2013

Predictive Models for *Escherichia coli* Concentrations at Inland Lake Beaches and Relationship of Model Variables to Pathogen Detection

Donna S. Francy
U.S. Geological Survey, dsfrancy@usgs.gov

Erin A. Stelzer
U.S. Geological Survey

Joseph W. Duris
Michigan Water Science Center

Amie M. G. Brady
U.S. Geological Survey

John H. Harrison
U.S. Geological Survey

See next page for additional authors

AEM
Journals.ASM.org

Predictive Models for *Escherichia coli* Concentrations at Inland Lake Beaches and Relationship of Model Variables to Pathogen Detection

Donna S. Francy,[a] Erin A. Stelzer,[a] Joseph W. Duris,[b] Amie M. G. Brady,[a] John H. Harrison,[a] Heather E. Johnson,[b] Michael W. Ware[c]

U.S. Geological Survey, Ohio Water Science Center, Columbus, Ohio, USA[a]; Michigan Water Science Center, Lansing, Michigan, USA[b]; U.S. Environmental Protection Agency, National Exposure Research Laboratory, Cincinnati, Ohio, USA[c]

Predictive models, based on environmental and water quality variables, have been used to improve the timeliness and accuracy of recreational water quality assessments, but their effectiveness has not been studied in inland waters. Sampling at eight inland recreational lakes in Ohio was done in order to investigate using predictive models for *Escherichia coli* and to understand the links between *E. coli* concentrations, predictive variables, and pathogens. Based upon results from 21 beach sites, models were developed for 13 sites, and the most predictive variables were rainfall, wind direction and speed, turbidity, and water temperature. Models were not developed at sites where the *E. coli* standard was seldom exceeded. Models were validated at nine sites during an independent year. At three sites, the model resulted in increased correct responses, sensitivities, and specificities compared to use of the previous day's *E. coli* concentration (the current method). Drought conditions during the validation year precluded being able to adequately assess model performance at most of the other sites. *Cryptosporidium*, adenovirus, *eaeA* (*E. coli*), *ipaH* (*Shigella*), and *spvC* (*Salmonella*) were found in at least 20% of samples collected for pathogens at five sites. The presence or absence of the three bacterial genes was related to some of the model variables but was not consistently related to *E. coli* concentrations. Predictive models were not effective at all inland lake sites; however, their use at two lakes with high swimmer densities will provide better estimates of public health risk than current methods and will be a valuable resource for beach managers and the public.

Current bacterial indicator methods used to monitor recreational water quality take 18 to 24 h before results are available. For example, in Ohio, a recreational water quality advisory is posted if the previous day's *Escherichia coli* concentration is above the single-sample bathing-water standard of 235 CFU per 100 ml (http://www.odh.ohio.gov/odhprograms/eh/bbeach/beachmon.aspx). Because bacterial concentrations might change overnight and even throughout the day (1, 2), water quality advisories may not reflect the current public health risk. Due to this time lag issue, water resource managers are seeking solutions that provide near-real-time estimates of recreational water quality (3).

Predictive models are recommended by the U.S. Environmental Protection Agency (EPA) to improve the timeliness and accuracy of recreational water quality assessments (4). Predictive models use rapid or easily measured environmental and water quality variables to yield the probability that the state standard will be exceeded or to estimate densities of bacterial indicators, such as *E. coli*. Predictive models have been used to provide near-real-time assessments ("nowcasts") of recreational water quality at Great Lakes beaches and are used as the basis for posting advisories at three Lake Erie beaches in Ohio (http://www.ohionowcast.info), three Lake Michigan beaches in Illinois (http://www.lakecountyil.gov/Health/want/Pages/SwimCast.aspx), and two Lake Michigan beaches in Wisconsin (http://www.wibeaches.us/). These models are also used to predict levels of *E. coli* in recreational rivers, including the Cuyahoga River in Ohio (http://www.ohionowcast.info) and the Chattahoochee River in Georgia (http://ga2.er.usgs.gov/bacteria/default.cfm).

Although predictive models have been used at coastal beaches, little work has been done to develop and test their use in inland recreational lakes and reservoirs. Inland water bodies are popular swimming and boating destinations throughout the United States. For example, in the Ohio State Park system, there are 78 designated swimming beaches, the majority of which are inland lakes (5). Alum Creek State Park, near Columbus, OH, and included in this study, receives over 2,000,000 visitors annually, similar to visitation rates at several Lake Erie beaches.

In spite of widespread use of inland recreational waters, there is also a paucity of information on the occurrence of pathogens that cause disease in these waters. Data on pathogens at inland beaches are needed in order to establish the link between results from predictive models and the density of pathogens that increase human health risk. In 2007 and 2008, pathogens associated with outbreaks of illness acquired from ambient recreational waters in the United States included *E. coli* O157:H7, *Shigella*, *Cryptosporidium*, and norovirus (6). In recreational epidemiological studies, diarrhea and respiratory ailments are the common reported health outcomes, and it is believed that these may be associated with a variety of unidentified enteric viruses (7). Avian species, such as gulls, which are commonly found at beaches, have been known to carry pathogens that can infect humans, such as *Campylobacter* spp. (8) and *Cryptosporidium* and *Giardia* (9, 10, 11). While ruminant species, such as cows and deer, are the primary reservoir of pathogenic *E. coli*, these pathogens have also been found in humans, swine, and other domestic and wild animals as host organisms (12). Markers of pathogenic *E. coli* have been

Received 28 September 2012 **Accepted** 21 December 2012
Published ahead of print 4 January 2013
Address correspondence to Donna S. Francy, dsfrancy@usgs.gov.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.02995-12.

doi:10.1128/AEM.02995-12

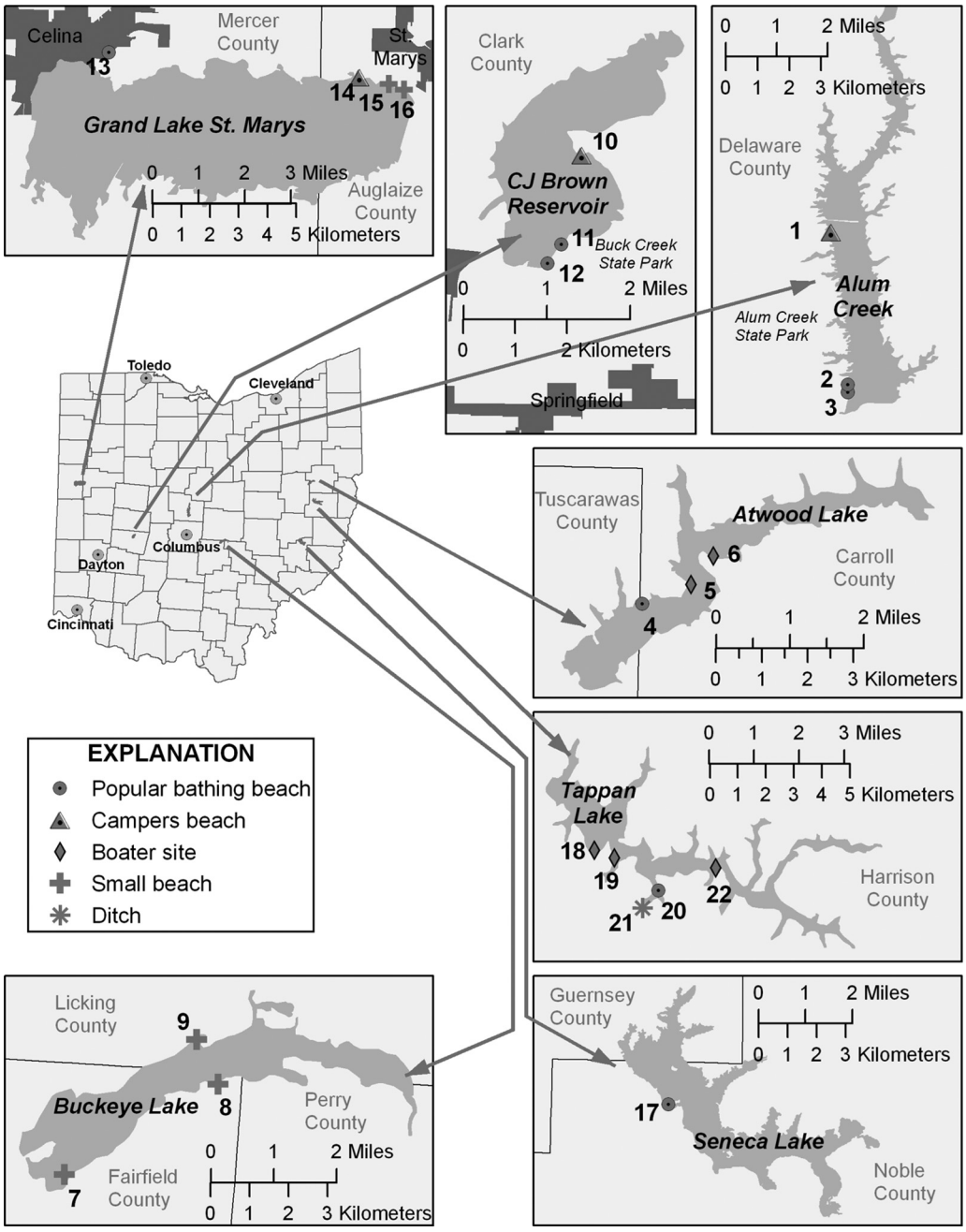


FIG 1 Inland lake sampling sites, 2010 to 2012.

found in river systems that can influence beach environments (13). *Salmonella* species are recognized for having a very large host range that includes humans, birds, and most other warm-blooded animals (14), but gulls and sewage are recognized as important sources of *Salmonella* in recreational waters (15). Unlike pathogenic *E. coli* and *Salmonella*, *Shigella* species are almost exclusively associated with human hosts (16), and thus only direct or indirect human fecal inputs would be sources of *Shigella* at beaches.

This article describes the results of research by the U.S. Geological Survey (USGS), in cooperation with local and state agencies, to determine if predictive models can be used to provide near-real-time assessments of water quality at inland recreational waters that are more accurate than current methods. Sampling was done 4 days/week at eight inland recreational lakes over three recreational seasons in Ohio to develop and validate models for future implementation of nowcast systems. At five sites, a subset of samples was analyzed for bacterial, protozoan, and viral pathogens to begin to understand the link between *E. coli* concentrations, environmental and water quality variables, and health risk from pathogens at inland lakes.

MATERIALS AND METHODS

Site descriptions. The study was done at 22 sites at eight inland recreational lakes in Ohio (Fig. 1 and Table 1). These included eight sites on

TABLE 1 Summary statistics of *Escherichia coli* concentrations at inland lake sites, 2010 to 2012

Site no.	Short name	Sampling yrs	No. of sampling days	Daily *E. coli* concn, 2010–2012 (MPN/100 ml)			% of days bathing-water standard was exceeded in:		Model developed from 2010-2011 data	Model validated in 2012
				Median	Minimum	Maximum	2010 and/or 2011	2012		
1	Alum Campers	2010–2012	143	10	1	2,400	7.7	5.8	Yes	Yes
2	Alum North	2010–2012	144	25	1	2,400	8.7	0	Yes	Yes
3	Alum Central	2010–2012	144	22	1	2,400	6.5	1.9	Yes	Yes
4	Atwood Main	2010–2012	190	41	1	2,400	25	25	Yes	Yes
5	Atwood Islands	2010	66	9	<1	360	3.0	NA[a]		
6	Atwood Cove	2010-2011	131	8	<1	>2,400	4.6	NA		
7	Buckeye Brooks	2010-2011	94	73	5	>2,400	23	NA	Yes	
8	Buckeye Fairfield	2010-2011	95	51	3	980	13	NA	Yes	
9	Buckeye Crystal	2010-2011	95	120	20	>2,400	31	NA	Yes	
10	Buck Creek Campers	2010–2012	78	4	<1	110	0	0		
11	Buck Creek North	2010–2012	149	31	1	>2,400	12	17	Yes	Yes
12	Buck Creek South	2010–2012	150	44	1	3,300	15	19	Yes	Yes
13	Eastview	2010	35	5	<1	330	2.8	NA		
14	GLSM Campers	2010–2012	134	49	<1	>2,400	30	3.7	Yes	Yes
15	GLSM West	2010–2012	97	110	4	>2,400	33	33	Yes	
16	GLSM East	2010–2012	135	42	1	>2,400	20	3.7	Yes	Yes
17	Seneca	2011-2012	111	29	1	2,400	8.3	22		
18	Tappan South	2010	65	1	<1	60	0	NA		
19	Tappan Bontrager	2010	65	5	<1	100	0	NA		
20	Tappan Main	2010–2012	190	52	2	4,900	22	20	Yes	Yes
21	Tappan ditch	2010-2011	14	480	34	>2,400	NA	NA		
22	Tappan Beall	2010	65	2	<1	86	0	NA		

[a] NA, not applicable.

popular beaches, three beaches located at campgrounds ("camper's beach"), five sites accessible only by boat ("boater's site"), five small beaches on canal lakes, and a ditch tributary to one of the popular beaches. At Alum Creek State Park (sites 2 and 3) and Buck Creek State Park (sites 11 and 12, located on CJ Brown Reservoir), two sampling sites were established because of the extended length of each beach. The canal lakes, Buckeye Lake and Grand Lake St. Marys (GLSM), are shallow man-made reservoirs constructed in the early 19th century for the Miami and Erie Canal, which connected the Ohio River with Lake Erie. One site (site 21) was included to determine concentrations of pathogens in a ditch that flows into Tappan Main (site 20); the ditch receives treated effluent from a wastewater package plant. Potential sources of fecal contamination at all beaches include birds and other wildlife, swimmers, domestic animals, and storm water runoff. Effluents from septic tanks are potential sources at Buck Creek, and treated wastewater is a potential source at Tappan Main; otherwise, no other point sources have been identified. Alum, Buckeye, Buck Creek, and GLSM are State Park beaches operated by Ohio Department of Natural Resources, Tappan and Atwood recreational sites are operated by the Muskingum Watershed Conservancy District, and Eastview is operated by the City of Celina, OH. Official USGS site names, identification numbers (which correspond to latitudes and longitudes), site descriptions, and agencies responsible for sampling are listed in Table S1 in the supplemental material.

Sample collection and frequency. Data were collected during the recreational seasons (May to September) of 2010 and/or 2011 for development of predictive models, for pathogens in 2011, and for validation of predictive models in 2012.

Samples for *E. coli*, turbidity, and bacterial pathogens (bacterial virulence genes and *Campylobacter*) were collected using the standard grab-sampling technique (17) at 0.6- to 1-m water depths in areas used for swimming. A 500-ml, 1-liter, or 3-liter sterile polypropylene sample bottle was filled with water about 0.3 m below the water's surface and immediately placed on ice. For predictive model development and validation, data were collected 4 days/week (including weekends). The USGS in Columbus, OH (Alum and Buckeye sites), a USGS student in Celina, OH (GLSM sites), and the Clark County Combined Health District (Buck Creek sites) sampled between 6 and 10 a.m. with consistent sampling times at each site. The Muskingum Watershed Conservancy District (MWCD) varied the order of lake sampling and sampled from 6 a.m. to 2 p.m. at the Atwood, Seneca, and Tappan sites. In 2011, afternoon sampling was added at four Alum and Buck Creek sites to determine temporal differences in water quality.

