Spring 2016

Relationships Between Fecal Indicator Bacteria and Environmental Factors at Edgewater Beach

Anna M. King
The University of Akron, annamking.ak@gmail.com

Please take a moment to share how this work helps you through this survey. Your feedback will be important as we plan further development of our repository.

Follow this and additional works at: http://ideaexchange.uakron.edu/honors_research_projects

Part of the Environmental Microbiology and Microbial Ecology Commons

Recommended Citation
King, Anna M., "Relationships Between Fecal Indicator Bacteria and Environmental Factors at Edgewater Beach" (2016). Honors Research Projects. 331.
http://ideaexchange.uakron.edu/honors_research_projects/331

This Honors Research Project is brought to you for free and open access by The Dr. Gary B. and Pamela S. Williams Honors College at IdeaExchange@UAkron, the institutional repository of The University of Akron in Akron, Ohio, USA. It has been accepted for inclusion in Honors Research Projects by an authorized administrator of IdeaExchange@UAkron. For more information, please contact mjon@uakron.edu, uapress@uakron.edu.
Relationships Between Environmental Factors and Fecal Indicator Bacteria at Edgewater Beach

Anna King

Spring Semester, 2016
The University of Akron

Honors Research Project: Biology
Abstract

To provide more timely estimates of indicator bacteria concentrations in recreational waters, environmental agencies are using predictive models to supplement conventional bacteria enumeration methods. As a tool to develop predictive models, correlational relationships between variables can be examined, to determine the statistical significance of explanatory variables. Prior research at Lake Erie beaches suggests that environmental conditions such as average wave height, rainfall, turbidity, and water temperature may demonstrate a strong relationship to concentrations of fecal indicator bacteria *Escherichia coli* and *Enterococcus faecalis* (enterococci). These relationships were examined at Edgewater Beach in Cleveland, Ohio, using statistical correlation tests on data collected from field measurements and analyses of lake water samples during the 2015 recreational season. *E. coli* concentrations were determined by analyzing samples prepared using EPA Method 9223B for Colilert-24®/Quanti-tray®. Enterococci concentrations were quantified by analyzing cultures prepared by EPA Method 1600 for membrane filtration of lake water samples. Additionally, ANOVA tests were used to examine the relationship between concentrations of fecal indicator bacteria and wave height. Results of the correlation tests demonstrated that water turbidity had the most significant correlation to *E. coli* and enterococci concentrations. Results of the ANOVA tests indicated that mean concentrations of bacteria associated with the highest wave heights was significantly greater than mean concentrations at lower wave heights. A weaker correlation was demonstrated in the relationships of bacteria concentrations to water temperature and antecedent rainfall. These results can provide insight on exploring potential variables to use in future regression models for examining indicator bacteria concentrations at Edgewater Beach.
Background

Spanning much of Ohio’s northern border, the Lake Erie coast has emerged as an area of great economic potential. As of 2014, annual visitor spending at Ohio’s Lake Erie region was estimated to surpass $10.7 million, comprising nearly a third of the state’s tourism revenue (1). In light of this, recreational activities such as boating, hunting, and swimming offer a promising opportunity for beach tourism in Northeast Ohio. However, a key challenge faced by this region echoes that of other beaches located along the Great Lakes: the risk of patrons’ exposure to waterborne pathogens from fecal contamination. This issue is especially important in Ohio, as water quality trends from 2009 to 2013 indicated that the state’s Great Lakes beaches consistently reported some of the highest percentages of samples with bacteria levels in exceedance of safety thresholds, compared to other states’ beaches bordering the Great Lakes (2).

