Receiver-Operating Characteristics Analysis of Fecal Indicator Bacteria and Pathogens

BY

BURCU YAVUZ
B.S., University of Chicago, 2005

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Master of Science in Public Health
in the Graduate College of the
University of Illinois at Chicago, 2012

Chicago, Illinois

Committee:
Samuel Dorevitch, Chair
Rachael Jones
Peter Scheff

DEDICATION

I would like to dedicate this thesis to managers of recreational water across the United States and abroad who are forced to make sense of out of information that very often does not. I hope what follows makes those decisions easier.

ACKNOWLEDGMENTS

This thesis could not have been accomplished without the continued support of Sam Dorevitch, Rachael Jones, and Peter Scheff. I am tremendously grateful for their indefatigable patience and guidance throughout the process. I have learned so much from the insights into data analysis, writing, and the art of scientific inquiry.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF ABBREVIATIONS

AF	Amplification Factor
ATCC	American Type Culture Collection
AUC	Area Under the Curve
APHA	American Public Health Association
BA	Barely Acceptable
CAWS	Chicago-Area Waterway System
CCE	Calibrator Cell Equivalents
CHEERS	Chicago Health Environmental Exposure and Recreation Study
CFU	Colony Forming Units
CSO	Combined-sewer Overflow
CV	Coefficient of Variation
$\Delta\Delta$CT	Delta Delta Cycle Threshold
ECM	Estimated Correlation Matrix
EPA	Environmental Protection Agency
FIB	Fecal Indicator Bacteria
FN	False Negative
FP	False Positive
GI	Gastrointestinal
GUW	General Use Waters
HCGI	Highly-Credible Gastrointestinal Illness
mL	Milliliter

LIST OF ABBREVIATIONS (continued)

MPN	Most Probably Number
MWRDGC	Metropolitan Water Reclamation District of Greater Chicago
NTAC	National Technical Advisory Committee
PFU	Plaque-Forming units
qPCR	Quantitative Polymerase Chain Reaction
RU	Relatively Unpolluted
ROC	Receiver-Operating Characteristics
SPC	Sample-Processing Control
SSM	Single Sample Maximum
STV	Statistical Threshold Value
TN	True Negative
TP	True Positive
UIC	University of Illinois—Chicago
WTP	Wastewater Treatment Plant

SUMMARY

A receiver-operating characteristics (ROC) curve is a graphical plot of the "sensitivity" (or true positive, TP, rate) against "1-specificity" (or false positive, FP, rate) of test with regard to some true condition of interest. The area under a ROC curve can be interpreted as the probability that a given test correctly ranks a TP diagnostic condition higher than a true negative (TN) diagnostic condition, providing a metric for evaluating the diagnostic quality of a given test. Using ROC analysis, this paper evaluates the diagnostic value of six indicator tests with regard to two waterborne pathogens. Specifically, we compare the diagnostic value of culture-based assays for enterococci and *E. coli*, somatic coliphage, F+ coliphage, and quantitative polymerase chain reaction (qPCR)-based assays for enterococci and *E. coli* with regard to the presence or absence of *Giardia* and *Cryptosporidium*. The results of this study indicate that qPCR based assays of enterococci and *E. coli* are superior to culture-based assays of the same organisms with regard to diagnosing the presence of *Giardia*.

1. BACKGROUND SECTION

1.1 Early Observations of Swimming Related Illness

In 1892, a group of soldiers bathed in the Danube River; ten were stricken with typhoid fever (Jaeger, 1892). This event, as reported of Jaegar, was one of the earliest documented cases in which recreational surface water activities led to the spread of disease. It is a public health concern that has persisted through subsequent decades.

Scientific investigations about recreational water quality began in the United States as early as the 1920s when scientists linked the heavily polluted waters of New Haven Harbor to a series of typhoid fever outbreaks. In 1923, the New Haven Health Demonstration published a study showing 61 cases of typhoid were undoubtedly the result of swimming in harbor waters that, at the time, received discharges of crude sewage approaching nearly 20 million gallons per day (Platt, 1923). Winslow and Moxon corroborated their work by actually measuring the levels of fecal bacteria in those waters (Winslow and Moxon, 1928). At around the same time, the American Public Health Association's (APHA) Committee on Bathing Beaches surveyed more than 2000 physicians in order to compile data about the prevalence of disease at bathing beaches (APHA, 1922). A dearth of epidemiological evidence made it impossible for the committee to form a consensus about a possible association between water quality and adverse health outcomes (APHA,1922). This, combined with a reluctance to spread panic among the public, meant that the medical committee declined proposing any standards for water quality then and later in 1936, 1940, and 1955 (Dufour, 2007). Although the APHA could not recommend microbiological standards, uncertainty about the safety of public bathing sites continued in the United States and abroad. Health officials and physicians became concerned

that sewage polluted water could transmit the most devastating diseases of that period: poliomyelitis, typhoid fever, and paratyphoid fever.

1.1.1 Gross Contamination of Recreational Water.

Resort towns became epicenters of concern about water quality. For municipalities near large bodies of water, the release of untreated waste into local beaches had at first been a practical alternative to the expensive and infeasible treatment processes (Robson, 1956). However, as water quality conditions deteriorated the impact of such outfalls became an important public health concern. In 1954, Moore used bacteriological surveys to study the effects of relocating of a sewage outfall that discharged pollution into an adjacent swimming beach (Moore, 1954). This UK study at a seaside resort in North Devon was conducted due to fears that the fouling of local beaches would severely impact the tourist economy at the site (Moore, 1954). At such venues, questions about water quality were not only important in the context of general public safety but were necessarily linked to economic survival. Moore pointed out that the financial success of a beach town depended on the appeal of local recreational water activities and that local authorities had economic justification for minimizing pollution at swimming beaches (Moore, 1954). Moore traced the path of fecal pollution from an outfall pipe and mapped the dissipation of sewage contamination against factors such as tide and beach topography. This study illustrates an important turning point in water quality management: the use of fecal bacterial indicators (FIB) by scientists, engineers, and local legislators began because those tests were a convenient way to trace the flow of raw sewage releases into recreational waterways.

Tracking fecal pollution, rather than waterborne pathogens, became the paradigm for water quality assessment and persisted even as the practice of directly discharging raw sewage into recreational waters diminished. While information about fecal pollution in public waters was an important step in water quality management, it was not enough to inform public health policy. Without epidemiological studies, it was not known if, and in what ways, measures of FIB related to the risk of swimming-associated illness. Therefore, public policy decisions about water quality were either established at arbitrary levels or not established at all.

1.1.2 The First Major Epidemiological Studies

Eventually, as gross contamination of beaches with fecal waste lessened and aesthetic standards of beaches improved, state authorities required a more salient metric of water quality: health outcomes. One of the earliest prospective cohort studies was conducted by Albert Stevenson, from the United States Public Health Service, concerning the health impacts of swimming in Chicago's Lake Michigan beaches, the Ohio River at Dayton, Kentucky, and finally sections of the Long Island Sound at New Rochelle and Mamaroneck, New York (Stevenson, 1953). Incidences of illness among swimmers and non-swimmers were calculated from self-reported health surveys collected from participants after a two month study period (Stevenson, 1953); water quality was assessed based on historical records and laboratory examination using the most probable number (MPN) method[1] to determine the number of

[1]The MPN method for determining coliform count depended on growing bacteria in a liquid culture medium containing lactose—the growth of microbes created a significant amount of gas that, when captured in an inverted tube, was used to estimate the bacterial count. It is worth noting that this method had high variability and tended to overestimate coliform counts when compared to standards.

"coliforms"[2] in 100 mL of water sample. At each region, two bathing sites were chosen for study—one with relatively good water quality and one with poor water quality.

While methodology problems precluded the possibility of establishing a dose-response relationship between water quality levels and adverse health outcomes, Stevenson found that illness rates among swimmers were significantly higher than among non-swimmers (Stevenson, 1953). Notwithstanding this result, only two circumstances showed that swimming in relatively polluted water posed a higher risk of disease than non-polluted water (Stevenson, 1953). When bacterial counts were higher than 2300 coliforms per 100 mL, swimming in Lake Michigan led to higher incidences of all illnesses; at that coliform level, swimming in the Ohio River (as compared to swimming in a nearby freshwater pool) lead to higher incidences of gastrointestinal (GI) illness only (Stevenson, 1953). Stevenson determined that half of reported illnesses were eye, ear, nose, and throat conditions, while GI illness made up one-fifth of reported illnesses, and skin irritations as well as other diseases the remainder (Stevenson, 1953). Although there was sufficient evidence to indicate some risk of adverse health outcomes, Stevenson could not offer a quantifiable description of that risk nor endorse strict bacterial water quality standards (Stevenson, 1953).

The UK Committee on Bathing Beach Contamination of the Public Health Laboratory Service conducted another early investigation about the health effects of swimming in water contaminated with fecal waste (Moore et al., 1959). For a period of five years, the committee reviewed bacteriological and sanitary surveys for forty popular swimming beaches in England

[2]Microbes from the "coliform" group were more specific for the fecal waste of warm-blooded animals. This subset became known as the "fecal coliform" group while the term "coliforms" was replaced with the phrase "total coliforms" to distinguish the two tests (NATC, 1968).

and Wales. Moore and colleagues also made an early attempt to isolate pathogenic organisms from sea water—*Salmonellae*, *Staphylococcus aureus*, *Shigellae*, *Mycobacterium tuberculosis*, and poliovirus. Their review of 150 cases of poliomyelitis did not point to any connection with swimming in sewage polluted water, though there was some evidence that contaminated water might contribute to an increase in cases of parathyroid fever (Moore et al., 1959). Water quality analysis by Moore and colleagues showed that densities of Salmonella (the causative agent of typhoid and paratyphoid fever) were not high enough to deliver an infectious dose through swimming (Moore et al., 1959). Low microbial concentrations and technical limitations prevented the isolation of other pathogens (Moore et al., 1959). Ultimately, it was the opinion of the committee that, aside from direct contact with pathogenically contaminated fecal matter, sewage discharges into public swimming waters posed no real public health risks and that water quality criteria could be evaluated on the basis of aesthetic measures (Moore et al., 1959).

Although neither group found epidemic level disease among swimmers, Stevenson (1953) was able to conclusively identify an excess of illness among swimmers as compared to non-swimmers. The differences between the data-gathering methods employed by Stevenson and those used by Moore et al. clarify the seeming contradiction. In the Stevenson study, families were instructed to record swimming activity on a calendar in tandem with any symptoms of eye, ear, nose, and throat ailments, GI illness, or skin irritations. In contrast, Moore et al. conducted a retrospective study. They arrived at epidemiological conclusions from medical records of acute cases of serious illnesses (typhoid fever, paratyphoid fever, and poliomyelitis) that had eventually reached the attention of public health officials well after the

condition had progressed. The authors based the lack of association between swimming and morbidity on the recollections of participants about how often and where they swam in the weeks prior to the onset of disease. Instances of disease were evaluated against the seasonal coliform means at the site where swimming occurred, effectively ignoring the importance of daily fluctuations in water quality. Surprisingly, Moore et al. stress that water quality strongly depends on day-to-day sewage releases, tidal patterns, beach topography, time of day, and season.

Strict microbiological standards of bathing beaches can be relaxed without serious detriment to public health—both the 1959 study by Moore et al. and 1953 study by Stevenson reached essentially this same conclusion from seemingly contradictory results. The Stevenson study showed only slight increases in mild illness from swimming activity while Moore et al. ruled out sea water as a vehicle for serious disease (Moore et al., 1959; Stevenson, 1953). Stevenson compared illness rates at North and South beaches in Chicago and found that 2,300 MPN/100mL as a predictor of high rates of GI Illness. On the other hand, Moore et al. (1959) determined water quality from mean seasonal coliform counts; that study could not link specific instances of illness with precise microbiological densities in swimming water or even specific instances of swimming (Moore et al., 1959; Stevenson, 1953). Neither study concluded a need for strict microbiological standards of water quality. Later epidemiological studies, using stricter definitions of exposure, rigorous documentation of morbidity, and specific information about water quality would not be so sanguine about the health effects of swimming in contaminated water. Nevertheless, these two studies raise important questions: Is there an endemic rate of sporadic disease for swimming in natural bodies of water (as implied by

Stevenson's findings) that is distinct in etiology and frequency from outbreaks of severe disease that Moore et al. looked for? Likewise, is the endemic rate, perhaps caused by different microbes than those causing outbreaks of severe disease, predicted by microbe density? If so, how can public health authorities predict the waxing and waning of such risks?

1.1.3 Early Standards

Despite the lack of consensus about the risks of contaminated water on recreational swimmers, concern among state authorities and the public about the consequences of swimming in contaminated water continued (Dufour, 2007). By 1963, 38 states had water quality standards and all of them based the standard on the numbers of total coliforms enumerated by the MPN method (Senn, 1963). For these states, coliform based standards for water quality ranged from 50 per 100 mL to 2400 per 100 mL and nearly half of all water quality standards used 1000 coliforms per 100 mL as the maximum acceptable density (Senn, 1963; Dufour, 2007). The frequent use of the 1000 coliforms per 100 mL was primarily based on a 1932 study conducted along the shore of Connecticut's Long Island Sound. This exhaustive study evaluated water quality at 922 sampling stations, established at approximately 1000 foot intervals for nearly 180 miles of shore (Scott, 1932). Using the MPN method, four samples (one for each stage of the tide) were taken by a boat in 2 to 6 feet of water at each station (Scott, 1932). The study found that roughly 7% of the total distance they tested fell into the poorest classification of water quality, Class D, with a coliform count greater than 1000 coliforms per 100 mL (Scott, 1932). Without any clear national guidelines, state level water quality managers adopted this value because it was attainable—so long as water did not fall into the poorest quality classification it would be remain open to swimming.

Regulation for water quality finally came under the national purview when in 1968 the Federal Water Pollution Control Administration commissioned the National Technical Advisory Committee (NTAC) to examine water quality standards for a multitude of water uses—one of these was recreational swimming (NTAC, 1968). The NTAC report specifically points out the variable relationship between total coliforms and fecal pollution as the main limitation in using that standard to assess water quality. In lieu of total coliform test, the NTAC recommended fecal coliforms as being both more closely linked to the fecal contamination of warm-blooded animals and as being a practical substitute for total coliforms. Although the limitations of the Public Health Service study (Stevenson, 1953) were acknowledged, the NTAC used the data generated by Stevenson to propose that adverse health outcomes were not likely to occur unless the total coliform count exceeded 2300 coliforms per 100 mL. Because a study of the Ohio River found that fecal coliforms consisted of about 18% of the microbes in a typical total coliform sample, the NTAC study promulgated the 400 fecal coliform per 100 mL (approximately 18% of 2,300) as an acceptable threshold for water quality. After including a safety factor, the NTAC's final recommendation was that a minimum of five samples taken over 30-day period should not exceed 200 fecal coliforms per 100 mL and 90% of samples taken during any 30-day sampling period should be lower than 400 fecal coliforms per 100 mL. Ultimately this national recommendation of 200 fecal coliforms per 100 mL (to be used for both marine and fresh water) was only slightly less restrictive than the 1000 total coliforms per 100 mL standard (Dufour, 2007).

The recommendation threshold of 200 fecal coliforms per 100 mL by the NTAC was based on two important assumptions. First, since the Public Health Service was unable to

identify an excess of illness below 2300 total coliforms per 100 mL of water in Lake Michigan, the public at large would not be in danger of illness below this threshold (Stevenson, 1953). Second, because 18% of total coliforms isolated from the waters of the Ohio River were fecal coliforms, this proportion would also apply to total coliforms in Lake Michigan. The serious limitations of Public Health Service's study (discussed previously) notwithstanding, Stevenson himself noted that the 2300 coliform per 100 mL should not be taken conclusively citing the limitations of data collection (Stevenson, 1953). The second assumption contradicts one of the NTAC's own premises—namely, that fecal coliforms are a variable portion of total coliforms (NTAC, 1968). As pointed out by Cabelli (1983), at best the 200 fecal coliform per 100 mL was based on detectable risk and at worst a standard that hinged on attainability.

1.2 The Biological Underpinning of Health Risks from Swimming

Quantifying the risk of health outcomes from recreational water activities has been complicated by the fact that documented cases of swimming-acquired illness are largely the result of outbreak conditions. It is these rare events that gain the attention of medical professionals and public health administrators (Cabelli, 1978). Typically, disease is the result of contact with grossly contaminated water. Attack rates are high, symptoms are recognized by medical professionals, and the disease etiology is quickly understood (Cabelli, 1978). Limitations in surveillance systems are such that, generally, only outbreak conditions are recognized. However, epidemiological studies like that of Stevenson showed that endemic transmission may account for an even larger portion of the disease burden (Eisenberg et al., 2002; Stevenson, 1953).

One of the only major outbreaks of enteric disease directly attributable to swimming was the 1974 episode of Shigellosis in Dubuque, Iowa (Cabelli, 1978; Rosenberg et al., 1976). Rosenberg et al. investigated forty-five culture-positive cases of shigellosis in 29 families. Symptoms included GI illness, abdominal pain, fever, chills, and vomiting. Nearly 96% of those afflicted consulted a physician and 40% were hospitalized, with one person having surgery for acute appendicitis. Ultimately, Rosenberg and colleagues found that the 31 confirmed cases of shigellosis had resulted from swimming in a small stretch of the Mississippi River where fecal coliform counts were nearly 90 times higher than regulations permitted (fecal coliform count of 17,500 organisms per 100 mL). At the time, water quality criteria were usually set to 200 fecal coliforms per 100 milliliters (Dufour, 2007). The case of Dubuque, Iowa reflected an outbreak of shigellosis resulting from extraordinarily high levels of pollution. Although cases like this one justified concerns about water quality, they could not elucidate the etiology and frequency of illness resulting from typical water quality conditions (Cabelli, 1978; Moore et al., 1959). More importantly, such studies could not help to establish policy based on acceptable levels of endemic risk.

Illness occurring from water recreation is usually self-limiting and mild, rarely requiring hospitalization (Cabelli, 1978). Hence, there is typically no evidence in hospitalization data, or other public health data sources, that swimming presents an acute public health threat, or indeed, any threat. It is unsurprising, then, that retrospective studies like that conducted by Moore et al. (1959) or Foster et al. (1971) failed to come up with any alarming rates of swimming-related illness, whether severe or mild. As Cabelli points out, the most commonly *reported* cases as of 1978 were leptospirosis, swimmer's itch, infections of *Aeromonas*

hydrophila and *Vibrio* species, and skin rashes associated with *P. aeruginosa*. Public health records did not corroborate Stevenson's findings about the prevalence of GI illness, eye infection, ear, nose, and throat infections, as well as respiratory illness—all of which represent the majority of recreational swimming illnesses (Stevenson, 1953; Cabelli, 1978; Prüss, 1998). Although Stevenson's study could not, in itself, inform water quality standards, it did demonstrate that persistent low-level disease—endemic disease—was observable. Stevenson showed that, with large sample sizes, epidemiologic studies could identify elevated risks of illness even in the absence of recognized outbreaks.

Water quality management had evolved dramatically in the first half of the twentieth century. Early public health officials, observing beach conditions that were aesthetically "revolting," sought to trace the path of raw sewage from the initial release into swimming waters until the final dissipation with the incoming tide (Moore, 1954). Fecal indicators were a sensible way to map the flow of fecal waste. Fecal waste, it was believed, carried pathogenic organisms into beach waters and it was these microbes that caused adverse health outcomes in swimmers. Thus, three simple premises—fecal indicators predict fecal waste, fecal wastes contain pathogens, and pathogens cause adverse health outcomes—formed the syllogistic backbone of water quality management. Given that these statements are true, the logical conclusion is that fecal indicators should predict adverse health outcomes. Large epidemiological studies, coupled with environmental sampling, identified variability in these health outcomes and in indicator densities. It then remained to be seen whether health effects could be associated in a measurable and consistent way with fecal indicator bacteria.

1.2.1 Fecal Indicators Predict Fecal Waste.

Fecal indicator bacteria are organisms found in the intestinal tract of humans and other warm-blooded animals. Although fecal indicators are reliably present in fecal waste, their concentrations in waste are highly variable. Once in the environment, indicators (as viable organisms) are dependent on a host of conditions that may depress or inflate their numbers in situ, without any changes in absolute quantities of fecal waste (Foster et al., 1971; Miescier and Cabelli 1982). Thus, fecal waste may be indicated where none exists or waters may be deemed safe even though contamination is known to be present (Hanes et al., 1964; Cohen and Shuval, 1972). Given these limitations, the choice of indicator species with which to trace fecal pollution has been debated since the inception of the technique. Total coliforms, fecal coliforms, enterococci[3], streptococci, staphylococci, *Escherichia* coli, *Pseudomonas aeruginosa*, *Clostriudium perfingens*, and other species have all been evaluated as fecal indicator candidates.

In 1949, Stevenson noted that all water quality standards promulgated were based on bacteriological densities of total coliforms. The coliform standard was used in 1953 and again in 1959 to evaluate water quality in seminal epidemiological studies both in the United States and abroad (Moore et al., 1959; Stevenson, 1953). However, as Stevenson points out in 1953, the ubiquity of coliform species in the environment, even in the absence of fecal pollution, created concerns that the standard would be unnecessarily restrictive. The report by the NTAC was the first to narrow the coliform group to bacteria specific for fecal waste from warm-blooded

[3] Enterococci were previously classified as belonging to the fecal streptococci subgroup, Lancefield's serological Group D. In 1984, genomic DNA analysis indicated a separate genus classification would be appropriate (Slanetz and Bartley, 1964; Schleifer and Kilpper-Bälz, 1984).

animals (NTAC, 1968). However, this report was by no means definitive; from the early studies in the 1950s to later investigations in the 1970s, the utility of multiple microorganisms were evaluated against the fecal waste-fecal indicator paradigm.

Moore (1954) sampled and measured levels of *Bacterium coli*[4] in order to investigate whether the relocation of an outfall pipe had positively affected the quality of water at a local beach. In a different technique, Robson chose to spike raw sewage with *Serratia marcescens* and *Serratis indica* before discharge as a way to track the pollutants "because the colonies produce a distinctive red color on nutrient agar" (Robson, 1956). In 1972, Cohen and Shuval published a study concerning the survival of coliforms, fecal coliforms, and fecal streptococci in the effluent-dominated waters of sewage treatment plants and rivers as well as a lake, and sources of drinking water in Israel. Their study stressed the variable efficacy of different indicators in tracking fecal pollution depending on such factors as temperature, season, dilution, and distance from the source and found that fecal streptococci were often the only microbial indicator reflecting the fecal origins of pollution distant from the source (Cohen and Shuval, 1972).

Cabelli et al. (1976) published a report evaluating *Pseudomonas aeruginosa* as a potential predictor of health outcomes in recreational water. However, the results of the research indicated that *P. aeruginosa* did not, in fact, have a positive correlation with fecal pollution. What's more, the indicator organism's etiology in non-fecal sources and multiplication in surface waters made it an unreliable indicator of fecal waste (Cabelli et al., 1976). In a 1980 study, Bison and Cabelli evaluated *Clostriudium perfingens* as a possible health-

[4] "*Bacterium*" is presently an invalid genus now reclassified as *Escherichia*.

effects indicator of water quality. This indicator, too, was deemed unreliable with regard to fecal pollution—the extended survival of spores made them a ubiquitous presence in all waters, polluted or not, and *C. perfingens* was frequently detected when sub-surface sedimentation and soils were disturbed (Bisson and Cabelli, 1980). *Bacteroides fragilis* was also found to be an inadequate detector of fecal pollution (Allsop and Stickler, 1985). Only tests for enterococci were consistently found to have some environmental stability. Enterococci were found to be "nutritionally fastidious"; these organisms did not multiply in surface waters and (along with the rest of the fecal streptococci group) were found to be abundant in human fecal waste and relatively sparse in the so-called "virgin" environments—i.e., conditions not affected by sewage or pollution (Hanes et al.,1964; Slanetz and Bartley, 1957; Slanetz and Bartley, 1964).

Environmental resilience and stability were the most important characteristic of indicator species. Researchers ruled out multiple organisms as useful indicators based on in situ increase or die-off. Continued concern about the environmentally-linked changes in indicator density prompted investigation of numerous environmental, chemical, physical, and biological variables, demonstrating that indicator survival is far from straightforward. Analysis of a stream bed free of human fecal contamination showed high values of *E. coli* and revealed that those microbial communities were persistent (Byappanahalli *et al., 2*003a). Other analyses found high concentration of both enterococci and *E. coli* in mats of macro-algae *Cladophora glomerata* along Lake Michigan shores (Whitman et al., 2003). Further investigation of this phenomenon not only showed that macro-algae could sustain these bacteria in a dry, dormant phase, but that the algal leachate supported in vitro multiplication of enterococci and *E. coli* (Whitman et al., 2003; Byappanahalli et al., 2003a). Sunlight was also an important variable. Obiri-Danso et

al. (1999) found that die-off of fecal indicator bacteria was higher in early morning hours where UV light intensity peaked. In 2002, Sinton et al. confirmed these findings. Whitman et al. (2004) demonstrated that during sunny days, densities of *E. coli* in Lake Michigan decreased by several orders of magnitude; they also found that the effects of wind speed on lake level, wave height, and turbidity significantly impacted changes in indicator density. Similar variability was found when looking at tidal patterns and sediment disturbances (Rosenfeld et al., 2006; Halliday and Gast, 2011).

Literature also shows that rainfall has significant effects on levels of fecal indicator bacteria (Noble et al., 2003). The detection of total coliform, fecal coliform, and enterococcus was found to differ considerably between wet weather and dry conditions (Noble et al., 2003). Noble and colleagues also showed that associations between indicators species also changed depending on rainfall. Modeling of total coliform, *E. coli*, and enterococci in an intertidal wetland were found to be substantially impacted by urban runoff and re-suspension of contaminated wetland sediments significantly contributed to an increase in the load of fecal indicator bacteria in intertidal waters (Sanders et al., 2005). In Santa Ana wetlands, increases were found in total coliforms, *E. coli* and enterococci in the days immediately following a dry-to-wet weather change (Evanson and Ambrose, 2006).

Water management policies place great emphasis on the indicator monitoring even though extensive evidence shows that indicator levels may have little to do with pollution. There exists an extensive body of research demonstrating the levels of fecal indicator bacteria do not relate exclusively to levels of human fecal pollution. Indicators may flourish in sediment beds and in animal waste; indicators can also remain viable for months on dry macro algae

(Fogarty et al., 2003; Whitman et al., 2003). It has been shown that indicator levels are decimated in the brightest parts of the day recharged in the night (Whitman et al., 2004). Temporal and spatial variability abound. Yet water quality managers continue to emphasize fecal indicator bacteria levels, even though these organisms are not in themselves pathogenic.

1.2.2 Fecal Waste Contains Pathogenic Organisms

Like indicators, human disease-causing pathogens are present in variable amounts in fecal waste—changing within communities and among populations. They are also subject to environmental exigencies, though typically different ones from indicator species (Miescier and Cabelli, 1982). This diverse group of microbes can include protozoa, bacteria, and viruses, many species of which have likely not been conclusively identified. Gerba and Smith (2005) point out that untreated sewage may contain 150 unknown enteric pathogens. Their presence in surface water can result from polluted rainwater runoff, wastewater discharge, and direct fecal contamination by animals or humans. Just as with indicator species, the concentration and viability of these pathogenic microbes are suspected to vary with rainfall, ultraviolet light intensity, temperature, water depth, water flow dynamics, and nutrient availability, as well as proximity to fecal pollutant sources (Chauret et al.,1999). With changes in the density and infectivity of microbial pathogens comes a concurrent change in risk for adverse health outcomes from exposure.

The hazards of fecal waste have long been understood. The most pathogenic organisms in human waste are typically *Campylobacter* spp., *Salmonella* spp. (non-typhoid), *Listeria*, *E. coli* O157:H7, *Cryptosporidium parvum*, *Giardia lamblia* (Gerba and Smith, 2005). Lipp et al. (2001) demonstrated that high densities of on-site disposal systems (i.e., septic tanks) in a coastal

community correlated well be detection of *Cryptosporidium*, *Giardia*, and enteroviruses in nearby beach water.

1.2.3 Pathogenic Organisms Cause Adverse Health Outcomes

Waterborne illnesses are typically attributed to a fecal source. Ingestion, dermal contact, and inhalation of water contaminated with fecal waste exposes individuals to pathogenic organisms and can cause significant burdens of disease in the public. Outbreak conditions highlight important pathogenic agents. In 1976, Rosenberg et al. investigated 45 confirmed cases of *Shigella sonnei* infection resulting from contact with grossly contaminated waters. Similarly, Bryan et al. (1974) implicated swimming in a contaminated fresh water lake with 14 cases of viral hepatitis A. In 1982, Kappus et al. used serological analysis to establish that an outbreak of Norwalk virus caused 103 cases of gastroenteritis. Evidence pointed to the contaminated waters of a swimming pool as well as person-to-person transmission (Kappus et al., 1982). Koopman and colleagues (1982) proposed that an outbreak of GI illness was also linked to Norwalk virus acquired by swimming a recreational park. Perhaps the most notorious example is the contamination of the municipal water supply of Milwaukee with *Cryptosporidium* which resulted in over 400,000 individuals affected by GI illness (MacKenzie et al., 1994).

1.3 Fecal Indicators Predict Adverse Health Outcomes

The difficulties in identifying a stable and consistent indicator of fecal waste underscore the dynamic mechanistic processes in the transmission of waterborne disease. Ideally, changes in the density of indicators organisms should reflect changes in quantities of fecal waste, which

in turn would dictate the pathogenicity of surface water. However, as researchers found, viable microbiological organisms are subject to the exigency of environmental conditions. The survival of both pathogens and indicators are highly dependent on a number of physical conditions. Apart from changes brought on by direct pollution, the concentration and viability of indicator species can vary dramatically as a result of rainfall, ultraviolet light intensity, temperature, water depth, water flow dynamics, and nutrient availability (Evanson and Ambrose, 2006; Miescier and Cabelli, 1982; Obiri-Danso et al., 1999; Rosenfeld et al. 2006; Savage and Hanes, 1971). Thus, the relationships between indicators and fecal waste, or fecal waste and pathogens, or even indicators and pathogens, could not be satisfactorily elucidated. In the absence of reliable information about such relationships, the US Environmental Protection Agency (EPA) sought to evaluate the proposition that fecal indicators predict illness based on a risk assessment model without attempting to evaluate microbiological underpinnings of that conclusion. The results of the prospective cohort studies have stipulated EPA policy and water quality guidelines to the present day.