Sampling methods for viral and protozoan pathogens included glass-wool filtration (18, 19) and manual ultrafiltration (20). Glass-wool filtration and manual ultrafiltration were chosen because they represented two types of filtration approaches used for concentrating pathogens: virus adsorption-elution (VIRADEL) and ultrafiltration, respectively. Glass-wool filters (special order from the USDA Agricultural Research Station, Marshfield, WI) concentrate microorganisms by charge interactions. The ultrafilters used were Rexbrane Membrane High-Flux, Rexeed-25S (Asahi Kasei Kuraray Medical Co., Ltd., Japan) with molecular cutoffs of 29,000 Da, surface areas of 2.5 m^2, and fiber inner diameters of 185 μm; they concentrate microorganisms by physical removal. Each sampling apparatus included a peristaltic pump that drew water through 9 m of sterile inlet tubing attached to the middle of a steel bar anchored to the lake or ditch bottom, where water depths were 0.6 to 1 m. On each sampling event, approximately 100 liters of water was sampled through both filters at lake sites. At the ditch site, 100 liters was filtered by ultrafiltration, but only 3 to 4 liters could be filtered through the glass-wool filter before clogging. After ultrafiltration, elution solution (0.01% Tween 80) was recirculated through the sampling apparatus in the field to remove microorganisms from the ultrafilter and collected into an eluate bottle. For glass-wool filtration, the elution step was done in the USGS Ohio Water Microbiology Laboratory in Columbus, OH (Columbus Laboratory).

Sampling events for pathogens included both rain events and dry days at five sites: Atwood Main, Buckeye Brooks, Buckeye Fairfield, Tappan Main, and Tappan ditch. Although a total of 31 samples were collected, they were not consistently analyzed for all pathogens. In addition to regular sampling for pathogens, five field blanks were collected and analyzed for all microorganisms, and seven replicates were collected and analyzed for bacterial pathogens. Replicates for protozoan and virus analyses were not included because of the low probability of a positive result. All field blanks were below the level of detection for bacterial, protozoan, and viral pathogens and *E. coli*. For bacterial pathogens, presence/absence results of the replicates were always in agreement.

Processing and analysis for bacteria. (i) *E. coli* and enterococci. Samples for bacterial indicators were processed or analyzed within 6 h of collection by the agency that collected the sample in a local laboratory using the Colilert Quanti-Tray/2000 method for *E. coli* (IDEXX Laboratories, Inc., Westbrook, ME) and the mEI agar method for enterococci (21). Sample processing and quality control procedures are described elsewhere (17).

(ii) Identification of *Shigella*, *Salmonella*, and pathogenic *E. coli* genes by enrichment and endpoint PCR. Twenty-two samples were analyzed for *Shigella* species, Shiga toxin-producing *E. coli* (STEC), and *Salmonella enterica* virulence genes. In a local laboratory, 100 ml of sample was plated using the mENDO agar method (22) within 6 h of sample collection. The resulting enrichment was enumerated, frozen, and shipped on dry ice to the USGS Michigan Bacteriological Research Laboratory in Lansing, MI (Lansing Laboratory) for further processing. After the plates were thawed for 15 min, the filters were folded in half four times and placed in a bead-beating tube with 0.65 g of 0.1-mm glass beads (Mo Bio Laboratories, Inc., Carlsbad, CA) with the open side facing down. Any liquid present on the plate was added to the bead-beating tube, and sterile deionized water was used to bring the total volume up to 1 ml. Samples were bead beaten for 2 min on high speed and then allowed to sit undisturbed for 5 min (to diminish foam). Bead-beating tubes containing the filters were stored at −70°C until DNA purification. Bead tubes were thawed, pulse vortexed, and further homogenized using a 200-μl pipette tip. DNA extraction was done by drawing off 100 μl for use in the Qiagen (Qiagen, Valencia, CA) DNeasy Gram-negative extraction protocol.

DNA extracted from the mENDO plate served as the template for several PCRs to identify specific toxin and virulence genes. *Shigella* species were identified using adapted methods of Islam et al. (23), targeting the invasion plasmid antigen H (*ipaH*) gene. *Salmonella enterica* was identified using methods adapted from the work of Chiu and Ou (24) to detect the invasion A (*invA*) and *Salmonella* plasmid of virulence (*spvC*) genes. Pathogenic Shiga-toxin producing *E. coli* (STEC) was identified by following the methods of Duris et al. (13) to detect the Shiga toxin 1 and 2 genes (stx_1 and stx_2), the intimin (*eaeA*) gene, and a generic 16S rRNA gene marker for *E. coli* in a four-gene multiplex PCR. *E. coli* O157 was detected using the methods of Osek (25) to detect the gene encoding the O157 surface protein (rfb_{O157}). The bovine-associated heat-labile toxin (LTIIa) and the human-associated heat-stable toxin (STh) were identified using methods adapted from the work of Jiang et al. (26). The porcine-associated heat-stable toxin (STII) was identified using methods adapted from the work of Khatib et al. (27). Details of all PCRs are listed in Table S2 in the supplemental material.

Standard quality assurance and control procedures were followed for all PCRs (28). Detection limits for PCRs were determined using serial dilutions of target chromosomal or plasmid DNA controls. For approximately every 20 samples of any given PCR, PCR positive controls near the detection limit and PCR negative controls (no template reactions) were included. If a reaction failed quality control tests for either of these controls, the reaction was repeated for all samples in the batch.

(iii) Identification of *Campylobacter jejuni* and *Campylobacter coli* by enrichment and endpoint PCR. Twenty-six samples were analyzed for *C. jejuni* and *C. coli* (*Campylobacter*). Selective enrichment for *Campylobacter* was done in the Lansing Laboratory by inoculating 14 ml of Bolton broth with Preston supplement (Oxoid, Cambridge, United Kingdom) with a 0.45-μm-pore-size mixed cellulose ester filter (Advantec MFS, Inc., Dublin, CA) through which 100 ml of sample water was passed (29). Samples were incubated for 4 h at 37°C and then transferred to a 41.5°C incubator for 48 h. After incubation, the growth was pelleted and the supernatant was decanted. The pellet was resuspended in 1 ml of 20% glycerol prepared in one-half-strength phosphate-buffered saline. Glycerol preparations were stored at −70°C until DNA extraction. Pellets from broth cultures were thawed at room temperature, and DNA was extracted using the Qiagen DNeasy Gram-negative extraction protocol. DNA extracted from the Bolton broth enrichment served as the template for a single PCR that detects a 16S rRNA gene fragment specific to *C. jejuni* and *C. coli*.

PCR was performed according to methods adapted from those of Inglis and Kalischuck (30). Details of the PCR are listed in Table S2 in the supplemental material. Quality assurance and quality control practices for *Campylobacter* PCR were the same as those performed for STEC, *Salmonella*, and *Shigella* PCR.

Processing and analysis for viruses and protozoa. (i) Postfiltration processing. Fourteen samples by manual ultrafiltration and 12 samples by glass-wool filtration were analyzed for *Cryptosporidium*, *Giardia*, adenovirus, enterovirus, and norovirus (protozoan and viral pathogens). The glass-wool filters and ultrafiltration eluates were transported to the Columbus Laboratory on ice and processed within 24 h of collection. Microorganisms were eluted from glass-wool filters by use of a beef extract and glycine solution and concentrated by polyethylene glycol (PEG) precipitation as described previously (18, 19). The final concentrate from the glass wool (volumes ranged from 145 to 230 ml) was split into aliquots for shipment for protozoan analysis and storage at −70°C for virus analyses. The ultrafiltration eluate was centrifuged at 3,300 × *g* for 30 min. The eluate pellet was resuspended with a sodium phosphate solution at a volume that completely dissolved the entire pellet (23.5 to 58 ml) for protozoan analysis. The remaining eluate supernatant (volumes ranged from 320 to 655 ml) from the ultrafiltration was flocculated with 40 g PEG and 5.7 g NaCl and processed and stored to obtain a final concentrate for virus analysis.

(ii) Analysis of viruses by qPCR and qRT-PCR. Viral RNA and DNA were extracted from the final concentrates using the QIAamp DNA mini-extraction kit (Qiagen, Valencia, CA) according to the manufacturer's instructions, except that the AL general lysis buffer was substituted for the AVL viral lysis buffer with the addition of carrier RNA (Qiagen, Valencia, CA). Samples were analyzed by use of quantitative PCR (qPCR) for adenovirus or quantitative reverse transcription-PCR (qRT-PCR) for enterovirus as described by Jothikumar et al. (31) and Gregory et al. (32). PCR inhibition was determined using matrix spikes by seeding the master mix with an extracted positive-control virus in a duplicate qPCR or qRT-PCR. The cycle threshold (C_T) of the sample was then compared to the C_T in the clean matrix control, which also used the same seeded master mix. Sample extracts were considered to be inhibited and were diluted and reanalyzed if the seeded test sample was >2 C_T cycles higher than the seeded clean matrix control.

The standard curves for molecular detection of adenovirus and enterovirus were created using virus stocks treated with Benzonase (Novagen, Madison, WI) as described elsewhere (18) except that the treated stocks were incubated overnight at 37°C as recommended by Novagen instead of 30 min at 37°C and 2 days at 4°C. Treated stocks were extracted, and the amount of viral DNA or RNA was measured by using PicoGreen or RiboGreen (Molecular Probes, Eugene, OR) using a spectrophotometer, and the number of genomic copies (gc) was calculated. After quantification, viral stocks were serially diluted using a 2% beef extract solution. Each standard point was extracted in duplicate and then analyzed by qPCR or qRT-PCR in duplicate along with each run. Replicate runs of the standard curve for adenovirus produced a dynamic range of 5.91 to 5.91E+06, an amplification efficiency of 99%, and an R^2 value of 0.985

and for enterovirus produced a dynamic range of 15.5 to 1.55E+07, an amplification efficiency of 96%, and an R^2 value of 0.998.

(iii) Immunomagnetic separation/immunofluorescence assay (IMS/FA) for protozoa. *Cryptosporidium* and *Giardia* were isolated and enumerated using EPA method 1623 with heat dissociation (33, 34). Processed samples were shipped overnight at 4°C from the Columbus Laboratory to the EPA National Exposure Research Laboratory, Cincinnati, OH. One IMS reaction was performed per sample. In highly turbid samples, an additional 10-ml deionized water rinse was added after the first IMS purification. The slides were stained with EasyStain G&C (BTF Pty. Ltd., North Ryde, Australia), following the manufacturer's protocol except that steps 3, 6, and 7 were omitted.

Environmental and water quality data. Personnel collected daily data for environmental and water quality variables expected to affect *E. coli* and pathogen concentrations.

(i) Field measurements. Upon arrival at the beach, the number of birds and swimmers were noted on field forms. For wave height measurements, a graduated rod was placed at the sampling location. Measurements of specific conductance and water temperature were done at the sampling location using a digital thermometer and/or *in situ* probe and standard USGS methods (35). In the laboratory, duplicate measurements of turbidity using the *E. coli* samples were made using a portable turbidimeter (model 2100P; Hach Company, Loveland, CO). Secchi disk measurements were made as an alternative indicator of water clarity at sites monitored by MWCD.

(ii) Sources of environmental data. Environmental data were obtained from the nearest airport weather station or agency gauge, and/or from radar (see Table S3 in the supplemental material). These environmental data were from locations that were within 25 miles from a study site, and most were within 10 miles. Airport rainfall and wind direction and speed data were obtained from the National Oceanic and Atmospheric Administration (NOAA) National Weather Service (NWS) forecast offices in Pittsburgh, PA, Cleveland, OH, and Wilmington, OH (http://www.erh.noaa.gov/). Hourly radar rainfall data from the NWS (http://water.weather.gov/precip/download.php) were compiled for single 4-km grids ("cells") surrounding a site and/or for 12 to 18 cells (multiple cells) that encompassed the drainage area to a lake. Data on rainfall, precipitation, stream stage or discharge, and water surface elevation were obtained from USGS or U.S. Army Corps of Engineers (USACE) stations through the USGS National Water Information System website (NWIS web) (http://oh.water.usgs.gov/). Solar radiation data were obtained from the Ohio Agricultural Research and Development Center Weather System (OARDC) (http://www.oardc.ohio-state.edu/newweather/).

(iii) Compiling data and calculating variables. Antecedent hourly rainfall data were compiled for the 24-h period ending at 7:00 a.m. for radar data or 8:00 a.m. for airport or agency rainfall. Using these data, the total rainfall for a 24-h period before daily sampling was calculated (R_{d-1}) consistently for all sites. Three radar rainfall variables were calculated: (i) the summed amount of radar rainfall in the previous 24 h in one cell (Radar1cell-R_{d-1}), (ii) the hourly maximum values among multiple cells divided by the number of cells for the previous 24 h (Radarxcell-av-R_{d-1}), and (iii) the sum from multiple cells for the previous 24 h (Radarxcell-sum-R_{d-1}). Data were then lagged 1 or 2 days to represent the amount of rainfall in the 24-h period 2 days (R_{d-2}) and 3 days (R_{d-3}) prior to sampling. Weighted rainfall variables were calculated from airport, agency gauge, or radar rainfall as described previously (3).

For stream stage and stream discharge, hourly data were compiled, and the mean value was calculated for the 24-h period up to 8:00 a.m. For water surface elevation, the instantaneous value at 8 a.m. near the time of sampling was used. For solar radiation, 5-min-interval data were compiled, and the summed value was calculated for 12 a.m. to 11:55 p.m. for the day previous to the day of sampling.

Antecedent hourly wind direction and wind speed data were compiled for the instantaneous value at 8 a.m. and for the 24-h period ending at 8 a.m. The 24-h wind variables were calculated by summing hourly wind vectors for the 24-h period preceding sampling and determining the direction and speed of the resultant vector. The instantaneous 8 a.m. and 24-h wind speed and direction variables were used to calculate alongshore and offshore wind components as described by the EPA (36). For some sites, wind directions were placed in categories by examining patterns in plots of *E. coli* concentrations as a function of wind direction. Site-specific wind codes were calculated by assigning the most weight to the range of wind directions associated with the highest *E. coli* concentrations. Processes affecting *E. coli* were also considered to ensure that the wind direction categories could be reasonably explained by physical processes.

Data management, statistical analysis, and modeling. Daily data on *E. coli* concentration, turbidity, wave height, specific conductance, water temperature, and protozoan pathogens were entered into the USGS NWIS website (http://nwis.waterdata.usgs.gov/oh/nwis/qwdata) using USGS site identification numbers (see Table S1 in the supplemental material).

Concentrations of *E. coli* were $\log_{10}$ transformed before any statistical testing and modeling was done. Concentrations of *E. coli* and field measurements and variables collected in the morning were compared to those collected in the afternoon by use of the signed-rank test, a nonparametric alternative to the paired *t* test, using the SAS 9.2 software program (SAS Institute Inc., Cary, NC). The relationships between the occurrence of pathogens and *E. coli* concentrations or some key explanatory variables were determined by use of the Wilcoxon rank-sum test using the statistical software package TIBCO Spotfire S+ 8.1 for Windows (Tibco Software Inc., Somerville, Mass.).

