haugland et al., 2005) or other units. By using known amounts (serial dilutions of bacteria cells) of DNA of an organism of interest, one first must create a standard curve to obtain the cell equivalents in your sample. Water quality measured by QPCR for enterococci was shown to predict swimming-associated health effects at two Great Lakes beaches (Wade et al., 2006) and be significantly related to results obtained by use of standard culture methods (Haugland et al., 2005).

Although rapid analytical methods are a promising solution, they are not yet ready for widespread use. IMS/ATP and QPCR, along with other rapid analytical methods, are currently being tested at southern California beaches (J.F. Griffith, Southern California Coastal Water Research Project, Costa Mesa, CA, pers. comm.). For QPCR, current limitations include the high cost of equipment and the lack of portability. Research is needed to establish uniform DNA standards, cell suspensions for standard curves, and interpretation procedures, as well as procedures to account for DNA-extraction efficiency and inhibition. The IMS/ATP method has been tested in a limited geographic area only, and the relation between RLUs and the standard culture method may need to be established for each location or type of beach. For both methods, studies are needed to establish the link between rapid analytical results with health risk.

Predictive models

While waiting for rapid analytical methods to be available for routine use, some beach managers have adopted predictive models. Predictive models use variable(s) that are easily and quickly measured as surrogates for concentrations of fecal-indicator bacteria. The models currently in use at bathing beaches include rainfall based alerts, deterministic models, and multivariable statistical models.

Rainfall-based alerts have been used by several communities for a number of years. In this type of system, the beach manager establishes a rainfall threshold by examining the relation between fecal-indicator bacteria concentrations and rainfall in historical data, either qualitatively or statistically. The beach is posted if the rainfall amount exceeds a beach-specific threshold. Officials with the City of Stamford, CT, have been using rainfall-based alerts for beaches on Long Island Sound for almost 20 years (Kuntz, 2006). Beaches are closed with $\geq$25 mm of rainfall under normal or above-normal rainfall conditions or $\geq$12.5 mm under drought conditions; beaches are also closed if floatables are observed in the water. In southern California, investigators found an increase in coastal marine bacteria concentrations associated with almost all rainfalls >6 mm and with every rainfall >25 mm (Ackerman and Weisberg, 2003). Officials in southern California issue warnings for the public to avoid recreational water contact for 3 days after rainfall events; San Diego County uses a rainfall threshold of >5 mm (http://sdcounty.ca.gov/deh/lwq/beachbay/index.html). Other communities in the USA that use rainfall advisories include Door County and Milwaukee, WI; Delaware (U.S. Environmental Protection Agency (USEPA), 1999); Myrtle Beach, SC; and Boston, MA—there are probably more.

Deterministic models use mathematical representations of the processes that affect bacteria concentrations. These processes include transport by currents; dieoff from sunlight, predation, etc.; and deposition and resuspension. Sources of bacterial contamination need to be known and quantified. The only deterministic model in use in the USA for beach-closure decisions that this author is aware of is the Regional Bypass Model (RBM) in the New York-New Jersey Harbor. The model simulates a discharge of a known quantity of bacteria from a sewage spill and calculates the water-quality response at beaches and shellfish harvesting areas (HydroQual, Inc., 1998; USEPA, 1999). Although not yet used for beach-closure decisions, hydrodynamic models are being developed for beaches at two locations in Lake Michigan. At both locations, bacterial water quality is known to be affected by the drainage from a nearby stream or ditch. Investigators are now able to track the movement of stream water and forecast lake circulation (http://www.glerl.noaa.gov/res/glcfs/bd/); adding the bacterial component is the next step.

