EPA United States
Environmental Protection
Agency

The EMPACT Beaches Project:
Results From a Study on Microbiological Monitoring in Recreational Waters
August 2005

The EMPACT Beaches Project

Results from a Study on Microbiological Monitoring in
Recreational Waters

Larry J. Wymer
Kristen P. Brenner
John W. Martinson
Walter R. Stutts
Stephen A. Schaub*
Alfred P. Dufour

U.S. Environmental Protection Agency
Office of Research and Development
National Exposure Research Laboratory
Cincinnati, OH 45268

* U.S. EPA, Office of Water, Health and Ecological Criteria Division, Washington, DC

Acknowledgements

Table of Contents

Index of Figures

Index of Tables

x

Overview of the Study Results

Quantitative guidelines with respect to indicator organism densities in recreational waters that have been established by the U. S. EPA have been supported by human health studies, notably those performed in the late 40's and early 50's by the U.S. Public Health Service, and in the 70's and early 80's by the U.S. EPA. Upper limits of 35 enterococci per 100 mL in marine water and 33 enterococci per 100 mL in freshwater were recommended for the geometric mean of at least five samples taken over a 30-day period [U.S. EPA, 1986]. An upper limit on the geometric mean of 126 *E. coli* per 100 mL was recommended for freshwater environments only, since *E. coli* failed to show a good relationship to swimming-associated illnesses in marine waters. A recommendation was also made that, for a designated public beach, no single sample exceed the 75th percentile of a lognormal distribution with these respective geometric means. Lacking site-specific log standard deviations, this implied an upper limit on any single sample of 104 enterococci per 100 mL in marine water (log standard deviation of 0.7), and 98 enterococci or 235 *E coli* per 100 mL in freshwater (log standard deviation of 0.4).

Providing numerical guidelines that are supported by scientific data was a great advance. Lacking, however, was scientific data of comparable quality to support monitoring schemes that would provide the most meaningful information about whether the guidelines are being met. The EMPACT (Environmental Monitoring for Public Access and Community Tracking) Beaches project has attempted to identify those characteristics of a beach environment that have significant impact on monitoring. This project examined five beach environments to determine the factors that most influence the measurement of beach water quality. Two ocean beaches, an estuarine beach, a Great Lakes beach, and a riverine beach were selected to provide as broad a representation of beach environments as possible. This approach was used to develop model protocols that can be applied to similar beaches.

Factors found to influence indicator levels

For each of the five study beaches, the greatest single determinant of microbial indicator level was found to be the depth zone, or, roughly, distance from the shoreline at which the sample was collected. Bacterial densities became substantially lower as one moved from ankle-deep to knee-deep to chest-deep water. This fact has important implications for sample design as well as for public health.

In contrast, this study found no significant difference in indicator level among samples that were taken at different depths below the surface, such as between those taken the standard 0.3 M from the surface and those taken from near the bottom. At three of the five beaches, no substantial systematic differences were found among samples taken at different points parallel to the beachfront, which spanned a distance of 60 M, as long as they came from water of the same depth. Minor "hot spots" and "cold spots" were found at two of the beaches, although the differences were not dramatic and were not consistently found at all times.

Significant declines in indicator densities from the morning to the afternoon (9:00 a.m. to 2:00 p.m.) were observed at four of the study beaches. This effect was seen only on sunny days at one freshwater beach, but was observed to be independent of sunshine at three others, a freshwater beach and two marine beaches. Indicator levels at the remaining beach, a West Coast marine environment, tended to be very low at all times.

From one day to the next, the geometric mean indicator density among all samples that were collected changed by a factor of 2 (doubling or halving) or more about half the time at each studied beach.

There was virtually zero correlation between the indicator level on any given day and the indicator level more than two days later. The geometric mean of five samples taken over a 29-day period (that is, on the same day of the week) did not accurately represent actual conditions on any given day during the 29-day period.