Data from 2010-2011 were used for exploratory data analysis and to develop site-specific predictive models for *E. coli*. These procedures are detailed by Francy and Darner (37) and were facilitated by use of beach modeling software (36). The software program, Virtual Beach, is a free tool available for building predictive models. The general steps in model development and selection using Virtual Beach were as follows. (i) After importing and validating the data set, compute alongshore and onshore wind components and $\log_{10}$ transform *E. coli* data. (ii) Transform explanatory variables using $\log_{10}$, inverse, square, and square root transformations. (iii) Examine the relationships between environmental and water quality variables and *E. coli* concentrations using Pearson's *r* correlation analysis and data plots. (iv) Select variables for model development that are significantly related to *E. coli* ($P < 0.05$) or show a pattern of a relation in the data plot. Select transformed variables if they improve the relation over the untransformed variable. (v) Rank the models by use of the predicted residual sums of squares (PRESS) statistic. (vi) Select a model that provides a compromise between having the lowest PRESS statistic, highest R^2 value, statistically significant variables, and fewest false negatives and false positives. The selected model should include variables that reasonably explain changes in *E. coli* concentrations and are relatively easy to measure. (vii) Complete model evaluation, such as checking residuals and outliers. (viii) The models predict the probability that the single-sample water standard will be exceeded. Establish threshold probabilities for posting advisories as described by Francy and Darner (37).

The model responses for the calibration data set (data used to develop the model, 2010-2011) and validation data set (data collected during an independent year, 2012) were evaluated in terms of the correct predictions, sensitivities, and specificities and compared to the use of the previous day's *E. coli* concentrations. A correct response was based on the actual *E. coli* concentration, measured by the culture method. The sensitivity was the percentage of exceedances of the bathing-water standard that were correctly predicted by the model. The specificity was the percentage of nonexceedances that were correctly predicted by the model. Correct responses, sensitivities, and specificities were also calculated using the previous day's *E. coli* concentration to predict the current day's *E. coli* concentration.

RESULTS

***E. coli* concentrations and differences between morning and afternoon samples.** Summary statistics for *E. coli* concentrations at 22 sites are listed in Table 1. *E. coli* concentrations ranged from <1 to 4,900 most probable number (MPN)/100 ml. Excluding Tappan ditch (site 21), which is not a swimming beach, median concentrations of *E. coli* were highest at Buckeye Crystal and GLSM West. The percentages of days that the standard was exceeded in 2012 were the same or nearly the same as those in 2010-2011 at Alum Campers, Atwood Main, the three Buck Creek sites, GLSM west, and Tappan Main. The standard was exceeded more often in 2010-2011 than in 2012 at Alum North, Alum Central, GLSM Campers, and GLSM East.

In addition to daily morning sampling during 2011, 30 afternoon samples were added at the Alum North and Central sites and 32 afternoon samples were added at the Buck Creek North and South sites. At Alum Creek, concentrations of *E. coli*, number of swimmers, wave height, and turbidity were statistically higher in afternoon samples than in morning samples ($P \leq 0.0004$, signed-rank test, data not shown), but the numbers of birds at the times of morning and afternoon samplings were not statistically different ($P = 0.2227$). For 8 out of 10 exceedances at Alum Creek, the *E. coli* single-sample bathing-water standard was exceeded in the afternoon sample but not in the morning sample (Fig. 2A). The standard was exceeded in 5.4% and 21.6% of the 30 morning and afternoon samples, respectively. At Buck Creek, concentrations of *E. coli*, number of swimmers, wave height, and turbidity were statistically higher in afternoon samples than in morning samples ($P < 0.05$; data not shown); in contrast, the number of birds was statistically higher in the morning samples than in the afternoon samples ($P = 0.0005$). At Buck Creek, the *E. coli* standard was exceeded in two morning samples (6.3%) and three afternoon samples (9.4%), with none of the five exceedances in concurrence (Fig. 2B). Combining the morning and afternoon results for each beach for Pearson's correlation analyses, the number of swimmers was significantly related to $\log_{10}$ *E. coli* concentrations at Alum Creek ($r = 0.56$) and Buck Creek ($r = 0.29$).

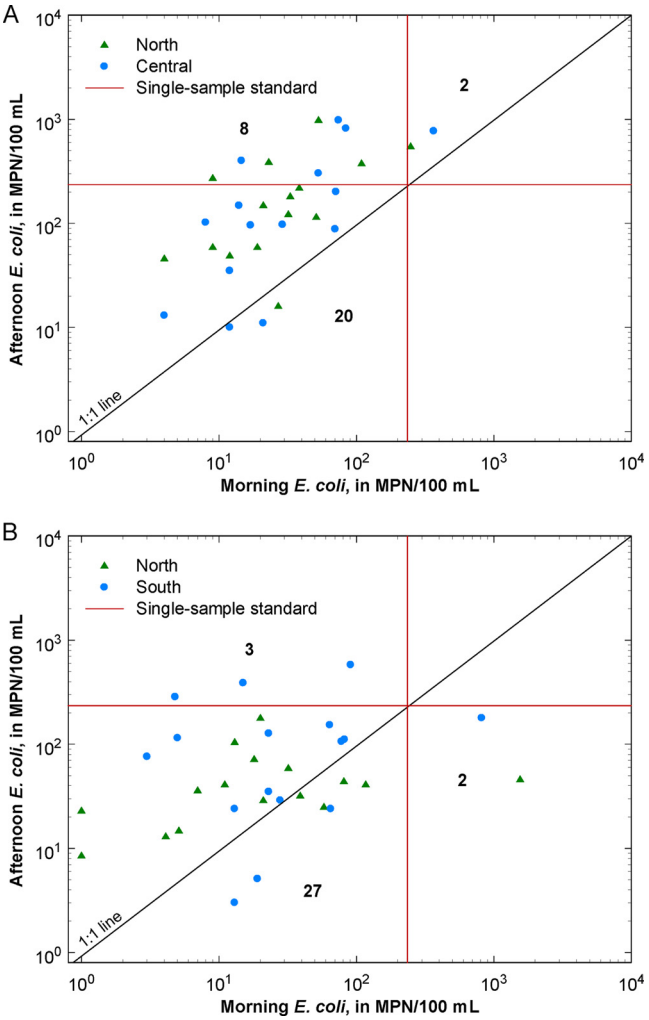


FIG 2 *E. coli* concentrations in morning and afternoon samples and comparisons to the single-sample bathing standard (235 MPN/100 ml) at Alum Creek (A) or Buck Creek (B).

Relationships of *E. coli* concentrations to environmental and water quality variables and predictive models at inland lake sites. Predictive models were developed using data collected during 2010-2011 for 13 out of 22 sampling sites (Table 1). Models were not developed for Tappan ditch because it is not a swimming beach, for Seneca because only 1 year of data was available, and for seven other sites because the *E. coli* standard was exceeded <5% of the time during 2010 or 2010-2011.

As a first step in predictive model development, Pearson's correlations between $\log_{10}$ *E. coli* concentrations (hereinafter "*E. coli* concentrations") and potential explanatory variables were determined. Table 2 presents a partial list of explanatory variables and includes those variables that were subsequently used in at least one model. Correlations that were significant ($P \leq 0.05$) are in bold and italics, and those used in models are shaded. Data are organized into four categories: field data, weather data from the NWS, radar rainfall data, and USGS and USACE gauge data.

Among the field measurements and observations, the overall highest correlation was found between *E. coli* and turbidity at Atwood Main ($r = 0.47$). It should be noted that the relation between the Secchi disk and *E. coli* at Atwood Main and Tappan Main ($r = -0.47$ and -0.23; data not shown) was the exact inverse of the relation between turbidity and *E. coli*. Day of the year was the field variable most often related (46%), and number of birds least often related (23%), to *E. coli*. Turbidity and/or water temperature was used in some models even though they were not always significantly related to *E. coli* (as shown by Pearson's *r* correlations) because plots of each of these variables versus *E. coli* concentrations indicated a positive trend (data not shown). Turbidity and water temperature were used in models six times, the highest frequency among all the variables in Table 2.

Weather data from the NWS nearest airport sites were used in models at eight sites. Rainfall data were used in only two models, although these data were related to *E. coli* at 54% and 62% of the sites. This was most likely due to collinearity between airport and radar rainfall data. Alongshore and offshore wind variables were significantly related to *E. coli* at 15 to 31% of the sites. Wind codes were compiled at the two Buck Creek and three GLSM beaches, where plots indicated patterns between wind directions and *E. coli* concentrations. For example, at Buck Creek North, higher *E. coli* concentrations were associated with winds from the southwest, west, and northwest, and these received a code of "1," while all

TABLE 2 Pearson's *r* correlations between $\log_{10}$ *E. coli* concentrations and explanatory variables at inland lake sites for daily sampling, 2010-2011[a]

Variable	Pearson's *r* correlation for beach site													% of sites for which variable was significant	No. of sites used in models
	Alum Campers	Alum North	Alum Central	Atwood Main	Buckeye Brooks	Buckeye Crystal	Buckeye Fairfield	Buck Cr North	Buck Cr South	GLSM Campers	GLSM West	GLSM East	Tappan Main		
Field measurements															
Day of yr	***−0.45***	0.11	0.00	−0.04	0.06	***−0.22***	***0.30***	***0.27***	***0.26***	***−0.27***	0.14	0.11	0.01	46	2
Turbidity	***0.41***	0.11	0.16	***0.47***	0.14	−0.08	−0.08	−0.15	0.00	***0.29***	0.12	***0.24***	***0.23***	38	6
Water temp	***−0.24***	0.15	0.07	***0.36***	0.20	−0.03	0.12	0.15	***0.25***	−0.10	0.08	0.18	***0.24***	31	6
Birds	0.15	0.01	0.00	−0.01	0.14	***0.25***	***0.31***	0.15	***0.37***	0.10	0.03	0.19	0.10	23	2
Swimmers	***0.22***	0.01	0.01	***0.39***	0.09	0.07	***0.28***	0.04	0.11	0.00	0.00	0.00	***0.20***	31	1
Wave height	0.02	−0.09	0.00	***0.36***	0.10	−0.18	***0.26***	0.13	−0.05	***0.32***	−0.01	***0.23***	0.11	31	1
Weather data from NWS															
Rainfall, R_{d-1}[a]	***0.30***	0.19	***0.34***	0.13	***0.27***	***0.36***	***0.31***	0.06	***0.24***	0.16	***0.25***	0.18	0.09	54	1
Rainfall, Rw48[e]	***0.35***	***0.21***	***0.34***	0.12	***0.31***	***0.39***	***0.34***	0.06	***0.21***	0.16	***0.23***	0.13	0.12	62	1
Wind alongshore, 8 a.m.	0.08	−0.08	0.02	0.01	−0.04	0.06	***−0.37***	−0.17	−0.14	0.06	***0.26***	0.08	−0.14	23	2
Wind alongshore, 24 h	−0.05	0.06	0.14	−0.01	***−0.27***	***0.29***	***−0.52***	***−0.31***	−0.11	−0.08	0.21	−0.02	***−0.29***	31	2
Wind offshore, 8 a.m.	0.00	−0.13	−0.09	0.12	−0.04	−0.11	***0.22***	0.10	0.08	***0.36***	−0.06	0.14	0.11	15	1
Wind offshore, 24 h	−0.12	−0.14	−0.05	0.13	0.04	***−0.25***	−0.06	0.00	−0.05	***0.28***	0.03	0.16	***0.22***	23	1
Wind code × wind speed, 8 a.m.	—	—	—	—	—	—	—	***0.20***	0.15	***0.42***	***0.26***	***0.35***	—	23	4
Radar rainfall															
Radarxcell-av-R_{d-1}[b,g]	***0.34***	0.16	***0.33***	***0.18***	***0.32***	***0.33***	***0.35***	0.15	***0.20***	0.20	***0.40***	***0.30***	0.16	69	2
Radarxcell-av-R_{d-2}[c,g]	***0.40***	0.13	0.11	0.07	0.20	0.10	***0.28***	0.06	0.05	0.03	0.13	−0.07	0.06	15	1
Radarxcell-av-R_{d-3}[d,g]	0.09	***0.31***	0.18	−0.08	0.01	−0.06	−0.07	0.00	0.00	0.01	***0.23***	0.04	0.08	15	2
Radarxcell-av-Rw48[e,g]	***0.46***	0.20	***0.35***	***0.19***	***0.33***	***0.39***	***0.42***	0.14	0.19	0.18	***0.39***	***0.21***	0.17	62	1
Radarxcell-av-Rw72[f,g]	***0.50***	***0.27***	***0.38***	0.16	***0.36***	***0.33***	***0.42***	0.12	0.17	0.17	***0.42***	0.19	***0.21***	69	1
Radarxcell-sum-R_{d-1}[b,h]	***0.36***	***0.24***	***0.38***	0.12	***0.26***	***0.37***	***0.35***	0.12	0.17	***0.22***	***0.40***	***0.29***	***0.19***	77	1
Radarxcell-sum-Rw48[e,h]	***0.52***	***0.30***	***0.40***	0.14	***0.26***	***0.40***	***0.40***	0.12	0.17	0.20	***0.37***	0.20	***0.20***	62	1
Radarxcell-sum-Rw72[f,h]	***0.56***	***0.36***	***0.42***	0.12	***0.31***	***0.33***	***0.40***	0.11	0.16	0.20	***0.41***	0.18	***0.23***	62	1
USGS or USACE gauge															
Rain gauge, R_{d-1}[b]	***0.41***	***0.33***	***0.45***	***0.18***	***0.27***	***0.27***	0.20	0.08	0.18	***0.23***	***0.29***	0.13	***0.19***	69	1
Rain gauge, R_{d-3}[d]	0.15	0.15	0.12	−0.02	−0.09	−0.05	−0.08	−0.12	−0.10	0.15	***0.23***	0.19	0.13	8	2
Rain gauge, Rw48[e]	***0.55***	***0.37***	***0.46***	***0.21***	***0.26***	***0.28***	0.20	0.10	0.14	***0.27***	***0.33***	0.12	***0.21***	69	2
Rain gauge, Rw72[f]	***0.58***	***0.39***	***0.45***	***0.18***	***0.23***	***0.26***	0.18	0.08	0.10	***0.29***	***0.36***	0.14	***0.23***	69	2
Discharge or stage, 24 h, lagged 1 day	0.18	−0.16	−0.17	0.00	—	—	—	***−0.32***	***−0.31***	0.20	0.00	−0.14	−0.02	15	2
Water surface elevation, 8 a.m.	***0.51***	***0.21***	***0.23***	—	***0.29***	0.16	***0.31***	0.04	0.07	—	—	—	—	38	1
Water surface elevation, change in 24 h	***0.46***	***0.30***	***0.38***	—	***0.24***	***0.30***	0.15	0.06	0.18	—	—	—	—	38	1

[a] Relations that were significant at $P < 0.05$ are in italics and bold. —, not determined. Variables used in selected models are shaded.
[b] R_{d-1} is the amount of rainfall in the 24-h period before sampling.
[c] R_{d-2} is the amount of rainfall in the 24-h period 2 days before sampling.
[d] R_{d-3} is the amount of rainfall in the 24-h period 3 days before sampling.
[e] Rw48 is the amount of rainfall in the 48-h period before sampling, with the most recent rainfall receiving the most weight and calculated as $(2 * R_{d-1}) + R_{d-2}$.
[f] Rw72 is the amount of rainfall in the 72-h period before sampling, with the most recent rainfall receiving the most weight and calculated as $(3 * R_{d-1}) + (2 * R_{d-2}) + R_{d-1}$.
[g] Radarxcell-av is the hourly maximum value among multiple 4-km cells divided by the number of cells for the time period specified.
[h] Radarxcell-sum is the sum from multiple 4-km cells for the time period specified.

other wind directions received a code of "0." Wind codes were not compiled at other beaches because no patterns were observed. The wind code multiplied by the wind-speed 8 a.m. variable was used in four models, the second-highest frequency among all the variables in Table 2.