1.3.1 The Environmental Protection Agency Studies for Marine and Freshwater

From 1972 to 1979 the US EPA conducted two sets of epidemiological and microbiological research projects to determine whether microbial indicators of fecal pollution predicted health impacts from swimming, one at marine bathing beaches and the other at freshwater locations. The EPA studies differed from previous epidemiological efforts in several important ways. First, swimming was an activity strictly defined by head exposure to water; second, controls were non-swimming but beach-going populations; finally, short trial periods were used (Cabelli, 1983). Earlier attempts by Stevenson and Moore had used seasonal means

of water quality and did not identify exposure groups according to total immersion in water (Stevenson 1953; Moore et al., 1959). Most importantly, the parameters of interest in the final model (incidence of disease versus water quality) were specifically investigated; the authors wished to know which were the “important” illnesses and the “best” indicators with regard to water quality (Cabelli, 1983). Using strict definitions of exposure groups and water quality data that matched exposure days, researchers developed a risk assessment model based on the rates of specific illness as a function of the densities of the most important indicators.

The study of marine bathing beaches was conducted at pairs of beaches in the New York City and Boston areas. In (or near) each of these cities, one relatively unpolluted (RU) and one barely acceptable (BA) (according to local guidelines) were studied (Cabelli et al., 1979). The researchers evaluated the health outcomes of the populations of swimmers and non-swimming controls who had engaged in weekend activities only. Using telephone surveys, they categorized any adverse health effects following beach attendance as GI, respiratory, or other. For GI symptoms, the authors sought to identify instances of highly credible GI illness (HCGI) by evaluating whether subjects experienced vomiting and diarrhea in concurrence with fever, nausea and stomachache (Cabelli et al., 1979). Respiratory symptoms were also subcategorized as sore throat; bad cough; chest cough; runny or stuffed nose; earache or runny ears; red, itchy or watery eyes (Cabelli et al., 1979).

The preliminary analysis found that swimming in BA water presented measureable risks of adverse health outcomes as opposed to not swimming, corroborating the findings of the 1953 Stevenson study (Cabelli et al., 1979). Researchers also found that GI illness rates were appreciably higher among children and low-middle socioeconomic individuals who had engaged

in swimming at Coney Island where water was deemed BA (Cabelli et al., 1979). Later research from the Lake Pontchartraln (Louisiana) and Boston Harbor (Massachusetts) showed that densities of enterococci were correlated with "highly credible" gastrointestinal illness and GI symptoms were strongly associated with distance from point sources of municipal wastewater (Cabelli et al., 1982). Although the majority of illnesses did not require hospitalization, the authors stress that they occurred at indicator levels within established guidelines—levels below which there was to be “zero risk” of adverse health outcomes (Cabelli et al., 1979). These results strongly implied that risk analysis of primary contact water recreation was possible based on a continuously increasing risk model and that typical safety thresholds had been set too high. The study also called into question the use of fecal coliforms as the most appropriate indicator of health outcomes for primary contact water recreation. The latter concern was critically important since the need to accurately trace fecal waste underpinned efforts to predict health outcomes.

Ultimately, the studies at marine bathing beaches concluded that there exists a strong linear relationship between HCGI and log10-transformed levels of enterococci in water; no other indicator had predictive value (Cabelli, 1983). Illness rates from swimming ranged between 0 and 28 illnesses per 1000 swimmers while concurrent quality conditions ranged from 3 enterococci per 100 ml to nearly 500 enterococci per 100 ml (Cabelli, 1983). The set of freshwater studies, conducted at Lake Erie and Keystone Lake, showed an association between GI illness and enterococci as well as *E. coli* (Dufour, 1984). For this study, the average rate of swimming-associated gastroenteritis ranged between 0 and 14 illnesses per 1000 swimmers (Dufour, 1984). Levels of *E. coli* had a mean of 72 per 100 ml and ranged from 18 to 250

organisms per 100 mL (Dufour, 1984). The mean for enterococci was roughly 20 per 100 ml and enterococci densities ranged from 6 to 80 per 100 ml. No other indicators had significant associations with swimming-related illness (Dufour, 1984).

In his 1983 study, Cabelli notes that the development of water quality criteria for swimming has followed three basic stages. In the first stage, the criteria were based on what was *attainable* for most recreational water bodies—standards feasible for roughly 90% of swimming waters (Scott, 1932). The second stage focused on *detectable* health outcomes. As of the NTAC report (1968), that value had been set by the Stevenson study to 2300 coliforms per 100 mL (or 400 fecal coliforms per 100 mL). After including a safety factor, the NTAC recommendation of 200 fecal coliforms per 100 mL was thought to eliminate risk (NTAC, 1968). The two reports by the EPA on marine and freshwater bathing beaches showed that the NTAC recommendation did not eliminate risk and that a standard based on *acceptable risk* could be derived from the rates of GI illness as predicted by levels of enterococci and *E. coli* (Cabelli, 1983; Dufour, 1984).

In 1986, the EPA promulgated the “Ambient Water Quality Criteria for Bacteria” based on the studies by Cabelli and Dufour. For marine water it was stipulated that “A geometric mean of 5 samples taken at equal time intervals over a 30-dayperiod shall not exceed 35 enterococci per 100 ml” (US EPA, 1986). In freshwaters the standard states “A geometric mean of 5 samples taken at equal intervals over a 30-day period shall not exceed 126 *E. coli* per 100 mL” and “A geometric mean of 5 samples taken at equal time intervals over a 30-day period

shall not exceed 33 enterococci per 100 mL" (US EPA, 1986). Single sample maximum (SSM[5]) values were also established for designated bathing beaches (US EPA, 1986). In freshwater, the SSM value for enterococci was set to 62 per 100 mL and for *E. coli* it was 235 per 100 mL; in marine water, the SSM for enterococci was 104 per 100 mL (US EPA, 1986). The standards were meant to be no more restrictive than the original 200 fecal coliform per 100 mL standard while at the same time more accurately quantifying the health of risks of swimming at the levels stipulated (US EPA, 1986).

In studying what could feasibly be used as predictors of water quality, researchers had bypassed the true etiological agents of waterborne illness. *Salmonella, E. coli*, *Cryptosporidium*, *Giardia*, enterovirus, norovirus, Shigella, streptococci, staphylococci, hepatitis virus, and many other organisms have all been identified as pathogenic organisms known to be the causative agents of disease outbreaks from swimming. Due to technical difficulties of isolating and accurately quantifying these organisms, they have not been reliable indicators of health effects and hence, water quality. Additionally, the causes of sporadic cases of recreational waterborne illness are not well characterized, so even if pathogen sampling and analysis were simplified, it is not clear which pathogens should be monitored for public health purposes. Although EPA guidelines had articulated risk, they did not address relationship between indicators and the true biological underpinning of waterborne illness.

1.3.2 Challenging the Environmental Protection Agency Standards

Several other prospective cohort studies have also evaluated the validity of using risk assessment to establish microbiological guidelines for water quality. Seyfried et al. conducted a

[5] SSM values were replaced with statistical threshold value (STV) by the EPA's draft criteria in 2011 (US EPA, 2011).

prospective cohort epidemiological-microbiological study at ten freshwater beaches in Ontario, Canada (Seyfried et al., 1985a; Seyfried et al., 1985b). The study design closely matched that of Cabelli and Dufour for beaches in the United States. The authors sought to identify health effects as predicted by fecal coliforms, fecal streptococci[6], coagulase-positive and coagulase-negative staphylococci, *Pseudomonas aeruginosa*, and heterotrophic bacteria (Seyfried et al., 1985a; Seyfried et al., 1985b). The results of their study confirmed the early observations of Stevenson and the corroborating evidence of Cabelli and Dufour that swimmers have a higher incidence of morbidity as compared to non-swimmers. Because the authors did not analyze enterococci, Seyfried et al. could not confirm the correlation of these microorganisms with health effects. However, they did find the deterministic model for the total staphylococcal versus total illness had correlation coefficients of 0.439 (Seyfried et al., 1985a; Seyfried et al., 1985b). The authors found that while fecal coliforms and fecal streptococci were correlated with total illness, staphylococcal deterministic model provided a more robust dose-response model (Seyfried et al., 1985a; Seyfried et al., 1985b).

Another freshwater study, a retrospective epidemiological and microbiological study by Ferley et al., evaluated the health effects from swimming in the Ardèche River in the south of France. The river, which receives untreated urban domestic sewage, is frequented by nearly 350,000 tourists each year. Study participants were chosen from visitors to the many "family camps" in the area and, after satisfying the inclusion criteria, participants were asked about the frequency, location, and duration of their swimming activities. Because of the proximity between the bathing beaches, participants swam at multiple locations and had combinations of

[6] Enterococci are a taxonomical subgroup of fecal streptococci.

exposures, creating a serious limitation in the ability of authors to infer associations. Results are also subject to recall bias since Ferley and colleagues recorded health outcomes after they had occurred. Exposure assessment was further complicated by the fact that researchers only took two water quality measurements each week. Notwithstanding these limitations, Ferley et al. were able to confirm that higher general morbidity was associated with higher levels of fecal streptococci and both coliform indicators. The relationship between fecal streptococci and "objective acute GI disease" was even stronger (correlation coefficient=0.62), though this was not true for the coliform groups. Like Stevenson, Cabelli, and Seyfried, Ferley et al. were able to confirm a higher incidence of GI illness among the exposed group. Moreover, Ferley et al. found a strong association between fecal coliform and skin disease (correlation coefficient = 0.67). In spite of the lack sensitivity, associations found by Ferley et al. suggest that indicators could be predictors of health effects.

A study of Hong Kong beaches in 1986 and 1987 by Cheung et al. also found risks associated with swimming (Cheung et al., 1990). This prospective cohort study closely followed the methodology of Cabelli et al. for marine bathing beaches—swimmers and non-swimmers were followed at nine Hong Kong beaches to determine whether simply engaging in primary contact water recreation carried a higher risk of morbidity (Cheung et al., 1990). Notably, Cheung et al. used relatively frequent sampling methods to assess water quality; each beach was sampled in three separate locations four times a day on study days giving researchers nearly twelve data points about water quality for each location-day. Cheung et al. found that there was a significant excess of illness among swimmers as compared to non-swimmers with the rate of HCGI symptoms 5 times higher for swimming than non-swimmers; rates of eye and

fever symptoms, 4 times higher; and total illness rates 2–3 times higher. Combined rates of swimming associated skin and HCGI symptoms provided the strongest correlation (r=0.73) with geometric mean *E. coli* densities, the only relationship from which a statistically significant linear relationship could be derived (Cheung et al., 1990). This finding contradicts Cabelli's conclusions about the importance of enterococci as an indicator and also refutes that study's claim that *E. coli* is not predictive of health impacts at marine bathing beaches (Cabelli, 1983). In concurrence with Seyfried's findings, Cheung et al. also showed that staphylococci were important indicators of health effects—specifically, ear and sore throat symptom rates—although this relationship was not strong enough to form a significant predictive model (Cheung et al., 1990).

While the US EPA acceptable illness rate is 19 cases of gastroenteritis symptoms per 1000 swimmers, the acceptable risk criterion for the beaches in Hong Kong is set at 15 gastroenteritis and skin symptoms per 1000 swimmers (Cheung et al., 1990). According to Cheung et al. (1990), this level of "acceptable risk" can be achieved by enforcing a geometric mean *E. coli* density of 180 *E. coli* for 100 mL as the maximum allowable threshold for swimmable water. It is noteworthy that this study demonstrated lower morbidity rates amongst Hong Kong beachgoers as compared to populations in the United States, even though Hong Kong beaches were more polluted based on *E. coli* densities. The authors stress that this fact may reinforce the need to establish location-specific targets for water quality that take into account endemic disease susceptibility and local indicator-pathogen relationships (Cheung et al., 1990). Cheung et al. also forward the idea that a dual indicator system would be more

protective; they suggest using threshold of 1000 staphylococci per 100mL to supplement the health effects information supplied by the *E. coli* levels (Cheung et al., 1990).

Another large prospective cohort study was conducted by Corbett et al. (1993) on the health effects of swimming at Sidney beaches. Investigators followed the methodology of Cabelli et al., recruiting beach goers and differentiating between swimmers and non-swimming controls (Cabelli, 1983;). Due to ethical concerns, researchers excluded children from the study, a fact that may have significantly affected their results. Corbett et al. attempted to improve upon previous study designs by sampling waters more frequently and associating exposures among swimmers with the precise water quality measurements. Corbett et al. were able to show a relationship between increasing densities of bacteria and increases in fever, respiratory ailments, and other symptoms. Researchers concluded that increases in illness could be associated with increasing levels of pollution and that a safe threshold for water quality is likely to exist. Despite these findings, the results of this study contradict previous work in a number of important ways. First, Corbett et al. found the fecal coliforms were better predictors of adverse health outcomes than fecal streptococci, refuting the conclusions of Cabelli et al. and Dufour et al. Researchers also found that GI symptoms were reported by only 4.1% of the subjects and the rates of GI illness did not increase with increasing counts of fecal bacteria (Corbett et al., 1993).

In forming the EPA water quality criteria, Cabelli et al. (1983) had asked the question "Which are the 'important' types of illnesses, and which is the 'best' indicator"? The answers to these two questions were gleaned from epidemiological research at a handful of specific locations. The studies for marine bathing beaches were conducted in New York City, at Coney

Island and Rockaways beaches; in Boston, at Revere and Nahant beaches; and at Lake Pontchartrain in the New Orleans area (Cabelli, 1983). The freshwater bathing beaches were on Lake Erie, in Erie, Pennsylvania; and at Keystone Lake about 60 miles east of Tulsa, Oklahoma (Dufour, 1984). Cabelli and Dufour found that the "important" type of illness was GI and that the "best" indicators were *E. coli* and enterococci. Based on the relationships found at these locations, recreational water quality standards were set for all waters in the United States (US EPA, 1986). However, continued epidemiological research has failed to rationalize the use of a uniform set of standards across all bodies of water. In Canadian waters, Seyfried et al. found that total staphylococcal correlated best with total illness rates (Seyfried et al., 1985a; Seyfried et al., 1985b). In a French river, Ferley et al. concluded that fecal streptococci and "objective acute GI disease" presented the strongest relationship (Ferley et al., 1989). Water quality at Hong Kong beaches, as shown by Cheung et al., could best be evaluated on the basis of the relationship between swimming associated skin and HCGI symptom rates and geometric mean *E. coli* densities, directly contradicting Cabelli's conclusions about the importance of enterococci as an indicator of marine water (Cabelli, 1983; Cheung et al., 1990). In Australia, Corbett et al. could not even confirm an increase of GI illness with increasing levels of pollution (Corbett et al., 1993). Empirical evidence across international waters showed that EPA's 1986 water quality criteria would likely result in different rates of illness in differing locations.

1.3.3 Re-evaluation of Environmental Protection Agency Data

A re-evaluation of EPA data showed that location specific differences in risk assessment models were evident even in the EPA study (Fleisher, 1991). The analysis concluded that the EPA water quality criteria had been reached by inappropriately grouping location data and

extrapolating those results to all water bodies. Specifically, Fleischer points out the inappropriate clustering of data points that limited the regression analysis to only 18 data points (Cabelli, 1983; Fleisher, 1991). Such clustering, it was argued, inappropriately grouped data from typical marine waters in New York City and Boston with brackish waters in a Lake Pontchartrain, Louisiana (Fleisher, 1991). Using actual numbers of swimmers and non-swimmers, (effectively reversing the clustering) Fleischer built a logistic regression model which predicted the probability of an individual contracting a disease based on a number of variables (Fleisher, 1991). In the first model it was shown that enterococci predicted GI illness (Fleisher, 1991). This model was significantly improved by the introduction of a location term and the addition of an interaction term which accounted for the effects of location on enterococci density (Fleisher, 1991).

Using this finalized model, research showed that increasing densities of enterococci predicted increases in GI illness at the marine beaches but did not predict the same in the brackish location; in Lake Pontchartrain the health effects remained unchanged in relation to changing enterococci densities (Fleisher, 1991). More importantly, Fleischer showed that the pooled dataset and clustered points used by the EPA study predicted 19 GI illnesses per 1000 persons based on an enterococci density of 35 organisms per 100 mL. At that enterococci density, the re-analysis predicted 24, 82, and 36 GI illnesses per 1000 persons at New York City, Boston, and Lake Pontchartrain, respectively (Fleisher, 1991). Fleischer's study also suggested that maximum allowable indicator densities could be increased in New York City and Lake Pontchartrain without excessive increase in risk to the swimmer . Overall, the re-analysis

presented by Fleischer et al. questioned the use of a single allowable maximum density of indicator species to govern all recreational water bodies in the United States.

Re-evaluation of EPA data by Fleischer showed that even amongst the locations studied, risks varied considerably and that data had been inappropriately grouped. The results called into question the specific relationship between microbial indicators of recreational water quality and health outcomes forwarded by the EPA study. However, the critique of EPA standards went even further. Finding large discrepancies of health risks among differing locations, Fleischer questioned whether a single microbial water quality standard would be appropriate for all bodies of water citing the dynamic relationship of pathogens and indicators in situ and the influence of environmental conditions (Fleisher, 1991). This paper and others began to raise an important concern in predicting waterborne illness from recreational swimming: the natural history of disease cannot be the same across all locations since endemic disease differ between locations and populations (Fleisher, 1991; Eisenberg et al., 2002).

1.4 Endemic Risk

National water quality standards are based on the simple premise that fecal indicators predict adverse health outcomes. These criteria are promulgated for all bodies of water because the relationship between indicator level and disease risk is believed to be broadly generalizable across locations and populations. Cabelli et al. (1983b) argued that outside of the context of disease outbreaks, there may be a constant indicator to pathogen ratio in wastewater. Along with Dennis and Wolman (1959), they posit that the input of pathogenic organisms into waste is the result of a predictable level of endemic illness in the population.

According to the authors, this assumption holds true for recreational waters that receive waste from large municipal wastewater treatment plants (Cabelli et al., 1983b). Waste from larger populations, they reason, would not be highly impacted from the pathogenic contributions from a small number of highly sick individuals (Cabelli et al., 1983b). Implicit in this assumption is the idea that that illness, susceptibility, and immunity are also constant across populations. Thus, water quality standards appropriate for one location could be extrapolated to all recreational water bodies.

1.4.1 Meta-analysis of Epidemiological Data

Meta-analyses of epidemiological data by Wade et al. (2003) suggest that a *static* and *consistent* relationship between a specific bacterial indicator of water quality and a specific health effect does not exist. Their examination of a large body of epidemiological work shows substantial variability in the relationship of indicator species and health effects. Wade et al. (2003) found significant heterogeneity between of 27 epidemiological studies and concluded that no single indicator could predict illness consistently "in all environments at all times." The results of epidemiological risk assessment, they argue, are dependent on multiple variables in multiple categories ranging from methodology to environmental conditions. Wade et al. investigated variability caused by water source, adjustment for covariates, study design, length of follow-up period, swimming definition, and geographic location. They found that several factors contributed to variability in relative risk selection of the control group (non-swimmers versus swimmers) as well as the relative health of the study population. Although Wade et al. found evidence supporting the use of enterococci and *E. coli* as water quality standards, the authors could not support the specific standards promulgated by the EPA. In another meta-

analysis, Prüss also supported the general use of indicator organisms as predictors of health outcomes (Prüss, 1998). However, this study, too, showed that epidemiological work pointed to location-specific relationships between indicators and health risk.

1.4.2 Location-Specific Risks

Epidemiological studies suggest that primary contact water recreation carries with it some measurable risk of adverse health outcomes. However, the precise nature and magnitude of that risk have not been agreed upon. The varying results by Cabelli and Dufour, Seyfried, Ferley, Cheung, Corbett, and others showed that, despite extensive research, empirical evidence could not support the use of uniform water quality standards across locations. While indicators may be predictive of health outcomes, the indicator threshold which differentiates a reasonably safe water-body from a hazardous one changes depending on specific conditions, unique locations, and the susceptibility of a population. Demographic disparities across location such as age and gender, as well as the presence or absence of immune-compromised individuals, mutations, and immunological adaptability can all create significant differences in endemic rates of disease across populations (Eisenberg et al., 2002). In other words, disease transmission is highly dependent on a population's number of susceptible, infected, diseased, and immune individuals.

1.5 A New Gold Standard of Water Quality Assessment

Emphasis on indicator organisms as predictors of water quality largely eclipsed pathogen monitoring efforts. Early-on, managers of water quality sought to trace the path of fecal waste using methodology that was both fast and cost-effective. The initial candidates for

FIB, total and fecal coliforms, were chosen because they were related to fecal waste and had the necessary economic and technological virtues. On the other hand, pathogens organisms have never been a good candidate for fecal indication. These microorganisms are difficult to enumerate because of their low concentrations in water and available methods were correspondingly slow, costly, and imprecise. Relationships between fecal indicators and pathogens are also dependent on complex chemical and biological conditions that have not been adequately addressed. Continued investigation of the health effects model has made it increasingly clear that morbidity is largely dependent on the endemic characteristics of disease transmission and risk assessment models are not broadly applicable. This fact, coupled with improvements in the detection and enumeration of pathogens, revived the importance of etiological agents of disease. Bacterial indicators of water quality had been evaluated against health effects; it remained to be seen how well indicators correlated with pathogens.

1.5.1 The Relationship between Indicators and Pathogens

Several studies have sought to evaluate the relationship between fecal indicator bacteria and pathogenic organisms. Like the health effects model, specific indicators had to be evaluated against specific pathogens and under specific conditions. Two investigations are examined here—both concerning fresh water rivers contaminated with treated wastewater. In 1989, Geldenhuys and Pretorius measured the relationship between enteric viruses and total and fecal coliforms, fecal streptococci, and coliphages in a stream polluted with treated domestic effluent. They found that physical and chemical parameters typically affected the survival of indicator species and pathogens differently. Temperature was determined to be the single most important factor influencing the number of enteric viruses ($r = -0.64$). Of the

indicators tested, only numbers of coliphage organisms remained predictive of viruses with changing temperatures (Geldenhuys and Pretorius, 1989). Lemarchand and Lebaron (2003) focused on the relationships between *Cryptosporidium, Salmonella* spp., enterococci, and fecal coliforms in water treatment. Sampling was performed at nine wastewater treatment plants (WTPs) and seven rivers. The authors were able to find a significant correlation between *Cryptosporidium* and *Salmonella spp.* in WTP influent, but not in effluent. Instead, effluent waters showed a relationship between *Cryptosporidium* oocysts and fecal coliform (Lemarchand and Lebaron, 2003). Analysis of river water showed a significant correlation between *Cryptosporidium* oocysts and fecal coliforms (r= 0.83) and another relationship between *Cryptosporidium* oocysts and enterococci (r = 0.71) (Lemarchand and Lebaron, 2003). Both investigations concluded that differences in the transport and survival of enterococci and *Cryptosporidium* oocysts, as well as physical and chemical parameters of the water, govern such relationships.

1.5.2 Differences in Methodology

The complex dynamics of both indicator and pathogen survival are brought to bear when pathogens are predicted by culture-based assays for indicator organisms. These tests necessarily require some synchronous pattern between the survival rates of both groups of organisms in order for a predictive relationship to exist. Due to the complexities of microbe survival in situ, this is rarely the case. However, the problem is circumvented with the use of rapid molecular assays that measure genomic material of given organism. Assays such as quantitative polymerase chain reaction (qPCR) measure viable and non-viable organisms. Hence, relationships between indicator bacteria and pathogen species are less impacted by the

survival characteristics of indicators. Moreover, qPCR-based assays provide rapid, same-day results for water quality management, making predictions timelier.

1.5.3 Assessing the Accuracy of Indicator Diagnosis

The usefulness of various indicator species has been evaluated against any number of conditions, in countless locations, and against a multitude of health effects and pathogenic organisms. Yet no consensus has been established regarding any one relationship. Public waters are still diagnosed as clean or polluted based on levels of FIB although the general diagnostic utility of bacterial indicators of water quality has not been confirmed. In other words, it is unknown whether fecal indicator bacteria are a "good" test of water quality. A "good" test is defined as a test that typically performs better than chance alone at correctly predicting a "true" condition. In order to evaluate the performance of a diagnostic test, the true condition must also be defined on the basis of some gold standard. In the case of FIB, there already exist two gold standards that have been used to validate FIB results, health effects and pathogens. It is beyond the scope of this study to discuss the diagnostic efficacy of indicator species using health effects as a gold standard. The low dose for infectivity of pathogens and the relatively low concentration of these organisms, even in WTPs, warrants the use of the pathogen detection limit as a metric to differentiate true risk levels in recreational water quality.

Harwood et al. (2005) used precisely this method to assess the diagnostic value of FIB at correctly predicting the presence or absence of pathogens. Their analysis of influent and effluent from six wastewater reclamation facilities in the United States suggested that there was no association between any combination of indicators and pathogens (Harwood et al., 2005). At each plant, samples were obtained a minimum of five times from influent and three

stages of effluent (biological treatment, filter effluent, and disinfected effluent) and were analyzed for enteric viruses, *Giardia*, *Cryptosporidium*, as well as the indicators coliphages, enterococci, fecal coliforms, total coliforms, and *Clostridium perfringens* (Harwood et al., 2005). In their analysis, the authors used binary logistic regression analysis to test whether indicator organisms were predictive of pathogens. Using a detection limit of 0.2 colony-forming units (CFU) per 100 mL for bacteria and 10 plaque-forming units (PFU) per 100 mL for coliphages, the authors ascertained the frequency of TP associations, in which detection of indicators corresponded with detection of pathogens; and TN correlations, where non-detection of indicators concurred with non-detection of pathogens. The authors found that, at their respective limits of detection, the percentage of results correctly predicted by indicators did not exceed 50%. Although discriminate analysis, involving clustering of all six indicator organisms, did provide predictive power, no single indicator could accurately categorize the “presence” or “absence” of pathogens (Harwood et al., 2005).

Despite some limitations, the analyses by Harwood et al. constitute an important step in the evaluation of indicator-pathogen relationships. Harwood et al. assessed the efficacy of indicator species in predicting the presence or absence of pathogens at one single threshold: the detection limit of the indicator. At this specific threshold, they could not find a predictive relationship between any combination of indicators and pathogens. However, the use of the indicator limit of detection as an important threshold of concern is not justified. The safety of water—whether it is swimming water, reclaimed water, or even drinking water—is not determined at the detection limit. Such a stringent safety criteria would be far too restrictive since, as Harwood points out, the prevalence of indicators is many orders of magnitude higher

than pathogens—indeed, indicators are frequently detected even when fecal contamination is absent (Harwood et al., 2005). With regard to swimming waters, public safety is assessed at the EPA national water quality criteria: 30-day geometric means and SSM indicator values ranging from 33 to 161 CFU per 100 mL. The results of Harwood et al. (2005) only show that, at the detection limit (0.2 CFU per 100 mL for *E. coli* and enterococci) the predictive power of indicator species with respect to pathogens is as good as the toss of a coin, 50–50. At that threshold, indicators do not provide information about pathogen presence or absence. However, the study inadvertently raises an important question. Are indicator species able to predict pathogens correctly at *any* threshold? If so, what is that threshold?

1.6 The Receiver Operating Characteristics Curve

Diagnostic tests evaluate two separate distributions—the distribution of observations in which a test condition is confirmed as present and the distribution of observations in which the test condition is confirmed as absent. The distribution of confirmed positive values will either be correctly identified by the test as positive (true positive, TP) or incorrectly identified as negative (false negative, FN). Similarly, the distribution of confirmed negative values can be correctly classified as negative (true negative, TN) or incorrectly identified as positive (false positive, FP). In the context of water quality, a TP response reflects conditions in which a positive FIB classification (i.e., water is not swimmable, polluted) matches the detection of a pathogen. The proportion of correct positive classifications taken over the total number of TP observations is often referred to as sensitivity. On the other hand, a TN response reflects conditions where a negative FIB classification (i.e., water is swimmable, clean) concurs with a

pathogen non-detect. The proportion of correct negative classifications taken over the total number of TN observations is often referred to as specificity.

Relative proportions of TP, FN, TN, and FP can be understood as accuracy—these parameters change depending on the threshold of classification. For a "good" diagnostic test,[7] higher thresholds for positive classification will make it increasing difficult to misclassify TN values as positive. This is because very few observations in the distribution of TN values will yield an exceedingly high test result. As the rate of FN classification decreases, the rate of TN classification will increase and specificity will improve. However, this increase in specificity comes at a cost. When specificity alone is maximized, sensitivity is minimized. A high threshold for positive classification will necessarily misclassify TP values as incorrectly negative. This exchange between sensitivity and specificity is also true if the situation were reversed. Decreasing the threshold for positive classification will decrease the frequency of FP classifications while TP classifications will be maximized, and sensitivity will be bought at the cost of specificity. Therefore, sensitivity and specificity must be maximized as a pair in order to find an optimal threshold.

A perfect test will never yield a TN test value high enough to be considered positive—there will be no FPs (false alarms) (Metz, 1978). Conversely, there will not be TP observations with test values low enough to be considered negative—there will be no FNs (misses) (Metz, 1978). Stated another way, in a perfect test, the two distributions of TP and TN values will not overlap with regard to test criteria (Metz, 1978). Such a test rarely exists. In the case of

[7] Given a test that is only as good as chance at correctly identifying a true condition, changing the threshold will not change sensitivity and specificity. Both will remain at 0.50, indicating that the likelihood of correct classification equals the likelihood of incorrect classification.

bacterial indicators of water quality, empirical evidence strongly suggests that there will be some overlap in the diagnostic results for the two distributions. Some portion of the TN distribution will yield uncharacteristically high values; conversely, the distribution of TP values may contain observations where the diagnostic test reported uncharacteristically low values. The inherent integrity of a given test is established by the amount of overlap between these two distributions and cannot be altered (Metz, 1978). However, a test can be improved by evaluating rates of sensitivity and specificity at all diagnostic thresholds concurrently and selecting the threshold which performs the best (Metz, 1978).