Findings from this study with regard to spatial and temporal variation suggest the following implications for a beach monitoring program:

- Time and location of sampling must be carefully considered.
  - Depth zones from which samples are collected are likely to have great effect on the resulting estimate of indicator density. Sampling in knee- to waist-deep water would seem to offer a reasonable, but still conservative, approach to estimating water quality, particularly given that health effects are based on quality of water at waist-depth or deeper.
  - Sampling at 0.3 M below the surface is justified based on exposure considerations. This study failed to discern differences at lower collection depths.
  - Sampling in the morning will likely be a conservative practice, in addition to perhaps being a convenient time to sample.
  - Sampling should be performed as close as practical to the day on which a decision is to be made regarding beach closure or advisement. Preferably this should be the day before, given current conditions of a one day turnaround for the results.

- A number of samples should be collected from different points in knee- to waist-deep water.
  - Results from the EMPACT Beaches Study can be used as a guideline for the initial determination of sample size requirements. For the beaches in this study, sample sizes of from 3 to 6 would be adequate to allow for 95% certainty of detecting a 0.3 log exceedence from the action level for the geometric mean, equivalent to a health risk of 3 to 3.5 cases of HCGI per 1000 swimmers.
  - As data are gathered, the sampling plan should be refined and ultimately based on a variance estimate that is uniquely associated with the subject beach.

- "Hot spots" within the bathing area, which may be known *a priori* or discovered as a result of sampling, should be considered as separate strata for sampling purposes.
  - These should be sampled independently and weighted appropriately in the final result.
  - If indicator density within the problem area is very different from that elsewhere, sampling should be limited to this area. A proposed rule-of-thumb is a two-fold difference.
  - Lesser differences would warrant a combined, stratified estimate, weighting the "hot spot" in proportion to its extent relative to the rest of the beachfront.

- Composite sampling may be used as a cost-efficient technique, enabling better sample coverage at minimally increased cost.
  - In the initial stages of a monitoring program, composite sampling should *not* be used in order to develop data that is necessary in estimating the appropriate sampling variance for a particular beach.
  - A composite sample estimates an arithmetic mean, which would require adjustment in order to equate this to standards based on a geometric mean.
  - If resources are available to collect the data necessary for developing a predictive model for the change in indicator density over a 24 hour period, this will likely result in much-improved assessment of water quality, given the 1-day lag in obtaining results from membrane filtration assays.

Introduction

Protecting the health of swimmers who use surface waters for recreation has been of interest to public health officials in the United States since 1930. It was well recognized at that early date that human excreta discharged to surface waters posed a health hazard to those who used the water for recreation. Although the relationship between swimming-associated health effects and feces-contaminated water used for swimming had not been defined, microbial limits based on coliform bacteria were used in many states, particularly when there was physical evidence of sewage contamination. The limiting values selected by responsible authorities were based more on attainment rather than on risk of illness. Thus, there was little uniformity among states regarding what level of coliforms constitute waters safe for swimming. Several states chose 1,000 coliforms per 100 ml as a measure of good quality water, but there was not much uniformity among states regarding what level of coliforms was a safe level. There was, however, a general understanding that fecal contamination of surface water posed a risk to those exposed to the water, and that the risk might be limited by setting a level of fecal contamination above which exposure would be unacceptable. The manner in which water samples were taken, the frequency of sampling, and the number of samples were usually not described in the early literature. It is probable, however, that multiple samples from several points along a beach were not taken. Early monitoring schemes seldom involved several samples. Even now the Environmental Protection Agency (EPA) recommends at least 5 samples taken over a 30-day period, a sampling scheme that was first suggested in 1968. That number of samples will not characterize a body of water as to its true quality. Similarly, weekly sampling, no matter how accurate, is an unlikely scheme to alert risk managers of poor quality water between sample dates. Taking one daily sample might solve the latter problem, but still would not provide an accurate characterization of the water.