Radar rainfall data were used in models at nine sites, and six radar variables from multiple cells were significantly related to *E. coli* at more than 60% of the sites. Single-cell radar rainfall data were compiled for Atwood Main, Buck Creek North and South, and Tappan Main, but these variables were not significantly related to *E. coli* at any of the sites (data not shown).

Three rainfall variables from USGS or USACE rain gauge sites were significantly related to *E. coli* at 69% of the sites. Mean discharge or stage for the past 24 h was not used in any models, although these variables were significantly related to *E. coli* at four sites (data not shown). Once again, these variables were most likely excluded from the models because of collinearity with other variables, such as radar rainfall. The mean discharge for the past 24 h lagged 1 day, however, showed a significant negative correlation to *E. coli* at the two Buck Creek sites and was used for those models. Solar radiation (the sum from the previous day) was not significantly related to *E. coli* at the two beaches where these data were available (Alum and Buck Creek; data not shown).

The selected best models are presented in the supplemental material (see "Equations for the selected best models for each inland lake site"). Model adjusted R^2 values, threshold probabilities, and responses from the calibration data set are presented in Table 3. Adjusted R^2 values ranged from 0.19 at GLSM West to 0.56 at Alum Campers. Threshold probabilities were set based on the calibration data set and represented a compromise between reducing false negatives and maintaining a relatively high percentage of correct responses. An example of setting the threshold probability for Buck Creek is presented in the supplemental material (see "Determining probabilities and establishing a threshold probability for issuing advisories"). All sensitivities were set at ≥50%, with specificities of >82%. Among the selected models,

TABLE 3 Selected models for nowcasting at inland lakes and responses using calibration data set, 2010-2011

Site for model	Adj. R^2 value[a]	Threshold probability[b]	No. of observations	No. of exceedances[c]	% correct	Sensitivity (%)	Specificity (%)
Alum Campers	0.56	20	84	7	94.0	85.7	94.8
Alum North	0.24	30	83	8	95.2	50.0	100.0
Alum Central	0.30	45	87	6	96.6	66.7	98.8
Atwood Main	0.41	30	121	28	85.1	67.9	90.3
Buckeye Brooks	0.25	40	67	19	83.6	63.2	91.7
Buckeye Crystal	0.21	40	89	27	74.2	55.6	82.3
Buckeye Fairfield	0.45	30	89	12	86.5	50.0	92.2
Buck Creek North	0.22	19	99	12	84.8	58.3	88.5
Buck Creek South	0.33	29	102	15	84.3	60.0	88.5
GLSM Campers	0.35	37	82	24	80.5	62.5	87.9
GLSM West	0.19	38	76	28	81.6	71.4	87.5
GLSM East	0.28	43	76	17	85.5	52.9	94.9
Tappan Main	0.20	34	120	24	84.2	50.0	92.7

[a] Fraction of the variation of *E. coli* concentrations that is explained by the model (Adj., adjusted).
[b] Established by examining the calibration data set to maximize correct responses.
[c] Number of days the Ohio single-sample bathing water standard of 235 CFU/100 ml was exceeded.

the highest percentage correct was found for Alum Central (96.6%), the highest sensitivity for Alum Campers (85.7%), and the highest specificity for Alum North (100%).

Validation of predictive models. Models for nine beaches were validated in 2012. The three Buckeye Lake sites were not included in the 2012 validation because of low R^2 values in 2 of the 3 models, a low percentage correct at Buckeye Crystal, and reduced swimmer density. GLSM West was not included in the 2012 validation because of a low R^2 value and because this beach was seldom used by swimmers.

The model responses during the validation year were compared to use of the previous day's *E. coli* concentration, the current method for assessing recreational water quality (Table 4). At Buck Creek North and South and Tappan Main, use of the model resulted in an increase in correct responses, sensitivities, and specificities compared to use of the persistence model. This was not the

TABLE 4 Nowcast model responses compared to use of previous day's *E. coli* concentration (persistence model) during validation in 2012[a]

Site	Model used	No. of observations	No. of exceedances[b]	% correct	Sensitivity (%)	Specificity (%)
Alum Campers	Nowcast	49	3	91.8	0.0	97.8
	Persistence	38	2	89.5	0.0	94.4
Alum North	Nowcast	49	0	100.0	0.0	100.0
	Persistence	38	0	100.0	0.0	100.0
Alum Central	Nowcast	49	1	91.8	0.0	93.8
	Persistence	38	1	97.4	0.0	100.0
Atwood Main	Nowcast	52	13	65.4	23.1	79.5
	Persistence	41	10	78.0	50.0	87.1
Buck Creek North	Nowcast	45	8	**80.0**	**62.5**	**83.8**
	Persistence	34	7	64.7	14.3	77.8
Buck Creek South	Nowcast	46	9	**73.9**	**55.6**	**78.4**
	Persistence	34	9	52.9	11.1	68.0
GLSM Campers	Nowcast	48	2	66.7	100.0	65.2
	Persistence	39	2	92.3	0.0	97.3
GLSM East	Nowcast	48	2	79.2	0.0	82.6
	Persistence	40	2	95.0	0.0	97.4
Tappan Main	Nowcast	52	10	**76.9**	**40.0**	**85.7**
	Persistence	42	9	69.0	33.3	78.8

[a] Model responses that could be evaluated as improved over those of the persistence model are in bold and shaded.
[b] Number of days the Ohio single-sample bathing water standard of 235 CFU/100 ml was exceeded.

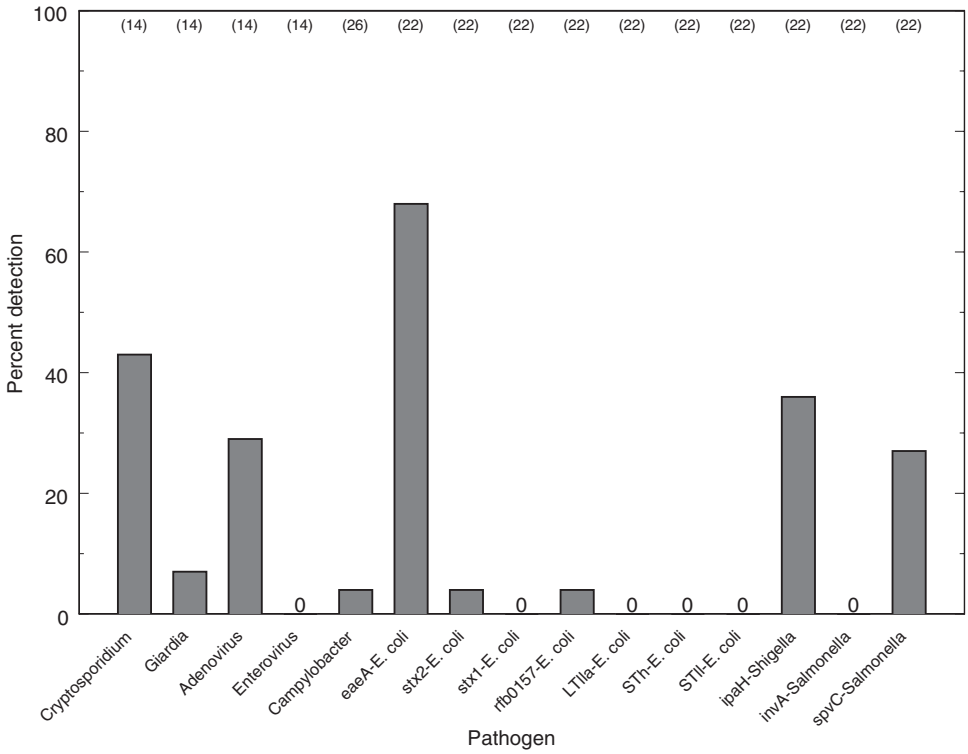


FIG 3 Percentages of detections of protozoan and viral pathogens and bacterial pathogen genes at Buckeye Lake, Tappan, and Atwood sites, 2011. The numbers of samples are in parentheses.

case for the other beaches; however, at four of the sites (Alum North, Alum Central, GLSM Campers, GLSM East), there were too few exceedances during 2012 (Table 1) to adequately assess model performance.

Pathogens in lake water samples. Concentrations of *E. coli*, protozoan and viral pathogens, presence or absence results for bacterial pathogens, values for key explanatory variables related to *E. coli* concentrations, and probability outputs for applicable predictive models are presented for 31 samples in Table S4 in the supplemental material.

The percentages of detections of pathogens in samples collected from Buckeye Lake and MWCD sites (Tappan and Atwood) are shown in Fig. 3. Enterovirus, *E. coli* stx_1, LTIIa, STh, and STII, and *Salmonella invA* were not found in any samples. *Cryptosporidium* was found in 43% of samples, and only one sample (7%) was positive for *Giardia*. Concentrations of *Cryptosporidium* and *Giardia* ranged from <0.1 to 2 oocysts or cysts/10 liters (see Table S4 in the supplemental material). Adenovirus was identified in 29% of samples, with concentrations ranging from <1.2 to 39 gc/liter (see Table S4). Five out of six detections of protozoan pathogens were done by use of ultrafiltration, whereas 3 out of 4 detections of adenovirus were done by use of glass-wool filtration (see Table S4). For the 16S rRNA genes marker for *Campylobacter*, only one sample (4%) was positive. The *eaeA* marker for pathogenic *E. coli* was the most frequently detected bacterial pathogen gene, being identified in 68% of samples.

Pathogen gene marker data representing pathogenic *E. coli*, *Shigella*, and *Salmonella* were collected for 22 samples. Three bacterial pathogen gene markers (for *E. coli*, *eaeA*; for *Shigella*, *ipaH*; for *Salmonella*, *spvC*) were identified in more than 20% of the samples, allowing a more robust statistical data analysis. Modeled parameter variables were split into two categories based on the presence or absence of each gene. Median values for model variables and probabilities for each group are shown in Table 5. Median rainfall and turbidity values were significantly higher ($P < 0.1$) and specific conductance was significantly lower in samples having the *eaeA E. coli* pathogen gene than in those lacking the gene. Samples containing the *spvC* marker for pathogenic *Salmonella* had higher median concentrations of *E. coli*, while samples containing the *ipaH* gene of pathogenic *Shigella* had significantly lower median concentrations of *E. coli*, than those lacking the gene. Samples having the *ipaH* gene of *Shigella* had significantly higher specific conductance values and higher (positive) alongshore and near-shore winds. Despite samples possessing the *eaeA* and *spvC* genes having similarly higher median rainfall values than those lacking the genes, possession of the *eaeA* gene of *E. coli* by samples was unrelated to the model probability of *E. coli*, while samples possessing the *spvC* gene had a significantly higher model probability of *E. coli*.

DISCUSSION

Although previous studies have documented the development of predictive models for Great Lakes beaches (38, 39) and ocean beaches (40), this was the first study to systematically investigate the use of predictive models at multiple inland recreational beaches. Predictive models were developed for 13 out of the 21 beach sites initially included in the current study. Models were not developed for seven sites because the *E. coli* single-sample bathing-water standard was exceeded <5% of the time, making them poor candidates for predictive modeling, and at one site because only 1

TABLE 5 Median values of water quality variables in the presence or absence of selected pathogen detection at inland lakes sites, 2011[a]

Model variable[b]	Median value of variable and associated *P* value								
	eaeA (*E. coli*)			*ipaH* (*Shigella*)			*spvC* (*Salmonella*)		
	Absent	Present	*P* value	Absent	Present	*P* value	Absent	Present	*P* value
E. coli (MPN/100 ml)	36	122	0.53	210	37	0.06	38	430	0.02
Specific conductance (μS/cm)	341	288	0.07	287	314	0.09	310	299	0.88
Turbidity (NTU)	23.1	29.7	0.08	30.0	22.9	0.13	28.9	28.5	0.97
Airport rain, 24 h (in.)	0.01	0.19	0.03	0.11	0.08	0.28	0.02	0.53	0.04
Rainfall, Radarxcell-sum-R_{d-1} (in.)	0.00	7.45	0.02	6.63	0.71	0.20	1.13	12.09	0.04
Rainfall, rain gauge, R_{d-1} (in.)	0.00	0.35	0.07	0.35	0.01	0.12	0.02	0.66	0.05
Water temp (°C)	27.30	26.50	0.50	26.85	26.90	0.97	26.25	28.70	0.08
Wind alongshore, 24 h (mph)	1.87	−0.07	0.50	−1.62	2.19	0.00	−0.74	1.56	0.30
Wind offshore, 24 h (mph)	1.22	−0.87	0.11	−0.83	1.34	0.07	0.77	−0.83	0.56
Model probability (%)	15.80	25.40	0.50	15.80	18.40	0.94	15.00	45.20	0.02

[a] The *P* value is the result of the Wilcoxon rank-sum test comparing the median value for each variable in samples with the pathogen absent to those in samples with the pathogen present.
[b] For variables, see footnotes for Table 2. NTU, nephelometric turbidity units; mph, miles per hour.

year of data was available. Previous work has shown that at least 2 years of data are needed to develop predictive models and that models work best at moderately contaminated beaches (39).

The variables used in models at inland lakes in the current study had some commonalities with those used in models at coastal beaches. In the current study, the variables used most often in models were radar rainfall (10 times), wind variables (10 times), rainfall from an airport or other agency gauge (9 times), turbidity (6 times), and water temperature (6 times). Similar to the present study, investigators used turbidity and radar and/or airport rainfall in models for two Lake Erie beaches (3) and these same variables plus wind direction at another Lake Erie beach (41). Wave height and day of the year were important predictors for *E. coli* at Lake Erie beaches (3) but were seldom used in models in the current study (≤2 times). This was to be expected, since smaller lakes have less fetch than the Great Lakes and thus lower wave heights and less influence from waves. At several urban Lake Michigan beaches, investigators found that winds influenced a nearby river's impact on beaches and thereby developed separate models for different prevailing wind directions incorporating variables for wave height, turbidity, and rainfall (38). At another Lake Michigan beach, the best-fit model contained measurements of winds, rainfall, solar radiation, lake level, water temperature, and turbidity (42). Because they expected different factors to influence Southern California ocean beaches on dry and wet days, Hou et al. (43) developed separate models for these two conditions. The important variables in models were rainfall and stream discharge (wet days only), tides, water temperature, winds, visitor numbers, waves, and solar radiation (dry days only) (43). In the present study, the day of the year, number of swimmers, wave height, discharge from a nearby stream, and water surface elevation were seldom used in models (<2 times). However, in the present study, the numbers of swimmers were related to *E. coli* concentrations when afternoon samples were included. The models for inland beaches in the present study and those for Great Lakes beaches in past studies showed the importance of selecting site-specific variables that address local geography, nearby stream discharge, runoff potential, wind direction patterns relative to the beach, contamination dilution, and local versus watershed-wide rainfall amounts. Inland water bodies are very different in terms of hydrology and water quality than ocean or Great Lakes beaches, and these differences need to be considered when including variables in site-specific models.