The study by Harwood et al. assessed the sensitivity and specificity of indicator species at a single point. However, it is possible to evaluate the sensitivity and specificity of a test at any threshold (Metz, 1978). At each threshold, proportions of TP, FN, TN, and FP positive values will change. Using these proportions, the performance of a test can be thought of as the rate of TP responses versus the rate of FP responses (negative values can also be used) (Metz, 1978). The two rates can be conveniently graphed as X-Y coordinates for every possible threshold of a given test resulting in a curve which stretches from approximately (0, 0) to (1, 1) (Metz, 1978). This graphical plot is known as a receiver-operating characteristics (ROC) curve (Metz, 1978).

2. OBJECTIVES

The primary objective of this study is to evaluate the diagnostic strength of microbial indicators of water quality with respect to the presence or absence of protozoan pathogens. The objective will be obtained using data from the Chicago Health Environmental Exposure and Recreation Study (CHEERS). CHEERS data include densities of the indicator microbes, enterococci, *E. coli*, somatic coliphage, and F+ coliphage and the pathogens *Cryptosporidium parvum* and *Giardia lamblia* (heretofore referred to as *Cryptosporidium* and *Giardia*, respectively). The approach uses receiver-operating characteristics (ROC) curve analysis, which compares the density of indicator microbes to the probability of pathogen presence. The area under the curve (AUC) is a quantitative measure of the ability of the indicator microbe to differentiate TN (pathogen absence) and TP (pathogen presence) conditions.

The primary objective is supported by four specific aims:

- Aim A. Statistically test for differences in diagnostic value of an indicator microbe for *Giardia* and *Cryptosporidium*.
- Aim B. Quantify diagnostic value of the indicator microbes as a function of environmental and hydrometerologic conditions. Specifically, each indicator-pathogen pair relationship will evaluated with respect to (1) location specific differences between the Chicago Area Waterways System (CAWS) and General Use Waters (GUW); (2) temporal variations between AM and PM measurements; (3) DRY and WET weather conditions; (4) BRIGHT and DARK sunlight conditions; (5) position with respect to wastewater treatment plants (WTP) (ABOVE vs. BELOW); and finally, (6) the presence or absence of a combined-sewer overflow (CSO) event.

- Aim C. Statistically test for differences in diagnostic value of the indicator microbes for protozoan pathogen presence as a function of environmental and hydrometerologic conditions.
- Aim D. Identify the indicator microbe density threshold that maximizes sensitivity and specificity of pathogen presence and absence prediction.

To the author's knowledge, no research has evaluated the diagnostic value of indicator species with regard to pathogens using ROC analysis. We hope to fill a knowledge gap about the use of ROC analysis to evaluate the diagnostic value of tests of water quality and the importance of existing water quality standards.

3. SETTING

The Chicago Area Waterways System (CAWS) is a 126km heavily engineered waterway which connects parts of the Chicago River to the Calumet Rivers via the Cal-Sag Channel and diverts these waters into the Lower Des Plaines; ultimately, the flow from these waterways drain into the Mississippi River Basin. Effluent from three wastewater treatment plants (WTPs) in the Metropolitan Water Reclamation District of Greater Chicago (MWRDGC) dominates the CAWS flow, making up 50-100% of the flow volume (Rijal et al., 2009; Rijal et al., 2011). This effluent begins as a combination of raw sewage and rainwater runoff in the Chicago area's combined sewers and undergoes primary sedimentation treatment as well as secondary aerobic digestion at the WWTPs; however, this wastewater it is not disinfected before discharge into the CAWS. Effluent from the WWTPs can contain pathogens concentrations orders of magnitude greater than general use waters (GUW) like Lake Michigan, which do not receive wastewater discharge. As a result, public access to CAWS waterways are limited to secondary contact recreational activities (e.g. boating, canoeing, kayaking) and indigenous wildlife, while GUW locations are open to full contact recreation (e.g. swimming).

Increasing recreational use of the CAWS prompted the Chicago Health Environmental Exposure and Recreation Study (CHEERS), a prospective cohort study designed to determine health risk from secondary contact recreation in the CAWS (Dorevitch, 2011). From August 2007 to July 2009, CHEERS surveyed more than 11,000 participants in order to compare health outcomes between those who engaged in limited contact recreation on the CAWS and those participants who had contact with the GUW, as well as control groups which had no water contact (Dorevitch, 2011). In order to determine a causal relationship between health

outcomes and microbial water quality, the density of indicator microbes and protozoan pathogens were measured on days and at locations in which participants engaged in water recreation. Sampled areas on the CAWS included the Cal-Sag Channel, the North and South Branches of the Chicago River, the Main Stem of the Chicago River, and the North Shore Channel. Samples were also collected at 6 hour intervals at locations approximately 3km above and 3km below the North side and Calumet WTPs on days when participants were recruited at any location in the CAWS. The GUW sites included Chicago beaches as well as the Des Plaines, DuPage, and Fox Rivers; inland freshwater bodies like Busse Lake and Tampier Lake as well as the Skokie Lagoons were also designated GUW sites. Certain sites that did not clearly fall into either category were labeled as "Other." Altogether, there were 18 locations identified as CAWS, 39 identified as GUW, and 4 identified as OTHER. In this study microbial water quality data from CHEERS is used to explore the relationship between water quality indicator species and the pathogens *Giardia* and *Cryptosporidium*.

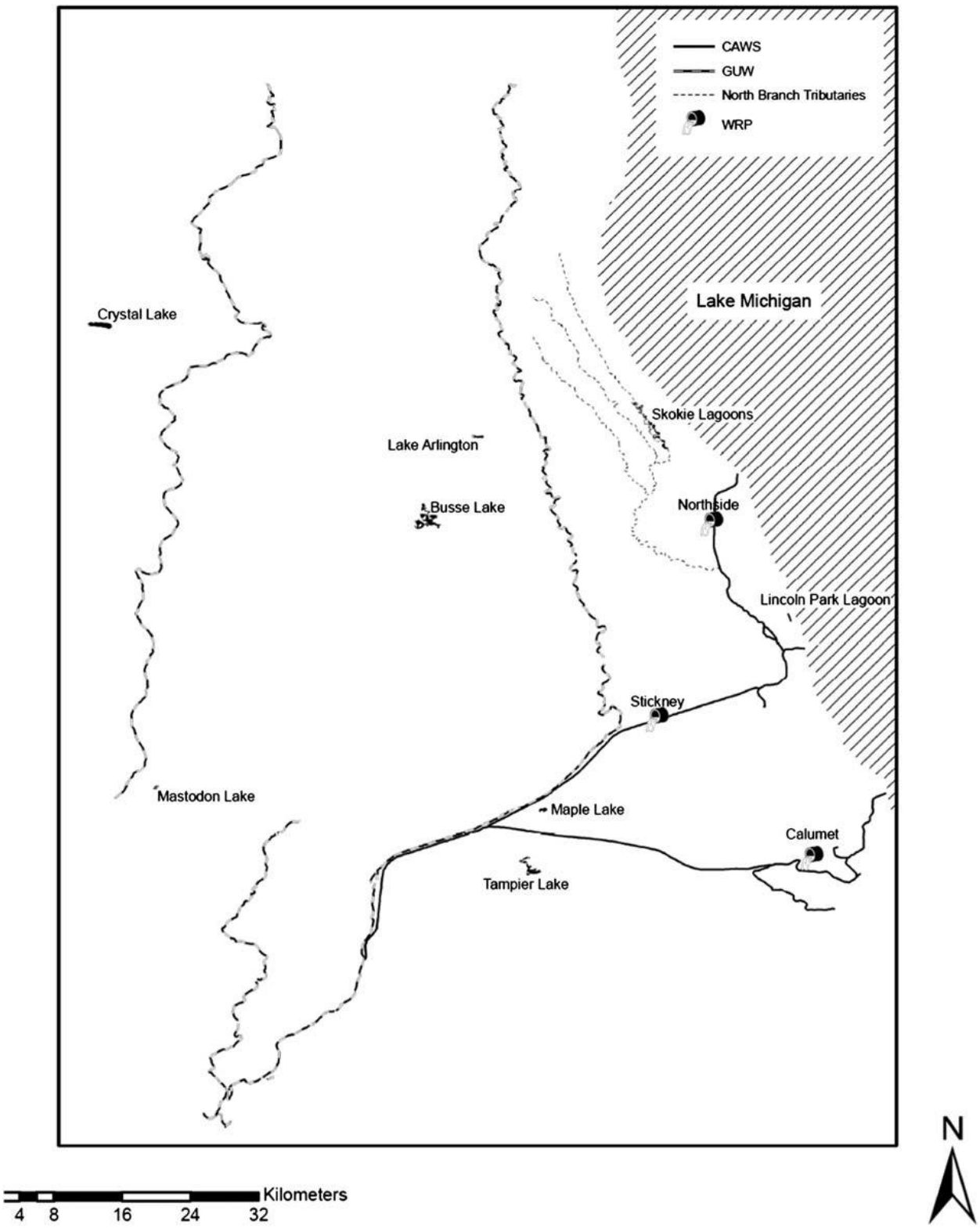

Figure 1. CHEERS sampling locations.

4. METHODS

4.1 Sampling

Due to weather constrictions, sampling was only performed during the summer season–typically from May until September. Water collection points at each location were determined by available access points such as piers, boat launches, and shallow beach water. Sampling procedures followed US EPA guidelines for direct "grab" sampling and large-volume sampling. Direct sampling of surface water (at a depth of 10cm) was performed by holding a sterile 2 liter container in the prevailing current at a 45 degree angle; these samples were collected at 2-hour intervals, 1–4 times per day. While direct sampling provided a large enough volume to conduct assays of indicator microbes, greater sample volumes were necessary to capture pathogens like *Giardia* and *Cryptosporidium*. Continuous flow centrifugation (CFC)—a method that concentrates protozoan parasite (oo)cysts—was used to capture *Giardia* and *Cryptosporidium* at quantifiable concentrations (Zuckerman and Tzipori, 2006). This process was performed 1–2 times per day at 6-h intervals, depending on the duration of participant recruitment in the epidemiologic study. Analyses for indicators species and pathogens were not be performed on the same water sample. Therefore, water for indicator analysis and water for pathogen analysis, though taken on the same day and in the same location, were acquired separately. While direct samples were generally collected within a one-hour window of large-volume sampling, not all observations for indicator species could be matched to a pathogen concentration within a two-hour window at the same day and in the same location.

Differences in location group were evaluated between samples taken on the CAWS (N=195) and GUW (N=98). Samples taken from the location group OTHER were excluded from

location-group analysis. Time-of-day comparisons were made between samples taken during AM and PM hours, although no AM-PM comparisons were made between samples collected on the same day and in the same location. While the original sample set contained N=324 observations, data for rainfall, CSO events, and sunlight were available for only N=309 observations. Inches of rainfall in 24, 48, 72, and 96 hours prior to each round of sampling were obtained from the Illinois State Water Survey (http://www.isws.illinois.edu/data.asp). Samples were categorized as WET if they were collected within 96 hour of a rain event and DRY if rain had not occurred in the 96 hours prior to sample collection. Data reported by the MWRDGC to the Illinois EPA provided information about CSO events. Samples were categorized as CSO if it was collected within 96 hours of a CSO event and NO CSO if there was no CSO discharge in the 96 hours prior to sample collection. An overflow event (active or passive) anywhere in a channel (either in the CAWS or OTHER subgroup, N=226) was defined as a CSO event for the channel. Samples dichotomized as either ABOVE or BELOW a WTP (CAWS only, N=195). Finally, cumulative solar radiation was measured by the Illinois EPA at the Jardine air monitoring station. Samples were labeled as BRIGHT if they were collected at a time when cumulative sunlight exceeded 1.86 (Langleys/minute)—that value corresponds to the 25^{th} percentile for solar irradiation measurements; otherwise, the samples were categorized as DARK.

4.2 Microbiological Analysis

Microbial analyses were conducted using established methods for the pathogens *Giardia* and *Cryptosporidium* as well as indicator species *Escherichia coli* and enterococci, somatic coliphage, and male coliphage. All sampling and data analysis techniques were subject

to quality-control measures which included field blanks, field splits, and spiked samples for recovery. Field blanks for *E. coli* and enterococci were taken at every sampling location by filling a standard collection bottle with two liters of sterile buffer; this was done to insure that water samples were not inadvertently contaminated on site. Additionally, each two-liter sample for these indicators was each split among three containers. Two splits were analyzed identically to evaluate methodological precision. The third split was spiked in the field with a certified quantity of indicator cells in the form of BioBalls (BTF, Pty, Ltd., North Ryde, Australia). For protozoan pathogens *Giardia* and *Cryptosporidium*, one blank sample per sampling date was prepared at the University of Illinois—Chicago (UIC) School of Public Health water laboratory by running 20 L of sterile water through the CFC system. Ten percent of samples were collected as splits. The first split was processed as a sample and the second split was spiked with a known quantity of (oo)cysts. Also at the UIC water laboratory, samples for coliphage analysis were spiked by pipetting 1 mL spike material for male-specific F+ coliphage and 1 mL for somatic coliphage into the 500 mL sample bottle. These spike materials for coliphage were prepared by Scientific Methods, Inc. (Granger, Indiana) and contained exact concentration levels and expiration dates. Testing laboratories were blinded as to the spike density.

After collection, all water samples were put on ice and transported to a commercial laboratory[8] for analysis. Following EPA methodological requirements, *E. coli* and enterococci were held for no longer than 6 hours before analysis. For these indicators, 87% of the 5,430 samples acquired met this requirement. Once at the testing laboratory, microbiological analysis of *E. coli* was performed according to US EPA Method 1603 with modified membrane-

[8] Indicator samples were analyzed by Microbac, Inc. in Merrilville, Indiana while pathogens were sent for analysis to Scientific Methods, Inc. in Granger, Indiana.

Thermotolerant *E. coli* (mTEC) agar (BD,product # 214880) (US EPA, 2006). Enterococci were analyzed using US EPA method 1600 with modified membrane-Enterococcus Indoxylb-D-Glucoside (mEI) agar (BD, Product 214881) (US EPA, 2006). In order to generate at least one plate in the methodological detectable range, at least five dilutions for both indicator species was performed. The limit of detection for both bacterial indicators was 1 CFU per 100 mL.

For coliphage samples, holding time before analysis was restricted to 48 hours according to EPA requirements. Of the 3,534 coliphage samples acquired, 95% met this requirement. Both F+ coliphage and somatic coliphage were analyzed according to US EPA method 1602 (US EPA, 2006). The limit of detection for F+ coliphage was 1 organism per 100 mL while the limit of detection for somatic coliphage was 10 organisms per 100mL.

Samples were prepared for qPCR analysis at facilities in the University of Illinois—Chicago. Under a laminar flow hood, 100 mL aliquots of each sample were dispensed into disposable filter funnels and vacuum filtered through 47-mm Millipore™ Isopore™ polycarbonate membrane filters, pore size 0.4 µm. Using sterile forceps, the filters were folded and inserted into 2 mL extraction tubes pre-loaded with 212–300 mm Sigma acid-washed glass beads (Cat# S0205-50, GeneRite). The tubes were stored at -80°C until they could be transported to Mycometrics, Inc (Monmouth Junction, New Jersey) for analysis by real-time qPCR. *Enterococci* were analyzed according to EPA Method A (US EPA, 2010). The analysis of *E. coli* by qPCR followed the methodology described by Chern and colleagues (Chern et al. 2011).

Calibrator cells for enterococci and *E. coli*—ATCC 29212 and ATCC 25922, respectively—were purchased from American Type Culture Collection (ATCC) and cultured on site at Mycometrics. Calibrator stocks were adjusted such that a 10 µL aliquot of enterococci would

contain 1 x 10^5 cells while a 10 µL of *E. coli* stock would contain 5 x 10^5 cells (Cat# 0205-50, GeneRite, New Jersey). Each time a set of unknowns were processed, 3 calibrator samples for each organism was prepared by spotting 10 µL aliquots of stock solution directly onto a clean polycarbonate filter, pre-loaded into bead tubes. Each of the three calibrator standards was run in duplicate for a total of six reactions and served as positive controls. Method blanks (negative controls) were prepared by filtering 100 mL of sterile water through a polycarbonate filter. In preparation for qPCR, 590 µL of extraction buffer was added to each tube containing a filter (unknowns, calibrator samples, and method blank). Sample buffer contained a quantity (0.2 mg mL^{-1}) of salmon testes DNA to serve as a sample processing control (SPC) (Sigma-Aldrich #D1626) to account for extraction efficiency and inhibition of the qPCR amplification.

First, the tubes were disturbed to allow glass beads to lyse the cells and release total DNA. The resulting crude homogenate was centrifuged at 12,000 x g for 1 min to precipitate the filter and any cellular fragments allowing 400 µL of supernatant to be transferred to clean, low-DNA binding, 1.7 mL microcentrifuge tubes (Cat# S1203-17-PK, GeneRite, New Jersey). This supernatant was again centrifuged at 12,000 x g for 5 minutes and 350 µL of clarified supernatant was transferred to another 1.7 mL tube. A 50 µL of aliquot of this clarified crude extract was removed from each sample and diluted five-fold in 200 µL of elution buffer (Cat# S2401-50, GeneRite, New Jersey). This final sample of DNA—spiked with internal standard, clarified, and diluted—was analyzed by qPCR.

In qPCR, the point at which target genomic DNA becomes detectable is known as the cycle threshold (CT). The final quantitation of an unknown target effectively involves two normalizations of this CT value. For every processed sample, the recovery of target DNA is

normalized against the internal standard by subtracting the CT value for the SPC from the CT value for the target sequence—this is ΔCT. The second normalization is achieved by subtracting the ΔCT value of the calibrator cells from the CT value of the unknown sample—this is ΔΔCT. That value (ΔΔCT) represents the normalized difference in the number of CTs between an unknown sample and a calibrator. The absolute number of calibrator cells represented by this value is found by raising the amplification factor (AF) to - ΔΔCT. For a perfect cycle of polymerase chain reaction, the absolute quantity of DNA should exactly double and the AF would equal 2. In a real laboratory environment, this perfect amplification is not achieved; AF is a value slightly less than 2 and is determined from the slope of a calibration curve. The value $AF^{-\Delta\Delta CT}$ is multiplied by the average number of calibrator cells and the final number of cells determined to be in a given unknown sample is reported as calibrator cell equivalents (CCE). The AF value for the target (enterococci) and reference (SPC) were roughly equal in this analysis (96% and 93%, respectively). Samples that showed substantial inhibition (that is, the difference between the CT values for the SPC between an unknown and calibrator sample was greater than or equal to 3) were excluded from data analyses.

For protozoan samples that time was 72 hours and 99% of 1,292 protozoan pathogen samples arrived in less than 72 hours. Analyses of *Giardia* and *Cryptosporidium* (oo)cysts were performed according to US EPA Method 1623 (US EPA, 2005). Twenty-liter samples were centrifuged at UIC, and CFC samples were sent out for analysis. In order to concentrate and purify (oo)cysts, pellets were processed by immunomagnetic separation with Dynabeads GC-Combo (Dynal Cat. No. 730.12). After being placed on slide wells, the (oo)cysts were stained and fluorescently-labeled with Aqua-Glo (TM) monoclonal antibodies G/C Direct (Waterborne

Inc., Cat. No. A100FLR, New Orleans, Louisiana). Fluorescent and differential contrast microscopy allowed identification of (oo)cysts based on size, shape, color, and morphology. The limit of detection for both *Giardia* and *Cryptosporidium* was one (oo)cyst per 20 liters, as one (oo)cyst could be reported for a 20L sample.

4.3 Quantitative Methodology

General statistical analyses were performed using SAS software version 9.2 (SAS Institute Inc., Cary, North Carolina). Detection frequencies of pathogens were calculated for each indicator test and within each category strata. Arithmetic means, geometric means, and medians and other descriptive measures were calculated for every indicator within the general analysis and for each category of analysis. Statistical differences in categorical strata were calculated between median values using the Wilcoxon-Mann-Whitney test, shown in Table V. Finally, Spearman rank correlation coefficients were calculated for all pairs of indicators and pathogens. Statistical significance was evaluated at α=0.05 level except in multiple comparisons in which a family-wise error rate was calculated using the Šidák correction α=0.000427.

MedCalc software was used for ROC analysis (MedCalc 12.3.0.0 for Windows, MedCalc Software, Mariakerke, Belgium). A ROC curve is the plot of "sensitivity" (or TP rate) against "1-specificity" (or FP rate) at every available threshold. Bamber (1975) describe the equivalence of the AUC with the Mann-Whitney statistic applied to the two samples $\{X_i\}$ and $\{Y_j\}$. Supposing that X_i represents that set of TN values ranging from i=1 to m and Y_j represents the set of TP values from j=1 to n, the Mann-Whitney statistic is a generalized U-statistic that can be calculated as follows by a kernel function:

$$\widehat{\emptyset} = \frac{1}{mn} \sum_{j=1}^{n} \sum_{i=1}^{m} \boldsymbol{\varphi}(\boldsymbol{X_i}, \boldsymbol{Y_j}), \qquad \text{(Equation 1)}$$

where,

$$\varphi(X, Y) = \begin{bmatrix} 1 & Y < X \\ \frac{1}{2} & Y = X \\ 0 & Y > X \end{bmatrix}$$

To calculate $\widehat{\emptyset}$, TN[9] observations (with regard to either *Giardia* or *Cryptosporidium*) are ranked from lowest to highest based on the test results for a given indicator. A rank score is assigned to every single observation in the TN dataset depending on the relative magnitude of that TN value with regard to each observation in the TP dataset. Every observation in the TP dataset that is higher than the TN value adds a score of 1, every observation that is equal to the TN observation adds a score of ½, and every TP observation lower than the TN observation adds a score of 0. For instance, if a TN observation is lower than 8 TP observations, tied with 1 TP observation, and higher than 2 TP observations, that TN observation would be assigned a rank score of 8.5. If the next higher TN observation is lower than 7 TP observations, tied with 2 TP observations, and higher than 1 TP observation that TN observation is assigned a rank score of 8, and so on. The sum of the rank scores divided by the value "mn" is the probability that a given diagnostic test can correctly rank a TN value lower than a TP value ($\widehat{\emptyset}$). A perfect classification test would score *m* TP cases correctly for every TN value *n*; thus, the probability of correct classification would be 1. The Mann-Whitney U statistic has been shown to be equal to

[9] The classification score can be tabulated for correct classification of positive or negative.

the area falling under the points comprising an empirical ROC curve when calculated by the trapezoidal rule (Bamber, 1975).

In the present analysis, statistical comparisons between AUC values are conducted for (1) each indicator between the two pathogens; (2) each indicator between categorical strata; and (3) between different indicators tests for the same pathogen. Comparisons between data are not mathematically justified without consideration of inter-test correlation as determined by Kendall Tau b analysis (DeLong et al., 1988). For independent sample sets, this correlation is minimal and does not require attention when comparing the AUC. The AUC values for a given indicator between two different pathogens are considered independent values as are AUC values of each indicator test between categorical strata. For these analyses, the Hanley and McNeil method is used to calculate the standard error of the AUC (Hanley and McNeil, 1982). This statistical test does not take into consideration inter-test Kendall Tau b correlations (Hanley and McNeil, 1982).

Diagnostic tests performed on the same population of samples are correlated; hence, the data are not independent. DeLong et al. (1988) presented a method for calculating the correlation between pairs of areas from an estimated correlation matrix (ECM). The correlation coefficient is used to transform the difference in AUCs between a pair of tests into a z-statistic. For large samples sizes, the z-distribution approximates the student's *t*-distribution, and z-statistic can be interpreted as the student *t*-statistic (DeLong et al., 1988). The z-statistic, then, is used to test the null hypothesis that there is no difference between the areas under the curve for the two tests being studied (DeLong et al., 1988). DeLong method is used for the calculation of the standard error of the AUC for the general analysis and for comparisons between

dependent sample groups (DeLong et al., 1988). The DeLong method is limited by the fact that, for inter-test comparison, all data must be paired perfectly (DeLong et al., 1988). In other words, sample sizes must be equal for any given comparison. This does not present a problem for the comparison between the culture based assays and coliphages since each diagnostic test was applied to the same 324 samples, all matched for date, location, and hour. However, indicator comparisons which included qPCR assays were limited to only those observations that could be matched with every sample in those assays groups.

Binomial exact confidence intervals for the AUC are calculated for the general analysis. Finally, the optimal diagnostic cutoff is determined for each indicator-pathogen pair in each category using the Youden Index (Youden, 1950). For the above analyses, statistical significance was evaluated at $\alpha=0.05$ level except in multiple comparisons in which a family-wise error rate was calculated using the Šidák correction $\alpha=0.000427$.

5. DESRCIPTIVE RESULTS AND GENERAL RELATIONSHIPS BETWEEN MICROBES

5.1 Data Set Preparation and Independence of Observations

From August 2007 to July 2009, CHEERS measured water quality in the Greater Chicago region, including the CAWS and water bodies designated as GUW (Dorevitch, 2011). The original CHEERS data set comprised 2,389 date-location-hours of water quality observations, of which only 446 observations included measured *Giardia* and *Cryptosporidium* (oo)cyst densities and valid culture-based measurements of *E. coli, enterococci*, *somatic, coliphage*, *and F+ coliphage* collected within a 2-hour window of the pathogen measurement. This analysis is limited to those 446 observations since the present analysis focuses on the ability of the indicator bacteria to predict protozoan pathogens. The 2-hour window for linkage of indicator bacteria measurements to pathogen measurements was implemented to preserve sample size. Within these 446 observations, 122 days contained more than one (usually only two) measurements taken at the same location (Dorevitch, 2011). Clustered data points are not necessarily independent and may disproportionately impact the area under the curve as compared to independent data points. As a result, microbe densities measured two or more times at the same location on the same day were averaged in order to avoid data clusters. This averaging left 324 observations available for the final analysis. In addition, *E. coli* and enterococci densities were measured by qPCR at the same day and location, and within 2-hours of protozoan pathogen measurements in 79 and 187 instances, respectively.

5.2 Descriptive Statistics

The final dataset of 324 observations are taken from 42 unique locations. Samples were obtained from CAWS locations in 195 instances, while GUW measurements were made 98 times. The location group OTHER accounted for the final 31 observations. Hence, the AUC reflects the variability in sampling locations and hydrometerologic conditions. While relationships between indicators and pathogens might be weaker in this cumulative analysis, the AUC reflects this heterogeneity. Detection frequencies for both pathogens are shown in Table I. The distributions of detected and non-detected pathogens determine the TP and TN classification rates at each indicator threshold and are important considerations in ROC analysis. Samples sizes for culture-based analysis (enterococci by culture, *E. coli* by culture, somatic coliphage, and F+ coliphage) are equal and have identical detection frequencies for the two pathogens. Densities of enterococci and *E. coli* as determined by qPCR analysis are only available for a smaller sub-set of observations and detection frequencies of pathogen values in this sub-set differ accordingly.

Table I
DETECTION FREQUENCIES OF PATHOGENS FOR EACH INDICATOR

	Enterococci (cx)	*E. coli* (cx)	S. Coliphage	F+ Coliphage	Enterococci (qPCR)	*E. coli* (qPCR)
N	324	324	324	324	187	79
Giardia (+)	253	253	253	253	141	60
Giardia (-)	71	71	71	71	46	19
Detection (%)	78.1	78.1	78.1	78.1	75.4	75.9
Crypto. (+)	168	168	168	168	68	24
Crypto. (-)	156	156	156	156	119	55
Detection (%)	51.9	51.9	51.9	51.9	36.4	30.4

Descriptive statistics of the indicator microbes are listed in Table II. The geometric mean concentrations are back-transformed after logarithmic transformation. The minimum values reflect detection limits for each of the organisms—one tenth the lowest detectable concentration for the organism. For all microbes, the mean densities are greater than the median densities, indicating that the data are not normally distributed. This finding was confirmed by the D'Agostino-Pearson, Chi-square, Kolmogorov-Smirnov tests of normality for all six indicators and both pathogens. This is consistent with the general finding that the concentrations of microbes and pollutants in the environment are described by a lognormal distribution. The coefficient of variation (CV) suggests moderate differences in the variability for each microbe. Indicator bacteria measured by qPCR are present in higher concentrations; this finding makes sense in light of the fact the qPCR quantifies genomic targets from both viable and non-viable bacteria while culture methods only quantify the former.

Table II

DESCRIPTIVE STATISTICS FOR CULTURE-BASED ASSAYS OF ENTEROCOCCI AND *E. COLI*, SOMATIC COLIPHAGE, F+ COLIPHAGE, AND QPCR-BASED ASSAYS OF ENTEROCOCCI AND *E. COLI*

	Enterococci (cx)	*E. coli* (cx)	S. Coliphage	F+ Coliphage	Enterococci (qPCR)	*E. coli* (qPCR)
Unit	CFU	CFU	PFU	PFU	CCE	CCE
N	324	324	324	324	187	79
Minimum	0.100	0.100	1.00	0.100	0.100	0.100
Maximum	18500	29300	37300	779	130000	1190000
A. Mean	614	1930	719	30.6	10800	35400
G. Mean	140	271	42.8	1.72	2330	1760
Median	176	301	48.2	1.00	3800	3500
SD	1600	4100	2330	77.1	18200	147000
CV	261	212	324	252	168	415

5.3 Arithmetic Mean, Geometric Mean, and Median Values of Indicators

Diagnostic validity of the six available indicators was examined in the context of several important variables: location group (CAWS vs. GUW), time (AM versus PM), rainfall (DRY vs. WET), position with respect to wastewater treatment plants (WTP) (ABOVE vs. BELOW), intensity of sunlight (BRIGHT vs. DARK), and the presence or absence of a CSO event (CSO versus NO CSO). Table III shows the arithmetic mean of indicator values for each categorical pair.

Table III
ARITHMETIC MEANS OF SIX INDICATOR TESTS FOR THE GENERAL ANALYSIS AND EACH CATEGORICAL STRATA.

Category	Enterococci (cx)	*E. coli* (cx)	S. Coliphage	F+ Coliphage	Enterococci (qPCR)	*E. coli* (qPCR)
	CFU	CFU	PFU	PFU	CCE	CCE
ALL	614	1930	719	30.6	10800	35400
CAWS	617	2780	1060	47.4	14300	63600
GUW	594	243	88.0	3.93	3260	1930
AM	929	2570	832	36.1	14900	46300
PM	395	1490	641	26.7	8850	31700
DRY	478	663	405	10.5	10500	9720
WET	625	2350	838	36.6	11300	54000
ABOVE	218	756	64.2	3.25	9750	16800
BELOW	786	3640	1480	66.1	16200	82300
BRIGHT	389	1790	716	27.0	11400	35900
DARK	1080	2670	801	38.7	9180	50700
CSO	1100	3520	829	56.4	12600	148000
NO CSO	407	2280	1070	35.7	14800	19100

Statistical differences were evaluated between geometric means of categorical strata (indicated in bold, Table IV). The geometric mean of each indicator was higher in CAWS samples than in GUW samples. Although measurements taken in the AM and PM showed fewer significant differences for indicators, differences between DRY and WET conditions as well as ABOVE and BELOW a WTP were significant for every indicator except enterococci by qPCR.