In the late 1940's and early 1950's a series of studies were carried out by the U.S. Public Health Service to determine the relationship between swimming-associated health effects and bathing water quality. The results of those studies were reported by Stevenson (1953). This data was used by the National Technical Advisory Committee (NTAC), commissioned by the Federal Water Pollution Control Administration in 1968 to establish guidelines for recreational waters. The NTAC recommended that a geometric mean of 200 fecal coliforms per 100 ml from 5 samples collected over a 30-day period be used as an upper limit which was believed to be protective of public health for swimmers [NTAC, 1968]. In the early 1970's the U.S. Environmental Protection Agency initiated a series of epidemiological studies to determine the relationship between swimming-associated illness and water quality as measured with multiple microbial indicators that are commonly found in feces [Cabelli, 1983; Dufour, 1984]. These studies showed that, of all the microbes examined, enterococci had the best relationship to swimmers' illness in marine waters. Thus, as the water quality became poorer, *i.e.*, the enterococci density increased, the swimming-associated illness rate increased. In fresh waters *E. coli* and enterococci were found to be effective for relating water quality to swimmer illness. EPA recommended an upper limit on geometric mean enterococci density of 35 per 100 ml for marine waters and 33 per 100 ml for fresh waters [U.S. EPA, 1986]. The means were to be calculated from at least 5 samples taken over a 30-day period. *E. coli* was not effective in marine waters, but did show a good relationship to swimmer-associated illness in freshwater environments. An upper limit of 126 *E. coli* per 100 ml was recommended for fresh waters [U.S. EPA, 1986]. These epidemiological studies conducted by the U.S. Public Health Service and the U.S. Environmental Protection Agency resulted in guidelines that limited the number of fecal indicator bacteria in a water body and, thereby, the number of swimming-associated illnesses to a tolerated level.

Providing numerical guidelines that are supported by scientific data was a great advance; however, these guidelines cannot stand alone. Monitoring schemes that will provide meaningful information

3

about whether the guidelines are being met must also be available. The manner in which monitoring is carried out at any particular beach is not likely to be the same for a marine beach with tidal cycles as for a freshwater beach, where other factors will significantly affect the level of indicator bacteria.

Since microbes almost universally have been recommended as the measure of water quality for recreational waters some of the characteristics that limit their usefulness should be addressed. Bacterial indicators frequently die off more rapidly than some of the pathogens found in feces [Fattal *et al.*, 1983]. When this happens, potential pathogens associated with feces might be present and pose a health hazard. Another issue with indicator microbes is that they usually take up to 24 hours to grow to visible colonies on solid media or turbidity in liquid media. This lengthy time lag creates a situation where exposed individuals are informed of risk of illness long after the fact [Bartram, 1999].

Physical factors may affect the measurement of water quality in ways that are not controllable. For instance, microbial indicators are affected by solar radiation, which cause many indicators to die off at variable rates [Fujioka and Narikawa, 1982], and neither the amount nor the intensity of sunlight is constant. These have a tendency to change from hour to hour and day to day. Rainfall is another factor that significantly influences the measurement of indicator microbes in beach waters [Olivieri *et al.*, 1977], but it is more predictable than other variables. Wind direction may also affect the quality of beach waters either by driving contamination toward the beach or away from the beach. Swimmer density can influence the quality of beach water as well. Swimmers shed high numbers of indicator bacteria from their bodies. Tub experiments have shown the millions of indicator bacteria are shed in very short periods of time [Smith and Dufour, 1993]. In marine waters, tides significantly affect the measurement of indicator bacteria, in addition to other factors mentioned [Cabelli *et al.*, 1974].

It is difficult to determine which of the above factors has the greatest effect on the monitoring of beach waters. It is likely that each beach will have its own characteristics based on predominant factors active at any one time. The EMPACT (Environmental Monitoring for Public Access and Community Tracking) Beaches project has attempted to define which characteristics are most significant with regard to monitoring approaches. This project examined five beach environments to determine the factors that most influence the measurement of beach water quality. Two ocean beaches, an estuarine beach, a Great Lakes beach, and a riverine beach were selected to provide as broad a representation of beach environments as possible. This approach was used to develop model protocols that can be applied to similar beaches.