A unique example of a site-specific variable can be found in the present study. The two Buck Creek sites are located on CJ Brown Reservoir, controlled by a USACE dam directly south of the beaches, with a USGS gaging station downstream from the dam. The mean discharge (flow) at the gaging station for the past 24 h, lagged 1 day, was negatively related to *E. coli* concentrations and was used in models for the two Buck Creek sites. The mean discharge as a negative coefficient was not used in any other models in the current study or in past studies. A negative correlation to *E. coli* indicates that when *E. coli* concentrations were higher, less water was moving through the dam. Under these low-flow conditions, the higher *E. coli* concentrations may be attributed to greater influences from local sources, such as septic systems, bathers, and wildlife.

In the current study at Alum Creek, the *E. coli* single-sample standard was exceeded much more often in afternoon samples (21.6%) than in morning samples (5.4%). This did not occur at the Buck Creek sites, where the percentage of exceedance was only slightly higher in the afternoon (9.4%) than in the morning (6.3%). At Alum Creek and Buck Creek beaches, the concentrations of *E. coli*, number of swimmers, wave height, and turbidity were statistically higher in the afternoon than in the morning. This is in contrast to the findings of other researchers, where bacterial indicator concentrations were higher in the morning than in the afternoon at a California ocean beach (1) and at a Lake Michigan beach (2). The swimmers may have a stronger influence on water quality at inland lake beaches than they have at coastal beaches because of less water and smaller amounts of dilution in inland lakes. The increased *E. coli* concentrations in the afternoon in the current study may have been from swimmer shedding and/or from resuspension of *E. coli* from bottom sediments. Gerba (44) conducted a literature review, modeled pathogen shedding, and concluded that persons of all ages shed fecal indicators and pathogens into recreational waters. In a study at Lake Erie beaches (45), bottom sediments from bathing areas contained *E. coli* and were identified as a potential source of resuspended *E. coli* for the water column. The models developed from samples collected in the morning in the present study may underestimate health risks at times when many swimmers are present in inland lakes.

Although models can perform fairly well when predicting responses to data used to develop them, a better test of a model is to predict responses during an independent, validation year (37). In the current study at inland lake beaches, nine models were validated and compared to use of the previous day's *E. coli* concentration (persistence model) during a validation year. Model results at several beaches could not be adequately evaluated because there were far fewer *E. coli* exceedances during the validation year (2012) than during the calibration years (2010-2011) (Table 1). This may have occurred because of climatic conditions that were different in 2010-2011 from those in 2012. For example, in central Ohio, where Alum Creek Reservoir is located, the area was rated as very moist in 2011 but was rated as being in moderate drought during 2012 (http://www.ncdc.noaa.gov/sotc/drought/). This highlights the importance of collecting data for development and validation of models during multiple years in order to include the variety of weather and water quality conditions that occur from year to year. Development of new models with 2012 data may improve model performance at the Alum Creek sites. At Atwood Main, overall correct responses, sensitivity, and specificity for the model were lower than those found for the persistence model (Table 4). Further examination of the model responses revealed that false positives were dominant early in the season and false negatives were dominant later in the season. Two subseason models (before and after July 15), therefore, may work better at Atwood Main. At three beaches (Buck Creek North, Buck Creek South, and Tappan Main), the nowcast model provided more-accurate responses than the persistence model during 2012 (Table 4, bolded responses), and these are good candidates for a nowcast system in 2013. At two Great Lakes beaches that are part of the Ohio Nowcast (http://www.ohionowcast.info), the models provided correct responses (84.2 and 74.4%), sensitivities (54.9 and 56.8%), and specificities (89.6 and 80.3%) that were in the same range as those found at these three sites, except that a lower sensitivity was found at Tappan Main (40%). Most of the false model responses at Tappan (9 out of 12) were found after July 22, indicating that two subseason models may provide better predictions.

A considerable number of published reports of studies of coastal recreational beaches describe the occurrence of pathogens (7, 46). Only a few of these types of studies have been done at inland recreational beaches, and many of these were done among compilations of different types of inland waters. For example, *Cryptosporidium* was detected in 22% and *Giardia* was detected in 47% of non-effluent-dominated Chicago-area waters that included river, Lake Michigan harbor and beach, and inland lake sites (47). In the present study, *Cryptosporidium* was found in 43% of inland water samples, but *Giardia* was found in only one sample (7%). Low levels of *Cryptosporidium* and *Giardia* were found in recreational lakes in Amsterdam (48), similar to levels found in the present study. A large-scale survey at 25 freshwater recreational and water supply sites in New Zealand showed that *Campylobacter* and human adenoviruses were most likely to cause human waterborne illness in recreational freshwater users (49). In the present study, adenovirus was found in 29% of samples, but *Campylobacter* was found in only one sample (4%). While bacterial pathogens have been identified as sources of outbreaks from recreational contact with water at inland lakes (50) and extensive studies were done looking at pathogens in various sources and inputs to recreational waters (51), there are only sporadic reports detailing the occurrence of bacterial pathogens at inland lake beaches (52, 53).

The data for three bacterial pathogen gene markers (for *E. coli*, *eaeA*; for *Shigella*, *ipaH*; for *Salmonella*, *spvC*) were used to identify relationships between the presence of the genes and model variables or *E. coli* concentrations. When the data for all beaches were combined, rainfall, conductivity, turbidity, water temperature, wind, and model probability were related to the presence/absence of at least one of the genes. *E. coli* concentrations were significantly higher in samples where the *spvC* (*Salmonella*) gene was present but not for the other two genes. These findings illustrate the relationships that different pathogens can have with environmental variables and with *E. coli*. To our knowledge, there are no other published studies that have examined bacterial pathogen occurrence in the context of environmental variables. At two Lake Michigan beaches, Wong et al. (7) demonstrated that predictive models of virus pollution were best described using wind speed, wind direction, and water temperature and traditional indicators did not generally address viral risks.

The current study showed that models could be used to provide near-real-time assessments at some recreational inland beaches and that some of the variables for inland lake sites were similar to those used at coastal beaches. Predictive models were not effective at all inland lake sites; however, their use at two lakes with high swimmer densities will provide better estimates of public health risk than current methods and a valuable resource for beach managers and the public. In implementing nowcast systems for inland lakes, beach managers should continue to be vigilant in monitoring water quality from year to year, refining models as needed, and working to understand the processes that affect fecal contamination at beaches. The variables used in the models at inland lakes were related to detection of some pathogen genes; more work needs to be done, however, to examine the relationships between explanatory variables and pathogens at inland recreational beaches.

ACKNOWLEDGMENTS

We thank Mark Swiger and staff (Muskingum Watershed Conservancy District), Rick Miller and Dan Chatfield and staff (Clark County Combined Health District), and Melissa Taylor and staff (Ohio Department of Natural Resources) for help with project planning and sampling. We are grateful to Mike Sudman and staff (City of Celina Water Plant) for the use of their laboratory facilities.

Support for this study was provided by the Ohio Water Development Authority, U.S. Geological Survey Cooperative Water Program, and by the U.S. Environmental Protection Agency through its Office of Research and Development.

This publication has been reviewed by the U.S. Environmental Protection Agency but does not necessarily reflect agency views. No official endorsement by the U.S. Environmental Protection Agency should be inferred. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

REFERENCES

1. **Boehm AB, Grant SB, Kim JH, Mowbray SL, McGee CD, Clark CD, Foley DM, Wellman DE.** 2002. Decadal and shorter period variability of surf zone water quality at Huntington Beach, California. Environ. Sci. Technol. **36**:3885–3892.
2. **Whitman RL, Nevers MD.** 2004. *Escherichia coli* sampling reliability at a frequently closed Chicago beach: monitoring and management implications. Environ. Sci. Technol. **38**:4241–4245.
3. **Francy DS, Bertke EE, Darner RA.** 2009, Testing and refining the Ohio Nowcast at two Lake Erie beaches—2008. US Geological Survey open-file

report 2009-1066. US Geological Survey, Reston, VA. http://pubs.usgs.gov/of/2009/1066/.
4. **US Environmental Protection Agency.** 2012. Recreational water quality criteria. EPA-820-F-12-058. Office of Water, US Environmental Protection Agency, Washington, DC. http://water.epa.gov/scitech/swguidance/standards/criteria/health/recreation/index.cfm.
5. **Ohio Department of Natural Resources.** 2008. Ohio State Parks 2008 annual report. Ohio Department of Natural Resources, Columbus, OH. http://www.dnr.state.oh.us/portals/2/annualreports/2008annualreport.pdf.
6. **Hlavsa MC, Roberts VA, Anderson AR, Hill VR, Kahler AM, Orr M, Garrison LE, Hicks LA, Newton A, Hilborn ED, Wade TJ, Beach MJ, Yoder JS.** 2011. Surveillance for waterborne disease outbreaks and other health events associated with recreational water—United States, 2007-08. MMWR Surveill. Summ. **60**(SS-12):1–63.
7. **Wong M, Kumar L, Jenkins TM, Xagoraraki I, Phanikumar MS, Rose JB.** 2009. Evaluation of public health risks at recreational beaches in Lake Michigan via detection of enteric viruses and human-specific bacteriological marker. Water Res. **43**:1137–1149.
8. **Jones K.** 2004. *Campylobacter* and wild birds. Health Hyg. **25**:11–12.
9. **Graczyk TK, Majewska AC, Schwab KJ.** 2008. The role of birds in dissemination of human waterborne enteropathogens. Trends Parasitol. **24**:55–59.
10. **Jellison KL, Lynch AE, Ziemann JM.** 2009. Source tracking identifies deer and geese as vectors of human-infectious Cryptosporidium genotypes in an urban/suburban watershed. Environ. Sci. Technol. **43**:4267–4272.
11. **Plutzer J, Tomor B.** 2009. The role of aquatic birds in the environmental dissemination of human pathogenic *Giardia duodenalis* cysts and *Cryptosporidium* oocysts. Hung. Parasitol. Int. **58**:227–231.
12. **Ishii S, Meyer KP, Sadowsky MJ.** 2007. Relationship between phylogenetic groups, genotypic clusters, and virulence gene profiles of *Escherichia coli* strains from diverse human and animal sources. Appl. Environ. Microbiol. **73**:5703–5710.
13. **Duris JW, Haack SK, Fogarty LR.** 2009. Gene and antigen markers of Shiga-toxin producing *E. coli* from Michigan and Indiana river water: occurrence and relation to recreational water quality criteria. J. Environ. Qual. **38**:1878–1886.
14. **Waage AS, Vardund T, Lund V, Kapperud G.** 1999. Detection of low numbers of Salmonella in environmental water, sewage and food samples by a nested polymerase chain reaction assay. J. Appl. Microbiol. **87**:418–428.
15. **Schoen ME, Ashbolt NJ.** 2010. Assessing pathogen risk to swimmers at non-sewage impacted recreational beaches. Environ. Sci. Technol. **44**:2286–2291.
16. **Gupta A, Polyak CS, Bishop RD, Sobel J, Mintz ED.** 2004. Laboratory-confirmed shigellosis in the United States, 1989–2002: epidemiologic trends and patterns. Clin. Infect. Dis. **38**:1372–1377.
17. **Myers DN, Stoeckel DM, Bushon RN, Francy DS, Brady AMG.** 2007. Fecal indicator bacteria. U.S. Geological Survey techniques of water-resources investigations, book 9, chapter A7, section 7.1. U.S. Geological Survey, Reston, VA. http://pubs.water.usgs.gov/twri9A/.
18. **Lambertini E, Spencer SK, Bertz PD, Loge FJ, Kieke BA, Borchardt MA.** 2008. Concentration of enteroviruses, adenoviruses, and noroviruses from drinking water by use of glass wool filters. Appl. Environ. Microbiol. **74**:2990–2996.
19. **Francy DS, Stelzer EA, Bushon RN, Brady AMG, Mailot BE, Spencer SK, Borchardt MA, Elber AG, Riddell KR, Gellner TM.** 2011. Quantifying viruses and bacteria in wastewater—results, interpretation methods, and quality control. U.S. Geological Survey scientific investigations report 2011-5150. US Geological Survey, Reston, VA. http://pubs.usgs.gov/sir/2011/5150/.
20. **Francy DS, Bushon RN, Brady AMG, Bertke EE, Kephart CM, Likirdopulos CA, Mailot BE, Schaefer FW, III, Lindquist HDA.** 2009. Comparison of traditional and molecular analytical methods for detecting biological agents in raw and drinking water following ultrafiltration. J. Appl. Microbiol. **107**:1479–1491.
21. **US Environmental Protection Agency.** 2006. Method 1600—enterococci in water by membrane filtration using membrane-enterococcus indoxyl-β-D-glucoside agar (mEI). EPA-821-R-06-009. Office of Water, US Environmental Protection Agency, Washington, DC.
22. **US Environmental Protection Agency.** 1991. Test methods for *Escherichia coli* in drinking water—test method 1105. EPA-600-4-91-016. Office of Water, Research and Development, US Environmental Protection Agency, Cincinnati, OH.
23. **Islam MS, Hasan MK, Miah MA, Sur GC, Felsenstein A, Venkatesan M, Sack RB, Albert MJ.** 1993. Use of the polymerase chain reaction and fluorescent-antibody methods for detection of viable but nonculturable *Shigella dysenteriae* type 1 in laboratory microcosms. Appl. Environ. Microbiol. **59**:536–540.
24. **Chiu CH, Ou JT.** 1996. Rapid identification of *Salmonella* serovars in feces by specific detection of virulence genes, *invA* and *spvC*, by an enrichment broth culture-multiplex PCR combination assay. J. Clin. Microbiol. **34**:2619–2622.
25. **Osek J.** 2003. Development of a multiplex PCR approach for the identification of Shiga toxin-producing *Escherichia coli* strains and their major virulence factor genes. J. Appl. Microbiol. **95**:1217–1225.
26. **Jiang SC, Chu W, Olson BH, He Choi J-WS, Zhang J, Le JY, Gedalanga PB.** 2007. Microbial source tracking in a small Southern California urban watershed indicates wild animals and growth as the source of fecal bacteria. Appl. Microbiol. Biotechnol. **76**:927–934.
27. **Khatib LA, Tsai YL, Olson BH.** 2003. A biomarker for the identification of swine fecal pollution in water, using the STII toxin gene from enterotoxigenic *Escherichia coli*. Appl. Microbiol. Biotechnol. **63**:231–238.
28. **US Environmental Protection Agency.** 2004. Quality assurance/quality control guidance for laboratories performing PCR analysis on environmental samples. EPA-815-B-04-001. US Environmental Protection Agency, Washington, DC.
29. **Baylis CL, MacPhee S, Martin KW, Humphrey TJ, Betts RP.** 2000. Comparison of three enrichment media for the isolation of *Campylobacter* spp. from foods. J. Appl. Microbiol. **89**:884–891.
30. **Inglis GD, Kalischuk LD.** 2003. Use of PCR for direct detection of *Campylobacter* species in bovine feces. Appl. Environ. Microbiol. **69**:3435–3447.
31. **Jothikumar N, Cromeans TL, Hill VR, Lu X, Sobsey MD, Erdman DD.** 2005. Quantitative real-time PCR assays for detection of human adenovirus and identification of serotypes 40 and 41. Appl. Environ. Microbiol. **71**:3131–3136.
32. **Gregory JB, Litaker RW, Noble RT.** 2006. Rapid one-step quantitative reverse transcriptase PCR assay with competitive internal positive control for detection of enteroviruses in environmental samples. Appl. Environ. Microbiol. **72**:3960–3967.
33. **Ware MW, Wymer L, Lindquist HD, Schaefer FW, III.** 2003. Evaluation of an alternative IMS dissociation procedure for use with method 1622: detection of Cryptosporidium in water. J. Microbiol. Methods **55**(3):575–583.
34. **US Environmental Protection Agency.** 2005. Method 1623: *Cryptosporidium* and *Giardia* in water by filtration/IMS/FA. EPA-815-R-05-002. Office of Water, US Environmental Protection Agency, Washington, DC.
35. **Wilde FD, ed.** Variously dated. Field measurements: U.S. Geological Survey techniques of water-resources investigations, book 9, chapter A6, sections 6.1, 6.3, and 6.7. US Geological Survey, Reston, VA. http://pubs.water.usgs.gov/twri9A6/.
36. **US Environmental Protection Agency.** 2012. Exposure assessment models—Virtual Beach. Center for Exposure Assessment Modeling, US Environmental Protection Agency, Athens, GA. http://www.epa.gov/ceampubl/swater/vb2/index.html.
37. **Francy DS, Darner RA.** 2006. Procedures for developing models to predict exceedance of recreational water-quality standards at coastal beaches. U.S. Geological Survey, techniques and methods, 6-B5. US Geological Survey, Reston, VA. http://pubs.usgs.gov/tm/2006/tm6b5/.
38. **Nevers MB, Whitman RL.** 2005. Nowcast modeling of *Escherichia coli* concentrations at multiple urban beaches of southern Lake Michigan. Water Res. **39**:5250–5260.
39. **Francy DS, Darner RA, Bertke EE.** 2006. Models for predicting recreational water quality at Lake Erie beaches. U.S. Geological Survey scientific investigations report 2006-5192. US Geological Survey, Reston, VA. http://pubs.usgs.gov/sir/2006/5192/.
40. **Stidson RT, Gray CA, McPhail CD.** 2011. Development and use of modeling techniques for real-time bathing water quality predictions. Water Environ. J. **26**:7–18.
41. **Zimmerman TM.** 2006. Monitoring and modeling to predict *Escherichia coli* at Presque Isle Beach 2, City of Erie, Erie County, Pennsylvania. US Geological Survey scientific investigations report 2006-5159. US Geological Survey, Reston, VA.
42. **Olyphant GA, Whitman RL.** 2004. Elements of a predictive model for