Techniques such as multiple linear regression (MLR) are used to develop multivariable statistical models (hereinafter "statistical models") on the basis of relations between fecal-indicator bacteria concentrations and variables known or suspected to affect their concentrations at a particular beach. Unlike deterministic models, the sources of fecal contamination do not need to be identified–unknown contamination sources are common situations at many beaches. Statistical models are being developed and tested in many areas of the USA; however, they are used for beach closure or advisory decisions at only three locations, all at Great Lakes beaches. These model-based advisories are the Swim Advisory Forecast Estimate (Project SAFE), SwimCast, and the Ohio Nowcast. For Project SAFE, a predictive model is used to forecast bacterial concentrations at five Lake Michigan beaches in northwest Indiana (http://www.glsc.usgs.gov/projectSAFE.php). The model has been incorporated into local management programs since 2005, and includes the variables wave height, turbidity in a nearby ditch and in the lake, rainfall, and chlorophyll concentrations. For SwimCast (M. Pfister, Lake County Health Dept., Waukegan, IL, pers. comm.), automated meteorological and water-quality sensors are installed in or near the surf zone at four Lake Michigan beaches in Lake County, IL (http://www.co.lake.il.us/health/ehs/SwimCastDataAP.asp). Models may include the variables air and water temperature, wind speed and direction, rainfall, relative humidity, wave height, lake level, and insolation (light energy). The specific variables that are used in SwimCast are most strongly associated with the fecal-indicator bacteria. Two models (Forest Park Beach and Illinois Beach State Park) have been used since 2005 for beach-closure decisions after model creation in 2004. Two additional models were developed in 2006 and 2007 (Rosewood Beach and Waukegan South Beach), respectively; the Rosewood Beach model was operational in 2007. Wave height is a variable that has been associated with the fecal-indicator bacteria at all four beach locations, whereas 24 and (or) 36-hour rainfall is associated with fecal-indicator bacteria at only two beach locations because of differences in drainage inflows. For the SwimCast, beach water quality is monitored daily in the morning and in the afternoon, five days per week, and the models can distinguish between morning and afternoon conditions. Additionally, predictions can be made on an hourly basis, a beneficial feature because the fecal-indicator bacteria concentrations and hydrometeorological conditions change at these beaches throughout the day. For the Ohio Nowcast, scientists and local agencies have been working to develop predictive models for five Lake Erie beaches; currently, beach advisories on the nowcast are based on a predictive model for one of those beaches (http://www.ohionowcast.info/index.asp). Details of the nowcast are presented below as an example of an operational beach statistical model.

The Ohio Nowcast

The Ohio Nowcast has been operational since 2006 for Huntington, Bay Village, OH, a suburb of Cleveland. A model based on data from 2000−2006 was developed for use in the nowcast in 2007. The model included the variables log turbidity, wave height, rainfall (radar and airport), and day of the year as predictors for *E. coli* concentrations. The output from the model was the probability that the Ohio single-sample bathing-water standard for *E. coli*, 235 colony-forming units per 100 ml (CFU 100 ml^{-1}), would be exceeded. A threshold probability of 30% was established for Huntington on the basis of historical data. The threshold was the probability associated with too great a risk to allow swimming—a probability that would warrant the issuance of a bathing-water advisory.

From May 21 through September 3, 2007, the daily nowcast for Huntington was operated by the Cuyahoga County Board of Health (CCHB) with assistance from the USGS. Samples were collected by the CCBH at Huntington by 9:00 a.m., seven days week. The CCBH compiled data on rainfall from the National Weather Service over the Internet, measured wave height and turbidity at the beach, and entered data into a Fortran program that computed the output from the model. The probability of exceeding the bathing-water standard was posted on the Ohio Nowcast web site by 9:30 each morning, along with current water-quality conditions. During 2007, the Huntington model resulted in a similar percentage of correct responses (82.7%) than that found by use of the previous day's *E. coli* concentrations (80.2%) (current method); this percent difference was smaller than expected and not as great as that found during 2006 (http://www.ohionowcast.info/ohionowcastfindoutmore.htm). The model did, however, result in a greater sensitivity (53%) than that found by use of the current method (31%) but with similar specificity. (The sensitivity is the percentage of events where the model correctly predicted exceedance of the bathing-water standard among days where the standard was exceeded. The specificity is the percentage of events where the model correctly predicted non-exceedance of the standard.) Investigators will continue to work on improving the predictive ability of the Huntington model in 2008. For example, the 2000−2007 data were split into two or three segments by date. A two-segment split resulted with data placed into two groups: (1) May 21–July 23 and (2) July 24–Sept 3. After examining the grouped data, it was noted that the percentages of false negatives decreased over time and the percentages of false positives increased over time (Francy and Darner, 2008). A false negative occurs when the model predicts that the standard will not be exceeded, but cultural results indicate that the *E. coli* concentration was actually greater than the standard. In contrast, a false positive occurs when the model predicts that the standard was exceeded, but bacteria-culture results indicate that the concentration was less than the standard. The pattern of false negatives and positives at Huntington would lead one to believe that a bi-phase seasonal model may be suitable for use in the nowcast in 2008.