Study Design

Sites

Beaches selected for this study are all located in or near partner cities in the Environmental Monitoring for Public Access and Community Tracking (EMPACT) grants program. The study was jointly conducted by the EPA Office of Water, Office of Science and Technology, and Office of Research and Development with the goal of helping communities collect, manage, and present environmental information to their residents. The following were the participating sites in this study (see Figure 1):

- West Beach, Indiana Dunes National Lakeshore, Ogden Dunes, Indiana, a freshwater beach on the shores of Lake Michigan. The EMPACT partner city is Gary.
- Belle Isle Park, Detroit, Michigan, a freshwater beach on the Detroit River between Lake St. Clair and Lake Erie.
- Wollaston Beach, Quincy, Massachusetts, a marine beach in Quincy Bay. The EMPACT partner city is Boston.
- Imperial Beach, Imperial Beach, California, a marine beach on the Pacific Ocean. The EMPACT partner city is San Diego.
- Miami Beach Park, Bowley's Quarters, Maryland, an estuarine beach on Chesapeake Bay near Middle River. The EMPACT partner city is Baltimore.

Figure 1. EMPACT study beaches.
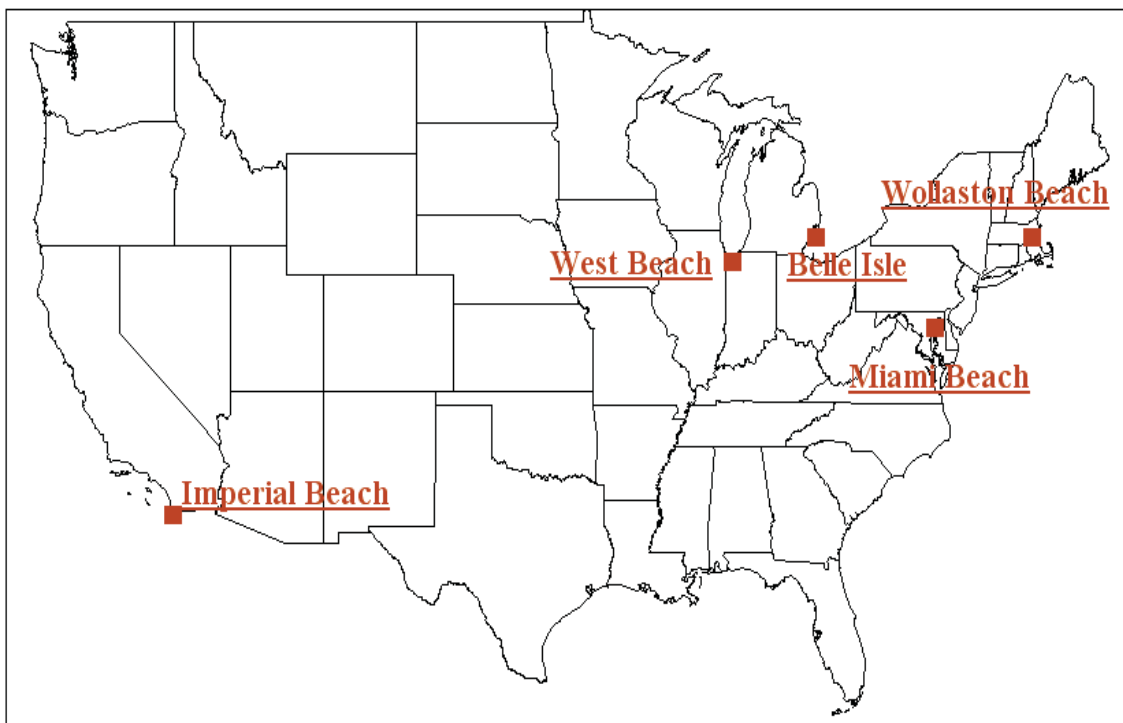
Sampling design

Samples were collected from each of nine locations within the limits of the beach's bathing area, each location being determined by an associated *transect* and *zone,* as illustrated in Figure 2. A transect is defined as an imaginary line through a fixed point on the beach and forming a right angle with the shoreline. A zone is defined as a contour line of equal water depth. As illustrated in Figure 2, each sampling location is defined by the intersection of transect and zone on a grid comprising three transects and three zones projected on the water's surface. A random point along the shoreline within the recognized beach area was selected to define the first transect (the leftmost transect in Figure 2). The middle transect was, then, determined as the parallel to the first transect at a distance of 20 meters, and the remaining transect, as the parallel to the middle transect at an additional distance of 20 meters, or 40 meters from the first transect.