determining beach closures on a real time basis—the case of 63rd Street Beach Chicago. Environ. Monit. Assess. **98**:175–190.
43. **Hou D, Rabinovici SJM, Boehm AB.** 2006. Enterococci predictions from partial least squares regression models in conjunction with a single-sample standard improve the efficacy of beach management advisories. Environ. Sci. Technol. **40**:1737–1743.
44. **Gerba CP.** 2000. Assessment of enteric pathogen shedding by bathers during recreational activity and its impact on water quality. Quant. Microbiol. **2**:55–68.
45. **Francy DS, Gifford AM, Darner RA.** 2003. *Escherichia coli* at Ohio bathing beaches—distribution, sources, wastewater indicators, and predictive modeling. Water-resources investigations report 02-4285. US Geological Survey, Columbus, OH. http://oh.water.usgs.gov/reports/Abstracts/wrir02-4285.html.
46. **Xagoraraki I, Kuo DH-W, Wong K, Wong M, Rose JB.** 2007. Occurrence of human adenoviruses at two recreational beaches of the Great Lakes. Appl. Environ. Microbiol. **73**:7874–7881.
47. **Dorevitch S, Doi M, Hsu F-C, Lin K-T, Roberts JD, Liu LC, Gladding R, Vannoy E, Li H, Javor M, Scheff PA.** 2011. A comparison of rapid and conventional measures of indicator bacteria as predictors of waterborne protozoan pathogen presence and density. J. Environ. Monit. **13**:2427–2435.
48. **Schets FM, van Wijnen JH, Schijven JF, Schoon H, de Roda Husman AM.** 2008. Monitoring of waterborne pathogens in surface waters in Amsterdam, The Netherlands, and the potential health risk associated with exposure to *Cryptosporidium* and *Giardia* in these waters. Appl. Environ. Microbiol. **74**:2069–2078.
49. **Till D, McBride G, Ball A, Taylor K, Pyle E.** 2008. Large-scale freshwater microbiological study: rationale, results and risks. J. Water Health **6**:443–460.
50. **Yoder JS, Hlavsa MC, Craun GF, Hill V, Roberts V, Yu PA.** 2008. Surveillance for waterborne disease and outbreaks associated with recreational water use and other aquatic facility-associated health events—United States, 2005-2006. MMWR Surveill. Summ. **57**:1–29.
51. **Water Environment Research Federation.** 2010. Quantification of pathogens and sources of microbial indicators for QMRA in recreational waters. PATH2R08. Water Environment Research Federation, Alexandria, VA.
52. **Keene WE, McAnulty JM, Hoesly FC, Williams LP, Hedberg K, Oxman GL, Barrett TJ, Pfaller MA, Fleming DW.** 1994. A swimming-associated outbreak of hemorrhagic colitis caused by Escherichia coli O157:H7 and Shigella sonnei. N. Engl. J. Med. **331**:579–584.
53. **Casas V, Miyake J, Balsley H, Roark J, Telles S, Leeds S, Zurita I, Breitbart M, Bartlett D, Azam F, Rohwer F.** 2006. Widespread occurrence of phage-encoded exotoxin genes in terrestrial and aquatic environments in Southern California. FEMS Microbiol. Lett. **261**:141–149.

TABLE S1 Inland lake study sites and agencies responsible for sampling, 2010–12

Site No.	Site name	USGS Site Identification Number[a]	Short name	Site description	Lake size (acres)	Maximum Depth (ft)	Sampled by
1	Alum Creek Reservoir at State Park Camper Beach	401409082584000	Alum Campers	Campers	3,269	60	USGS Columbus
2	Alum Creek Reservoir at State Park Beach North	401131082581300	Alum North	Popular bathing beach			USGS Columbus
3	Alum Creek Reservoir at State Park Beach center	401126082581300	Alum Central	Popular bathing beach			USGS Columbus
4	Atwood Lake Main Beach	403219081155500	Atwood Main	Popular bathing beach	1,540	30	MWCD[b]
5	Atwood Lake Boaters nr Islands	403232081151000	Atwood Islands	Boaters			MWCD
6	Atwood Lake Boaters nr cove	403252081144900	Atwood Cove	Boaters			MWCD
7	Buckeye Lake at State Park Brook's Park Beach	395405082310000	Buckeye Brooks	Small beach	2,873	14	USGS Columbus
8	Buckeye Lake at State Park Fairfield Beach	395520082281500	Buckeye Fairfield	Small beach			USGS Columbus
9	Buckeye Lake at State Park Crystal Beach	395557082283800	Buckeye Crystal	Small beach			USGS Columbus
10	CJ Brown Reservoir at State Park Camper Beach	395801083434700	Buck Creek Campers	Campers	1,970	50	CCCHD[c]
11	CJ Brown Reservoir at Main Beach North	395705083440100	Buck Creek North	Popular bathing beach			CCCHD
12	CJ Brown Reservoir at Main Beach South	395653083441200	Buck Creek South	Popular bathing beach			CCCHD
13	Lake at Eastview Park Beach at Celina	'403304084323300	Eastview	Popular bathing beach			USGS in Celina
14	Grand Lake at State Park Campers Beach	403242084262500	GLSM Campers	Campers	12,896	16	USGS in Celina
15	Grand Lake at State Park West Beach at St. Marys	403235084253900	GLSM West	Small beach			USGS in Celina
16	Grand Lake at State Park East Beach at St. Marys	403229084251600	GLSM East	Small beach			USGS in Celina
17	Seneca Lake at Swimming Beach near Senecaville	395433081250100	Seneca	Popular bathing beach	3,550	31	MWCD
18	Tappan Lake Boaters at South Shore	402013081122700	Tappan South	Boaters			MWCD
19	Tappan Lake Boaters at Bontrager Bay	402004081115700	Tappan Bontrager	Boaters			MWCD
20	Tappan Lake at Main Swimmers Beach	401926081105100	Tappan Main	Popular bathing beach	2,350	34	MWCD
21	Unnamed ditch tributary to Tappan Lake	401910081111200	Tappan ditch	Ditch			MWCD
22	Tappan Lake Boaters at Beall Bay	401951081092400	Tappan Beall	Boaters			MWCD

[a] Corresponds to the latitude and longitude plus two additional numbers (usually 00 unless there are multiple locations at the same site)

[b] Muskingum Watershed Conservancy District

[c] Clark County Combined Health District

TABLE S2 List of polymerase chain reaction targets, conditions[a], and detection Limits.

Gene	Primers 5´-3´	Detection Limit (ng/µL)	Control DNA	Primer (µM)	Polymerase	PCR Buffer	MgCl2 (mM)	BSA (ug/uL)	Initial Denaturing Temp. (°C)	Initial Denaturing Time (min: sec)
Shiga-toxin producing *E. coli* Multiplex PCR										
stx1	F: ACACTGGATGATCTCAGTGG	0.1	ATCC 35150	0.05	ABI Amplitaq Gold	Buffer II	3	0.1	95	10:00
	R: CTGAATCCCCCTCCATTATG			0.05						
stx2	F: CCATGACAACGGACAGCAGTT	0.1	ATCC 35150	0.05						
	R: CCTGTCAACTGAGCAGCACTTTG			0.05						
eaeA	F: GTGGCGAATACTGGCGAGACT	0.1	ATCC 35150	0.05						
	R: CCCCATTCTTTTTCACCGTCG			0.05						
E. coli 16s rDNA	F: GGAAGAAGCTTGCTTCTTTGCTGAC	0.1	ATCC 35150	0.025						
	R: AGCCCGGGGATTTCACATCTGACTTA			0.025						
Pathogenic *E. coli*										
rfbO157	F: CGTGATGATGTTGAGTTG	0.01	ATCC 35150	0.1	ABI Amplitaq Gold	Buffer II	2	0.1	95	10:00
	R: AGATTGGGTTGGCATTACTG			0.1						
STII	F:GCATCTATGTTCGTTTTTTCTATTG	0.001	cloned fragment	0.5	Promega GoTaq	Green Flexi	2	--	95	1:00
	R:GCAACCATTATTTGGGCG			0.5						
STh	F: CSCTCAGGATGCTAAACCAG		ATCC 35401	0.4	ABI Amplitaq Gold	Buffer II	2.5	0.1	96	3:00
	R: TTAATAGCACCCGGTACAAGC			0.4						
LTIIa	F:GGGTGTGCATTTCAGCGAC	0.001	cloned fragment	0.5	Promega GoTaq	Green Flexi	2	--	95	1:00
	R:CGTCCACCCGGAATATACCA			0.5						
Campylobacter jejuni and *Campylobacter coli*										
Campy 16s rDNA	F: ATCTAATGGCTTAACCATTAAAC	0.01	ATCC 33291	0.5	Promega GoTaq	Green Flexi	3	--	95	10:00
	R: GGACGGTAACTAGTTTAGTATT			0.5						
Shigella										
ipaH	F: GTTCCTTGACCGCCTTTCCGATAC	0.01	ATCC 9290	0.25	ABI Amplitaq Gold	Buffer II	3	0.1	94	5:00
	R: GCCGGTCAGCCACCCTC			0.25						
Salmonella										
invA	F: ACAGTGCTCGTTTACGACCTGAAT	0.1	ATCC 14028	0.25	Promega GoTaq	Green Flexi	1.5	1	94	5:00
	R: AGACGACTGGTACTGATCGATAAT			0.25						
spvC	F: ACTCCTTGCACAACCAAATGCGGA	0.1	ATCC 14028	0.5	Promega GoTaq	Green Flexi	1.5	0.3	94	5:00
	R: TGTCTTCTGCATTTCGCCACCATCA			0.5						

[a]All concentrations are reported as final concentrations in 15 uL reactions, all reactions used 0.2 mM dNTPs

[b]Touchdown annealing from 65 °C to 55 °C for the first 10 cycles, final 20 cycles 55 °C annealing temperature

TABLE S2 List of polymerase chain reaction targets, conditions[a], and detection Limits.

		Denaturing		Annealing		Extension		Final Extension		
Gene	Cycles	Temp. (°C)	Time (Min: Sec)	Temp. (°C)	Time (Min)	Temp. (°C)	Time (Min)	Temp. (°C)	Time (Min)	Reference
Shiga-toxin producing *E. coli* Multiplex PCR										
stx1 stx2 eaeA E. coli 16s rDNA	35	95	0:30	56	0:40	72	1:30	72	5:00	Duris et al., 2009
Pathogenic *E. coli*										
rfbO157	30	94	0:30	53	0:30	72	0:30	72	5:00	Osek, 2003
STII	40	95	0:30	61	0:30	72	0:30	72	6:00	Khatib et al., 2003
STh	30	94	0:30	touch-down[b]	0:30	72	1:30	72	7:00	Jiang et al., 2007
LTIIa	35	95	0:30	57	0:30	72	0:30	72	6:00	Jiang et al., 2007
Campylobacter jejuni and *Campylobacter coli*										
Campy 16s rDNA	35	95	0:30	61	1:30	72	1:00	72	10:00	Inglis and Kalischuck, 2003
Shigella										
ipaH	35	94	1:00	60	0:30	72	1:00	72	5:00	Islam et al., 1993
Salmonella										
invA	30	94	0:30	56.5	0:50	72	0:30	72	5:00	Chiu and Ou, 1996
spvC	30	94	0:30	53.5	0:50	72	1:00	72	10:00	Chiu and Ou, 1996

[a]All concentrations are reported as final concentrations in 15 uL reactions, all reactions used 0.2 mM dNTPs

[b]Touchdown annealing from 65 °C to 55 °C for the first 10 cycles, final 20 cycles 55 °C annealing temperature

TABLE S3 Sources of environmental data for predictive models for each lake[a]

Site Nos.	Lake	Rainfall and wind speed and direction	Stream stage or discharge	Water surface elevation (lake or reservoir)	Solar radiation
1-3	Alum Creek	•NWS, Columbus Ohio State University Airport •USGS 03228805, Alum Creek at Africa •Radar, 17 cells	•USGS 03228805, discharge	•USACE Alum Cr Reservoir nr Westerville (USGS 03228804)	•OARDC Delaware Station
4-6	Atwood Lake	•NWS, New Philadelphia, Harry Clever Field •USGS 03120500, McGuire Cr. at Leesville •Radar, 10 cells and 1 cell	•USGS 03120500, stage		
7-9	Buckeye Lake	•NWS, Newark Heath Airport •USGS 395417082314200, Buckeye Lk nr Millersport •Radar, 15 cells		•USGS 395417082314200, Buckeye Lake near Millersport	
10-12	Buck Creek	•NWS, Cox Dayton International Airport •USGS 395726083445400, Rain gage at CJ Brown Reservoir Dam •Radar, 12 cells and 1 cell	•USGS 03267900, Mad River at Paris Pike at Eagle City, discharge	•USACE Clarence J. Brown Reservoir nr. Springfield (USGS 03268090)	•OARDC Western Station
13-16	Grand Lake St. Marys	•NWS, Lima Allen County Airport •USGS 04180988, St. Marys River at Rockford •USGS 403233084342200, Weather Station at Celina Water Plant •Radar, 18 cells	•USGS 04180988, discharge		
18-22	Tappan Lake	•NWS, New Philadelphia, Harry Clever Field •USGS 03120500, McGuire Cr. at Leesville •Radar, 12 cells and 1 cell	•USGS 03120500, stage		