Table IV

GEOMETRIC MEANS OF SIX INDICATOR TESTS FOR THE GENERAL ANALYSIS AND EACH CATEGORICAL STRATA,WITH STATISTICAL DIFFERENCES DETERMINED USING THE T-TEST FOR TWO SAMPLES WITH UNEQUAL VARIANCES

Category	Enterococci (cx)	*E. coli* (cx)	S. Coliphage	F+ Coliphage	Enterococci (qPCR)	*E. coli* (qPCR)
	CFU	CFU	PFU	PFU	CCE	CCE
ALL	140	271	42.8	1.72	2330	1760
CAWS	**199**	**650****	**110**	**4.41**	**6000**	**10200**
GUW	**48.3**	**34.5**	**3.74**	**0.26**	**237**	**141**
AM	**195****	324**	49.5**	1.89**	**4400****	2380
PM	**111**	240	38.7	1.62	**1710**	1590
DRY	**46.2****	**93.1****	**14.5****	**0.658**	1410**	**600****
WET	**176**	**368**	**57.2**	**2.18**	2550	**3490**
ABOVE	**90.8**	**121**	**5.04**	**0.248**	3610**	**8440**
BELOW	**277**	**1330**	**405**	**14.9**	7440	**11000**
BRIGHT	**111****	246**	42.0**	1.61**	2090**	1590**
DARK	**199**	346	36.0	1.55	3210	3980
CSO	**312**	804**	152**	5.81**	6240	**24100****
NO CSO	**190**	609	112	3.32	6160	**5390**

*Statistically significant differences at α=0.05 are indicated in bold.

**Variances are not statistically different by the Folded F-test and pooled t-test is used.

Finally, the Wilcoxon-Mann-Whitney test was used to evaluate the differences in median values for categorical pairs, shown in Table V.

Table V

COMPARISONS OF MEDIAN BETWEEN CATEGORIES OF SAMPLING CHARACTERISITICS*

Category	Enterococci (cx)	*E. coli* (cx)	S. Coliphage	F+ Coliphage	Enterococci (qPCR)	*E. coli* (qPCR)
	CFU	CFU	PFU	PFU	CCE	CCE
ALL	176	301	48.2	1.00	3800	3500
CAWS	**201**	**862**	**293**	**8.90**	**9800**	**10200**
GUW	**51.6**	**31.6**	**1.00**	**0.100**	**363**	**747**
AM	**188**	310	50.0	1.00	**7100**	8220
PM	**156**	297	47.4	1.00	**3030**	3200
DRY	**42.7**	**115**	**6.88**	**0.195**	4520	**2830**
WET	**217**	**465**	**160**	**1.00**	3870	**6100**
ABOVE	**94.8**	**152**	**2.47**	**0.100**	6100	5790
BELOW	**333**	**1930**	**650**	**25.0**	11500	12000
BRIGHT	**155**	310	55.0	0.937	3450	3340
DARK	**180**	260	20.0	1.00	6750	7800
CSO	**368**	915	378	7.42	7000	**17800**
NO CSO	**194**	674	274	4.03	9990	**6100**

*Statistically significant differences at α=0.05 are indicated in bold.

5.4 Spearman Rank Correlation between Absolute Densities of Indicators and Pathogens

Spearman rank correlation coefficients were calculated between all six indicators and pathogens (Table VI). Strong relationship between indicators and the pathogen *Giardia* were evident ($0.4 < p < 0.7$), particularly for the qPCR based assays of enterococci and *E. coli*. The relationship between indicators and the pathogen *Cryptosporidium* were substantially weaker ($p < 0.4$) . Between indicators, the strongest correlations were observed between the two somatic coliphages, and between the culture-based assay for *E. coli* and somatic coliphage as well as F+ coliphage.

Table VI

SPEARMAN RANK CORRELATION COEFFICIENTS BETWEEN ALL INDICATORS AND PATHOGENS[*]

		Giardia	*Cryptospo-ridium*	Enterococci (cx)	*E. coli* (cx)	Somatic Coliphage	F+ Coliphage	Enterococci (q)	*E. coli* (q)
	Units	cysts/20L	oocysts/20L	CFU/100 mL	CFU/100 mL	PFU/100 mL	PFU/100 mL	CCE	CCE
N	Units	324	324	324	324	324	324	187	79
Giardia	cysts/20L	◊	◊	◊	◊	◊	◊	◊	◊
Cryptosporidium	oocysts/20L	0.401	◊	◊	◊	◊	◊	◊	◊
Enterococci (cx)	CFU/100mL	0.417	0.292	◊	◊	◊	◊	◊	◊
E. coli (cx)	*E. coli* CFU/100mL	0.607	0.338	0.663	◊	◊	◊	◊	◊
Somatic Coliphage	PFU/100mL	0.632	0.405	0.585	0.745	◊	◊	◊	◊
F+ Coliphage	PFU/100mL	0.622	0.334	0.523	0.715	0.850	◊	◊	◊
Enterococci (q)	CCE	0.689	0.190[**]	0.492	0.638	0.564	0.535	◊	◊
E. coli (q)	CCE	0.624	0.01[***]	0.502	0.594	0.562	0.533	0.594	◊

[*] All samples statistically significant at α=0.001, unless otherwise indicated.

[**] Statistically significant at α=0.01.

[***] Not statistically significant.

5.5 Absolute Changes in Indicator Densities across Time

The finding that mean indicator microbe densities are higher in the AM hours than PM hours (Table III) is explored further in Figure 2, which compares indicator densities with hour of sampling. Microbe densities appear to be slightly lower in the late afternoon relative to mid-afternoon. As pointed by Obiri-Danso et. al. and Rosenfeld et al., sunlight can result in significant die-off in viable indicator organisms during day; after a period of die-off, absolute densities of indicators can rebound during the night (Obiri-Danso et al., 1999 and Rosenfeld et al., 2006). This decrease in the absolute densities of indicators is more apparent for the culture-based assays than qPCR assays, which quantify viable organisms.

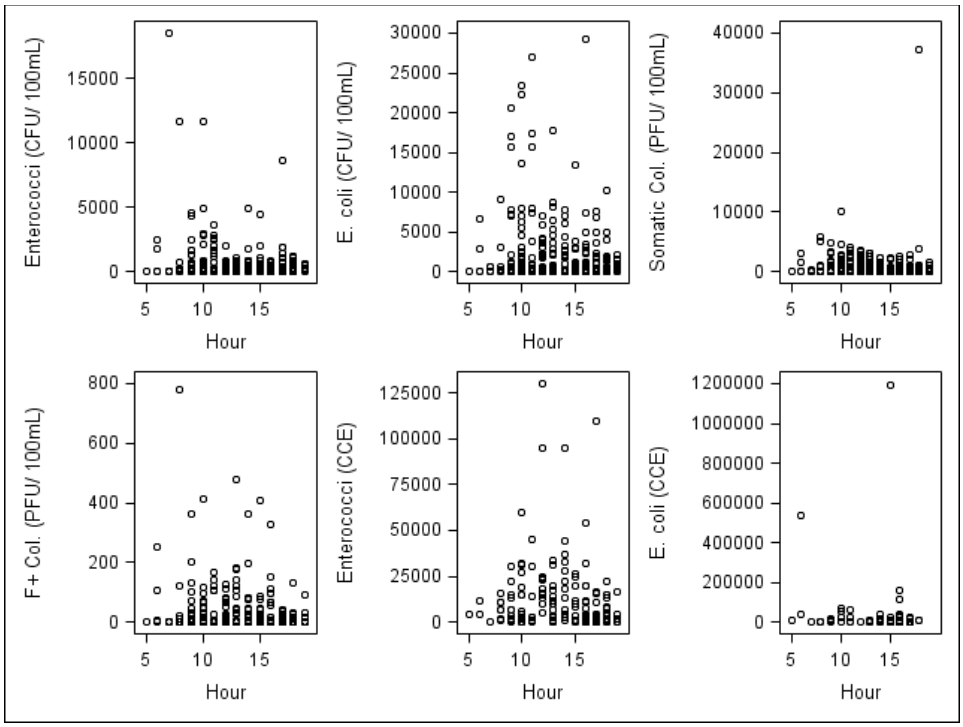

Figure 2. Absolute indicator densities plotted against sampling hour.

5.6 Absolute Changes in Indicator Densities as a Function of Rain

The finding in Tables III, IV, and V that some microbe densities are higher in WET conditions than DRY conditions is further explored in Figure 3, which compares indicator density with the log10-transformed magnitude of cumulative rainfall within the 96 hours prior to sample collection. Densities of enterococci by culture, *E. coli* by culture, F+ coliphage, and enterococci by qPCR show an upward trend with increasing amounts of rain. The effect of rain on the densities of somatic coliphage, and *E. coli* by qPCR is less clear. Significant increases in densities of FIB after a rain event have been associated with urban runoff as well as the disruption of bacteria in sediment (Noble et al., 2003; Sanders et al., 2005; Ahn et al., 2005).

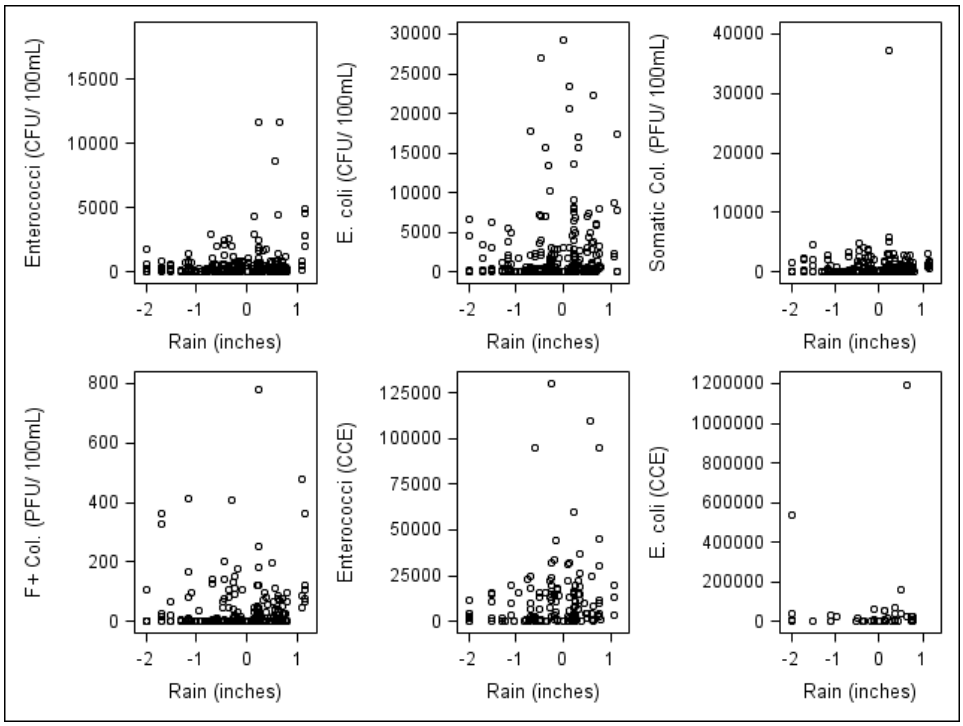

Figure 3. Absolute indicator densities plotted against log10-transformed magnitude of rain.

5.7 Absolute Changes in Indicator Densities as a Function of Combined-Sewer Overflows

Figure 4 displays indicator densities as a function of the log10-transformed magnitude of the most recent CSO event (N=226). Although there is some indication of upward trends for F+ coliphage and enterococci by culture, the scatter diagrams do not conclusively show increases in indicator density for increasing magnitudes of CSO. Water samples were diluted to quantify indicators by culture and this likely reduced the precision of measurements in the extremely high ranges observed following CSO events. Thus, indicator densities, while much higher following CSOs, may not be substantially different as a function of CSO magnitude. Moreover, magnitude data for CSO events are based on the duration of flow from a gated outfall, an imprecise measurements which limits accuracy for CSO magnitudes.

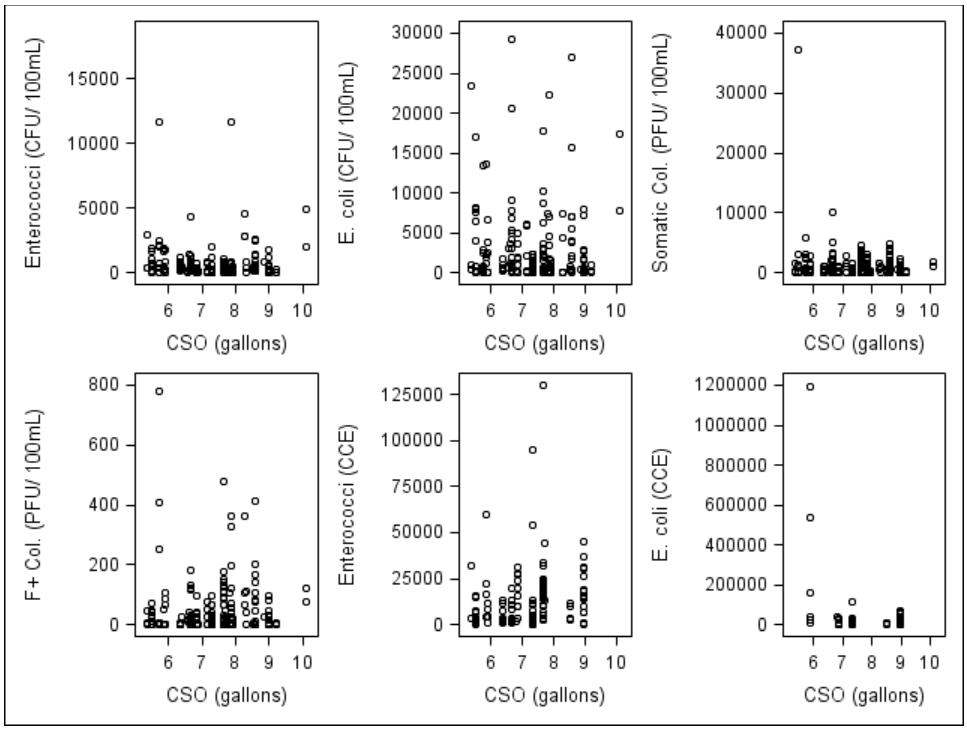

Figure 4. Indicator densities plotted against the log10-transformed magnitude of combined sewer overflow events

5.8 Detection Frequencies for *Giardia* and *Cryptosporidium*

This ROC analysis focuses on the correct prediction of the presence or absence of the protozoan pathogens *Giardia* and *Cryptosporidium*. Table VII shows the frequencies of detection for the two pathogens in each category of analysis. Detection frequencies for culture-based assays are grouped since those tests were linked to the same set of pathogen data. Detection frequencies for the two qPCR assays, available for only a limited number of observations, are listed separately. Several observations are apparent regarding the detection frequency of *Giardia* cysts and *Cryptosporidium* oocysts. First, more pathogens were detected

in the CAWS than in GUW location groups. Second, *Giardia* was detected more frequently than *Cryptosporidium*. Third, the detection of *Cryptosporidium* oocysts in the subset of samples with qPCR measures of indicators is different than the overall *Cryptosporidium* detection frequency. Fourth, although detection did vary with AM vs. PM or BRIGHT vs. DARK, (oo)cyst detection was more frequent at locations BELOW WTPs (vs. ABOVE).

Table VII

DETECTION FREQUENCIES OF *GIARDIA* AND *CRYPTOSPORIDIUM* (OO)CYSTS FOR EACH CATEGORY OF ANALYSIS, CULTURE-BASED ASSAYS ARE GROUPED WHILE QPCR ASSAYS ARE PRESENTED SEPARATELY

	GIARDIA			*CRYPTOSPORIDIUM*		
Category	% Detection for Culture-based Values	% Detection for Subset of Enterococci (q) Values	% Detection for Subset of *E. coli* (q) Values	% Detection for Culture-based Values	% Detection for Subset of Enterococci (q) Values	% Detection for Subset of *E. coli* (q) Values
ALL	78.1	75.4	75.9	51.9	36.4	30.4
CAWS	91.3	88.9	97.6	62.1	39.8	35.7
GUW	46.9	39.3	40.0	24.5	17.9	16.7
AM	82.0	75.4	90.0	62.4	39.3	45.0
PM	75.4	75.4	71.2	44.5	34.9	25.4
DRY	74.2	59.3	66.7	58.1	33.3	37.5
WET	78.1	77.6	78.7	51.8	36.8	25.5
ABOVE	86.2	87.5	91.7	55.2	34.4	16.7
BELOW	93.4	89.5	100	65.0	42.1	43.3
BRIGHT	76.6	76.2	73.2	51.8	37.3	30.4
DARK	78.7	68.3	80.0	60.0	34.1	26.7
CSO	87.1	80.0	100	61.4	44.4	50.0
No CSO	93.6	96.5	97.1	64.7	44.2	34.3

Rank correlation and independence between tests of *Giardia* and *Cryptosporidium* were evaluated in each category of analysis in order to determine the appropriate test of comparison between ROC curves of the two pathogens. The DeLong method for comparing the area under dependent ROC curves accounts for the Kendall tau b correlation between sets of data (DeLong et al., 1988). Therefore, that measure of association was calculated for all pairs of *Giardia* and *Cryptosporidium* data in order to determine whether the relationship was strong enough to warrant the use of the DeLong method. Additionally, the degree of independence was evaluated between the detection frequencies of the two pathogens.

Table VIII shows Kendall rank correlation coefficient (τ) computed from continuous density values for pathogens as well as the Chi-squared test of independence between detection frequencies for each pathogen. Kendall correlation coefficients and odds ratios between concordant and discordant pairs for the continuous dataset showed limited positive association between the two pathogens for nearly all categories (τ = -0.0148, -0.476). The subset of samples for the DRY sub-group were the only data showing moderate correlation between *Giardia* and *Cryptosporidium* (τ = 0.476). In most groups, the odds of concordance between ranks of *Giardia* and *Cryptosporidium* were twice as likely as the odds of discordance. The Chi-squared test between dichotomous (presence vs. absence) data showed statistical independence between detection frequencies of the two pathogens in every category except NO CSO. Because detection frequencies between the two pathogens in that category failed to reach statistical independence, it was the only category omitted from the statistical comparison of ROC curves between *Giardia* and *Cryptosporidium*. In all other categories, both the Kendall τ and the Chi-square test validate the use of the Hanley and McNeil comparison.

Table VIII

EVALUATION OF CORRELATION AND INDEPENDENCE BETWEEN *GIARDIA* AND *CRYPTOSPORIDIUM*

Category	N	Kendall Tau (τ)[a]	Correlation p-Value	Odds Ratio pC/pD	Chi-Square Test [b]	Chi-Square p-Value
ALL	324	0.315	<0.0001	1.9197	62.1	P < 0.0001
CAWS	195	0.197	0.0002	1.4907	17.7	P < 0.0001
GUW	98	0.346	0.0001	2.0581	18.9	P < 0.0001
AM	133	0.249	<0.0001	1.6631	19.5	P < 0.0001
PM	191	0.371	<0.0001	2.1797	38.8	P < 0.0001
DRY	247	0.476	<0.0001	2.8168	43.8	P < 0.0001
WET	62	0.291	<0.0001	1.8209	21.0	P < 0.0001
ABOVE	58	-0.0148	0.8640	0.9708	4.98	P = 0.0257
BELOW	137	0.166	0.0041	1.3981	9.87	P = 0.0017
BRIGHT	222	0.369	<0.0001	2.1696	50.6	P < 0.0001
DARK	75	0.226	0.0042	1.5840	12.2	P = 0.0004
CSO	70	0.304	0.0002	1.8736	13.6	P = 0.0002
NO CSO	156	0.2726	0.0593	1.7495	1.83	P = 0.1768

[a] Kendall Tau (τ) rank correlation coefficient between *Giardia* and *Cryptosporidium*.

[b] Chi-square statistic between detection frequencies of *Giardia* and *Cryptosporidium*.

5.9 Pathogen True Negative and True Positive Distributions for All Indicators

The area under a ROC curve can be interpreted as the probability that a given test will correctly rank a TP observation higher than a TN observation (DeLong et al., 1988). A perfect diagnostic test will rank every TP observation higher than every TN observation, correctly delineating the two distributions. Because indicator tests of pathogens are imperfect, the range of indicator values assigned to TP and TN observations will overlap. This overlap is visualized by creating frequency histograms for the log10-transformed indicator densities of TN and TP distributions for each pathogen and plotting those histograms on the same axis as shown in Figures 5 to 16.

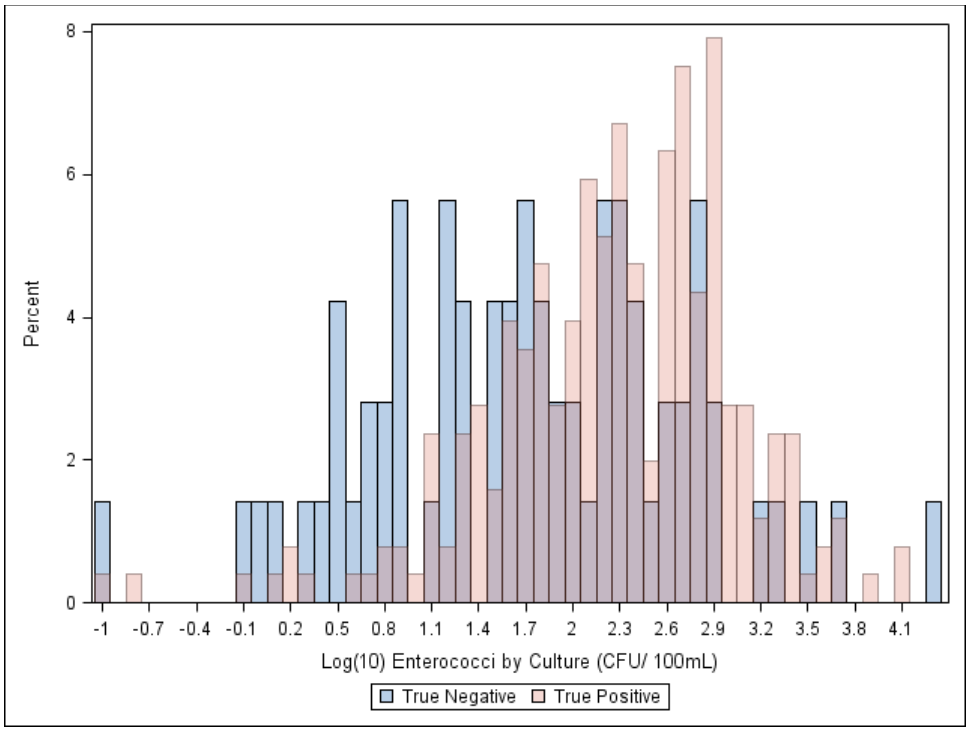

Figure 5. True negative and positive distributions of *Giardia* plotted against enterococci (cx).

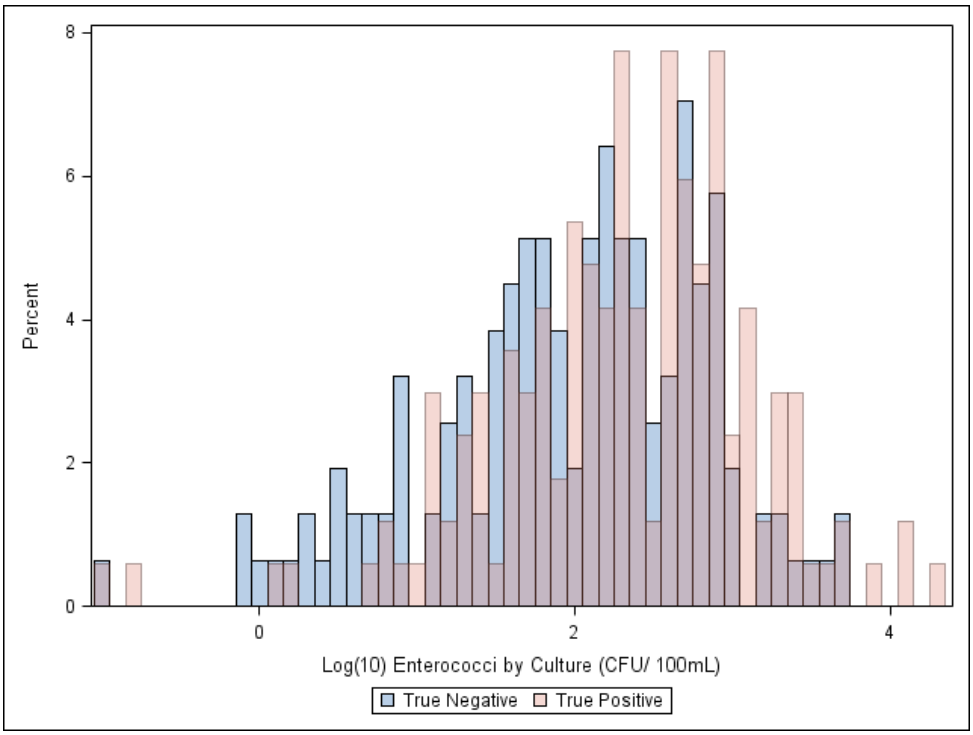

Figure 6. True negative and positive distributions of *Cryptosporidium* plotted against enterococci (cx).

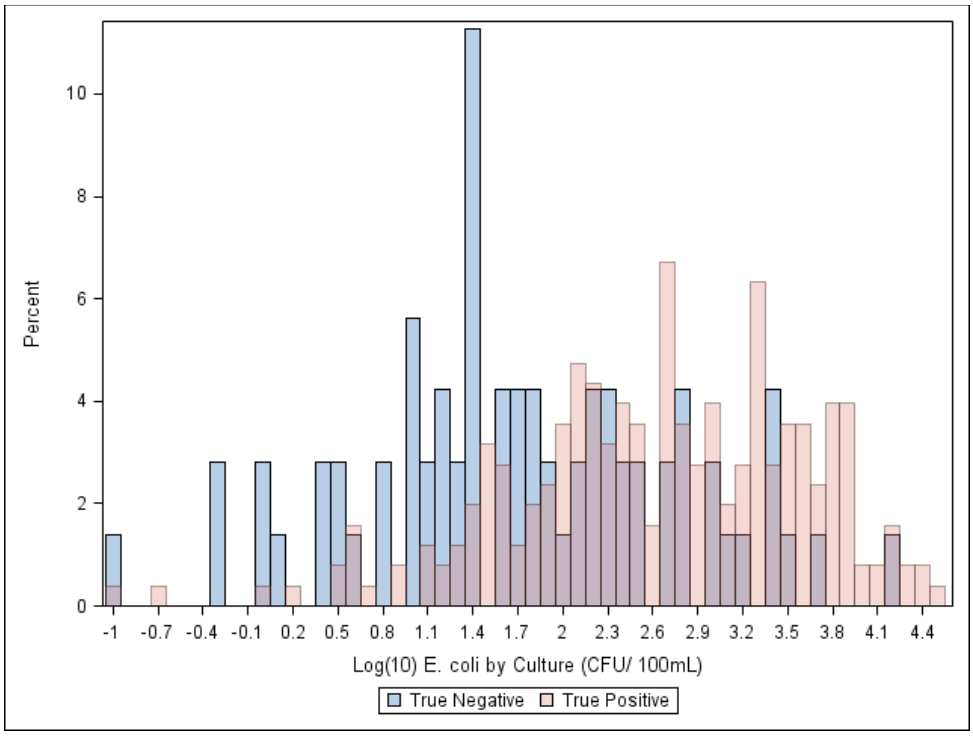

Figure 7. True negative and positive distributions of *Giardia* plotted against *E. coli* (cx).

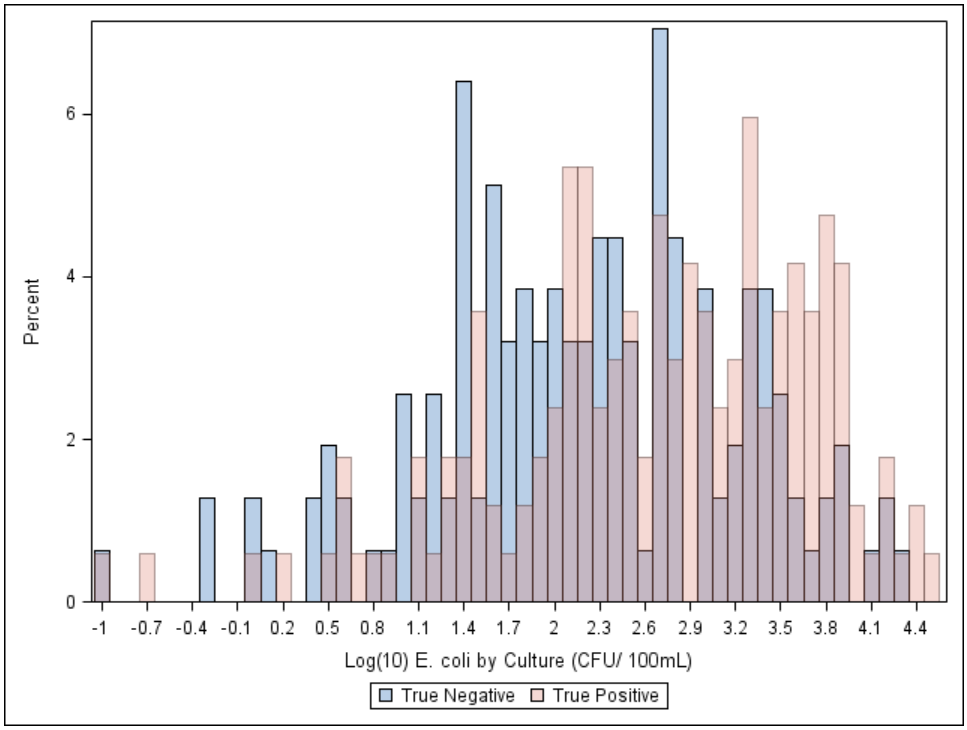

Figure 8. True negative and positive distributions of *Cryptosporidium* plotted against *E. coli* (cx).