Figure 2. Beach sampling grid.
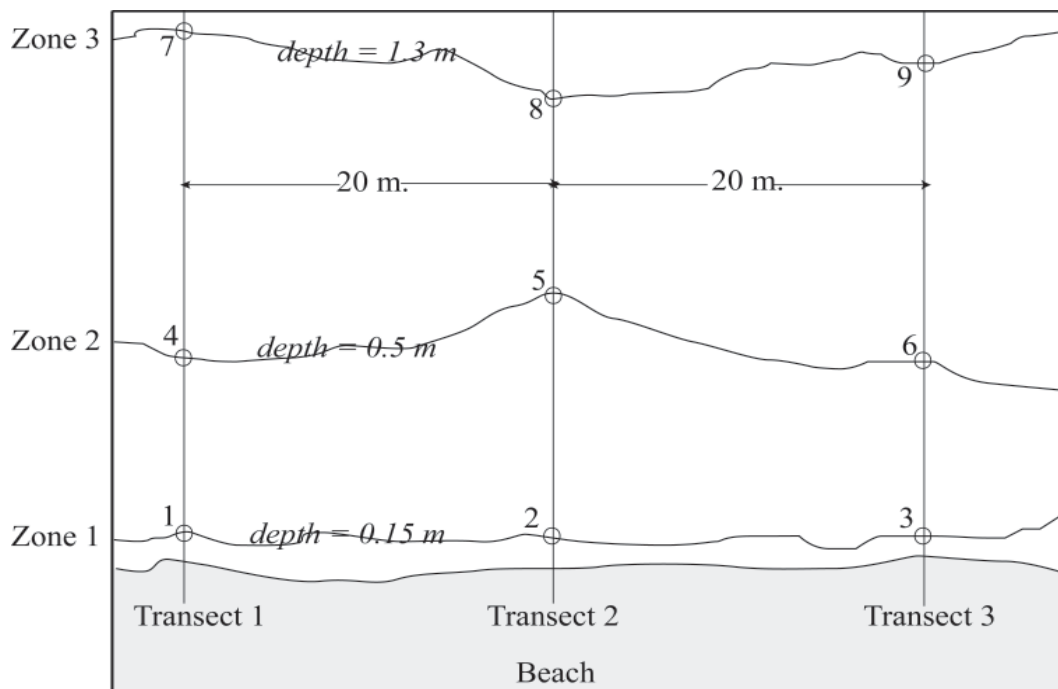
Intersections of sampling zones with these transects were determined by the depth of the water at the time of sampling. Zone 1 was determined at each transect as the point at which the water first attained a depth of 0.15 meters ("ankle-deep"); zones 2 and 3 were determined as those points along each of the transects where the water first attained depths of 0.5 meters ("knee-deep") and 1.3 meters ("chest-deep"), respectively. Buoys demarcating the swimming area in Belle Isle were so located that the water within this boundary never attained a depth of 1.3 meters; at this beach, therefore, the buoys themselves, at which the water was approximately 1 meter deep or about waist level, demarcated zone 3. The bathing area at Miami Beach Park was similarly restricted, the water depth along zone 3 being about 1.1 meter. Low tide at Wollaston Beach resulted in shallows extending quite far out from the beach. A 100 meter limit was established on the distance an individual would go to collect samples, resulting in waist-deep water at zone 3 at these times.

Water samples were collected from a depth of 0.075 meters below the surface in ankle deep water and 0.3 meters below the surface in knee-deep and chest-deep water. These sampling depths are shown schematically in Figure 3. Note that the actual sample locations (1-9 in Figure 2) may vary from one visit to the next, in terms of both absolute geographical location and relative distance from shore, in accordance with variations in water level and/or changing contour of the floor, particularly at the marine beaches.

Figure 3. Cross-section of beach sampling area.