From May 21 through Sept 10, 2007, data were collected by the Northeast Ohio Regional Sewer District (NEORSD) for testing of a predictive model for possible future inclusion in the Ohio Nowcast. Samples were collected at two beaches in Cleveland, OH—Edgewater and Villa Angela. The Edgewater model included the variables log turbidity, wave height, radar rainfall, and lake level, with a threshold probability of 29%. The Villa Angela model included the variables log turbidity, airport rainfall, and water temperature, with a threshold probability of 32%. During 2007, the Edgewater model resulted in a greater percentage of correct responses (82.1%) than that found by use of the current method (66.7%); the model also resulted in both higher sensitivity and specificity than the current method. These results show that predictions based on the Edgewater model can be used to issue advisories through the Ohio Nowcast system in 2008. In contrast, at Villa Angela, the model resulted in correct responses only 61.3% of the days monitored; this percentage was lower than that achieved by use of the current method (74.6%). At Villa Angela, other options will be examined, such as the use of a rapid analytical method for *E. coli* for issuing beach advisories.

At Lakeshore Beach in northeast Ohio (Ashtabula, OH), model results were similar to those at Villa Angela (E.E. Bertke, USGS, Columbus, OH, pers. comm.). At both Villa Angela and Lakeshore, the bathing-water standard was exceeded on the majority of days sampled. It appears, therefore, that predictive models do not work as well at beaches where the standard is exceeded on the majority of days. In fact, models do not work well at relatively clean beaches either (for example, <5 exceedances per season), because of the lack of data on days when the standard is exceeded. Models seem to work best at moderately contaminated beaches. Regardless, predictive modeling is a dynamic process wherein water-resource managers must continue to collect additional data to improve the predictive ability of the models from year to year.

Although recent attention has been focused on coastal beaches, it is important to keep in mind that many recreational activities take place in recreational rivers. The Cuyahoga River is a valuable recreational resource in northeast Ohio, especially within the Cuyahoga Valley National Park (CVNP). Recreational water quality within CVNP has been a concern to park managers since the inception of the park (Brady, 2007). Management would like to encourage use of the river by canoeists and waders when the water quality, based on *E. coli* concentrations, is within acceptable limits. To reach this goal, scientists and park employees have been testing the use of rapid assessments of recreational water quality. Data were collected during 2004−2006 to develop predictive models for river sites. Of all the variables tested—*E. coli* by IMS/ATP, turbidity, streamflow, and rainfall—turbidity alone was found to be the best predictor of *E. coli* concentrations at one river site. During 2007, the turbidity model produced more correct responses (81%) than the current method (62%), and it had a higher sensitivity and specificity (A.M.G. Brady, USGS, Columbus, OH, pers. comm.). Adding the turbidity model for the CVNP site to the Ohio Nowcast is a possible next step.