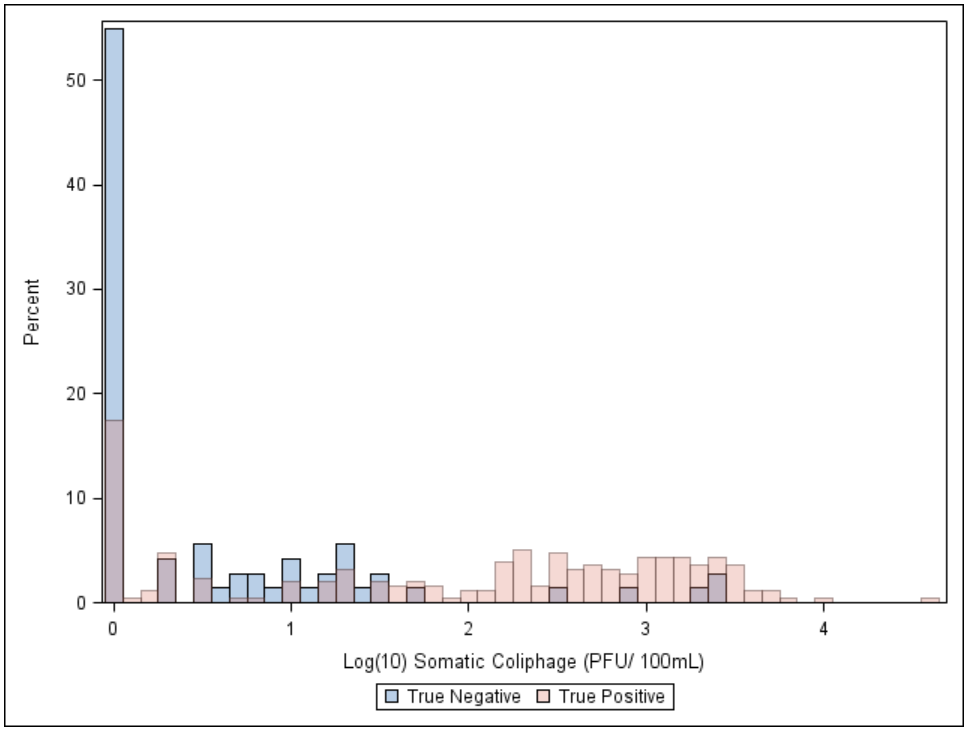

Figure 9. True negative and positive distributions of *Giardia* plotted against somatic coliphage.

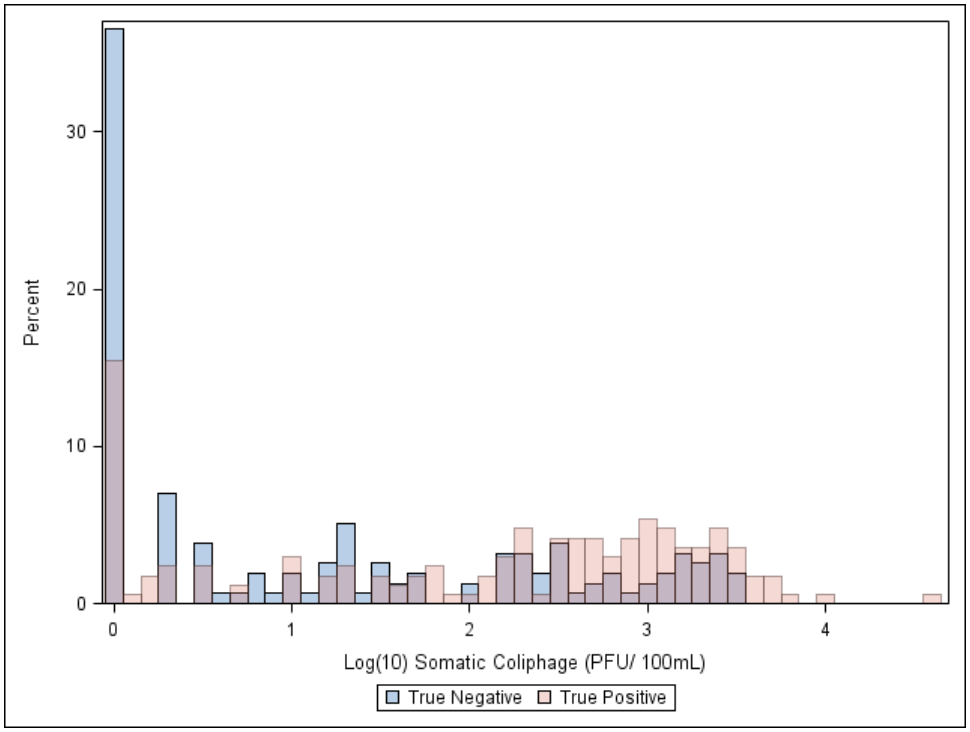

Figure 10. True negative and positive distributions of *Cryptosporidium* plotted against somatic coliphage.

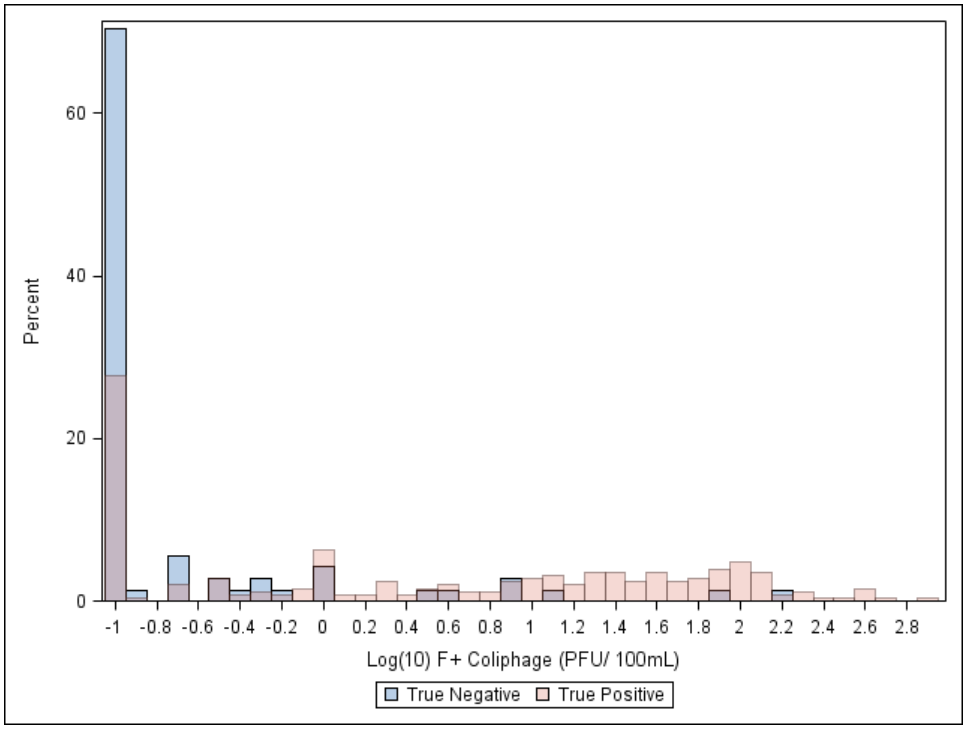

Figure 11. True negative and positive distributions of *Giardia* plotted against F+ coliphage.

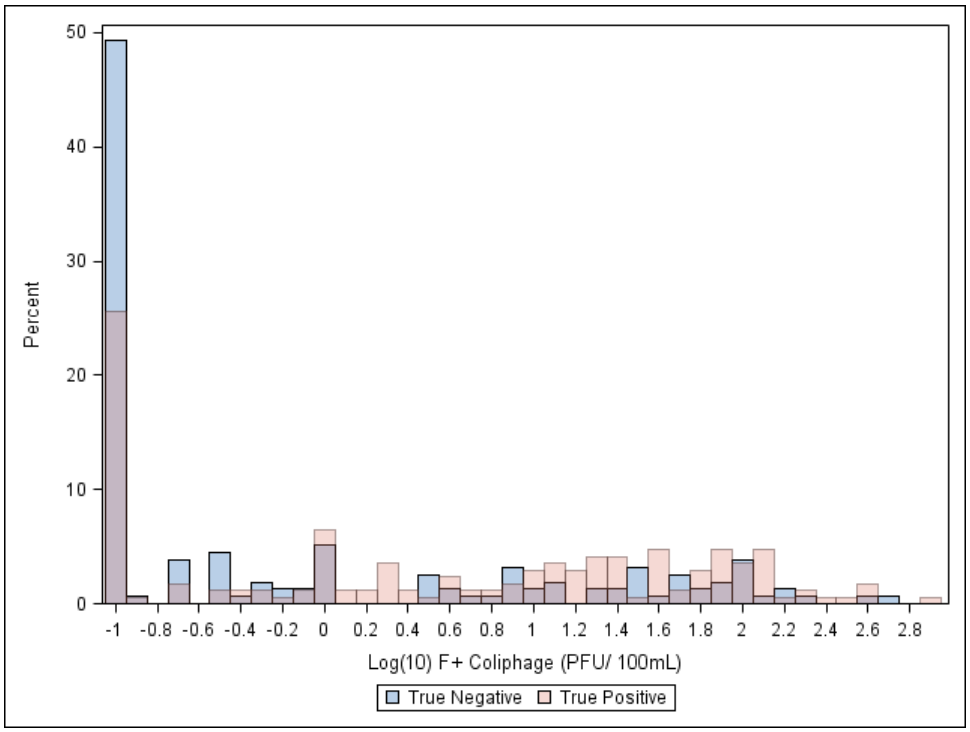

Figure 12.True negative and positive distributions of *Cryptosporidium* plotted against F+ coliphage.

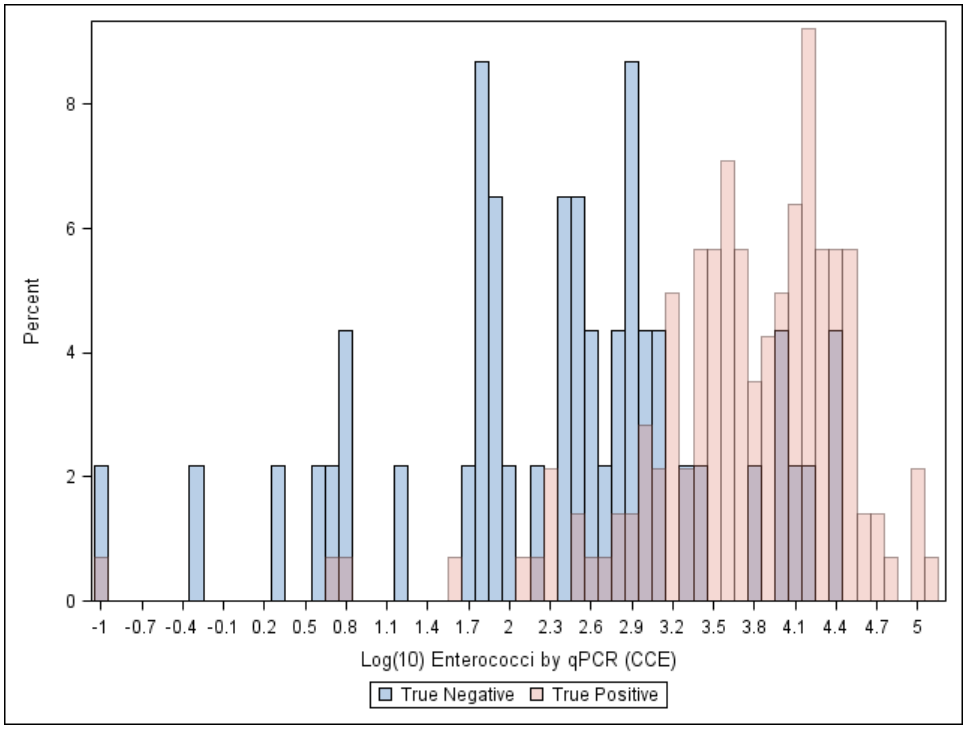

Figure 13. True negative and positive distributions of *Giardia* plotted against enterococci (q).

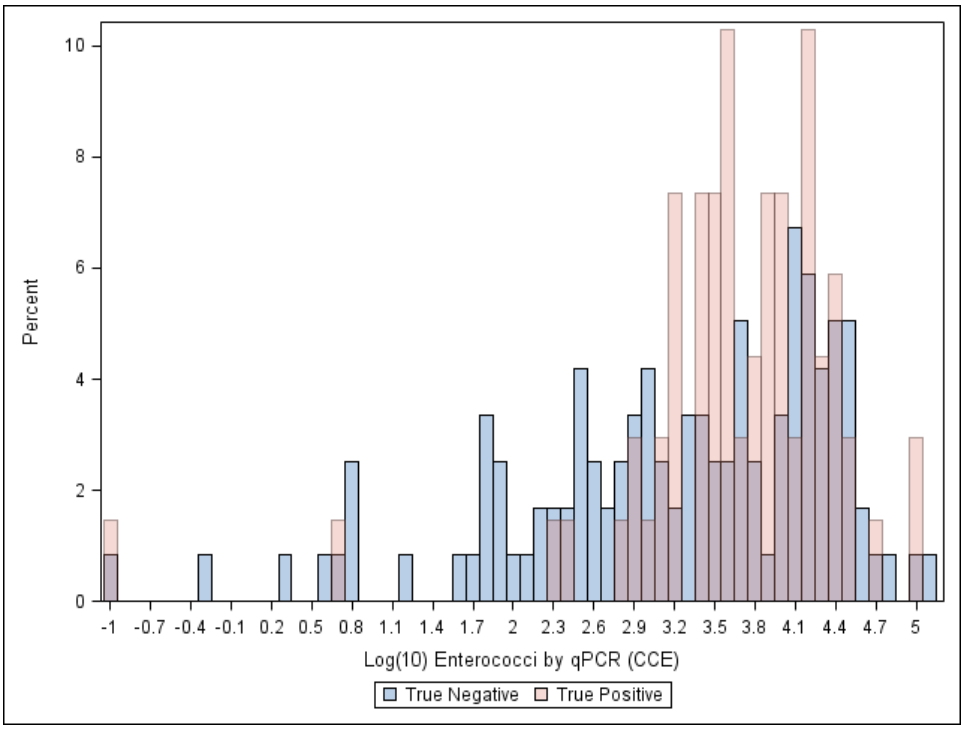

Figure 14. True negative and positive distributions of *Cryptosporidium* plotted against enterococci (q).

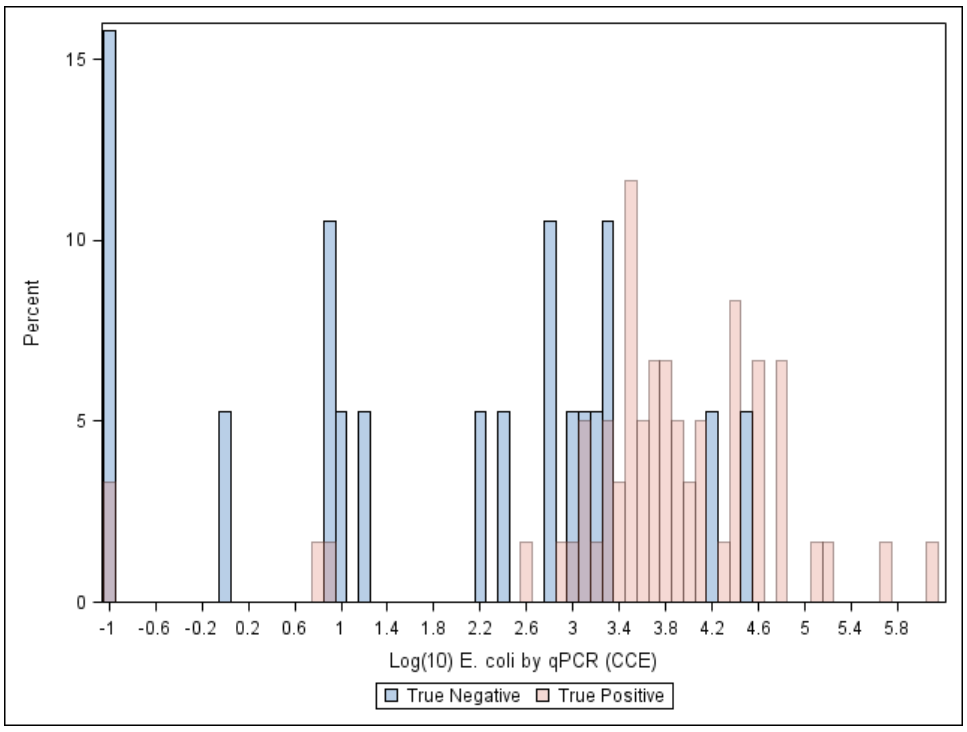

Figure 15. True negative and positive distributions of *Giardia* plotted against *E. coli* (q).

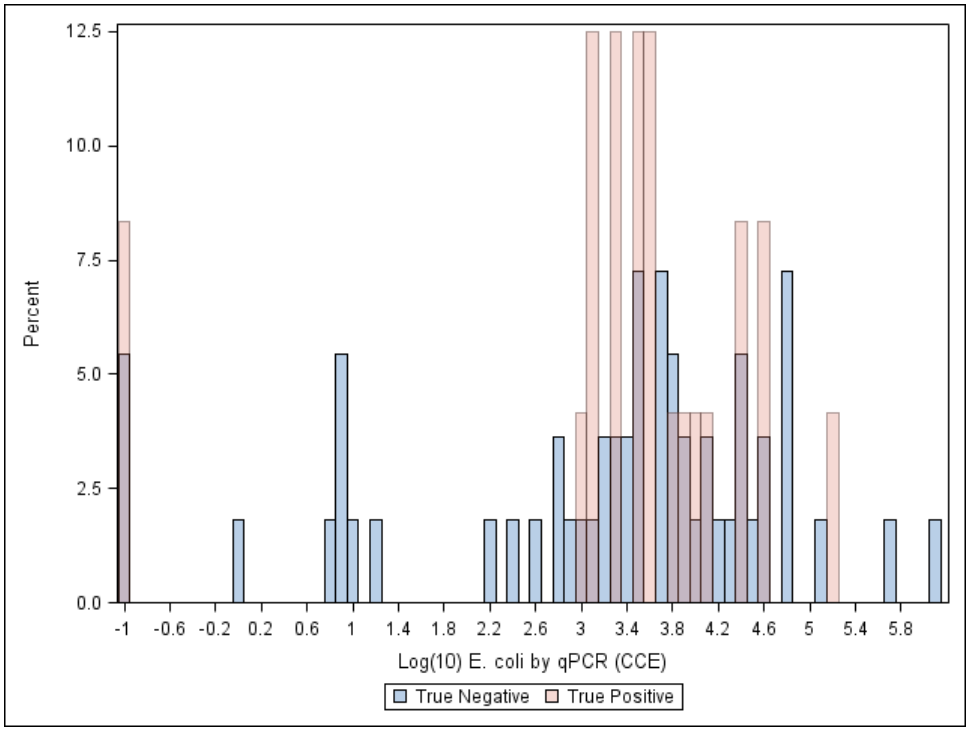

Figure 16. True negative and positive distributions of *Cryptosporidium* plotted against *E. coli* (q).

Figures 5-16 qualitatively illustrate the ability of the indicator microbes to differentiate pathogen presence and absence: Good ability to differentiate is indicated by distributions that do not share the same range and/or show distinct distributions. Inspection of the figures reveals that while indicator microbe densities are generally higher when the water is positive for *Cryptosporidium* than negative for *Cryptosporidium*; indicator microbe densities are consistently and more substantially increased when the water is positive for *Giardia* than negative for *Giardia*. In general, the distributions for *E. coli* and enterococci measured by culture show less separation for the presence and absence of pathogens, than the coliphages and *E. coli* and enterococci measured by qPCR. The diagnostic strength of indicators and the comparisons of diagnostic strength are formally evaluated by ROC analysis.

6. RECEIVER-OPERATING CHARACTERISTICS ANALYSIS

6.1 Receiver-Operating Characteristics Curves for Constructed for All Available Data Points

Each indicator-pathogen test was evaluated in context of the thirteen categories discussed in Section 4.1 (the complete dataset and six category pairs). More than 500 ROC curves were generated for every combination of indicator, category, and available sample size. The diagnostic value of indicators was examined between the two pathogens. Diagnostic differences within categorical pairs were also examined. Diagnostic strength was also compared among indicators within each category and the optimal diagnostic cutoff was determined for each indicator-pathogen pair.

Figure 17 and Figure 18 show the ROC curves generated for *Giardia* and *Cryptosporidium*, respectively for the best available sample size of each of the six indicator tests. The AUCs for each test group and the sample sizes used for the construction of ROC curves are listed in Table IX. Informally, ROC curves that are closer to top-left corner of the graph are stronger diagnostic tests for a given pathogen; conversely, the closer a curve is to the chance diagonal (Y=X, AUC=0.5), the weaker the test (Fawcett, 2006). Figure 17 shows that no test of *Giardia* was below the chance diagonal. The areas under the ROC curve for indicator tests of *Cryptosporidium* appear noticeably smaller than tests of *Giardia* (Figure 18). At several thresholds, the qPCR-based assay for *E. coli* performed worse than chance alone at correctly ranking samples positive for *Cryptosporidium* higher than samples that are negative for that pathogen. On the other hand, F+ coliphage and somatic coliphage emerge as possibly important predictors of *Cryptosporidium*. The 95% confidence intervals for the AUCs show that ROC curves generated for *Cryptosporidium* were barely above the chance diagonal (Table IX).

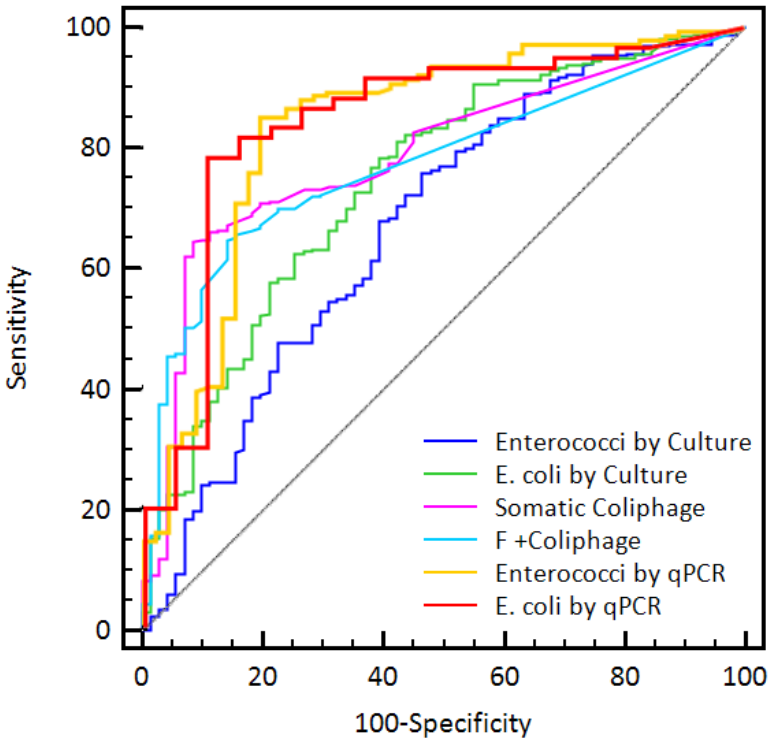

Figure 17. ROC curves for six indicator tests of *Giardia* for the general analysis.

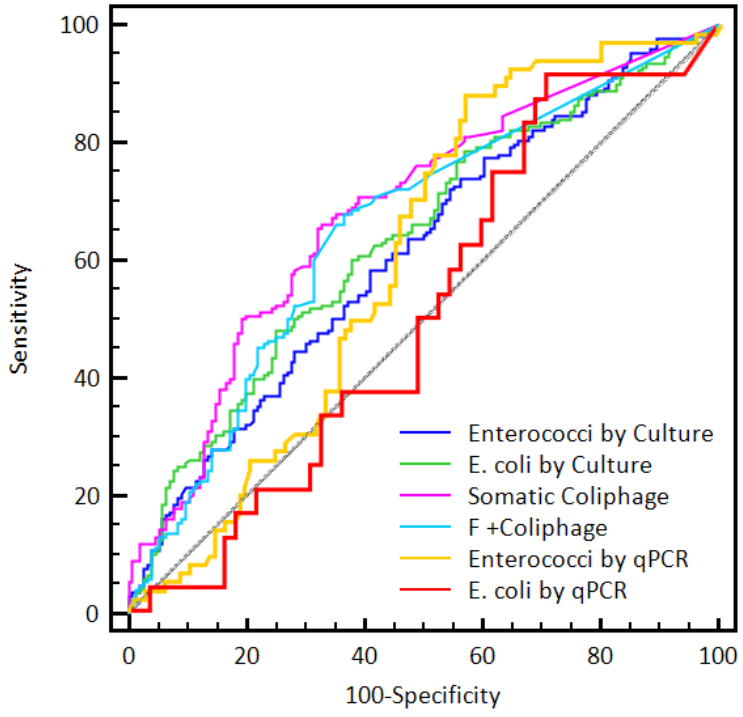

Figure 18. ROC curves for six indicator tests of *Cryptosporidium* for the general analysis.

The AUCs listed in Table IX show that qPCR-based tests of enterococci and E. coli performed better than other classification models for correctly ranking samples in which *Giardia* was detected higher than samples in which *Giardia* was not detected. The 95% confidence bounds for both these tests are greater than any other modality, with the upper confidence limits approaching perfect classification. Confidence bounds for tests of *Cryptosporidium* are considerably closer to the value 0.5, implying that they may be only slightly better than chance alone at differentiating samples positive or negative distributions.

Table IX
ROC ANALYSIS FOR THE COMPLETE DATA SET FOR BOTH PATHOGENS

GIARDIA

	AUC	Standard Error[a]	95% CI[b]	N
Enterococci by Culture	0.672	0.0384	0.618 to 0.723	324
E. coli by Culture	0.745	0.0336	0.693 to 0.791	324
Somatic Coliphage	0.788	0.0281	0.740 to 0.831	324
F+ Coliphage	0.773	0.0261	0.724 to 0.818	324
Enterococci by PCR	0.837	0.0376	0.776 to 0.887	187
E. coli by PCR	0.846	0.0567	0.747 to 0.917	79

CRYPTOSPORIDIUM

	AUC	Standard Error[a]	95% CI[b]	N
Enterococci by Culture	0.613	0.0311	0.558 to 0.667	324
E. coli by Culture	0.636	0.0307	0.581 to 0.688	324
Somatic Coliphage	0.684	0.0294	0.631 to 0.734	324
F+ Coliphage	0.655	0.0298	0.600 to 0.706	324
Enterococci by PCR	0.605	0.0411	0.531 to 0.676	187
E. coli by PCR	0.511	0.068	0.396 to 0.626	79

[a] DeLong method for calculating standard error is used (DeLong *et al.*, 1988).

[b] Confidence intervals are calculated using the binomial exact method.

6.2 Analysis Receiver-Operating Characteristics Analysis of Indicators and Pathogens for General Use Waters

Special attention should be paid to the performance of indicators within general use waters, as general use waters are designated as appropriate for all recreational uses, including swimming. Table X shows the AUC values for each indicator test with respect to the two pathogens in general use waters.

Table X
AREAS UNDER THE CURVE FOR INDICATORS AND PATHOGENS—GUW

GIARDIA				
	AUC	Standard Error[a]	95% CI[b]	N
Enterococci by Culture	0.585	0.0585	0.482 to 0.684	98
E. coli by Culture	0.616	0.0578	0.512 to 0.712	98
Somatic Coliphage	0.667	0.0515	0.565 to 0.759	98
F+ Coliphage	0.659	0.046	0.557 to 0.752	98
Enterococci by PCR	0.783	0.0652	0.653 to 0.882	56
E. coli by PCR	0.720	0.103	0.527 to 0.867	30
CRYPTOSPORIDIUM				
	AUC	Standard Error[a]	95% CI[b]	N
Enterococci by Culture	0.543	0.0706	0.439 to 0.644	98
E. coli by Culture	0.523	0.0698	0.420 to 0.625	98
Somatic Coliphage	0.634	0.0637	0.530 to 0.729	98
F+ Coliphage	0.597	0.0623	0.493 to 0.695	98
Enterococci by PCR	0.757	0.0952	0.623 to 0.861	56
E. coli by PCR	0.620	0.161	0.425 to 0.790	30

[a] DeLong method for calculating standard error is used (DeLong *et al.*, 1988).

[b] Confidence intervals are calculated using the binomial exact method.

Enterococci qPCR is a strong diagnostic test for *Giardia* and *Cryptosporidium* (AUC=0.720 and AUC=0.620, respectively). The lower confidence limit for this test was greater than any other indicator for both pathogens. Somatic coliphage and *E. coli* by qPCR also appear to be important indicators of *Giardia* (AUC=0.667 and AUC=0.783, respectively). Nevertheless, a simple comparison of raw AUC values between the complete dataset (which includes the present 98 observations) and this subset of values shows that every single AUC calculated from GUW data is lower than its counterpart in the analysis of all data points combined. There are two exceptions to this relationship—both the qPCR-based assay for enterococci and the qPCR-based assay for *E. coli* performed better in the GUW subset than in the complete dataset.

6.3 Areas Under Curve for each Indicator-Category-Pathogen.

Table XI shows the area under the curve for each indicator and each category for the pathogen *Giardia,* where the best category for the prediction of *Giardia* presence for each microbe is indicated in bold. For example, enterococci best predicts *Giardia* presence when water quality was measured in PM hours. With the exception of a few categories (CAWS, BELOW, and CSO PRESENT) enterococci by qPCR and *E. coli* by qPCR are the best predictors of *Giardia* presence in any given category. Conversely, enterococci by culture and *E. coli* by culture are the least predictive. Table XI shows the AUCs for tests of *Cryptosporidium*: In general, the AUC for *Cryptosporidium* are lower than the AUC for *Giardia*. For all six indicators, the best AUC for *Giardia* is higher than the best AUC for *Cryptosporidium*.