In recreational waters, fecal contamination originates from a variety of human and non-human sources. Most human-sourced fecal pathogens enter the environment via raw sewage, stormwater, and groundwater discharges. These discharges drain into the surface waters near beach sites by way of outlet pipes and nearby tributaries (3), (4). Non-human sources of fecal pollution include waste from mammals and birds congregating along the shoreline. These sources of pathogens can vary over time from changes in land usage, effluent fluctuations, and severe weather events (3), (4). To minimize the public’s risk of illness, near-shore waters at recreational beaches are monitored by environmental agencies during annual swimming seasons. These agencies use standard methods established by the United States Environmental Protection Agency (USEPA) to measure concentrations of *Escherichia coli* and *Enterococcus faecalis* (enterococci) in samples collected from beach sites of interest (5). The use of these bacteria as indicator organisms is supported for several reasons. A key advantage of testing specifically for these bacteria is that a more precise indication of pathogenic fecal bacteria can be attained, as opposed to more general tests for total coliforms, which could potentially detect bacteria that are not pathogenic or fecal in origin (4). Additionally, multiple epidemiological studies suggested that concentrations of *E. coli* in waters show a strong positive correlation to gastrointestinal illnesses in humans (6). Finally, culturing and enumeration methods for *E. coli* and enterococci can be more efficient and cost-effective than other methods (7). Although enumeration of indicator bacteria provides a baseline estimate of waterborne pathogens, it should be recognized that their presence does not absolutely confirm the presence of pathogens, and their presence is an estimation of additional fecal pathogens that are likely to reside in the sample (4). In light of this, it is important to be aware that it is possible for indicator organisms to have a survival profile that differs from additional pathogens existing in the sample (8).

Standards defining safe levels of fecal indicator bacteria in recreational waters arose from the Clean Water Act, which was established in 1972. This enactment appointed the USEPA as the regulatory authority on water quality monitoring practices used by environmental agencies throughout the United States. In 1986, the use of *E. coli* and enterococci as indicator organisms was introduced, and safety thresholds based on their concentrations were established. The safety threshold for *E. coli* concentrations in recreational freshwater was established at a geometric mean of 126 CFUs per 100 mL in multiple samples in a 30-day interval, and 235 CFUs per 100 mL in single samples. For enterococci concentrations in recreational freshwater, the safety threshold was established at a geometric mean of 33 CFUs per 100 mL in multiple samples in a 30-day interval, and 104 CFUs per 100 mL in single samples (5). These recommendations were based on the results of epidemiological studies conducted by the USEPA, which found the lowest incidence of gastrointestinal illness was reported by swimmers exposed to beach water with bacteria concentrations at or below those used to define safety thresholds (5). If concentrations of *Escherichia coli* and enterococci are detected to be above safe levels, water quality advisories and beach closures are issued, which notify the public to avoid contact with the water. Advisories remain in effect until the bacteria concentrations are determined to be within a safe range (9).
For beach monitoring purposes, the most conventional methods used to quantify concentrations of viable fecal indicator bacteria are membrane filtration and plate counts cultured on m-TEC (membrane-thermotolerant *E. coli*) and mE1 (*Enterococcus* Indoxyl-β-D-Glucoside) agar. In general, these methods are favored because they require relatively less equipment, funding, and training to perform than more advanced methods (10). Colilert® tests, which identify *E. coli* and total coliforms based on their expression of β-D-galactosidase, are favored for similar reasons. In beach monitoring, an ongoing issue encountered with traditional culturing methods is the time it takes to get results (10). Using standard methods, it takes between 18 and 24 hours to culture colonies of *E. coli* and enterococci for enumeration, which means results are generally not available until one day after sample collection. Because water quality can fluctuate significantly over the course of a day (e.g. from sunlight exposure or severe weather events), the validity of the water quality advisories founded on enumeration results can be compromised.

To supplement the limitations of culture-based methods, environmental agencies have developed real-time predictive models to forecast bacteria concentrations at specific beaches (11). These models use multiple linear regression and other statistical methods on existing datasets to examine relationships between various explanatory variables and bacteria concentrations. Models are adapted to individual beaches by combining the most statistically significant predictors into multiple linear regression equations, which can be used to forecast daily bacteria concentrations or estimate the probability that bacteria concentrations will exceed the safety threshold (12). In Ohio, the model output is then posted to the Ohio Nowcast website early in the day, before culture-based enumeration results become available. As a preliminary step, correlational relationships between variables can be examined prior to developing predictive regression models, to determine the statistical significance between variables.