○ = sampling location
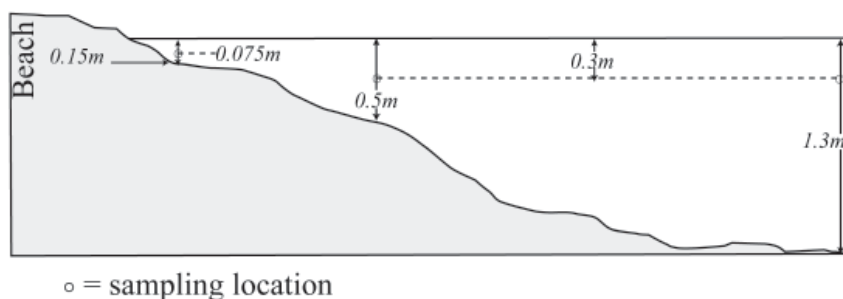
Water samples were collected in 500-mL pre-sterilized, polypropylene bottles. Care was taken in collecting samples near the bottom to avoid introducing additional sand or other solids into the samples. Following collection, samples were maintained on ice until analysis.

Sampling schedule

On each day, one of four sampling plans, referred to as "basic", "hourly", "replicate", and "depth" sampling, was followed according to the schedules shown in Figures 4a and 4b. The actual schedule at West Beach most closely followed the planned schedule, which was to have been identical at all five beaches. It was anticipated, and was, in fact, the case in all cities, that events such as inclement weather and rough water would force revision of the schedules. At Belle Isle and Imperial Beach, major activities necessitated changes in schedule and cancellation of certain sampling visits. The total number of sampling visits and total samples collected under each of these plans is summarized in Table 1 for each beach. Following are descriptions of each sampling plan.

Basic sampling
On each basic sampling day, samples were collected on two separate visits, at 9:00 a.m. and 2:00 p.m. A single sample was collected at each of the nine sampling locations (1 through 9 in Figure 2) on each sampling visit.

Hourly sampling
On each hourly sampling day, sampling was performed on ten separate occasions, hourly on the hour, the first set being collected at 9:00 a.m. and the last at 6:00 p.m. A single sample was collected at each of the nine sampling locations on each occasion.

Replicate sampling

On each replicate sampling day, samples were collected on two separate visits, at 9:00 a.m. and 2:00 p.m. On each sampling visit, ten replicate water samples were collected at the central sampling location (location 5 in Figure 2), and two replicate samples were collected at each of the eight remaining locations. Initially, because it was anticipated that its location on a river would make the collection of true replicates difficult, Belle Isle was to be an exception to the general replicate sampling scheme, with two samples being collected at all nine sampling points. However, after the first two replicate sampling visits, ten replicates were taken from the central location at this beach as well, when it became apparent that water flow was negligible within the protected confines of the bathing area.

Depth sampling

On each depth sampling day, samples were collected on two separate visits, at 9:00 a.m. and 2:00 p.m., with samples being collected at each of the nine sampling locations in Figure 2, the same as on a basic sampling visit. In addition to the standard sampling depths (0.3 m for knee and chest deep water), a sample was collected at a depth of 0.425 meters (0.075 meters, or 3 inches, from the bottom) at each of the three sampling points in the knee-deep zone. At each of the three chest deep zone sampling points, two additional samples were collected, one from mid-depth (normally, 0.65 meters from the surface) and another at a depth of 0.075 meters, or 3 inches, from the bottom. Thus, a total of nine additional samples were collected beyond the nine samples collected in the basic sampling plan.

Table 1.   Number of sampling visits and total number of samples collected.

Location	Basic sampling		Hourly		Replicate		Depth	
	Visits	Samples	Visits	Samples	Visits	Samples	Visits	Samples
West Beach	69	610	139	1162	16	416	8	144
Belle Isle	75	671	80	705	16	334	8	138
Wollaston	71	638	138	1242	16	407	8	144
Imperial Beach	68	612	140	1254	16	416	8	144
Miami Beach Park	72	643	140	1257	16	416	8	144