Table XI

THE AREA UNDER ROC CURVE OF INDICATORS GENERATED FOR THE PRESENCE OF *GIARDIA* CYSTS FOR ALL DATAPOINTS AND SIX DICHOTOMOUS CATEGORIES OF COMPARISON*

Category	Enterococci (cx)	*E. coli* (cx)	S. Coliphage	F+ Coliphage	Enterococci (qPCR)	*E. coli* (qPCR)
ALL	0.672	0.745	0.788	0.773	0.837	0.846
CAWS	0.582	0.613	0.757	0.717	0.668	n/a**
GUW	0.585	0.616	0.667	0.659	0.783	0.720
AM	0.561	0.635	0.755	0.725	0.819	0.972
PM	**0.728**	**0.803**	0.813	**0.802**	0.868	0.819
WET	0.691	0.768	0.798	0.793	0.809	0.866
DRY	0.556	0.640	0.728	0.658	**0.932**	0.793
ABOVE	0.575	0.523	0.519	0.514	0.589	n/a**
BELOW	0.536	0.592	**0.832**	0.767	0.711	n/a**
BRIGHT	0.682	0.760	0.790	0.786	0.844	0.798
DARK	0.604	0.702	0.788	0.730	0.905	**1.000**
NO CSO	0.556	0.521	0.780	0.714	0.593	n/a**
CSO	0.638	0.683	0.680	0.707	0.799	n/a**

*The highest available AUC for each indicator is denoted in bold.

**Not available due to limited numbers of *Giardia* non-detect values.

Table XII

THE AREA UNDER ROC CURVE OF INDICATORS GENERATED FOR THE PRESENCE OF *CRYPTOSPORIDIUM* OOCYSTS FOR ALL DATAPOINTS AND SIX DICHOTOMOUS CATEGORIES OF COMPARISON*

Category	Enterococci (cx)	*E. coli* (cx)	S. Coliphage	F+ Coliphage	Enterococci (qPCR)	*E. coli* (qPCR)
ALL	0.613	0.636	0.684	0.655	0.605	0.511
CAWS	0.542	0.547	0.589	0.574	0.526	0.723
GUW	0.543	0.523	0.634	0.597	**0.757**	0.620
AM	0.540	0.573	0.667	0.652	0.551	0.859
PM	0.648	**0.675**	0.703	0.662	0.640	0.671
WET	0.666	0.655	0.697	0.675	0.578	0.560
DRY	0.540	0.626	0.680	0.607	0.685	0.533
ABOVE	0.579	0.558	0.531	0.522	0.602	**0.900**
BELOW	0.511	0.549	0.578	0.547	0.506	0.742
BRIGHT	**0.673**	0.671	**0.711**	**0.695**	0.585	0.584
DARK	0.519	0.559	0.656	0.581	0.706	0.727
NO CSO	0.546	0.525	0.549	0.511	0.613	0.736
CSO	0.624	0.572	0.699	0.683	0.548	0.694

*The highest available AUC for each indicator is denoted in bold.

The ten highest AUCs for each pathogen are ranked in Table XIII for each indicator test and the category of analysis for each AUC is also noted. For *Giardia* cyst presence, 9 of the top 10 AUCs were obtained when *E. coli* and enterococci were measured by qPCR assay. There is no clear relationship among the categories which yield the best AUC for the pathogen *Giardia*. Although the two highest AUC values for *Cryptosporidium* are in the same range as AUC values of *Giardia*, the AUCs for this pathogen are noticeably weaker than those for *Giardia*. While the confidence bounds for AUC values generated from tests of *Giardia* do not fall below 0.700, every available AUC for indicators of *Cryptosporidium* show a lower confidence limit less than that value, with some AUCs having confidence bounds close to the chance diagonal.

Table XIII

TEN BEST AUCS FOR ALL INDICATOR CATEGORIES

GIARDIA

INDICATOR	AUC	STANDARD ERROR[a]	95% CI[b]	N	CATEGORY
E. coli by qPCR	1.000	0.0000	0.782 to 1.000	15	DARK
E. coli by qPCR	0.972	0.0278	0.785 to 1.000	20	AM
Enterococci by qPCR	0.932	0.0545	0.765 to 0.993	27	DRY
Enterococci by qPCR	0.905	0.0642	0.772 to 0.974	41	DARK
Somatic Coliphage	0.879	0.0984	0.712 to 0.968	31	OTHER
Enterococci by qPCR	0.868	0.0433	0.796 to 0.922	126	PM
E. coli by qPCR	0.866	0.0629	0.735 to 0.948	47	WET
E. coli by qPCR	0.846	0.0567	0.747 to 0.917	79	ALL
Enterococci by qPCR	0.844	0.0428	0.769 to 0.903	126	BRIGHT

CRYPTOSPORIDIUM

INDICATOR	AUC	STANDARD ERROR[a]	95% CI[b]	N	CATEGORY
E. coli by qPCR	0.900	0.1200	0.594 to 0.996	12	ABOVE
E. coli by qPCR	0.859	0.0913	0.631 to 0.972	20	AM
Enterococci by qPCR	0.757	0.0952	0.623 to 0.861	56	GUW
E. coli by PCR	0.742	0.0962	0.550 to 0.883	30	BELOW
E. coli by PCR	0.736	0.0941	0.560 to 0.870	35	CSO ABSENT
E. coli by PCR	0.727	0.1590	0.443 to 0.919	15	DARK
E. coli by PCR	0.723	0.0872	0.564 to 0.850	42	CAWS
Enterococci by Culture	0.717	0.1080	0.528 to 0.864	31	OTHER
Somatic Coliphage	0.711	0.0348	0.647 to 0.770	222	BRIGHT
Enterococci by qPCR	0.706	0.0888	0.544 to 0.838	41	DARK

[a] DeLong method for calculating standard error is used (DeLong *et al.*, 1988).

[b] Confidence intervals are calculated using the binomial exact method.

6.4 Differences in Diagnostic Strength between Pathogens

Table XIV lists the difference in AUCs of curves generated for the prediction of *Giardia* and *Cryptosporidium* for each indicator-test, for the general analysis (ALL) and each of 12 sub-categories. This analysis makes multiple comparisons using the same statistical inference; consequently, the Type I error rate increases every time the test is repeated for independent sample sets. In this case, the assumption of independence is only valid between categories of analysis and not between indicators. Therefore, the α level for categorical analyses was adjusted for the 12 independent comparisons using the Šidák correction; this reduced the level of statistical significance from $\alpha=0.05$ to $\alpha=0.000427$. Because indicator tests are dependent data groups, the Šidák correction was not required to adjust α levels by an additional factor of six. For this reason, the comparison of AUCs within the general analysis is statistically significant at $\alpha=0.05$.

Table XIV

THE DIFFERENCE IN THE AREA UNDER THE CURVE (AUC) BETWEEN *GIARDIA* AND *CRYPTOSPORIDIUM* FOR EACH INDICATOR FOR ALL DATAPOINTS AND SIX DICHOTOMOUS CATEGORIES OF COMPARISON [a*]

Category	Enterococci (culture)	*E. coli* (culture)	S. Coliphage	F+ Coliphage	Enterococci (qPCR)	*E. coli* (qPCR)
ALL	0.059	**0.109**	**0.104**	**0.118**	**0.232**	**0.335**
CAWS	0.04	0.066	**0.168**	**0.143**	0.142	n/a[***]
GUW	0.042	0.093	0.033	0.062	0.026	0.1
AM	0.021	0.062	0.088	0.073	**0.268**	0.113
PM	0.08	**0.128**	**0.11**	**0.14**	**0.228**[**]	0.148
DRY	0.016	0.014	0.048	0.051	**0.247**	0.26
WET	0.025	**0.113**	**0.101**	**0.118**	**0.231**[**]	**0.306**
ABOVE	-0.004	-0.035	-0.012	-0.008	-0.013	n/a[***]
BELOW	0.025	0.043	**0.254**	**0.22**	0.205	n/a[***]
BRIGHT	0.0009	0.089	0.079	0.091	**0.259**[**]	**0.214**
DARK	0.085	0.143	0.132	0.149	0.199	0.273
CSO	-0.068	-0.051	0.081	0.031	0.045	n/a[***]
NO CSO	n/a[****]	n/a[****]	n/a[****]	n/a[****]	n/a[****]	n/a[****]

[a] Hanley and McNeil test for independent ROC curves (Hanley and McNeil, 1982).

[*] Relationships that are less than or equal to α=0.05 are indicated in bold.

[**] Statistically significant relationships are assessed after Šidák correction, α = 0.00427.

[***]Not available due to limited numbers of *Giardia* non-detect values.

[****]Hanley and McNeil test could not be performed due to a lack of statistical independence.

With the exception of enterococci by culture, every indicator measured by culture methods was a statistically better test of *Giardia* than *Cryptosporidium* at α=0.05 in the general analysis (ALL). Because the level of statistical significance in the categorical analyses were lowered by the Šidák correction, the only statistically significant difference between tests of the two pathogens occurred for enterococci by qPCR in WET weather, BRIGHT sunlight, and PM hours. Nevertheless, there were important (though not significant) differences within each category for the prediction of *Giardia* relative to *Cryptosporidium*. These included location group CAWS, PM time of day, WET conditions, sample collection BELOW the water treatment plant, and BRIGHT sunlight.

General differences in diagnostic strength between the two pathogens were also evident for specific indicator tests; the p-value for these relationships was less than or equal to α=0.05 but did not reach statistical significance at α=0.00427. In six out of 13 categories, qPCR-based enterococci was a better test of *Giardia* then of *Cryptosporidium* (three categories of which were statistically significant difference at α=0.00427). Conversely, enterococci by culture did not show a diagnostic difference between the two pathogens under any condition or α level. Somatic coliphage and F+ coliphage were both better tests of *Giardia* than of *Cryptosporidium* in four out of twelve categories. Additionally, *E. coli* by culture and *E. coli* by qPCR were both better at correctly predicting the presence of *Giardia* over *Cryptosporidium* in WET conditions. In no subset of the data defined by sampling location or conditions was an indicator a statistically better test of *Cryptosporidium* than *Giardia*. However, it is unclear whether these differences reflect the greater predictive power of the indicators with regard to *Giardia*, or whether the difference is due to recovery for the pathogen *Cryptosporidium*.

6.5 Differences in Diagnostic Strength between Categories

This study also evaluated statistical differences within paired categories in order to determine if conditions existed during which the diagnostic value of indicators improved: Six categories were identified, each of which had two conditions. TABLE XV summarizes the differences of AUCs for each indicator between the category pairs. Differences that exceeded $\alpha=0.05$ are indicated in bold. However, because of the multiple comparison problem, the Šidák correction was employed and statistical significance is evaluated at $\alpha=0.00427$. At this level p-value, only one statistically significant difference occurred between categories: *E. coli* by qPCR was a better test of *Giardia* in DARK conditions than BRIGHT conditions.

Table XV

DIFFERENCE IN AUC FOR ROC CURVES BETWEEN DICHOTOMOUS STRATA OF 6 CATEGORIES FOR ALL INDICATORS[a*]

GIARDIA

Pair	Enterococci (culture)	*E. coli* (culture)	S. Coliphage	F+ Coliphage	Enterococci (qPCR)	*E. coli* (qPCR)
CAWS – GUW	-0.003	-0.003	0.090	0.058	-0.115	n/a[***]
AM – PM	-0.167	**-0.168**	-0.058	-0.077	-0.049	0.153
DRY – WET	**-0.135**	-0.128	-0.070	-0.135	0.123	-0.073
ABOVE – BELOW	0.039	-0.069	**-0.313**	-0.253	-0.122	n/a[***]
BRIGHT – DARK	0.078	0.058	0.002	0.056	-0.061	**-0.202**[**]
CSO – NO CSO	-0.082	-0.162	0.100	0.707	-0.206	n/a[***]

CRYPTOSPORIDIUM

Pair	Enterococci (culture)	*E. coli* (culture)	S. Coliphage	F+ Coliphage	Enterococci (qPCR)	*E. coli* (qPCR)
CAWS – GUW	-0.001	0.024	-0.045	-0.023	**-0.231**	0.103
AM – PM	-0.108	-0.102	-0.036	-0.010	-0.089	0.188
DRY – WET	-0.126	-0.029	-0.017	-0.068	-0.107	-0.027
ABOVE – BELOW	0.068	0.009	-0.047	-0.025	0.096	0.158
BRIGHT – DARK	**0.154**	0.112	0.055	0.114	-0.121	-0.143
CSO – NO CSO	0.078	0.047	0.150	**0.172**	-0.065	-0.042

[a] Hanley and McNeil test for independent ROC curves (Hanley and McNeil, 1982).

[*] Relationships that are less than or equal to α=0.05 are indicated in bold.

[**] Statistically significant relationships are assessed after Šidák correction, α = 0.00427.

[***]Not available due to limited numbers of *Giardia* non-detect values.

While the difference in diagnostic strength of indicator tests between categories did not reach statistical significance at the Šidák corrected α level, several relationships were suggested by large differences in AUC values (at α=0.05). The predictive power of enterococci by culture appeared to improve in WET weather than DRY weather (observations are labeled as WET if rained occurred any time during the 96 hours before sample retrieval). The culture-based assay for *E. coli* was a better indicator *Giardia* during AM hours. Somatic coliphage was a better indicator BELOW a wastewater treatment plant than ABOVE one. The qPCR based assay of *E. coli* was a better indicator of *Giardia* during periods of less intense sunlight. The qPCR-based enterococci assay performed better in the GUW than in the CAWS for the correct ranking of samples positive for *Cryptosporidium*. This was also true for F+ coliphage during conditions in which a CSO had occurred. These relationships suggest a pattern in which conditions favoring increased indicator density correspond with increases in the diagnostic strength of indicator tests. Indicators appear to be more predictive of pathogens in the first swell of pollution such as when occurs in WET weather, after a CSO event, or BELOW a WTP. Environmental and hydrometerologic conditions that promote indicator survival may positively influence the diagnostic strength of indicators tests that are culture-based. This hypothesis is evaluated by testing the difference in AUCs between high and low densities of indicators.

For every available data point within the present analysis, observations for each indicator test were sorted in order of increasing magnitude and the datasets was separated into subgroups of highest and lowest values. ROC curves were constructed for each sub-group and the differences in AUCs between HIGH and LOW indicator densities are summarized in Table XVI. Statistical testing of AUC differences show that HIGHER indicator densities were not

diagnostically better predictors of the presence or absence of pathogens. In fact, the only statistically significant relationships showed that tests of enterococci by qPCR performed better with regard to both pathogens in the sub-set of lower indicator test values. Generally, however, no consistent trends existed between indicator density and diagnostic strength. This is consistent with the earlier finding that no single environmental condition or sampling location points to consistent changes in predictive power of indicators. These findings suggest that diagnostic strength is relatively robust which regard to changes in environmental conditions, hydrometerologic changes, locations, and even absolute densities of indicators.

Table XVI

COMPARISON OF THE STRENGTH OF INDICATORS BETWEEN HIGH AND LOW INDICATOR DENSITIES

GIARDIA

	HIGH			LOW			COMPARISON	
Indicator	AUC	SE[a]	N	AUC	SE[a]	N	Diff.	p-Value[*]
Enterococci by Culture	0.548	0.0696	162	0.667	0.0482	162	-0.119	0.1598
E. coli by Culture	0.592	0.0686	162	0.673	0.0459	162	-0.081	0.3264
Somatic Coliphage	0.670	0.1030	162	0.558	0.0459	162	0.112	0.3206
Male Coliphage	0.693	0.0937	162	0.556	0.0460	162	0.137	0.1894
Enterococci by PCR	0.531	0.0870	93	0.819	0.0434	94	-0.288	**0.0031**
E. coli by PCR*	0.605	0.094	40	0.770	0.076	39	-0.165	0.173

CRYPTOSPORIDIUM

	HIGH			LOW			COMPARISON	
Indicator	AUC	SE[a]	N	AUC	SE[a]	N	Diff.	p-Value[*]
Enterococci by Culture	0.555	0.0454	162	0.556	0.0453	162	-0.001	0.9876
E. coli by Culture	0.586	0.0458	162	0.543	0.0463	162	0.043	0.5091
Somatic Coliphage	0.544	0.0484	162	0.535	0.0473	162	0.009	0.8942
Male Coliphage	0.504	0.0475	162	0.506	0.0482	162	-0.002	0.9764
Enterococci by PCR	0.567	0.0603	93	0.770	0.0556	94	-0.203	**0.0133**
E. coli by PCR	0.583	0.099	40	0.651	0.100	39	-0.068	0.629

[a] Hanley and McNeil method for calculating standard error is used (Hanley and McNeil, 1982).

[*] Statistically significant differences at α=0.05 are indicated in bold.

6.6 Difference in Diagnostic Strength between Indicators

A third objective of the present study was to compare the ROC curves between indicator tests for pathogens within the available categories. In the present analysis, indicator values were derived from splits of the same water sample and are therefore considered dependent observations. The DeLong comparison between dependent ROC curves necessitated perfect pairing between sample groups and resulted in three possible sets of comparisons *(DeLong et al., 1988)*. First, comparisons between only the four culture based assays (the largest sample group in any category) yielded 6 indicator-indicator pairs for each of two pathogens in 13 categories for a total of 156 comparisons. The dataset was then limited to only those observations with values for the culture based assays and enterococci by qPCR. The 10 indicator-indicator pairs, 13 categories, and 2 pathogens yielded 260 comparisons. Finally the sample size was limited further in order to include those observations that had values for *E. coli* by qPCR thereby including all six indicators in the comparison pool. This group of analyses resulted in the smallest sample size and created 15 indicator-indicator pairs in 13 categories for 2 pathogens for a total of 390 paired comparisons. Accounting for the number of indicators, pathogens, categories, and available sample sizes, 806 comparisons were made. Table XVII tallies only the statistically significant relationships at $\alpha = 0.05$ and notes the directionality of the comparison—because the comparisons are all dependent with respect to another a Bonferroni correction of α Type I error was not needed. Of the 91 significant pairs, there were only 6 circumstances in which the same statistically significant indicator-indicator held true for *Giardia* and *Cryptosporidium*.

Table XVII
FREQUENCY OF STATISTICALLY SIGNIFICANT DIFFERENCES IN AUC VALUES OF INDICATOR MICROBES PREDICTING *GIARDIA* AND *CRYPTOPSORIDIUM* PRESENCE/ABSENCE

Microbe Relationships, based on AUC	*Giardia*	*Cryptosporidium*	Both	Total
E. coli by culture < Enterococci by qPCR	5	0	0	5
E. coli by culture < F+ Coliphage	0	1	0	1
E. coli by culture < Somatic Coliphage	4	4	1	8
E. coli by qPCR < Enterococci by qPCR	4	1	1	5
E. coli by qPCR < F+ Coliphage	0	3	0	3
E. coli by qPCR < Somatic Coliphage	0	3	0	3
Enterococci by culture < *E. coli* by culture	9	0	0	9
Enterococci by culture < Enterococci by qPCR	13	1	1	14
Enterococci by culture < F+ Coliphage	5	1	1	6
Enterococci by culture < Somatic Coliphage	15	3	1	18
Enterococci by qPCR < F+ Coliphage	1	0	0	1
Enterococci by qPCR < Somatic Coliphage	0	1	0	1
F+ Coliphage < *E. coli* by qPCR	0	1	0	1
F+ Coliphage < Enterococci by qPCR	5	1	1	6
F+ Coliphage < Somatic Coliphage	4	3	0	7
Somatic Coliphage < *E. coli* by qPCR	0	1	0	1
Somatic Coliphage < Enterococci by qPCR	1	0	0	1
Somatic Coliphage < F+ Coliphage	1	0	0	1
Total	67	24	6	91

Overall, the culture-based test for enterococci was found to be a statistically inferior test of *Giardia* when compared with another indicator 47 times and was never statistically superior to any other indicator. Given the low AUC values for culture-based enterococci culture in nearly every category of analysis, this indicator microbe was the weakest test of *Giardia*. In contrast, enterococci by qPCR was statistically superior to other indicators in 31 comparisons—this assay proved to be a better test than enterococci by culture 14 times, and a better test than *E. coli* by culture 5 times. However, it should be noted that only 2 of the 31 scenarios showed enterococci by culture to be a better indicator of *Cryptosporidium*.

The two coliphage assays were markedly different. Somatic coliphage proved to be a strong predictor of pathogen presence, statistically exceeding the predictive power of other indicators in 37 scenarios; somatic coliphage was superior to enterococci by culture 18 times and *E. coli* by culture 8 times. Of the 37 statistically significant comparisons, 14 showed that somatic coliphage was a superior indicator of *Cryptosporidium*. Thus, somatic coliphage is a good predictor of both pathogens as compared to other assays. Conversely, F+ coliphage was generally a poor predictor of both pathogens and statistically inferior in 14 comparisons.

The culture-based assay for *E. coli* statistically weaker than another indicator in 14 comparisons and was the second weakest test behind enterococci by culture. However, unlike enterococci by qPCR, *E. coli* by qPCR did not prove to be a drastic improvement on the culture based assay of the same organism. The qPCR-based assay for *E. coli* proved to be statistically superior in only two comparisons and inferior in eleven comparisons. Of the comparisons in which *E. coli* by qPCR was the inferior test, 7 concerned the pathogen *Cryptosporidium*. These findings seem to conflict with the AUC rankings which show that *E. coli* by qPCR generated the

highest AUC values for both pathogens. The apparent paradox is explained by the small samples sizes for observations of *E. coli* by qPCR. Because sample sizes for this assay were substantially smaller than other sets of observations, standard error values were large; therefore, even large differences in AUC values between *E. coli* by qPCR and another indicator did not reach statistical significance. The comparative value of *E. coli* by qPCR is unclear at this stage due to sample size restrictions.

The relationships listed in Table XVII show that enterococci by culture is consistently the lease predictive diagnostic test for either pathogen. The culture-based assay for *E. coli* also appears to be a poor predictor of the two pathogens, though not as strikingly as enterococci by culture. The qPCR-based assay for enterococci appears to be the best indicator of *Giardia*, though not an important predictor of *Cryptosporidium*. Somatic coliphage is a relatively strong indicator of both pathogens. Larger sample sizes are required to evaluate whether *E. coli* by qPCR is statistically superior to other indicator tests. Clearer diagnostic differences exist between tests of *Giardia* (67 significant comparisons) than tests of *Cryptosporidium* (24 significant comparisons). This corroborates the AUC findings which show that all indicator tests are stronger tests of *Giardia*.

6.7 Optimal Thresholds

Optimization of the classification threshold is permitted by ROC analysis. In this study, sensitivity and specificity were weighted equally and the Youden index was used to evaluate that point at which the distance from perfect classification on an ROC curve (0,1) was minimized (Youden, 1950). Table XVIII lists the optimal thresholds of each indicator for *Giardia*

and *Cryptosporidium*; where available, EPA STVs are listed for comparison. Both tables provide the sensitivity and specificity of ROC thresholds and available criteria values. It should be noted that EPA criteria for culture-based values of enterococci and *E. coli* predict a specific level of GI illness among swimmers while the ROC optimal thresholds predict the presence or absence of pathogens on Chicago area waterways, including locations not fit for swimming. Thus, the two classification thresholds are not directly comparable. Instead, the differences in sensitivity and specificity with regard to pathogens are evaluated, as shown in Figures 19 to 24 for *Giardia* and Figures 25 to 30 for *Cryptosporidium*.

Table XVIII

OPTIMAL INDICATOR THRESHOLDS FOR ALL INDICATORS

GIARDIA

	AUC	Threshold[a]	Sensitivity	Specificity	Criteria	Sensitivity	Specificity
Enterococci (cx) (CFU/100 mL)	0.672	58.4	75.9	53.5	61.0[b]	72.5	50.0
E. coli (cx) (CFU/100 mL)	0.745	87.6	78.3	60.6	235[b]	80.0	57.5
S. Coliphage (PFU/100 mL)	0.788	31.6	64.4	91.6	10.0[c]	75.0	77.0
F. Coliphage (PFU/100 mL)	0.773	0.562	64.8	85.9	n/a	n/a	n/a
Enterococci (q) (CCE)	0.837	1200	85.1	80.4	1000[d]	88.5	75.0
E. coli (q) (CCE)	0.846	2020	78.3	89.5	n/a	n/a	n/a

CRYPTOSPORIDIUM

	AUC	Threshold[a]	Sensitivity	Specificity	Criteria	Sensitivity	Specificity
Enterococci (cx) (CFU/100 mL)	0.613	88.4	72.0	45.5	61.0[b]	75.8	39.3
E. coli (cx) (CFU/100 mL)	0.636	698	48.2	75.0	235[b]	63.3	56.1
S. Coliphage (PFU/100 mL)	0.684	50.0	65.5	68.0	10.0[c]	73.4	52.9
F. Coliphage (PFU/100 mL)	0.655	0.804	67.9	63.5	n/a	n/a	n/a
Enterococci (q) (CCE)	0.605	1240	88.2	42.9	1000[d]	88.2	39.7
E. coli (q) (CCE)	0.511	868	91.7	29.1	n/a	n/a	n/a

[a] Derived from ROC analysis.

[b] EPA criteria (US EPA, 1986)

[c] Limit of detection for somatic coliphage.

[d] US EPA, 2010.

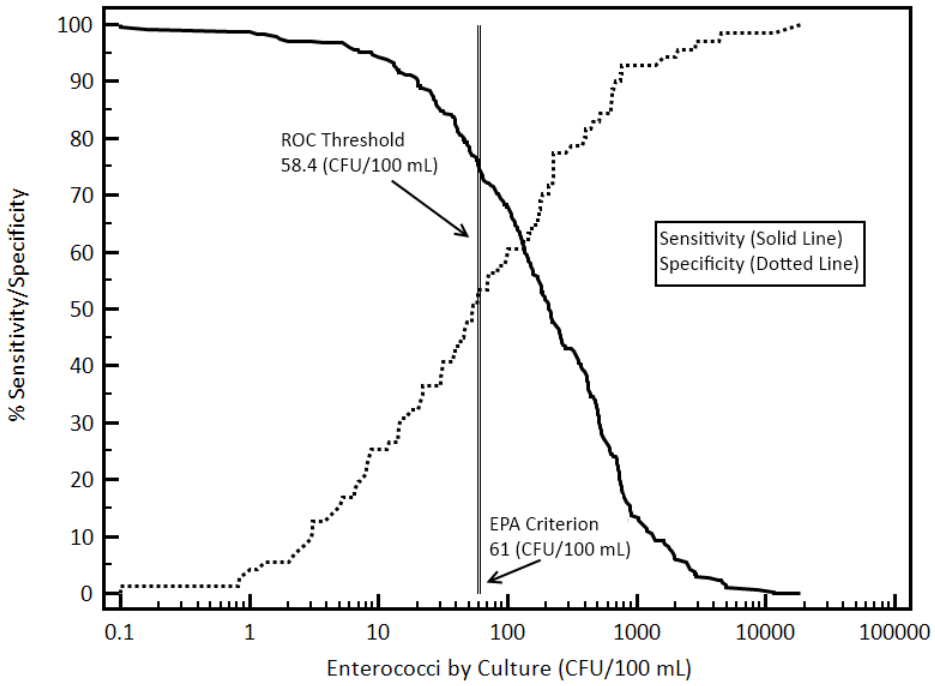

Figure 19. Sensitivity and specificity for culture-based densities of enterococci—*Giardia*.

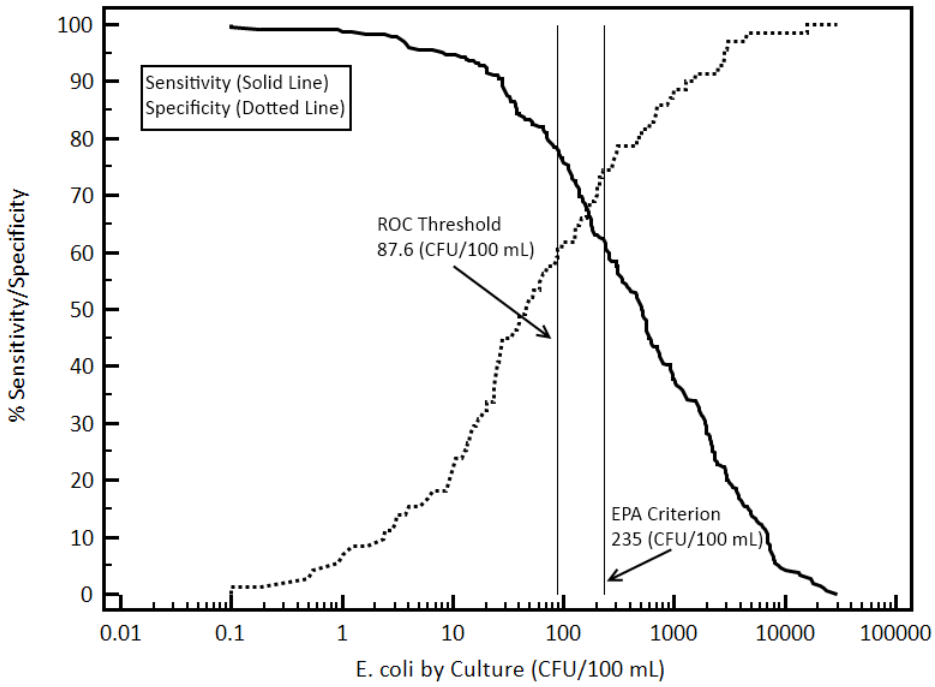

Figure 20. Sensitivity and specificity for culture-based densities of *E. coli*—*Giardia*.

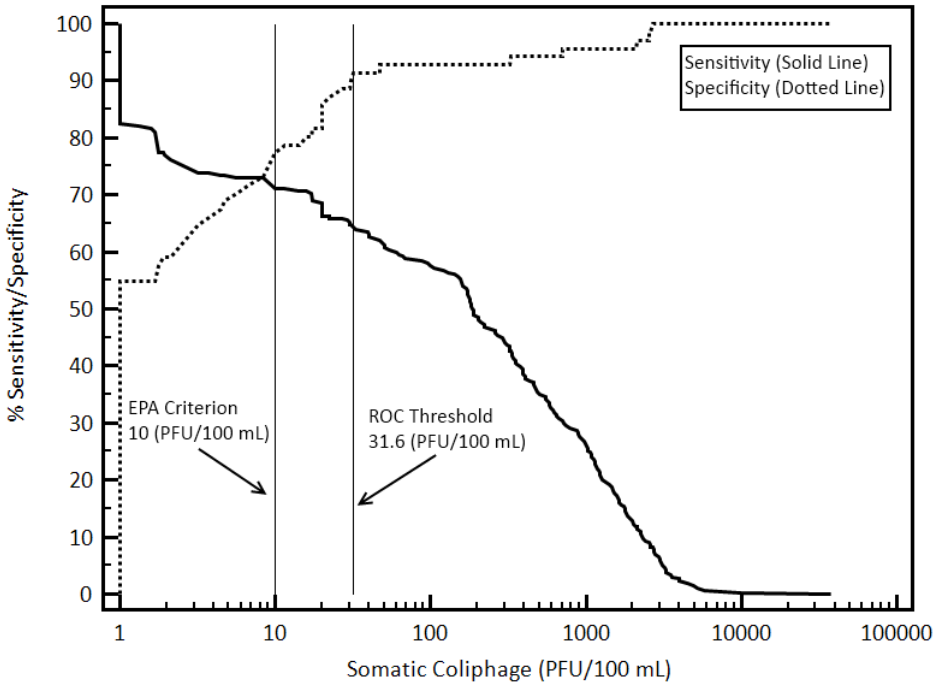

Figure 21. Sensitivity and specificity for densities of somatic coliphage—*Giardia*.