Beach monitoring literature emphasizes that care must be taken when choosing explanatory variables for analysis, because environmental factors can influence bacteria concentrations differently over time. For example, sources of indicator bacteria can vary from changes in land usage, treatment plant effluent fluctuations, sewage overflows, and severe weather events. Additionally, some environmental factors can be more or less significant to bacteria concentrations at certain beaches over others. This was demonstrated in a 2006 study on Lake Erie beaches by the U.S. Geological Survey, which found a stronger correlation between wind direction and *E. coli* concentrations at Lakeview and Villa Angela beaches than at Huntington, Lakeshore, or Edgewater beaches (13). Caution must also be used even when incorporating significant explanatory variables into predictive models, as the interaction of variables can introduce multicollinearity. For example, turbidity and wave height have shown strong associations to bacteria concentrations (14), but the potential for collinearity exists because turbidity is generally associated with wave height, due to sediment disruption from wave activity.

**Purpose**

The objective of this project was to examine the correlational relationships between fecal indicator bacteria and various environmental conditions at Edgewater Beach in Cleveland, Ohio. The location, methods, and data used in this project were acquired as part of an internship completed during the summer of 2015, with the microbiology laboratory at the Northeast Ohio Regional Sewer District. This laboratory was contracted to perform water sample collection, data collection, sample culturing, and enumeration for use in ongoing research by the USEPA (15). The variables examined in this project were concentrations of fecal indicator bacteria (*E. coli* and enterococci), wave height, rainfall, turbidity, and water temperature. Turbidity was chosen as a variable because research suggests a positive association with higher concentrations of bacteria such as *E. coli*, which live on the surface of sand and soil particles suspended in the lake water (16). Rainfall, wave height, and water temperature were selected as variables based on feasibility of access and results from water quality studies at other Lake Erie beaches, which suggested that these factors demonstrated were relevant to indicator bacteria concentrations (13), (17).
Additionally, it is important to note that the statistical tests used in this project differ from predictive regression models (such as the models used for Nowcast reporting) in that they aim to investigate the strength of associations between variables, rather than forecasting probable bacteria concentrations. By generating output such as correlation coefficients and F-statistic values, the statistical tests performed in this project can help illustrate which environmental conditions show the strongest relationship to bacteria concentrations at Edgewater Beach, and identify relevant variables to include in regression models.

Sample Collection

Water samples for bacteria culturing and turbidity analysis were collected during the 2015 recreational season (May 31st through September 4th) from six separate locations at Edgewater Beach, at coordinates of approximately 41° 29’ 21” N, 81° 44’ 24” W (Figure 1). Samples were collected once daily, Sunday through Thursday, between the hours of 7:00 a.m. and 9:00 a.m. Six 1-liter grab samples were collected according to USEPA guidelines at nearshore and offshore sites, and were composited for analysis. An additional 8-liter grab sample was collected offshore at a site on the east side of the beach, at a water depth of 1 meter. The nearshore sites were located at water depths of 0.3 meters, and the offshore sites were located at water depths of 1 meter. In general, sampling locations and times remained consistent throughout the sampling period. A few exceptions arose during high-wave conditions at the beach, which resulted in either no sample collection for that day, or sample collection at shallower depths. All samples were collected in sterile plastic bottles, transported to the lab on wet ice, and were processed within 6 hours of collection.

![Figure 1. The white circles denote the locations of sampling sites at Edgewater Beach in Cleveland, Ohio (22). Coordinates for the general location are approximately 41° 29’ 21” N, 81° 44’ 24” W.](image)

Data Collection

At each sampling site, measurements of environmental parameters were collected daily. Water temperature values were measured using a YSI 556 Multiparameter, and were recorded in degrees Celsius. Wave heights were measured in feet using a wave stick, averaging measurements taken from the east and west side of the beach. Turbidity was measured Nephelometric Turbidity Units (NTUs), using an HF Scientific Micro 100 Turbidimeter in the laboratory. Average rainfall was measured as the average amount of rain, in inches, that fell during a 24-hour period 1, 2, and 3 days before sampling, as recorded from data compiled from the National Oceanic and Atmospheric Administration (NOAA) weather station at the
Cleveland Hopkins Airport. *E. coli* concentrations were quantified as the most probable number (MPN) of cells per 100 mL of sample, and enterococci concentrations were quantified as the number of colony-forming-units (CFUs) per 100 mL of sample.