[a] USGS, U.S. Geological Survey; NWS, National Weather Service; USACE, U.S. Army Corps of Engineers; OARDC, Ohio Agricultural Research and Development Center

TABLE S4. Concentrations or detections of bacterial indicators, pathogens, and microbial source tracking (MST) markers in inland lake samples and associated predictive model variables, 2011 [a]

				Protozoan pathogens				Enteric viruses (gc/L)				Bacterial pathogens (presence=1, absence=0)											
				Cryptosporidium (oocysts/10 L)		Giardia (cysts/10 L)		Adenovirus		Enterovirus			STEC genes							Shigella	Salmonella	
Date	Time	Site	E. coli (MPN/ 100 mL)	Ultra-filter	Glass wool filter	Ultra-filter	Glass wool filter	Ultra-filter	Glass wool filter	Ultra-filter	Glass wool filter	Campy	eaeA	stx2	stx1	rfb0157	LTIIa	STh	STII	shig (ipah)	invA	spvC
7/11	9:19	Atwood AM	A 6	--	--	--	--	--	--	--	--	0										
7/31	8:47	Atwood AM	32	--	--	--	--	--	--	--	--	0										
7/31	12:41	Atwood PM	A 660	--	--	--	--	--	--	--	--	0										
8/27	14:00	Atwood PM	A 36	<0.2	<0.2	<0.2	<0.2	<2.2	E~b3.4	<11	<14		0	0	0	0	0	0	0	0	0	0
9/11	13:38	Atwood PM	210	0.2	<1.6	<0.1	<1.6	E~b39	<18	<91	<91	0	1	0	0	0	0	0	0	0	0	0
7/11	9:55	Buckeye Crystal	29	2	<0.2	<1	<0.2	<16	<1.2	<159	<12	0	0	1	0	0	0	0	0	1	0	0
7/17	7:05	Buckeye Crystal	58	--							--	0	0	0	0	0	0	0	0	1	0	1
7/24	7:10	Buckeye Crystal	130	--	--	--	--	--	--	--	--		1	0	0	0	0	0	0	1	0	1
7/30	6:55	Buckeye Crystal	920	--	--	--	--	--	--	--	--		1	0	0	0	0	0	0	0	0	1
8/8	8:22	Buckeye Crystal	24	--	--	--	--	--	--	--	--	0	1	0	0	0	0	0	0	0	0	0
8/22	7:12	Buckeye Crystal	650	--	--	--	--	--	--	--	--	0	1	0	0	0	0	0	0	0	0	1
8/30	8:10	Buckeye Crystal	A 740	<1.3	<0.3	<1.3	<0.3	<20	<1.8	<196	<18	0	1	0	0	0	0	0	0	0	0	0
9/13	10:50	Buckeye Crystal	39	<0.3	<0.4	<0.3	<0.4	<18	<2.5	<91	<25	0	1	0	0	0	0	0	0	1	0	0
6/28	11:51	Buckeye Fairfield	820	<1	0.2	<1	<0.2	<9.2	<1.2	<46	<6	0	1	0	0	0	0	0	0	0	0	0
7/17	7:38	Buckeye Fairfield	19	--	--	--	--	--	--	--	--	0	1	0	0	0	0	0	0	0	0	0
7/24	8:00	Buckeye Fairfield	210	--	--	--	--	--	--	--	--		1	0	0	0	0	0	0	0	0	1
7/25	10:07	Buckeye Fairfield	210	0.5	<0.2	<0.5	<0.2	<8.2	E~b4.7	<42	<16	1	0	0	0	0	0	0	0	0	0	0
7/30	7:35	Buckeye Fairfield	80	--	--	--	--	--	--	--	--		1	0	0	0	0	0	0	0	0	0
8/8	8:02	Buckeye Fairfield	26	--	--	--	--	--	--	--	--	0	1	0	0	0	0	0	0	0	0	0
8/22	8:00	Buckeye Fairfield	39	--	--	--	--	--	--	--	--	0	1	0	0	0	0	0	0	0	0	0
9/6	9:33	Buckeye Fairfield	A 17	<0.5	<0.4	<0.5	<0.4	<16	E~b10	<77	<11	0	1	0	0	0	0	0	0	1	0	0
6/29	10:23	Tappan Main	3	<0.4	<0.2	0.8	<0.2	<11	<6.9	<44	<69	0	0	0	0	0	0	0	0	1	0	0
7/11	10:54	Tappan Main	201	--	--	--	--	--	--	--	--	0										
8/7	13:17	Tappan Main	122	<0.3	<0.1	<0.3	<0.1	<28	<3.6	<139	<18	0	1	0	0	0	0	0	0	1	0	0
8/14	10:24	Tappan Main	96	--	--	--			--	--		0										
8/15	10:17	Tappan Main	21	0.3	<0.1	<0.3	<0.1	<3.6	<13	<36	<53	0										
6/29	13:51	Tappan ditch	34	<0.7	--	<0.7	--	<13	--	<128	--	0	0	0	0	0	0	0	0	1	0	0
7/11	11:16	Tappan ditch	344	--	--	--	--	--	--	--	--	0										
8/7	17:53	Tappan ditch	2000	1.2	--	<0.3	--	<118	--	<1,100	--	0	0	0	0	1	0	0	0	0	0	1
8/14	10:39	Tappan ditch	>2400	--	--	--	--	--	--	--	--	0										
8/15	13:10	Tappan ditch	ND	<0.4	<6	<0.4	<6	<14	<112	<139	<222	0										

[a] A, average of two replicates; R, value reported was detected in one field replicate, but not in the second replicate; --, not done; E, estimated value; ~, duplicates do not check (Ct values); b, value was extrapolated below lowest method range or instrument linear range

[b] The probability (in percent) of exceeding the single-sample bathing water standard of 235 CFU/100 mL. Probabilities above the established threshold probability (beach is posted with advisory) are italicized

TABLE S4. Concentrations or detections of bacterial indicators, pathogens, and microbial source tracking (MST) markers in inland lake samples and associated predictive model variables, 2011 [a]

			Environmental and water-quality variables												
			Rainfall				Airport Wind, 24 h		Field variables						Model output
Date	Time	Site	Airport rain, 24 h (in)	USGS gage, 24 h (in)	Radar, Ave Max 48 h weight (in)	Radar hourly total 24 h	Wind Alongshore 24 h	Wind Offshore 24 h	Water Temp (°C)	Conduc-tance	Wave height (in)	Turb (NTRU)	Birds	Swimmers	Probability[a]
7/11	9:19	Atwood AM	0.00	0.00	0.00	0.00	0.37	1.86	28.6	--	0.25	5.5	0	0	4.3
7/31	8:47	Atwood AM	0.01	0.00	0.03	0.00	-1.62	-0.78	29.0	200	0.5	10.3	0	0	10.2
7/31	12:41	Atwood PM	0.00	0.00	0.03	0.00	-1.62	-0.78	30.4	200	1.5	15.8	0	87	***63.0***
8/27	14:00	Atwood PM	0.01	0.01	0.00	0.00	-1.97	-0.74	26.0	201	4	16.9	4	20	20.6
9/11	13:38	Atwood PM	0.02	0.00	0.08	0.84	0.01	-0.9	24.4	202	6	12.5	30	0	4.3
7/11	9:55	Buckeye Crystal	0.00	0.00	0.00	0.00	1.87	0.32	29.1	315	0	22.3	31	0	18.4
7/17	7:05	Buckeye Crystal	0.00	0.00	0.00	0.00	2.31	2.36	27.3	312	0.25	24.3	0	0	13.8
7/24	7:10	Buckeye Crystal	1.00	1.79	--	22.96	1.99	-0.66	29.6	286	0	22.7	0	0	***75.0***
7/30	6:55	Buckeye Crystal	0.19	0.49	0.68	12.14	1.56	-3.24	29.5	287	0	29.4	0	0	***38.8***
8/8	8:22	Buckeye Crystal	0.29	0.35	0.56	7.45	1.46	-3.18	27.2	229	0	30.3	0	0	***44.4***
8/22	7:12	Buckeye Crystal	0.75	0.83	0.80	12.03	0.03	-2.4	25.5	286	0	27.5	3	0	***74.3***
8/30	8:10	Buckeye Crystal	0.00	0.00	0.00	0.00	-1.33	1.21	24.5	273	0.5	28	0	0	11.5
9/13	10:50	Buckeye Crystal	0.08	0.02	0.33	1.84	2.07	-1.41	26.5	288	0.5	33.2	4	0	***43.3***
6/28	11:51	Buckeye Fairfield	0.005	0.06	0.00	0.00	-2.55	-3.83	24.6	341	0	27.8	3	1	1.3
7/17	7:38	Buckeye Fairfield	0.00	0.00	0.00	0.00	-0.15	-3.3	26.5	343	0.25	29.7	130	0	5.7
7/24	8:00	Buckeye Fairfield	1.00	1.79	--	22.96	-1.93	-0.83	28.6	310	0.5	33.8	69	0	***45.2***
7/25	10:07	Buckeye Fairfield	0.03	0.69	1.33	5.81	-2.06	1.22	29.7	341	1	32.3	0	0	15.8
7/30	7:35	Buckeye Fairfield	0.19	0.49	0.68	12.14	-3.32	1.39	29.1	310	0.25	40.9	1	0	15.0
8/8	8:02	Buckeye Fairfield	0.29	0.35	0.56	7.45	-3.21	1.41	27.8	240	1	35.2	26	0	25.4
8/22	8:00	Buckeye Fairfield	0.75	0.83	0.80	12.03	-1.62	1.77	24.6	--	0.25	35.7	0	0	7.4
9/6	9:33	Buckeye Fairfield	0.07	0.02	0.47	1.42	5.46	5.15	22.0	305	7	36.5	4	0	1.6
6/29	10:23	Tappan Main	0.00	0.00	0.00	0.00	2.88	3.19	25.0	609	0	10.6	--	20	1.8
7/11	10:54	Tappan Main	0.00	0.00	0.00	0.00	-1.88	-0.3	29.7	616	0.25	12.9	31	0	29.6
8/7	13:17	Tappan Main	0.30	0.09	0.30	2.99			29.0	608	0.5	19.6	15	40	***41.8***
8/14	10:24	Tappan Main	0.875	1.37	1.24	15.12	-0.97	0.23	25.4	636	0	15.4	38	0	***46.7***
8/15	10:17	Tappan Main	0.49	0.03	0.84	4.94	2.46	1.53	25.0	601	3	13.8	8	0	22.8
6/29	13:51	Tappan ditch	0.00	0.00	0.00	0.00	--	--	24.6	554	--	23.1	0	0	--
7/11	11:16	Tappan ditch	0.00	0.00	0.00	0.00	--	--	26.4	572	--	34.6	50	0	--
8/7	17:53	Tappan ditch	0.30	0.09	0.30	2.99	--	--	28.8	461	--	32.2	0	0	--
8/14	10:39	Tappan ditch	0.875	1.37	1.24	15.12	--	--	22.0	464	--	77.8	0	0	--
8/15	13:10	Tappan ditch	0.49	0.03	0.84	4.94	--	--	--	530	--	20.2	0	0	--

[a] A, average of two replicates; R, value reported was detected in one field replicate, but not in the second replicate; --, not done; E, estimated value; ~, duplicates do not check (Ct values); b, value was extrapolated below lowest method range or instrument linear range

[a] The probability (in percent) of exceeding the single-sample bathing water standard of 235 CFU/100 mL. Probabilities above the established threshold probability (beach is posted with advisory) are italicized

Supplemental File 2. Equations for the selected best models for each inland lake site. Models were used to predict *E. coli* concentrations from measurements of environmental and water-quality variables. Equations were generated using Virtual Beach software (1).

Alum Creek State Park

Gage rainfall data from USGS 03120500, Alum Creek at Africa; radar rainfall from 17, 4-km cells

Alum Campers

$Log_{10}(E.\ coli) = 0.552 - 0.006*(DayofYear) + 0.252*\ (Radarxcell\text{-}av\text{-}R_{d\text{-}2})^{0.5} + 0.149*(GageRw72) + 1.549*Log_{10}(Turbidity)$

Alum North

$Log_{10}(E.\ coli) = -1.397 + 0.092*(WaterTemp) + 0.347*\ (Radarxcell\text{-}av\text{-}R_{d\text{-}3})^{0.5} + 0.060*(Turbidity) + 0.181*(GageRw48)$

Alum Central

$Log_{10}(E.\ coli) = 0.372 + 0.001*(WaterTemp)^2 + 0.012*(Radarxcell\text{-}av\text{-}Rw72)^2 + 0.003*(Turbidity)^2 + 0.267*(GageR_{d\text{-}1})$

Atwood Lake Main Beach

Airport rainfall data from New Philadelphia, Harry Clever Field; radar rainfall from 10, 4-km cells

$Log_{10}(E.\ coli) = -0.742 + 0.002*(WaterTemp)^2 + 0.275*\ (Turbidity)^{0.5} + 0.133*(AirportRw48 + Radarxcell\text{-}av\text{-}Rw48) + 0.072*\ (Swim\_no)^{0.5}$

Buckeye Lake State Park

Lake level data from USGS 395417082314200, Buckeye Lk nr Millersport; airport rainfall and wind data from Newark Heath Airport; radar rainfall from 15, 4-km cells

Buckeye Brooks

$Log_{10}(E.\ coli) = -1{,}600 + 1.787*(LakeLevel) + 6.082*(LakeLevelChange) + 0.142*(WaterTemp)$

Buckeye Crystal

$Log_{10}(E.\ coli) = 1.922 + 0.035*(AirportWindA\_comp24) + 0.313*(AirportRw48) - 0.044*(AirportWindO\_comp24)$

(Beach orientation is 18.4 degrees)

Buckeye Fairfield

$Log_{10}(E.\ coli) = 1.204 - 0.076*(AirportWindA\_comp24) + 0.079*\ (Radarxcell\text{-}sum\text{-}Rw48)^{0.5} + 0.040*\ (Birds\_no)^{0.5} + 0.520*\ (WaveHt)^{0.5}$

(Beach orientation is -126.67 degrees)

Buck Creek State Park

Discharge data from USGS 03267900, Mad River at Paris Pike at Eagle City; airport rainfall and wind data from Cox Dayton International Airport; radar rainfall from 12, 4-km cells

Buck Creek North

$Log_{10}(E.\ coli) = 0.600 + 0.950*(Radarxcell\text{-}av\text{-}R_{d\text{-}1})^{0.5} + 171.03*[1/(DischR_{d\text{-}2})] + 0.036*(AirportWindSp*WindCode)$

Buck Creek South

$Log_{10}(E.\ coli) = 0.632 + 134.45*[1/(Disch48)] + 0.924*(AirportR_{d\text{-}1})^{0.5} + 0.032*(Birds\_no) + 0.033*(AirportWindSp*WindCode)$