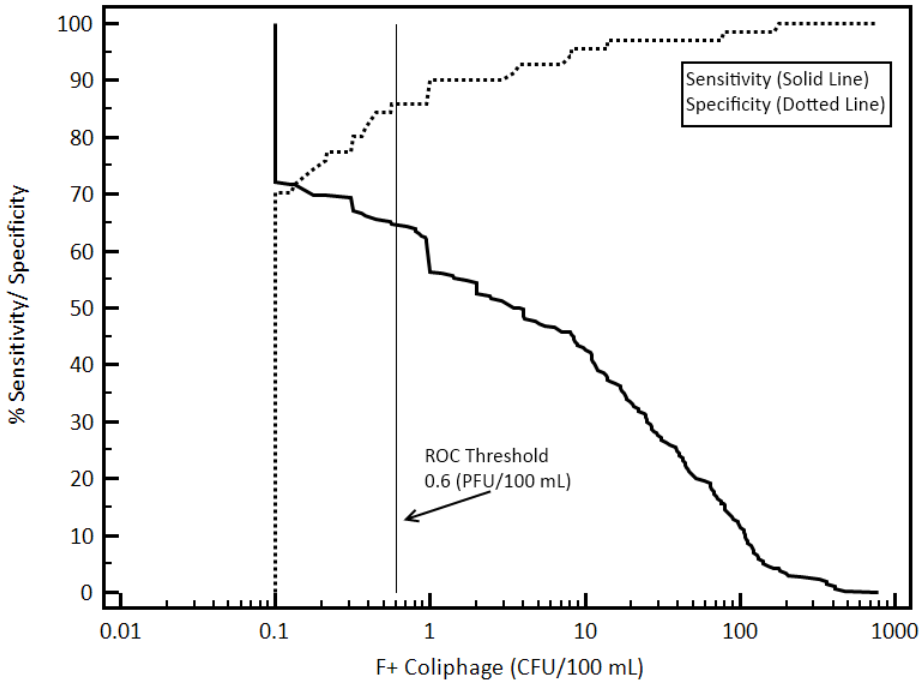

Figure 22. Sensitivity and specificity for densities of F+ coliphage—*Giardia*.

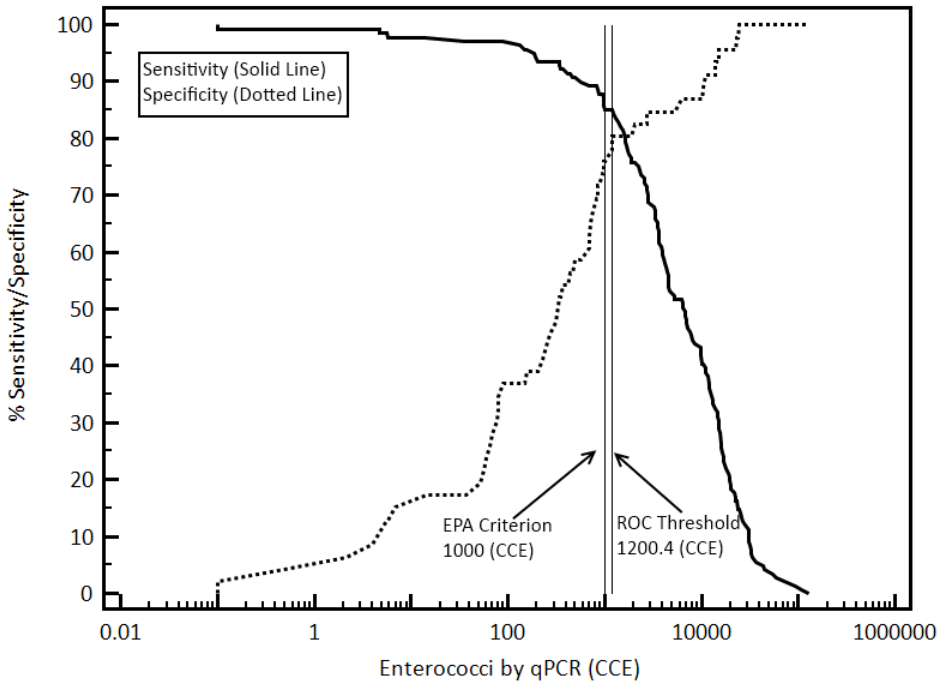

Figure 23. Sensitivity and specificity for qPCR-based densities of enterococci—*Giardia*.

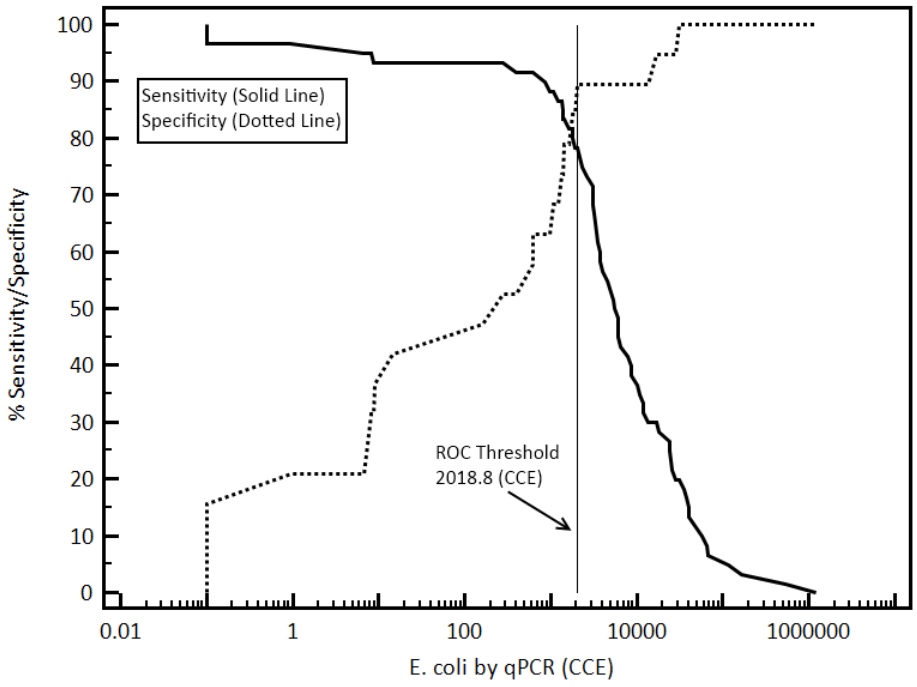

Figure 24. Sensitivity and specificity for qPCR-based densities of *E. coli*—*Giardia*.

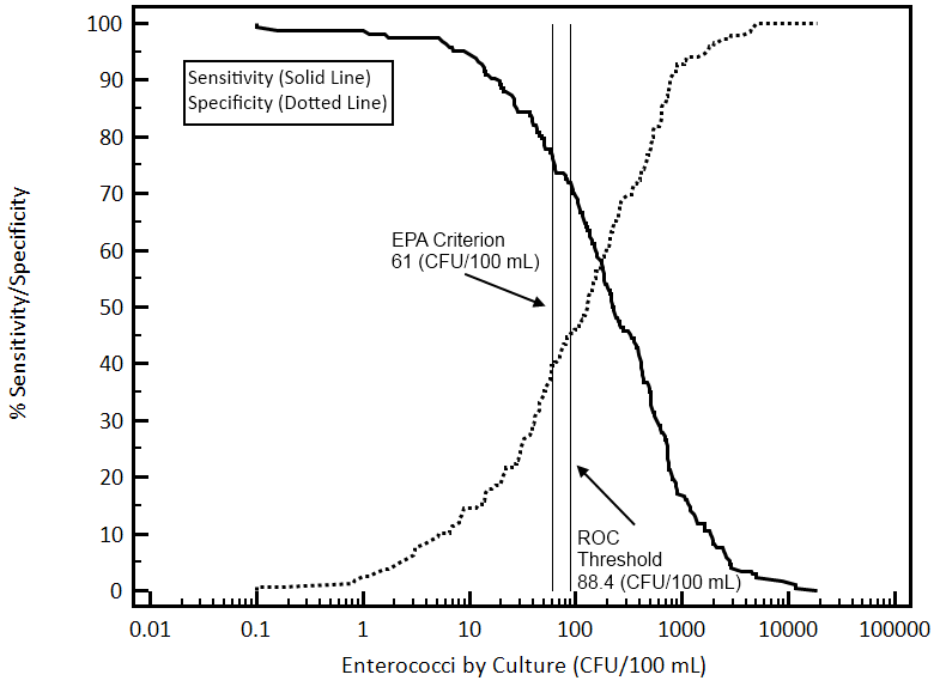

Figure 25. Sensitivity and specificity for culture-based densities of enterococci—*Cryptosporidium*.

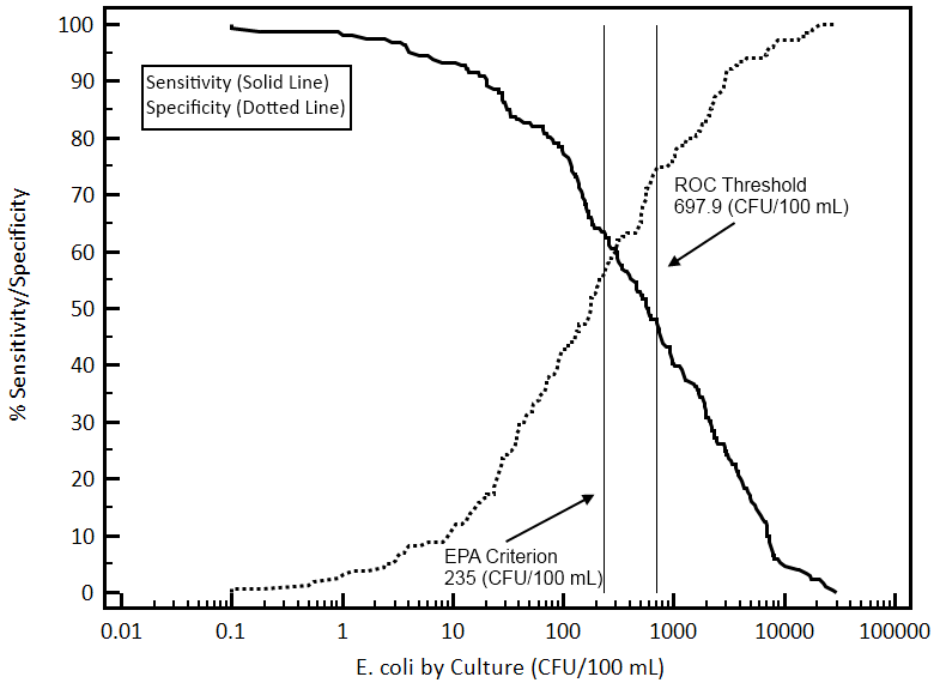

Figure 26. Sensitivity and specificity for culture-based densities of *E. coli—Cryptosporidium*.

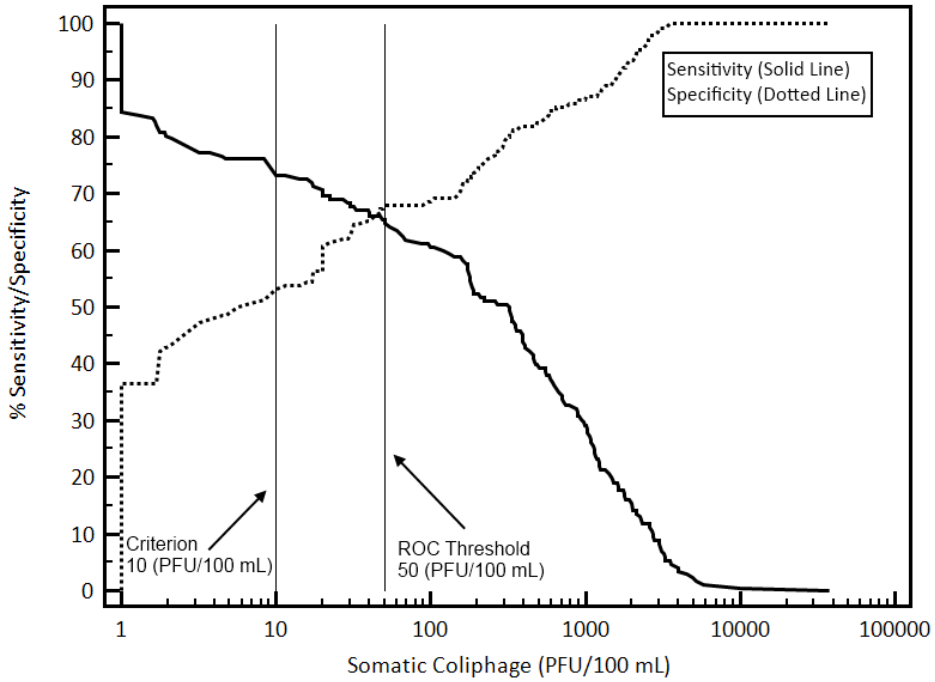

Figure 27. Sensitivity and specificity for somatic coliphage—*Cryptosporidium*.

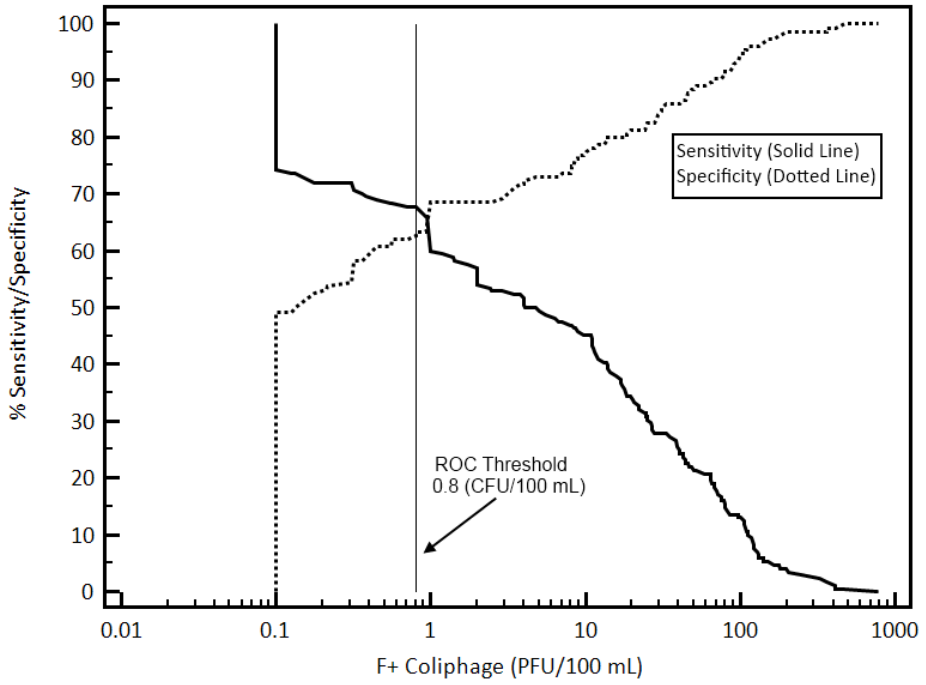

Figure 28. Sensitivity and specificity for F+ coliphage—*Cryptosporidium*.

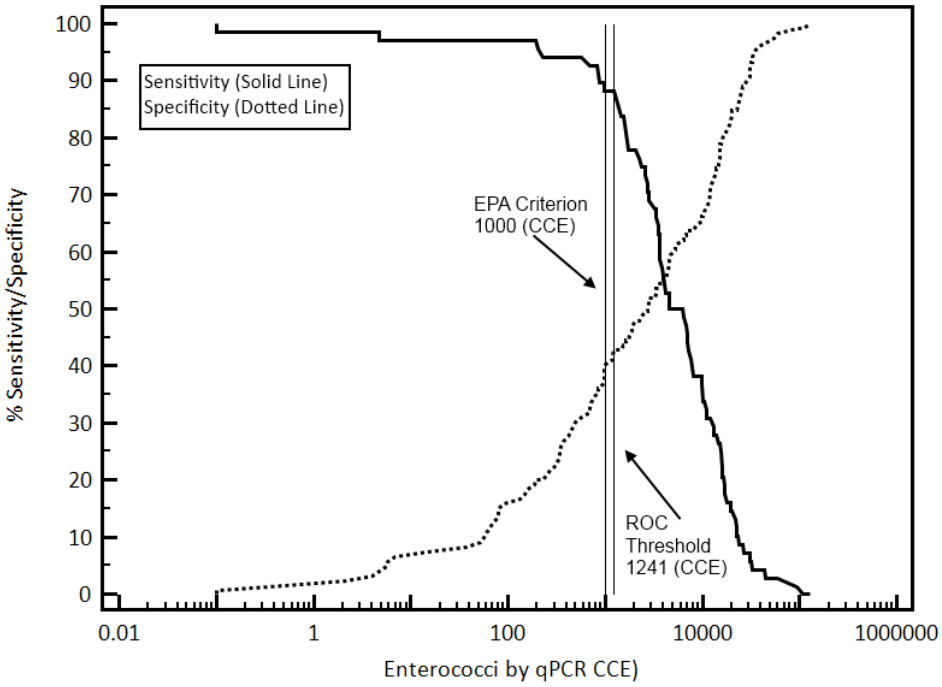

Figure 29. Sensitivity and specificity for qPCR-based densities of enterococci—*Cryptosporidium*.

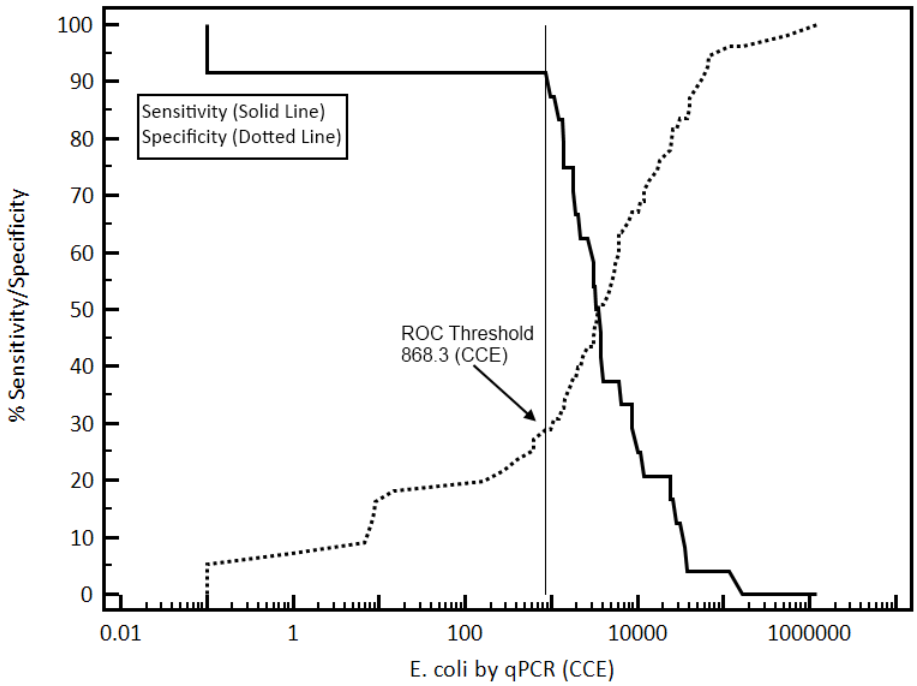

Figure 30. Sensitivity and specificity for qPCR-based densities of *E. coli*—*Cryptosporidium*.

In the general analysis, the optimal culture-based enterococci threshold for predicting *Giardia* presence was 58.4 CFU/100 mL, a value very close to the EPA fresh water quality criterion, 61 CFU/100 mL. Although the current EPA criterion is nearly optimized with regard to sensitivity and specificity, the AUC (0.672) suggests that enterococci by culture is a relatively weak predictor of *Giardia*. The confidence interval for this tests ranges from 0.618 to 0.723, indicating the test is only slightly better than chance alone at correctly identifying the presence of *Giardia*. The optimal prediction threshold for the culture-based enterococci test of *Cryptosporidium* yielded a value of 88.4 (CFU/100mL) with an AUC of 0.613. The difference in areas between the culture-based enterococci assays of the two pathogens is not statistically significant, and the sensitivity and specificity at the optimal threshold for *Giardia* (75.9 and 53.3, respectively) are not substantially different than the corresponding values for *Cryptosporidium* (72.0 and 45.5, respectively). Taking into the consideration the confidence intervals and the strength of other indicator tests, the culture-based enterococci assay is at best a weak to moderate predictor of both pathogens, even at the best threshold.

Then again, the optimal threshold for enterococci by qPCR (1200 CCE) corresponds to an AUC of 0.846 with a confidence interval of 0.747 to 0.917. The differences in AUCs show that enterococci by qPCR is a significantly better test of *Giardia* than the culture-based assay. Comparing the sensitivity and specificity of the optimal threshold for the qPCR-based assay (85.1 and 80.4, respectively) with the corresponding values for enterococci by culture, both for the optimal ROC threshold (75.9 and 53.5, respectively) and the EPA criteria (72.5 and 50.0, respectively) demonstrate the that differences are largely in specificity. With regard to *Cryptosporidium*, the general analysis shows enterococci by qPCR to be a significantly weaker

test of that pathogen—the optimal threshold of enteroccoci by qPCR is a sensitive (88.2) for the presence of *Cryptosporidium*, but not specific (42.9). Again, the difference is largely in specificity.

Comparison of the culture-based and qPCR-based assays for *E. coli* shows a similar pattern. The two tests performed at exactly the same sensitivity for the prediction of *Giardia*; however, the specificity of those two assays was markedly different. The specificity of the culture-based assay (60.6) was 32% lower than the specificity of the qPCR-based assay (89.5). This pattern was reversed for tests of *Cryptosporidium* presence. At the optimal ROC threshold, the culture-based *E. coli* test for *Cryptosporidium* was substantially more specific (75.0) than it was sensitive (48.2). On the other hand the qPCR-based assay was sensitivity (91.7) and not specific (29.1).

Somatic coliphage is a moderate-to-strong test for the presence of both *Giardia* and *Cryptosporidium*. With regard to *Giardia*, somatic coliphage is substantially more specific (91.6) test than it is sensitive (64.4). Specificity decreases substantially for the test of *Cryptosporidium* (from 91.6 to 68.0). The sensitivity and specificity of F+ coliphage follow a similar pattern between the two pathogens. This indicator test is also highly specific for *Giardia* (85.9) but not for *Cryptosporidium* (68.0). Thus, the optimal thresholds of both coliphage assays have nearly equal sensitivity rates for *Giardia* and *Cryptosporidium*, the specificity rates of both tests drop for the second pathogen.

Table XIX lists the optimal thresholds for every indicator test of *Giardia* within the general analysis (ALL) and for each category studied and Table XX shows the same for tests of tests of *Cryptosporidium*.

Table XIX

OPTIMAL THRESHOLDS FOR *GIARDIA* FOR ALL INDICATORS—*GIARDIA*

	Enterococci (cx)	*E. coli* (cx)	S. Coliphage	F+ Coliphage	Enterococci (qPCR)	*E. coli* (qPCR)
Units	CFU/100mL	CFU/100mL	PFU/100mL	PFU/100mL	CCE	CCE
Criteria	61	235	10	n/a	1000	n/a
ALL	58.43	87.64	31.62	0.56	1200.44	2018.82
CAWS	227.38	1567.00	20.00	8.12	1200.00	n/a
GUW	184.58	27.51	1.00	0.45	950.00	184.58
AM	31.50	907.62	20.00	8.12	2044.50	0.10
PM	99.01	87.64	31.62	0.56	830.74	2018.82
DRY	20.98	46.28	31.62	0.56	748.33	1788.85
WET	99.01	296.65	47.38	0.22	1200.44	2018.82
ABOVE	38.06	174.04	20.00	1.00	340.00	n/a
BELOW	226.31	1567.00	20.00	8.12	15000.00	n/a
BRIGHT	99.01	87.64	47.38	0.56	1200.44	2018.82
DARK	31.50	907.62	20.00	0.45	1200.00	615.13
CSO	226.31	3075.71	20.00	8.12	1200.00	n/a
NO CSO	71.69	174.04	20.00	1.00	340.00	n/a

Table XX

OPTIMAL THRESHOLDS FOR *GIARDIA* FOR ALL INDICATORS—*CRYPTOSPORIDIUM*

	Enterococci (cx)	*E. coli* (cx)	S. Coliphage	F+ Coliphage	Enterococci (qPCR)	*E. coli* (qPCR)
Units	CFU/100mL	CFU/100mL	PFU/100mL	PFU/100mL	CCE	CCE
Criteria	61	235	10	n/a	1000	n/a
ALL	88.43	697.93	50.00	0.80	1240.97	868.33
CAWS	256.36	3075.71	307.90	1.00	1240.97	3747.00
GUW	644.68	27.51	1.00	0.45	466.35	1064.03
AM	708.16	110.16	20.00	1.00	6300.00	3700.00
PM	155.16	681.12	47.38	0.80	830.74	1064.03
DRY	5.00	52.36	42.43	0.80	748.33	1788.85
WET	174.54	697.93	31.62	0.80	1240.97	625.70
ABOVE	81.85	23.22	20.00	1.00	16000.00	3747.00
BELOW	1033.00	3075.71	307.90	33.00	30740.85	3700.00
BRIGHT	174.54	309.62	47.38	0.95	830.74	1788.85
DARK	3.00	907.62	30.00	0.32	6500.00	3112.88
CSO	256.36	993.98	42.43	1.00	1240.97	11949.90
NO CSO	700.00	3000.00	307.90	98.23	24000.00	4000.00

Tables XIX and XX show considerable variability in the optimal thresholds of each indicator between paired categories. The optimal thresholds for the correct diagnosis of *Giardia* presence are substantially higher within CAWS locations or for samples collected during AM hours, WET weather, BELOW a wastewater treatment plant, in DARK conditions, and after a CSO event. These higher ROC thresholds correspond to the substantially higher densities of indicator microbes in those analysis categories. The findings follow a similar, though not identical pattern for *Cryptosporidium* (refer to Table XX). The relative magnitudes of optimal indicator thresholds for tests of *Cryptosporidium* between paired categories are highly irregular; this finding likely due to the lower diagnostic strength of tests of that pathogen.

7. DISCUSSION

For a given test, diagnostic strength with regard to a condition of interest is effectively evaluated by ROC analysis. With regard to water quality, indicators have been shown to be predictive of health outcomes, but their relationship with protozoan pathogens *Giardia* and *Cryptosporidium* has not been made clear. To the author's knowledge this is the only study to use ROC analysis in order to evaluate the diagnostic strength of indicator species with regard to the two pathogens of interest. The objectives of this study were five-fold:

1. Evaluate the general diagnostic strength of indicator microbes with regard to *Giardia* and *Cryptosporidium*.
2. Determine any difference in diagnostic strength of each indicator between the two pathogens.
3. Evaluate the diagnostic strength of indicators between locations groups, AM and PM sampling times, wet and dry weather conditions, presence and absence of CSO events, bright sunlight and low sunlight, and sampling location above and below a wastewater treatment plant.
4. Compare the diagnostic strength between indicators pairs within the general analysis and the categorical analyses.
5. Determine optimal indicator thresholds for the correct prediction of pathogen presence/absence.

Analysis of the general diagnostic strength of indicator microbes demonstrated that the qPCR-based assays for enterococci and *E. coli* were the best indicator microbes for the pathogen *Giardia*. Enterococci and *E. coli* assays based on qPCR were better able to discriminate *Giardia* presence under a variety of environmental conditions, as indicated by the consistency with which high AUC values were obtained (Table IX). The areas under the ROC curves for these tests were greater in magnitude than any other indicator in the general analysis and in each of the categorical subgroups, with the exception of NO CSO. Out of the ten best ROC curves (based on AUC values), nine were based on the two qPCR assays. The relative performance of indicator assays with regard to the pathogen *Cryptosporidium* is less clear. The coliphage assays yielded the highest AUC values for the general analysis. Within most categorical sub-groups, the coliphage assays had higher AUC values than the culture-based assays for enterococci and *E. coli*, but were not clearly superior to the qPCR-based assays of the same organisms. The former observation was partially corroborated by examining the top ten AUC values for *Cryptosporidium* in which qPCR assays, and *E. coli* by qPCR in particular, held eight of ten places. This informal evaluation of diagnostic strength between indicators was borne out in statistical analyses between indicators.

There were also clear differences in the diagnostic strength of individual indicators between pathogens. In the present analysis, indicator values were generally more informative about the presence of *Giardia* than of *Cryptosporidium*. Within the general analysis, five out of six indicators were statistically better able to diagnose the presence or absence of *Giardia* than of *Cryptosporidium*; enterococci by culture was the only exception. In 72 categorical analyses, only 7 instances showed tests of *Cryptosporidium* to have higher AUC values than tests of

Giardia and none of those were statistically significant at α=0.05. Conversely, there were 65 categorical analyses in which tests of *Giardia* yielded higher AUC values than tests of *Cryptosporidium*, with 17 of those differences being statistically significant α=0.05 and 3 category-specific comparisons (WET, PM, BRIGHT) in which enterococci by qPCR was a better test of *Giardia* at the Šidák corrected α=0.00427.