**Bacteria Culturing and Enumeration Methods**

After transport to the lab, the six 1-liter samples were composited and processed for analysis. The single 8-liter grab sample was processed and analyzed separately as an additional source of data. *E. coli* concentrations were determined using EPA Method 9223B for Colilert-24®/Quanti-tray® enumeration (18). In this procedure, the Colilert® reagent was added to 1x, 10x, and 100x dilutions of each sample, all of which were prepared in duplicate. The samples were then incubated at 35 ± 0.5 °C for approximately 24 - 28 hours. After incubation, the wells of the Quanti-tray® that yielded positive results were counted, and used to calculate the most probable number (MPN) of *E. coli* cells per 100 mL of sample (18). Enterococci concentrations were quantified using counts of colony-forming-units (CFUs) from cultures prepared using EPA Method 1600 for membrane filtration. In this procedure, 1x, 10x, and 100x dilutions of each sample were prepared in duplicate, and were passed through membrane filters with a pore size of 0.45 ± 0.02 µm (19). After filtration, the membranes were placed on petri dishes containing mEI agar, and incubated at 41 ± 0.5 °C for approximately 22 - 24 hours. After incubation, the membrane filters were visually inspected for CFUs typical of enterococci. The number of CFUs counted from each dilution that fell within an acceptable range (20 – 60 CFUs) were then used in the calculation for CFUs of enterococci per 100 mL of sample (19).

**Statistical Methods**

Correlations between bacteria concentrations, turbidity, and water temperature were analyzed using the Pearson’s Correlation test on data collected from the 6 1-liter composited samples. Relationships of bacteria concentrations to rainfall were analyzed using the Spearman’s Correlation test on data collected from the 8-liter sample. Wave height was analyzed using a one-way ANOVA test and Tukey’s Pairwise Comparison test on data collected from the 8-liter sample. The data collected from Edgewater Beach was summarized and analyzed using the “ANOVA”, “Tukey’s Test”, and “Correlation” applications in Excel® and Minitab®.

**Correlation Tests.** The Pearson’s correlation test was used to examine the relationship between *E. coli* and enterococci concentrations, turbidity, and water temperature. The variables satisfied the following conditions for Pearson’s correlation: they were quantitative measurements, displayed a linear relationship, and were normally distributed. The linear relationship between the variables is demonstrated in the scatterplots and trend lines shown in Figure 1A-D. To assure a normal distribution, the variables were transformed to a base-10 logarithmic scale. A normal distribution was validated with the Ryan-Joiner test for normality, and by using more than 30 randomly sampled observations (n = 66). A significance level (probability of observing a strong correlation by chance) of 0.05 was established.

The relationship between *E. coli* and enterococci concentrations and rainfall measurements were examined using the Spearman’s correlation test. This test was used because the rainfall data was not normally distributed, consisted of quantitative measurements, and displayed a monotonic relationship to bacteria levels, in which explanatory and response variables increase or decrease at different rates (20). The assumption of non-normality was validated by its failure of the Ryan-Joiner test for normality, and the heavily right-skewed distribution. This skewness was due to the prevalence of zero and near-zero values in the rainfall measurements (e.g. days with low or no rainfall). Transformations did not normalize the rainfall data, so the untransformed values were used, as they represented field measurements and not the absence of data. Outliers were removed, and a significance level of 0.05 was established.
ANOVA Tests. To examine the significance of differences between mean concentrations of *E. coli* and enterococci at different wave height conditions, a one-way analysis of variance (ANOVA) was used. Base-10 log-transformed *E. coli* and enterococci concentrations were grouped into three categories based on ranges of wave heights measured at the time of sampling (0.0 – 0.13 ft., 0.14 – 0.65 ft., and 0.66 ft. – 3.0+ ft., respectively for *E. coli*; and 0.0 ft. – 0.20 ft., 0.21 ft. – 0.60 ft., and 0.61 ft. – 3.0+ ft., respectively for enterococci). The mean log bacteria concentrations (assumed to represent geometric means) in each category were examined using the ANOVA test. To specify which means were significantly different, a Tukey Pairwise Comparison was used. To validate the ANOVA test, the assumptions of constant variance and normality for the response variables (log *E. coli* and enterococci concentrations) were considered. Constant variance for the different response categories was confirmed by calculating that the largest standard deviation of the responses in the three categories was not greater than twice the smallest standard deviation of the responses in the three categories. Although each response was measured from an independent random sample, it was revealed that the distribution of the enterococci response variables was skewed toward the lower end of the distribution range. As such, the dataset was not normally distributed, which reduces the validity of the ANOVA test. This violation originates from the number of replicate values present, and the small amount of observations in each group (ranging from n = 18 to n = 21).