Data collection

Microbiological analysis

Freshwater samples, collected at Belle Isle and West Beach were assayed for *Escherichia coli* by the modified mTEC agar membrane filter method [U.S. EPA 2000a], while the marine and estuarine water samples, collected at Imperial Beach, Wollaston Beach, and Miami Beach Park, were analyzed for enterococci by the mEI agar membrane filter method (Method 1600) [Messer and Dufour, 1998 U.S. EPA 1997b, U.S. EPA 2000a]. The methods are summarized in Table 2. Using volumes of 100, 10, and 1 mL of each sample, filtrations were begun within six hours of collection and completed within eight hours of collection [Code of Federal Regulations 2000]. When concentrations were very high on the previous day's sample(s), 0.1 mL volumes of the water samples were also analyzed. Bacterial counts,

volumes, and other critical information, such as analysis start time and incubation start and completion times, were recorded on specially-designed data sheets, faxed daily to EPA (NERL-Cincinnati), and entered into the EMPACT database for subsequent audits of data quality and data analysis.

Table 2. Summary of laboratory methods.

Method Used	Medium Used	Incubation Time and Temperatures (˚C)	Detection Limits (Colonies Per Plate)	Ideal # of Colonies Per Membrane
E. coli (EPA 1603)	mTEC2 agar	2 hours @ 35 ± 0.5 20-22 hours @ 44.5 ± 0.2	1 - 200	20 - 80
Enterococci (EPA 1600)	mEI agar	24 ± 2 hours @ 41 ± 0.5	1 - 200	20 - 60

The general microbiology laboratory quality control procedures used in this study have been previously described [American Public Health Assoc. 1998, Bordner *et al.* 1978, U.S. EPA 1997a, U.S. EPA 2000a]. In addition, each preparation of modified mTEC agar and mEI agar was pre-tested for performance with known positive (target) and negative (non-target) cultures, and membrane filter, phosphate-buffered dilution water, and agar controls were performed with each set of samples that was analyzed (For example, all of the samples taken at the nine locations on the three transects for a specific time and date would be considered a set of samples.)

Other data

Other data that were collected at each sampling visit are shown in Table 3. Hourly and high and low water tidal data were obtained from the National Ocean Service of the National Oceanic and Atmospheric Administration for the reference station nearest to each respective study beach.

Figure 4a. Freshwater beach sampling schedules.

West Beach

Sun.	Mon.	Tues.	Wed.	Thur.	Fri.	Sat.
July, 2000						1
2 D	3	4	5 D	6	7	8
9 R	10	11	12 R	13	14	15
16 H	17 H	18 H	19 H	20 H	21 H	22 H
23 R	24	25	26 R	27	28	29
30 H	31 H					

Sun.	Mon.	Tues.	Wed.	Thur.	Fri.	Sat.
August		1 H	2 H	3 H	4 H	5 H
6 R	7	8	9 R	10	11	12
13 D	14	15	16 D	17 PM	18	19
20 R	21	22 D	23 R	24	25	26
27	28	29	30	31		

Belle Isle

Sun.	Mon.	Tues.	Wed.	Thur.	Fri.	Sat.
July, 2000						1
2 D	3	4	5 D	6 AM	7	8
9 R	10	11	12 R	13	14	15
16 H	17 H	18 H	19 H	20 H	21 H	22 H
23 R	24	25	26 R	27	28	29
30 H	31 X					

Sun.	Mon.	Tues.	Wed.	Thur.	Fri.	Sat.
August		1 X	2 X	3 X	4	5
6 R	7	8	9 R	10	11	12
13 D	14	15	16 D	17	18	19
20 R	21	22	23 R	24	25	26
27	28	29	30	31		

Basic sampling day
H Hourly sampling day (9:00 a.m.–6:00 p.m.)
X No sampling visits on this day

R Replicate sampling day
D Depth sampling day
AM Only 9:00 am samples were collected on this day
PM Only 2:00 pm samples were collected on this day

10

Figure 4b. Marine and estuarine beach sampling schedules.