In the CHEERS data, matrix spikes of *Giardia* and *Cryptosporidium* (oo)cysts had similar recovery rates (20% and 27%, respectively) (Dorevitch, 2011). However, precision, determined by concordance of results between a split sample, was higher for *Cryptosporidium* (+/-200%) than for *Giardia* (+/- 50%) (Dorevitch, 2011). Krometis et al. provided a possible explanation for this finding, showing that recovery of *Cryptosporidium* oocysts were significantly negatively correlated with particles size 2 to 10 μm when using EPA Method 1623 (Krometis et al., 2009). Differences in the recovery between *Giardia* and *Cryptosporidium* would have particularly strong impacts on the ROC curves generated in the present analysis since true positive and true negative conditions are categorical variables based on the pathogen detection or non-detection. Nevertheless, the ROC curves provide important information about pathogen prediction in light of these technical limitations. The ROC analysis shows that indicators provide more accurate information about the presence of *Giardia* than about the presence of *Cryptosporidium* in our setting and with the sample processing (CFC) and analysis methods employed. This information is valuable to water quality managers who need to understand the degree to which indicators reflect true water pollution and which pollutants are being indicated.

The finding of the general analysis, that most indicators are significantly better diagnostic tests of *Giardia* than of *Cryptosporidium*, was corroborated by substantial differences in specificity at the optimal threshold of prediction for nearly every indicator microbe. With the exception of the culture-based *E. coli*, the optimal thresholds for five indicator microbes showed nearly identical sensitivity rates between tests of *Giardia* and *Cryptosporidium*, demonstrating only a substantial loss in specificity for the latter. This finding has important implications for water quality managers. Managing beach closures on the basis of the optimal *Giardia* threshold will ensure the best TP and TN decisions for nearly all tests. On the other hand, beach closures based on optimal *Cryptosporidium* thresholds will show nearly identical TP identification while increasing the rate of false positive identification. This raises an important question: is it more important to correctly identify swimming waters free of *Giardia*, or is it more urgent to correctly identify waters contaminated with *Cryptosporidium*? The former criterion would yield the best diagnostic accuracy for a given indicator test, but only with regard to the pathogen *Giardia*. Using the latter criterion would create an increase in the rate of false alarms for beach closures with regard to *Cryptosporidium*.

The dependence of indicator-pathogen relationships upon environmental conditions that may influence microbe influx and survival—location group, time of sample collection, rainfall condition, position with regard to wastewater treatment plant, sunlight level, and CSO events—was explored by fitting ROC curves to two strata within each category of analysis. Although conditions known to promote indicator survival (DARK, PM) or contribute to absolute indicator densities (CAWS, CSO, WET) did show higher average values for indicator densities and generally better AUCs values, there was only one relationship statistically significant at

α=0.000427 (enterococci by qPCR was a better test in DARK conditions). Nor were there any specific conditions that demonstrated the best AUC values for any given indicator. The results also showed that such changes did not significantly impact accurate prediction of pathogen presence or absence. ROC analysis provides an avenue by which to tailor water quality criteria to microbial conditions within a specific location. For instance, water locations with higher indicator densities might be better managed at a higher decision threshold with regard to pathogen presence/absence. This circumvents a problem posed by broadly-applied national criteria in which the same water quality standard is applied across water bodies with differing microbial ecologies. Consequently, location specific analyses of the type conducted here might provide specific information about a particular location—this information would hinge on pathogen levels rather than illness rates but the validity of those findings would at least be tailored to a given body of water.

Comparisons between relative indicator strengths might be the most definitive finding of this study. Currently, water quality managers typically employ EPA Method 1603 in order to evaluate water quality at a geometric mean (GM) density of 126 CFUs per 100 mL and a STV of 235 CFU per 100 mL. The optimal threshold as determined by ROC analysis for *Giardia* (87.6 CFU per 100 mL) does not substantially improve on the sensitivity and specificity of the EPA criterion value 235 CFU/100mL, while the ROC optimal threshold for *Cryptosporidium* (698 CFU per 100 mL) is both a weaker diagnostic threshold and significantly less protective for acute GI illness. Nevertheless, ROC analysis does provide an important alternative to the currently used method for the management of recreational freshwaters. Namely, enterococci by qPCR and to a lesser extent, *E. coli* by qPCR, provide a method for evaluating water quality that not only

yields same-day results but is also significantly more predictive for the pathogen *Giardia*. Enterococci by qPCR was a statistically better test than *E. coli* by culture in five separate analyses, including the general analysis. The qPCR-based assay for *E. coli* was the strongest test of *Giardia* based on AUC values, but the limited sample size for this test meant that comparisons with other indicators did not reach statistical significance. Somatic coliphage and F+ coliphage also appeared to be important tests of the two pathogens; in particular, somatic coliphage emerged as the only test which was a strong predictor of both pathogens and statistically better test than *E. coli* by culture. In fact the EPA freshwater criteria, dependent on viable enterococci and *E. coli* microbes and requiring 24 hours to process, were the weakest tests in the entire analysis.

The optimal ROC threshold for qPCR-based assays of enterococci for *Giardia* (1200 CCE) is a sensitive (85.1) and specific (80.4) test for that pathogen. At approximately that same threshold value (1240 CCE), enterococci is a sensitive (91.7), though not at all specific (29.1), test of *Cryptosporidium* as well. The best available threshold for the qPCR-based enterococci assay is also close to the EPA recommended criterion for water quality (1000 CCE, STV). Considering that this assay yields same-day results for water-quality, it appears to be a strong alternative to the currently used EPA water quality criteria for culture-based assays of enterococci and *E. coli*. A similar assessment can be made for *E. coli* by qPCR, though the optimal threshold for the two pathogens do not correspond as closely. Finally, the coliphage assays, though superior predictors of the two pathogens, do not yield same-day results for water quality as do the rapid qPCR based assays.

8. LIMITATIONS AND FUTURE WORK

Although the data made it possible to study diagnostic strength among the six indicators, limitations in sample size, and lopsided distributions of TP and TN samples for the two pathogens limited the analysis. Sample size determines, to a large degree, the width of confidence intervals, which broaden as sample size shrinks reducing the ability to statistical detect differences in AUCs. This limitation in power limited the evaluation the diagnostic superiority of the qPCR-based assay for *E. coli*, in particular. The sample size required to detect differences between two AUCs at chosen levels of Type I and Type II errors was calculated using the method of Hanley and McNeil for *Giardia* presence discrimination (Table XXI): the sample size was sufficiently large to detect statistical differences between AUCs for only five comparisons. In two other instances, statistically significant differences were detected even through the sample size estimated by the Hanley-McNeil method was not attained. Large differences in the AUCs between enterococci by culture and *E. coli* by culture should have reached significance for a sample size of 330; the actual sample size (324) is extremely close to this value and probably reflects differences in correlation analysis between the two methods. On the other hand, a similar difference in AUCs between enterococci by qPCR and *E. coli* by culture reached significance under the DeLong method despite having a significantly smaller sample size than required by Hanley and McNeil. It is unclear which statistical properties account for the difference in this assessment. Acquiring data with the aim of creating roughly equal numbers of samples between TP and TN events would improve the diagnostic evaluation of a given test. In the present analysis, the diagnostic evaluation of *E. coli* by qPCR with regard to *Giardia* was limited in several sample groups with no TN values for *Giardia*. Additionally,

conducting indicator tests of equal numbers of sampling observations would allow all data points for a given test to be included in the DeLong comparisons between dependent assays. Ultimately, increased sample sizes would increase statistical power, making it easier to distinguish diagnostic differences between tests.

Table XXI

SAMPLE SIZES NEEDED TO REACH STATISTICAL SIGNIFICANCE BETWEEN PAIRS OF INDICATORS

	Giardia (+)	N[a]	*Giardia* (-)	N[b]	$Area_1$	$Area_2$	N[c]	N[d]
enterocx~ ecolicx	0.388	253	0.050	71	0.672	0.745	330	324**
enterocx ~ somcoli	0.206	253	0.159	71	0.672	0.788	134	324*
enterocx~ malcoli	0.473	253	0.096	71	0.672	0.773	158	324*
enterocx ~ enteroq	0.457	141	0.077	46	0.685	0.837	68	187*
enterocx ~ ecoliq	0.410	60	-0.091	19	0.773	0.846	267	79
ecolicx ~ somcoli	0.248	253	0.219	71	0.745	0.788	793	324
ecolicx ~ malcoli	0.307	253	0.223	71	0.745	0.773	1810	324
ecolicx ~ enteroq	0.210	141	0.482	46	0.767	0.837	239	187**
ecolicx ~ ecoliq	0.371	60	0.841	19	0.852	0.846	15509	79
somcoli ~ malcoli	0.279	253	0.757	71	0.788	0.773	4009	324
somcoli ~ enteroq	0.092	141	0.601	46	0.812	0.837	1606	187
somcoli~ ecoliq	0.212	60	0.102	19	0.872	0.846	1503	79
malcoli~ enteroq	0.194	141	0.523	46	0.771	0.837	263	187
malcoli ~ ecoliq	0.497	60	0.806	19	0.846	0.846	∞	79
enteroq ~ ecoliq	-0.012	60	0.887	19	0.941	0.846	79	79*

[a] Number of TP *Giardia* observations.

[b] Number of TN *Giardia* observations.

[c] Sample required to reach statistical significance

[d] Actual sample size.

[*] Observation that reached statistical significance under both tests.

[**] Observation reached statistical significance only under DeLong.

Another major limitation of this study was the difference in measurement precision of *Giardia* and *Cryptosporidium*. Although the performance of each indicator test was compared between the two pathogens, the precision with which densities of pathogens were determined varied considerably according to CHEERS data (CHEERS final report). Future ROC analysis of indicators with regard to pathogens would do well to look for alternative means of determining the presence or absence of *Cryptosporidium* or omit this pathogen altogether in favor of another pathogen or group of pathogen whose presence or absence can be determined more precisely.

Final limitation of the current results reflects the equal importance assigned to sensitivity and specificity by the Youden Index. It is by no means clear whether public health officials would seek to preserve sensitivity over specificity with regard to decisions threshold based on indicator tests, or vice versa. Preserving sensitivity over specificity would be more protective of public health, with regard to pathogen presence/absence. However, disproportionate preference for correct positive responses would create a higher rate of false alarms—that is, close beaches in which pathogens were not present in the water. A cost-benefit analysis is an important aspect of determining the relative importance of sensitivity and specificity with regard to metric of public health.

This examination provides specific information about the performance of six distinct indicator tests for the presence or absence of two pathogens in surface water in and around the city of Chicago.

CITED LITERATURE

American Public Health Association. 1922. “Report of the Committee on Bathing Places.” *American Journal of Public Health* 12: 121–123.

Bamber, D. 1975. “The Area Above the Ordinal Dominance Graph and the Area Below the Receiver Operating Characteristic Graph.” *Journal of Mathematical Psychology* 12 (4): 387–415.

Bisson, J. W., and V. J. Cabelli. 1980. “Clostridium Perfringens as a Water Pollution Indicator.” *Journal of the Water Pollution Control Federation* 52 (2) (February): 241–248.

Bryan, J. A., J. D. Lehmann, I. F. Setiady, and M. H. Hatch. 1974. “An Outbreak of Hepatitis-A Associated with Recreational Lake Water.” *American Journal of Epidemiology* 99 (2) (February): 145–154.

Byappanahalli, M., M. Fowler, D. Shively, and R. Whitman. 2003. “Ubiquity and Persistence of Escherichia Coli in a Midwestern Coastal Stream.” *Applied and Environmental Microbiology* 69 (8) (August): 4549–4555.

Cabelli, V. J. 1978. “Swimming Associated Disease Outbreaks.” *Journal of the Water Pollution Control Federation* 50 (6): 1374–1377.

Cabelli, V. J. 1983. “Health Effects Criteria for Marine Waters.” EPA-600/1-80-031. U.S. Environmental Protection Agency, Health Effects Laboratory, Research. Triangle Park, NC.

Cabelli, V. J., H. Kennedy, and M. A. Levin. 1976. “Pseudomonas Aeruginosa-fecal Coliform Relationships in Estuarine and Fresh Recreational Waters.” *Journal of the Water Pollution Control Federation* 48 (2) (February): 367–376.

Cabelli, V. J., A. P. Dufour, M. A. Levin, L. J. McCabe, and P. W. Haberman. 1979. “Relationship of Microbial Indicators to Health Effects at Marine Bathing Beaches.” *American Journal of Public Health* 69 (7) (July): 690–696.

Cabelli, V.J., A.P. Dufour, L.J. McCabe, and M.A. Levin. 1982. “Swimming-associated Gastroenteritis and Water Quality.” *American Journal of Epidemiology* 115, (4): 606–616.

Cabelli, V. J., A. P. Dufour, L. J. McCabe, and M. A. Levin. 1983. “A Marine Recreational Water Quality Criterion Consistent with Indicator Concepts and Risk Analysis.” *Journal of the Water Pollution Control Federation* 55 (10) (October): 1306–1314.

Chang, K., and D. J. Stickler. 1985. “An Assessment of Bacteroides Fragilis Group Organisms as Indicators of Human Faecal Pollution.” *The Journal of Applied Bacteriology* 58 (1) (January): 95–99.

Chauret, C., S. Springthorpe, and S. Sattar. 1999. “Fate of *Cryptosporidium* Oocysts, *Giardia* Cysts, and Microbial Indicators During Wastewater Treatment and Anaerobic Sludge Digestion.” *Canadian Journal of Microbiology* 45 (3) (March): 257–262. doi:10.1139/w99-001.

Chern, E. C., S. Siefring, J. Paar, M. Doolittle, and R. A. Haugland. 2011. “Comparison of Quantitative PCR Assays for Escherichia Coli Targeting Ribosomal RNA and Single Copy Genes.” *Letters in Applied Microbiology* 52 (3) (March): 298–306.

Cheung, W. H. S., K. C. K. Chang, R. P. S. Hung, and J. W. L. Kleevens. 1990. “Health Effects of Beach Water Pollution in Hong Kong.” *Epidemiology & Infection* 105 (01): 139–162.

Cohen, J., and H. I. Shuval. 1972. “Coliforms, Fecal Coliforms, and Fecal Streptococci as Indicators of Water Pollution.” *Water, Air, & Soil Pollution* 2 (1): 85–95.

Corbett, S. J., G. L. Rubin, G. K. Curry, and D. G. Kleinbaum. 1993. “The Health Effects of Swimming at Sydney Beaches.” *American Journal of Public Health* 83 (12) (December): 1701–1706.

DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson. 1988. “Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: a Nonparametric Approach.” *Biometrics* 44 (3) (September): 837–845.

Dennis, J. M., and A. Wolman. 1959. “1955–56 Infectious Hepatitis Epidemic in Delhi, India.” *American Water Works Association* 51 (10) (October): 1288–1298.

Dorevitch, S. 2011. *The Chicago Health, Environmental Exposure, and Recreation Study (CHEERS)*. Environmental Monitoring and Research Division. Chicago, IL: Metropolitan Water Reclamation District of Greater Chicago.

Dufour, A. P. 1984. “Health Effects Criteria for Fresh Recreational Waters.” EPA-600/1-84-004, Office of Research and Development, US Environmental Protection Agency, Cincinnati, OH.

Dufour, A.P. 2007. “The Evolution of Water Quality Criteria in the United States, 1922–2003.” In Statistical Framework for Recreational Water Quality Criteria and Monitoring, edited by L.J. Wymer. West Sussex, England: John Wiley & Sons: 1–12.

Eisenberg, J. N., E. Y. W. Seto, A. W. Olivieri, and R. C. Spear. 1996. “Quantifying Water Pathogen Risk in an Epidemiological Framework.” *Risk Analysis* 16 (4): 549–563.

Eisenberg, J. N. S., M. A. Brookhart, G. Rice, M. Brown, and J. M. Colford. 2008. "Disease Transmission Models for Public Health Decision Making: Analysis of Epidemic and Endemic Conditions Caused by Waterborne Pathogens." *Environmental Health Perspectives* 110 (8) (June): 783–790.

Evanson, M., and R. F. Ambrose. 2006. "Sources and Growth Dynamics of Fecal Indicator Bacteria in a Coastal Wetland System and Potential Impacts to Adjacent Waters." *Water Research* 40 (3) (February): 475–486.

Ferley, J. P., D. Zmirou, F. Balducci, B. Baleux, P. Fera, G. Larbaigt, E. Jacq, B. Moissonnier, A. Blineau, and J. Boudot. 1989. "Epidemiological Significance of Microbiological Pollution Criteria for River Recreational Waters." *International Journal of Epidemiology* 18 (1) (March 1): 198–205.

Field, K. G., and M. Samadpour. 2007. "Fecal Source Tracking, the Indicator Paradigm, and Managing Water Quality." *Water Research* 41 (16) (August): 3517–3538.

Fleisher, J. M. 1991. "A Reanalysis of Data Supporting U.S. Federal Bacteriological Water Quality Criteria Governing Marine Recreational Waters." *Journal of the Water Pollution Control Federation* 63 (3) (May 1): 259–265.

Fogarty, L. R., S. K. Haack, M. J. Wolcott, and R. L Whitman. 2003. "Abundance and Characteristics of the Recreational Water Quality Indicator Bacteria Escherichia Coli and Enterococci in Gull Faeces." *Journal of Applied Microbiology* 94 (5): 865–878.

Foster, D. H., N. B. Hanes, and S. M. Lord. 1971. "A Critical Examination of Bathing Water Quality Standards." *Journal of the Water Pollution Control Federation* 43 (11) (November 1): 2229–2241.

Geldenhuys, J. C., and P. D. Pretorius. 1989. "The Occurrence of Enteric Viruses in Polluted Water, Correlation to Indicator Organisms and Factors Influencing Their Numbers." *Water Science Technology* 21: 105–109.

Gerba, C. P., and J. E. Smith. 2005. "Sources of Pathogenic Microorganisms and Their Fate During Land Application of Wastes." *Journal of Environmental Quality* 34 (1): 42–48.

Halliday, E., and R. J. Gast. 2011. "Bacteria in Beach Sands: An Emerging Challenge in Protecting Coastal Water Quality and Bather Health." *Environmental Science and Technology* 45 (2) (January 15): 370–379.

Hanes, N. B., W. B. Sarles, and G. A. Rohlich. 1964. "Dissolved Oxygen and Survival of Coliform Organisms and Enterococci." *Journal of the American Water Works Association* 56 (4) (April): 441–446.

Hanley, J. A., and B. J. McNeil. 1982. “The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve.” *Radiology* 143 (1) (April): 29–36.

Harwood, V. J., A. D. Levine, T. M. Scott, V. Chivukula, J. Lukasik, S. R. Farrah, and J. B. Rose. 2005. “Validity of the Indicator Organism Paradigm for Pathogen Reduction in Reclaimed Water and Public Health Protection.” *Applied and Environmental Microbiology* 71 (6) (June): 3163–3170.

Jaeger, H. 1892. “Die Aetiologie der infectiösen fieberhaften Icterus.” *Zeitschrift Für Hygiene Und Infectionskrankheiten* XII: 525–602.

Kappus, K. D., J. S. Marks, R. C. Holman, J. K. Bryant, C. Baker, G. W. Gary, and H. B. Greenberg. 1982. “An Outbreak of Norwalk Gastroenteritis Associated with Swimming in a Pool and Secondary Person-to-person Transmission.” *American Journal of Epidemiology* 116 (5) (November): 834–839.

Koopman, J. S., E. A. Eckert, H. B. Greenberg, B. C. Strohm, R. E. Isaacson, and A. S. Monto. 1982. “Norwalk Virus Enteric Illness Acquired by Swimming Exposure.” *American Journal of Epidemiology* 115 (2) (February): 173–177.

Lemarchand, K., and P. Lebaron. 2003. “Occurrence of Salmonella Spp. and *Cryptosporidium* Spp. in a French Coastal Watershed: Relationship with Fecal Indicators.” *FEMS Microbiology Letters* 218 (1): 203–209.

Lipp, E. K., S. A. Farrah, and J. B. Rose. 2001. “Assessment and Impact of Microbial Fecal Pollution and Human Enteric Pathogens in a Coastal Community.” *Marine Pollution Bulletin* 42 (4) (April): 286–293.

MacKenzie, W. R., N. J. Hoxie, M. E. Proctor, M. S. Gradus, K. A. Blair, D. E. Peterson, J. J. Kazmierczak, D.G. Addiss, K.R. Fox, J.B. Rose, J.P. Davis. 1994. “A Massive Outbreak in Milwaukee of *Cryptosporidium* Infection Transmitted Through the Public Water Supply.” *The New England Journal of Medicine* 331 (3) (July): 161.

Metz, C. E. 1978. “Basic Principles of ROC Analysis.” *Seminars in Nuclear Medicine* 8 (4) (October): 283–298.

Miescier, J. J., and V. J. Cabelli. 1982. “Enterococci and Other Microbial Indicators in Municipal Wastewater Effluents.” *Journal of the Water Pollution Control Federation* 54 (12) (December): 1599–1606.

Moore, B. 1954. “A Survey of Beach Pollution at a Seaside Resort.” *The Journal of Hygiene* 52 (1) (March): 71–86.

Moore, B., A. R. Blowers, P. B. Crone, A. E. A. Hughes, R. I. Hutchinson, and J. E. Jameson. 1959. "Sewage Contamination of Coastal Bathing Waters in England and Wales: A Bacteriological and Epidemiological Study." *The Journal of Hygiene* 57 (4) (December): 435–472.

National Technical Advisory Committee (NTAC). 1968. *Water Quality Criteria*. Washington, DC: Federal Water Pollution Control Administration, Department of the Interior.

Noble, R. T., D. F. Moore, M. K. Leecaster, C. D. McGee, and S. B. Weisberg. 2003. "Comparison of Total Coliform, Fecal Coliform, and Enterococcus Bacterial Indicator Response for Ocean Recreational Water Quality Testing." *Water Research* 37: 1637–1643.

Obiri-Danso, K., K. Jones, and N. Paul. 1999. "The Effect of the Time of Sampling on the Compliance of Bathing Water in New England with the EU Directive on Bathing Water Quality." *Journal of Coastal Conservation* 5 (1) (January): 51–58.

Platt, P. S. 1923. *Report on New Haven Health Center Demonstration*. New Haven, Connecticut: Connecticut Department of Health.

Prüss, A. 1998. "Review of Epidemiological Studies on Health Effects from Exposure to Recreational Water." *International Journal of Epidemiology* 27 (1) (February): 1–9.

Rijal, G., C. Petropoulou, J. K. Tolson, M. DeFlaun, C. Gerba, R. Gore, T. Glymph, T. Granato, C. O'Connor, L. Kollias, R. Lanyon. 2009. "Dry and Wet Weather Microbial Characterization of the Chicago Area Waterway System." *Water Science and Technolgy* 60, (7): 1847–1855.

Rijal, G., J.K. Tolson, C. Petropoulou, T.C. Granato, A. Glymph, C. Gerba, M.F. Deflaun, C. O'Connor, L. Kollias, and R. Lanyon. 2011. "Microbial Risk Assessment for Recreational Use of the Chicago Area Waterway System." *Journal of Water and Health* 9, (1): 169–186.

Robson, J. E. 1956. "A Bacterial Method for Tracing Sewage Pollution." *Journal of Applied Microbiology* 19 (2): 243–246.

Rosenberg, M. L., K. K. Hazlet, J. Schaefer, J. G. Wells, and R. C. Pruneda. 1976. "Shigellosis from Swimming." *The Journal of the American Medical Association* 236 (16) (October): 1849–1852.

Rosenfeld, L. K., C. D. McGee, G. L. Robertson, M. A. Noble, and B. H. Jones. 2006. "Temporal and Spatial Variability of Fecal Indicator Bacteria in the Surf Zone Off Huntington Beach, CA." *Marine Environmental Research* 61 (5) (June): 471–493.

Sanders, B. F., F. Arega, and M. Sutula. 2005. “Modeling the Dry-weather Tidal Cycling of Fecal Indicator Bacteria in Surface Waters of an Intertidal Wetland.” *Water Research* 39 (14) (September): 3394–3408.

Savage, H. P., and N. B Hanes. 1971. “Toxicity of Seawater to Coliform Bacteria.” *Journal of the Water Pollution Control Federation* 43 (5) (May): 854–861.

Schleifer, K. H., and R. Kilpper-Bälz. 1984. “Transfer of Streptococcus Faecalis and Streptococcus Faecium to the Genus Enterococcus Nom. Rev. as Enterococcus Faecalis Comb. Nov. and Enterococcus Faecium Comb. Nov.” *International Journal of Systematic Bacteriology* 34 (1) (January): 31–34.

Scott, W. J. 1932. “Survey of Connecticut’s Shore Bathing Waters.” *American Journal of Public Health* 22 (3) (March): 316–321.

Senn, C. L. 1963. “Coliform Standards for Recreational Waters. Progress Report of the Public Health Activities Committee.” *Journal of the Sanitary Engineering Division, Proceedings of the American Society of Civil Engineers* 89 (SA4): 57–94.

Seyfried, P. L., R. S. Tobin, N. E. Brown, and P. F. Ness. 1985a. “A Prospective Study of Swimming-related Illness. II. Morbidity and the Microbiological Quality of Water.” *American Journal of Public Health* 75 (9): 1071–1075.

Seyfried, P. L., R. S. Tobin, N. E. Brown, and P. F. Ness. 1985b. “A Prospective Study of Swimming-related Illness. I. Swimming-associated Health Risk.” *American Journal of Public Health* 75 (9) (September): 1068–1070.

Sinton, L. W., C. H. Hall, P. A. Lynch, and R. J. Davies-Colley. 2002. “Sunlight Inactivation of Fecal Indicator Bacteria and Bacteriophages from Waste Stabilization Pond Effluent in Fresh and Saline Waters.” *Applied and Environmental Microbiology* 68 (3) (March): 1122–1131.

Slanetz, L. W., and C. H. Bartley. 1957. “Numbers of Enterococci in Water, Sewage, and Feces Determined by the Membrane Filter Technique with an Improved Medium.” *Journal of Bacteriology* 74 (5) (November): 591–595.

Slanetz, L. W., and C. H. Bartley. 1964. “Detection and Sanitary Significance of Fecal Streptococci in Water.” *American Journal of Public Health* 54 (4) (April): 609–614.

Stevenson, A. H. 1949. “Water Quality Requirements for Recreational Uses.” *Sewage Works Journal* 21 (1) (January): 110–114.

Stevenson, A. H. 1953. “Studies of Bathing Water Quality and Health.” *American Journal of Public Health* (43): 529–538.

US EPA. 1986. “Ambient Water Quality Criteria for Bacteria—1986. Bacteriological Ambient Water Quality Criteria for Marine and Freshwater Recreational Waters.” EPA 440/5-84-002, Washington DC.

US EPA. 2005. “Method 1623: *Cryptosporidium* and *Giardia* in Water by Filtration/IMS/FA.” Washington, D.C..

US EPA. 2006. “Method 1600: Enterococci in Water by Membrane Filtration Using Membrane-Enterococcus Indoxyl-b-D-Glucoside Agar (mEI)”. EPA 821-R-06–009. Washington DC.

US EPA. 2006. “Method 1603: Escherichia Coli in Water by Membrane Filtration Using Modified Membrane-Thermotolerant Escherichia Coli Agar (Modified mTEC)”. EPA 821-R-06–011. Washington DC.

US EPA. 2010. “Method A: Enterococci in Water by TaqMan® Quantitative Polymerase Chain Reaction (qPCR) Assay”. 2010. Washington DC.

US EPA. 2011. “Draft Recreational Water Quality Criteria.” Office of Water 820-D-11-002. Washington DC.

Wade, T. J., N. Pai, J. N. S. Eisenberg, and John M. Colford Jr. 2003. “Do U.S. Environmental Protection Agency Water Quality Guidelines for Recreational Waters Prevent Gastrointestinal Illness? A Systematic Review and Meta-analysis.” *Environmental Health Perspectives* 111 (8) (June): 1102–1109.

Whitman, R. L., D. Shively, H. Pawlik, M. B. Nevers, and M. N. Byappanahalli. 2003. “Occurrence of Escherichia Coli and Enterococci in Cladophora (Chlorophyta) in Nearshore Water and Beach Sand of Lake Michigan.” *Applied and Environmental Microbiology* 69 (8): 4714–4719.

Whitman, R. L., M. B. Nevers, G. C. Korinek, and M. N. Byappanahalli. 2004. “Solar and Temporal Effects on Escherichia Coli Concentration at a Lake Michigan Swimming Beach.” *Applied and Environmental Microbiology* 70 (7) (July): 4276–4285.

Winslow, C. E. A., and D. Moxon. 1928. “Bacterial Pollution of Bathing Veach Waters in New Haven Harbor.” *American Journal of Epidemiology* 8 (3) (May): 299–310.

Zuckerman, U., and S. Tzipori. 2006. “Portable Continuous Flow Centrifugation and Method 1623 for Monitoring of Waterborne Protozoa from Large Volumes of Various Water Matrices.” *Journal of Applied Microbiology* 100, (6): 1220–1227.

VITA

NAME:	Burcu Yavuz
EDUCATION:	B.A., Human Development, University of Chicago, Chicago, Illinois, 2005
RESEARCH:	Environmental and Occupational Health Sciences Division, UIC School of Public Health; Laboratory of Dr. Samuel Dorevitch, 2010–2012
	Department of Chemistry, Northwestern University; Laboratory of Dr. Chad Mirkin, 2009–2010
	Children's Hospital of Oakland—Research Institute; Laboratory of Dr. Bruce Ames, 2006–2009
TEACHING:	Environmental and Occupational Health Sciences Division, UIC School of Public Health; Teaching assistant for Professor Margit Javor, 2012
	Environmental and Occupational Health Sciences Division, UIC School of Public Health; Teaching assistant for Professor Emily Ahonen, 2011
	The Princeton Review, Medical College Admission Test Training Program; Biology Instructor, 2010
PUBLICATIONS:	Suh, J. H., Kim, R., Yavuz, B., Lee, D., Lal, A., Ames, B. N., Shigenaga, M. K. "Clinical assay of four thiol amino acid redox couples by LC-MS/MS: utility in thalassemia." *J Chromatogr B Analyt Technol Biomed Life Sci*. 2009 Oct 15;877(28): 3418–3427.
	Ratcliffe, K. E., Tao, Q., Yavuz, B., Stoletov, K. V., Spring S.C., Terman, B. I. "Sck is expressed in endothelial cells and participates in vascular endothelial growth factor-induced signaling." *Oncogene*. 2002 Sep 12;21(41): 6307–6316.