Results

Correlation Tests. Based on the output of the correlation tests, it was revealed that water turbidity had the most significant correlation to *E. coli* concentrations. The resulting Pearson’s correlation coefficient (r) of 0.52 indicates a moderately positive correlation between turbidity and *E. coli* concentrations. Further, a low p-value 6.59e-6 suggests that this relationship is statistically significant (Figure 2A; Table 1). Similar results were shown in the correlation output for turbidity and enterococci concentrations (Figure 2B; Table 2), with a Pearson’s r of 0.57 and p-value of 3.16e-7. Taken together, these results indicate a significant relationship between turbidity and bacteria concentrations, supporting the use of turbidity as a variable in regression models for *E. coli* and enterococci concentrations at Edgewater Beach. Output from Pearson’s correlation tests relating bacteria concentrations to water temperature revealed a weak, negative correlation, with Pearson’s r values of -0.19 and -0.31, and p-values of 0.04 and 0.01 for *E. coli* and enterococci, respectively (Figure 2C; Table 1), (Figure 2D; Table 2).

Figure 2. Scatterplot matrix and Pearson correlation coefficients (r) for base-10 log transformations of *E. coli* concentrations, enterococci concentrations, turbidity, and water temperature in samples collected from Edgewater Beach.
Table 1. Summary of correlation coefficients and p-values examining the relationship between log<sub>10</sub> E. coli concentrations and explanatory variables.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Correlation Coefficients</th>
<th>p-values</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turbidity</td>
<td>0.52</td>
<td>6.59 e&lt;sup&gt;-6&lt;/sup&gt;</td>
<td>66</td>
</tr>
<tr>
<td>Temperature</td>
<td>-0.19</td>
<td>0.04</td>
<td>66</td>
</tr>
<tr>
<td>Rainfall in 24-hr pd. 1 day before sampling</td>
<td>0.48</td>
<td>1.00 e&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>66</td>
</tr>
<tr>
<td>Rainfall in 24-hr pd. 2 days before sampling</td>
<td>0.24</td>
<td>0.05</td>
<td>66</td>
</tr>
<tr>
<td>Rainfall in 24-hr pd. 3 days before sampling</td>
<td>0.09</td>
<td>0.47</td>
<td>66</td>
</tr>
</tbody>
</table>