Grand Lake St. Marys State Park

Airport wind data from Lima Allen County Airport; radar rainfall from 18, 4-km cells; gage rainfall from USGS 04180988, St. Marys River at Rockford

GLSM Campers

$Log_{10}(E.\ coli) = 2.317 + 0.054*(AirportWindSp*WindCode) - 0.00002*(DayofYear)^2 + 0.361*(GageRw72)^{0.5} + 0.000003*(Turbidity)^2$

GLSM West

$Log_{10}(E.\ coli) = 1.885 + 0.094*\ (AirportWindA\_comp\_Inst)^{0.5} + 0.947*(Radarxcell\text{-}av\text{-}R_{d\text{-}1}) + 0.624*(Radarxcell\text{-}av\text{-}R_{d\text{-}3}) + 1.668*(GageR_{d\text{-}3})^2$

(Beach orientation is 81.6 degrees)

GLSM East

$Log_{10}(E.\ coli) = 0.826 + 0.163*\ (AirportWindSp*WindCode)^{0.5} + 0.001*(WaterTemp)^2 + 1.402*(GageR_{d\text{-}3}) + 0.079*(Radarxcell\text{-}sum\text{-}R_{d\text{-}1})^{0.5}$

Tappan Lake Main Beach

Airport wind data from New Philadelphia, Harry Clever Field; radar rainfall from 12, 4-km cells

$Log_{10}(E.\ coli) = -1.210 + 0.050*(AirportWindO\_comp\_Inst) + 0.127*\ (Radarxcell\text{-}sum\text{-}Rw72)^{0.5} + 0.876*Log_{10}(Turbidity) + 0.064*(WaterTemp) - 0.082*(AirportWindA\_comp\_Inst)$

(Beach orientation is -177.05 degrees)

Key to acronyms and variables

nr, near
USGS, U.S. Geological Survey
USACE, U.S. Army Corps of Engineers

Field observations or measurements

E. coli: concentration of *Escherichia coli*, in most-probable number per 100 milliliters
Swim_no: the number of swimmers in the water at the time of sampling
Turbidity: turbidity of the sample, in nephelometric turbidity ratio units
WaterTemp: water temperature at time of sampling, in degrees Celsius
WaveHt: wave height as measured with a graduated rod, in feet
Birds_no: the number of birds on the beach at the time of sampling
Dayofyear: the number representing the date beginning with 1 for January 1 and 365 or 366 for December 31 (the latter being a leap year)

Rainfall from USGS gage ("Gage") or National Weather Service nearest airport site ("Airport")

R_{d-1}: the total rainfall, in inches, for the 24-h period before sampling
R_{d-2}: the total rainfall, in inches, for the 24-h period 2 days before sampling
R_{d-3}: the total rainfall, in inches, for the 24-h period 3 days before sampling
Rw48: the amount of rainfall, in inches, for the 48-h period before sampling, with the most recent rainfall receiving the most weight. Calculated as $(2*R_{d-1})+R_{d-2}$
Rw72: the amount of rainfall, in inches, for the 72-h period before sampling, with the most recent rainfall receiving the most weight. Calculated as $(3*R_{d-1}) + (2*R_{d-2}) + R_{d-3}$

Wind direction and speed from the National Weather Service nearest airport site

WindA_comp: a measure of the component of the wind velocity moving parallel to the shoreline, for the instantaneous (INST) value near 8 a.m. or a 24-hr vector up to 8 a.m. (24), calculated as:

Wind A = -wind speed * cosine ((wind direction – beach orientation) * π/180)

A positive value indicates winds are moving from right to left across the beach when looking toward the water from the shoreline (U.S. Environmental Protection Agency, 2012).

WindO_comp: a measure of the component of the wind velocity moving perpendicular to the shoreline, the instantaneous (INST) value near 8 a.m. or a 24-hr vector up to 8 a.m. (24), calculated as:

Wind O = wind speed * sine ((wind direction – beach orientation) * π/180)

A positive value indicates winds are moving from the water toward shore (U.S. Environmental Protection Agency, 2012).

WindCode: site-specific wind code calculated by assigning the most weight to the range of wind directions associated with the highest *E. coli* concentrations.

WindSp: wind speed, in miles per hour

Radar rainfall from National Weather Service

Radarxcell-av-R_{d-1}: hourly maximum radar rainfall among multiple cells divided by the number of cells for the 24-h period before sampling

Radarxcell-av-R_{d-2}: hourly maximum radar rainfall among multiple cells divided by the number of cells for the 24-h period 2 days before sampling

Radarxcell-av-R_{d-3}: hourly maximum among multiple cells divided by the number of cells for the 24-h period 3 days before sampling

Radarxcell-av-Rw48: hourly maximum radar rainfall among multiple cells divided by the number of cells for the 24-h periods 1 and 2 days before sampling. The most recent rainfall receives the most weight. Calculated as: (2*Radarxcell-av-R_{d-1}) + Radarxcell-av-R_{d-2}

Radarxcell-av-Rw72: hourly maximum radar rainfall among multiple cells divided by the number of cells for the 24-h periods 1, 2 and 3 days before sampling. The most recent rainfall receives the most weight. Calculated as: (3*Radarxcell-av-R_{d-1}) + (2*Radarxcell-av-R_{d-2}) + Radarxcell-av-R_{d-3}

Radarxcell-sum-R_{d-1}: the sum from multiple cells for the 24-h period before sampling

Radarxcell-sum-Rw48: the sum from multiple cells for the 24-h periods 1 and 2 days before sampling. The most recent rainfall receives the most weight. Calculated as: (2*Radarxcell-sum-R_{d-1}) + Radarxcell-sum-R_{d-2}

Radarxcell-sum-Rw72: the sum from multiple cells for the 24-h periods 1, 2 and 3 days before sampling. The most recent rainfall receives the most weight. Calculated as: (3*Radarxcell-sum-R_{d-1}) + (2*Radarxcell-sum-R_{d-2}) + Radarxcell-sum-R_{d-3}

Stream discharge and water surface elevation from a nearby USGS or USACE gage

DischR_{d-1}: the mean discharge, in cubic feet per second, for the 24-h period 2 days before sampling

LakeLevel: lake level at 8 a.m., in feet

LakeLevelChange: change in lake level (today – yesterday) at 8 a.m., in feet

References

1. **U.S. Environmental Protection Agency**. 2012. Exposure Assessment Models—Virtual Beach. Center for Exposure Assessment Modeling, U.S. Environmental Protection Agency, Athens, GA. http://www.epa.gov/ceampubl/swater/vb2/index.html.

Supplemental File 3. Determining probabilities and establishing a threshold probability for issuing advisories—an example at Buck Creek State Park, Ohio. The method described below was modified from Francy and Darner (1).

Two types of output may be produced by multiple linear regression models for determining bacterial water quality at recreational sites. The first is the predicted bacterial-indicator concentration. The second output—the probability that the single-sample maximum bathing-water standard (235 CFU/100 mL for *E. coli* at Ohio beaches) will be exceeded—was added because prediction intervals were shown to be fairly wide (2). This approach has been successfully applied to beaches that are part of the Ohio Nowcast at Great Lakes beaches (http://www.ohionowcast.info/). Predictive models applied through the Ohio Nowcast have provided more accurate predictions than the current method for assessing water quality (using the previous days' *E. coli* concentrations) (3).

The probability that the predicted value is greater than 235 CFU/100 mL is computed as the probability of Student's *t* being greater than *x,* with the degrees of freedom equaling the number of observations used in the regression minus the number of regression coefficients in the regression equation.

$x = (\log(235) - \hat{y}) / sep$

where
$\hat{y}$ is the regression estimate of the $\log_{10}$ *E. coli*, and
sep is the standard error of prediction of *y*.

This approach was applied to models for inland lakes beaches in the current study. For each model, a probability associated with too great a risk to allow swimming is determined—this is called the threshold probability. Threshold probabilities are determined by taking the dataset used to develop the model (calibration dataset) and finding the probability that is a reasonable balance between achieving a high number of correct responses and a low number of false negative responses. Computed probabilities that are less than the threshold indicate that bacterial water quality is most likely acceptable for swimming. Computed probabilities equal to or greater than the threshold probability indicate that the water quality is most likely not acceptable and that a water-quality advisory may be needed.

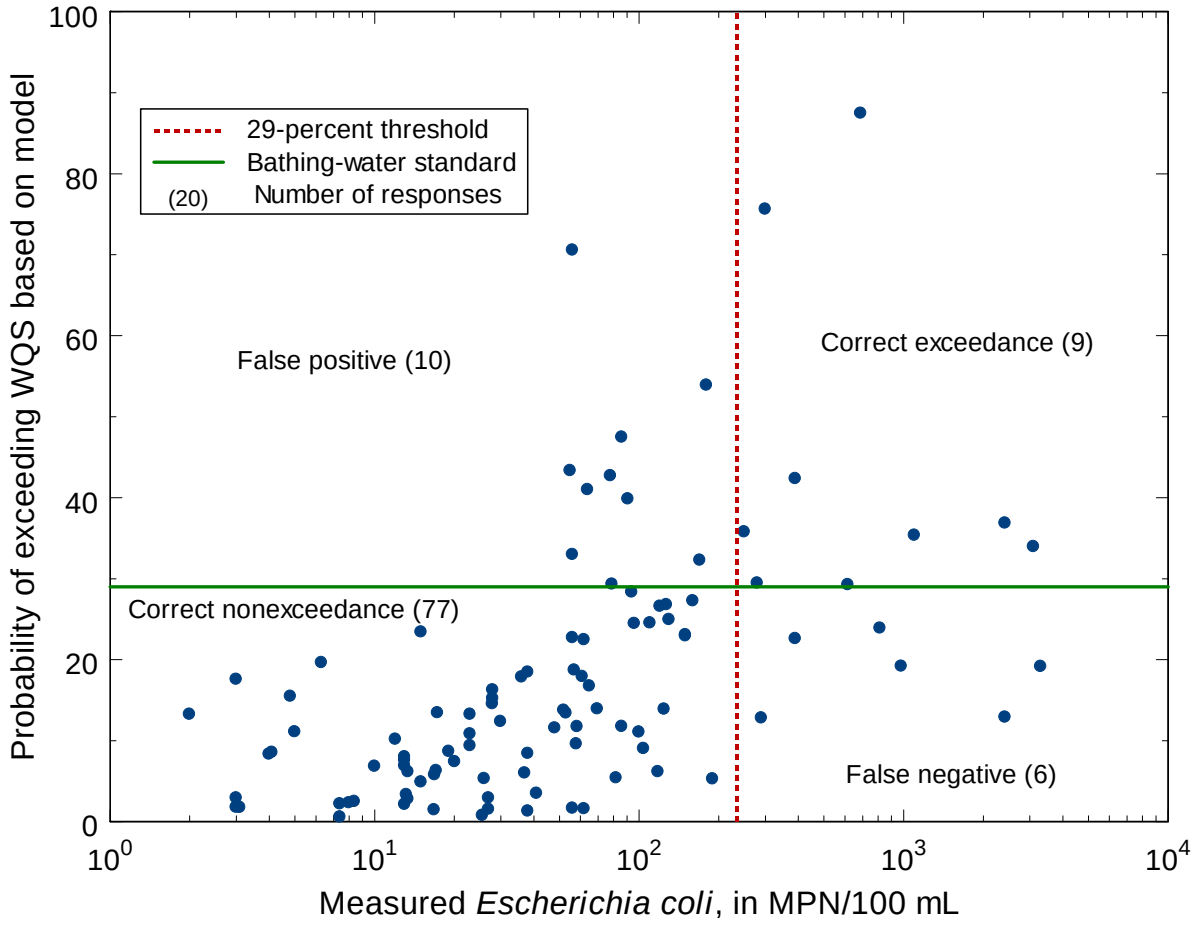
The following example was done using Virtual Beach (version 2.2) (4) and the calibration data set from Buck Creek South 2010–11. The probability of exceeding the standard for each data point was output and a threshold probability was set. This concept can be best explained by examining the plot for the Buck Creek 2010–11 model with a 29-% threshold (see the plot below) and then explaining the process used to determine the 29-% threshold. The plot is divided into four quadrants by a vertical line through 235 CFU/100 mL on the x-axis and a horizontal line through the threshold probability of 29. The four quadrants are

- Correct nonexceedance (specificity). *E. coli* concentration met the standard (was less than 235 CFU/100 mL), and the predicted probability of exceedance was below the

threshold. The specificity is the proportion of nonexceedances that are correctly predicted as being below the standard.

- False positive. *E. coli* concentration met the standard, but the predicted probability of exceedance was above the threshold.
- Correct exceedance (sensitivity). *E. coli* concentration exceeded the standard, and the predicted probability of exceedance was above the threshold. The sensitivity is the proportion of actual exceedances that are predicted correctly.
- False negative. *E. coli* concentration exceeded the standard, but the predicted probability of exceedance was below the threshold.

Our goals for good model performance are overall correct responses $\geq$ 80%, sensitivities $\geq$50%, and specificities $\geq$ 85%. By raising or lowering the horizontal line, one can determine the best threshold probability. This determination is somewhat subjective. Responses for different thresholds are listed in the table below the plot. In the example below, a threshold of 55 would have produced the highest number of correct responses (88, or 86.3%) but would also have produced 13 false negatives. False negative responses are especially troubling because the recreational water quality is determined to be acceptable when in fact the standard was exceeded. Thresholds between 40 and 50 do little to reduce the number of false negatives. Thresholds of 30 and 35, reduce the number false negatives, but still maintain sensitivities under 50%. Selecting a threshold of 29, maintains a high number of correct responses (86, or 84.3%), increases the sensitivity to 60%, and represents a compromise between false negative and false positive responses. Setting the threshold to a lower value than 29 increases the number of false positives without any further reduction to the numbers of false negatives.



Probability (%)	Total correct	False -	False +	Sensitivity (%)	Specificity (%)
55	88	13	1	13	99
50	87	13	2	13	98
45	86	13	3	13	97
40	84	12	6	20	93
35	86	9	7	40	92
30	85	8	9	47	90
29	86	6	10	60	89
28	85	6	11	60	87
25	82	6	14	60	84

FIG S1. Establishment of the threshold probability for 102 samples collected at Buck Creek State Park South, 2010–11.

References

1. **Francy DS, Darner RA**. 2006. Procedures for developing models to predict exceedance of recreational water-quality standards at coastal beaches. U.S. Geological Survey Techniques and Methods 6-B5. U.S. Geological Survey, Reston, VA. http://pubs.usgs.gov/tm/2006/tm6b5/.
2. **Francy DS, Darner RA**. 1998, Factors affecting *Escherichia coli* concentrations at Lake Erie public bathing beaches. U.S. Geological Survey Water-Resources Investigations Report 98-4241. U.S. Geological Survey, Reston, VA.
3. **Francy DS, Bertke EE, Darner RA**. 2009. Testing and refining the Ohio Nowcast at two Lake Erie beaches—2008. U.S. Geological Survey Open-File Report 2009-1066. U.S. Geological Survey, Reston, VA. http://pubs.usgs.gov/of/2009/1066/
4. **U.S. Environmental Protection Agency**. 2012. Exposure Assessment Models—Virtual Beach. Center for Exposure Assessment Modeling, U.S. Environmental Protection Agency, Athens, GA. http://www.epa.gov/ceampubl/swater/vb2/index.html.