Discussion and conclusions

Recognizing the need for accurate, rapid assessments of recreational water quality, communities in the USA have applied different solutions. Solutions are beach specific and include the use of rapid analytical methods and (or) predictive models. Predictive models include the simpler rainfall-based alerts, deterministic models that require data on sources of contamination, or more complicated multivariable statistical models. There is no single rapid-assessment solution that will work for every community and every beach. The type of rapid assessment chosen for a recreational resource depends on the technical expertise of available staff and willingness of the community to accept nontraditional methods.

Although predictive models have been shown to work well at some beaches and are becoming more widely accepted, there are several roadblocks to their widespread implementation. The USEPA BEACH Grant program (USEPA, 2007) has provided funding to develop and implement predictive models; but continued funding is needed to maintain and refine models. Not all agencies have the technical expertise to develop and maintain models or the management capability to ensure that models are used for beach advisory or closure decisions. Given that, what is needed is coordinated leadership on the regional or state level to provide oversight and technical expertise and ensure that high quality, consistent data are collected. Some agencies are hesitant to adopt new methods, especially methods that are more complicated and different from traditional methods. Training is needed; work is needed to test the use of different endpoints for predictive models, such as enterococci for marine and fresh waters. And last, but most important, there is a lack of supporting epidemiological data. Studies are therefore needed to link results from predictive models, to the rate at which recreational users are contracting gastrointestinal illnesses.

References

Ackerman, D., Weisberg, S.B., 2003, Relationship between rainfall and beach bacterial concentrations on Santa Monica Bay beaches. J. Water Health 1(2), 85–89.

Boehm, A.B., Grant, S.B., Kim, J.H., Mowbray, S.L., McGee, C.D., Clark, C.D., Foley, D.M., Wellman, D.E., 2002. Decadal and shorter period variability of surf zone water quality at Huntington Beach, California. Environ. Sci. Tech. 36(18), 3885–3892.

Bushon, R.N., Brady, A.M., Likirdopulos, C.A., Cireddu, J.V., 2009. Rapid detection of *Escherichia coli* and enterococci in recreational water using immunomagnetic separation/adenosine triphosphate technique. J. Applied Microbiol. 106, 432–441.

Brady, A.M.G., 2007. Rapid method for *Escherichia coli* in the Cuyahoga River. U.S. Geological Survey Open-File Report 2007–1210. Available at http://pubs.usgs.gov/of/2007/1210/pdf/ofr20071210.pdf.

Francy, D.S., Darner, R.A., 2008. Nowcasting beach advisories at Ohio Lake Erie beaches. U.S. Geological Survey Open-File Report 2008-1427.

Haugland, R.A., Siefring, S.C., Wymer, L.J., Brenner, K.P. Dufour, A.P., 2005. Comparison of *Enterococcus* measurements in freshwater at two recreational beaches by quantitative polymerase chain reaction and membrane filter culture analysis. Water Research 39, 559–568.

HydroQual, Inc., 1998. Modeling Evaluations and Users Guide. HydroQual, Inc., Mahwah, NJ.

Kuntz, J.E., 2006. Predictability of swimming prohibitions by observational parameters. Presented at the New England Interstate Water Pollution Control Commission Beach Closure Workshop, 2006 February 7. Available at http://www.neiwpcc.org/beachworkshop/index.asp.

Lee, J., Deininger, R.A., 2004. Detection of *E. coli* in beach water within 1 hour using immunomagnetic separation and ATP bioluminescence. Luminescence 19, 31–36.

Noble, R.T., Weisberg, S.B., 2005. A review of technologies for rapid detection of bacteria in recreational waters. J. Water Health 3, 381–392.

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

U.S. Environmental Protection Agency, 2007. Beaches Environmental Assessment and Coastal Health Act. Federal Register 72 (7), January 11, 2007, 1320–1325.

Wade, T.J., Calderon, R.L., Sams, E., Beach, M., Brenner, K.P., Williams, A.H., Dufour, A.P., 2006. Rapidly measured indicators of recreational water quality are predictive of swimming-associated gastrointestinal illness. Environ. Health Perspectives 114(1), 24–28.