Table 2. Summary of correlation coefficients and p-values examining the relationship between log<sub>10</sub> enterococci concentrations and explanatory variables.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Correlation Coefficients</th>
<th>p-values</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turbidity</td>
<td>0.57</td>
<td>3.16 e&lt;sup&gt;-6&lt;/sup&gt;</td>
<td>66</td>
</tr>
<tr>
<td>Temperature</td>
<td>-0.31</td>
<td>0.01</td>
<td>66</td>
</tr>
<tr>
<td>Rainfall in 24-hr pd. 1 day before sampling</td>
<td>0.59</td>
<td>1.00 e&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>66</td>
</tr>
<tr>
<td>Rainfall in 24-hr pd. 2 days before sampling</td>
<td>0.08</td>
<td>0.50</td>
<td>66</td>
</tr>
<tr>
<td>Rainfall in 24-hr pd. 3 days before sampling</td>
<td>0.25</td>
<td>0.04</td>
<td>66</td>
</tr>
</tbody>
</table>
ANOVA Tests. The single-factor ANOVA test for log E. coli concentrations at different wave heights generated a p-value of 0.195e-5, indicating a statistically significant relationship between the variables. Additionally, the critical f-value of 3.074 is less than the f-test statistic of 11.91, which demonstrates that the variation of the responses between the categories is greater than the variation of the responses within the categories. As such, the null hypothesis can be rejected, as the mean log E. coli concentration in the (0.66 ft. – 3.0 ft.) category is significantly higher than the others (Figure 3; Tables 3 – 4). This is further demonstrated in a Tukey pairwise comparison of means for log E. coli concentrations at different wave heights. Grouping information uses the Tukey method and 95% confidence intervals of log E. coli concentrations in each wave height category. Shown in Table 4, the means of the categories that do not share a letter are significantly different.

Table 3. Single-factor ANOVA output for mean log_{10} E. coli concentrations (MPN/100 mL) at different wave heights.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Squares</th>
<th>F-test statistic</th>
<th>P-value</th>
<th>F-critical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>4.479</td>
<td>2</td>
<td>2.239</td>
<td>11.91</td>
<td>1.95e-5</td>
<td>3.074</td>
</tr>
<tr>
<td>Within Groups</td>
<td>26.904</td>
<td>97</td>
<td>0.277</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>31.382</td>
<td>99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Tukey pairwise comparison of means, for mean log_{10} E. coli concentrations (MPN/100 mL) at different wave heights.

<table>
<thead>
<tr>
<th>Factor</th>
<th>n</th>
<th>Mean</th>
<th>Grouping</th>
<th>Std. Dev</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.66 – 3.0+ ft.)</td>
<td>43</td>
<td>2.18</td>
<td>A</td>
<td>0.657</td>
<td>(1.999, 2.532)</td>
</tr>
<tr>
<td>(0.14 – 0.65 ft.)</td>
<td>41</td>
<td>1.93</td>
<td>A</td>
<td>0.523</td>
<td>(1.751, 2.112)</td>
</tr>
<tr>
<td>(0.0 – 0.13 ft.)</td>
<td>36</td>
<td>1.53</td>
<td>B</td>
<td>0.362</td>
<td>(1.3415, 1.727)</td>
</tr>
</tbody>
</table>

Figure 3. Tukey pairwise comparison of means, for log_{10} E. coli (MPN/100 mL) at different wave heights, showing means and 95% confidence intervals of log E. coli concentrations in each wave height category.
The single-factor ANOVA test for log enterococci concentrations at different wave heights displayed a p-value of 0.046, which indicates a relationship of borderline significance. Additionally, the critical f-value of 3.172 is less than the f-test statistic of 3.254, which demonstrates that the variation of the responses between the categories is greater than the variation of the responses within the categories. As such, the null hypothesis can be rejected, as the mean concentration of enterococci in the (0.61 ft. – 0.30+ ft.) category is significantly higher than the mean concentration of enterococci (0.0 ft. – 0.20 ft.) category (Figure 4; Tables 5 – 6). Grouping information uses the Tukey method and 95% confidence intervals of log enterococci concentrations in each wave height category. Shown in Table 6, the means of the categories that do not share a letter are significantly different.

**Table 5. The single-factor ANOVA output for mean log<sub>10</sub> enterococci concentrations (CFU/100 mL) at different wave heights.**

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Squares</th>
<th>F-test statistic</th>
<th>P-value</th>
<th>F-critical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>3.633</td>
<td>2</td>
<td>1.817</td>
<td>3.254</td>
<td>0.046</td>
<td>3.172</td>
</tr>
<tr>
<td>Within Groups</td>
<td>29.585</td>
<td>53</td>
<td>0.558</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>33.218</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 6. Tukey pairwise comparison of means, for log<sub>10</sub> enterococci concentrations (CFU/100 mL) at different wave heights.**

<table>
<thead>
<tr>
<th>Factor</th>
<th>n</th>
<th>Mean</th>
<th>Grouping</th>
<th>Std. Dev</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.61 - 3.0+ ft.)</td>
<td>18</td>
<td>1.9</td>
<td>A</td>
<td>0.855</td>
<td>(1.514, 2.233)</td>
</tr>
<tr>
<td>(0.21 - 0.60 ft.)</td>
<td>18</td>
<td>1.3</td>
<td>A B</td>
<td>0.746</td>
<td>(0.951, 1.670)</td>
</tr>
<tr>
<td>(0.0 - 0.20 ft.)</td>
<td>21</td>
<td>1.3</td>
<td>B</td>
<td>0.686</td>
<td>(0.948, 1.614)</td>
</tr>
</tbody>
</table>

**Figure 4. Tukey pairwise comparison of means, for log<sub>10</sub> enterococci (CFU/100 mL) at different wave heights, showing means and 95% confidence intervals of log enterococci concentrations in each wave height category.**
Discussion

The results of this project showed that the most prominent relationship existed between water turbidity and fecal indicator bacteria, with p-values of $6.59 \times 10^{-6}$ and $3.16 \times 10^{-6}$, and correlation coefficients of 0.52 and 0.57 for *E. coli* and enterococci, respectively. Additionally, ANOVA test results indicated that mean log concentrations of *E. coli* and enterococci in the highest wave-height categories were significantly higher than the others, indicating relationships of significance. This may be due to the agitation of bacteria and sediments that occurs during high-wave conditions, which would have remained undisturbed in calmer conditions. Water temperature and rainfall did not show a substantial relationship to bacteria concentrations, illustrated by weakly significant p-values of 0.04 and 0.01, and low correlation coefficients of -0.31 and -0.19 for *E. coli* and enterococci, respectively. Collectively, these results suggest that turbidity and wave height would be stronger candidates to include as predictor variables in a regression model for Edgewater Beach, whereas water temperature and rainfall may be less relevant.

It is important to note that a key limitation of this study was the size of the datasets used. Because one season of data was used in the statistical tests for this project, the total number of observations used in the tests was approximately 66. Although this number of observations did not violate the assumptions of the correlation tests, it did decrease the validity of the ANOVA test examining wave height and enterococci, as the number of observations in each category ranged from 18 to 43. Literature on statistical analysis of similar data suggests increasing the size of the dataset to include at least two seasons of data may reduce the potential for error in the conclusions drawn from statistical tests (11). Another challenge encountered in this study was finding the most useful statistical test to analyze non-normal distributions of predictor variables (specifically rainfall and wave-height). In the future, it may be advantageous to consider the use of non-parametric statistical tests for non-normally distributed environmental data. For example, the Kruskal-Wallis test could be used as an alternative to the ANOVA test if transformed wave-height data continued to appear skewed and observations were less than 100. An additional improvement would be to find a source of rainfall data from a site nearer to the sampling location, as the data used in this project was taken from the NOAA weather station at the Cleveland Hopkins Airport, which was located approximately 12 miles southwest of the sampling location. These improvements can be incorporated into future correlation studies, to improve their utility in exploring potential variables to use in predictive models.
Acknowledgements

Much appreciation to Dr. Angela Hartsock, Dr. Jim Holda, Nichole Schafer, and Mark Citriglia for their time as readers of this paper. Additionally, many thanks to the helpful supervisors and colleagues in the Analytical Services Department at the Northeast Ohio Regional Sewer District, for the opportunity to work with them and learn about the methods involved in beach monitoring. Special thanks to Lindsey Venesky for her guidance in data organization and insights on statistical methods.


References


10. **Mesquita, Sandra and Noble, Rachel T.** *Recent Developments in Monitoring of Microbiological Indicators of Water Quality Across a Range of Water Types.* 2013. p. 33.


15. **The Northeast Ohio Regional Sewer District.** Enumeration and Modeling of Freshwater Pathogens - Edgewater Beach. 2015.


20. **Research Skills One, Correlation interpretation, Graham Hole v.1.0.** *Eight things you need to know about interpreting correlations.* n.d. p. 3.