Wollaston Beach

	Sun.	Mon.	Tues.	Wed.	Thur.	Fri.	Sat.
July, 2000							1
	2 (D)	3	4	5	6	7	8
	9 (R)	10	11	12 (R)	13	14 (D)	15
	16 (H)	17 (H)	18 (H)	19 (H)	20 (H)	21 (H)	22 (H)
	23	24 (R)	25	26 (R)	27	28	29
	30 (H)	31 (H)					

	Sun.	Mon.	Tues.	Wed.	Thur.	Fri.	Sat.
August			1 (H)	2 (H)	3 (H)	4 (H)	5 (H)
	6 (R)	7	8	9 (R)	10	11	12
	13	14 (D)	15	16 (PM)	17	18	19
	20	21	22	23 (R)	24	25	26
	27 (R)	28	29	30 (D)	31		

Imperial Beach

	Sun.	Mon.	Tues.	Wed.	Thur.	Fri.	Sat.
July, 2000							1
	2 (D)	3	4	5 (D)	6	7	8
	9 (R)	10	11	12 (R)	13	14	15
	16 (H)	17 (H)	18 (H)	19 (AM)	20 (PM)	21 (H)	22 (H)
	23 (R)	24	25	26 (R)	27	28	29
	30	1 (H)					

	Sun.	Mon.	Tues.	Wed.	Thur.	Fri.	Sat.
August			1 (H)	2 (H)	3	4	5
	6	7	8	9	10 (H)	11 (H)	12 (H)
	13 (H)	14 (H)	15 (H)	16 (D)	17	18	19 (D)
	20 (R)	21	22	23 (R)	24 (AM)	25 (PM)	26
	27 (R)	28	29	30 (R)	31		

Miami Beach Park

	Sun.	Mon.	Tues.	Wed.	Thur.	Fri.	Sat.
July, 2000							1
	2	3	4	5 (D)	6	7 (D)	8
	9	10	11	12 (R)	13	14 (R)	15
	16 (H)	17 (H)	18 (H)	19 (H)	20 (H)	21 (H)	22 (H)
	23	24	25	26 (R)	27	28 (R)	29
	30 (H)	31 (H)					

	Sun.	Mon.	Tues.	Wed.	Thur.	Fri.	Sat.
August			1 (H)	2 (H)	3 (H)	4 (H)	5 (H)
	6	7	8	9 (R)	10	11 (R)	12
	13	14	15	16 (D)	17	18 (D)	19
	20	21	22	23 (R)	24	25 (R)	26
	27	28	29	30	31		

Legend:

- ☐ Basic sampling day
- H Hourly sampling day (9:00 a.m.-6:00 p.m.)
- ▦ No sampling visits on this day
- R Replicate sampling day
- D Depth sampling day
- AM Only 9:00 am samples were collected on this day
- PM Only 2:00 pm samples were collected on this day

11

Table 3.   Data collected at each sampling visit (in addition to microbial data).

Measurement	Description & units of measurement	Units
Date and time	Date and time of day.	
Air temperature	Measured by thermometer at a fixed location every visit.	°C
Water temperature	Measured by thermometer at the central sampling location at appropriate depth for thermometer.	°C
Cloud cover	**S**unny, **M**ostly **S**unny (20-50% cloud cover), **C**loudy (50-70% cover), **M**ostly **C**loudy (70-99% cover), **O**vercast	S, MS, C, MC, O
Rainfall since last visit	Measured by rain gauge; collected each sampling visit.	Inches
Wind speed	Sustained speed measured by wind gauge.	Miles per hour
Wind direction	Compass direction to nearest semi-quadrant leeward on wind gauge.	N, NE, E, etc.
Current direction	Described in relation to shoreline facing out.	Descriptive: (onshore, right, *etc.*)
Wave height	Meter stick measurement at central sampling point.	Meters
Bather density in the water	Number of bathers in the water, within the sampling area.	Categorical: <20, 20-100, 100-200, >200
Bather density on the beach	Number of bathers on beach, within outer transects.	Categorical: <20, 20-100, 100-200, >200
Boats	Number of boats in the water within 500 m of sampling area.	Categorical: <20, 20-100, 100-200, >200
Animals	Animals within 20M of the sampling area in the water or on the beach; also includes number of fowl or other birds in the air near the sampling area.	Type & number of animals
Debris	Description of debris floating in the water or washed on shore within the sampling area.	Categorical; "None," "Very Little," "Little," "Lots;" and description
Total Suspended Solids (TSS)	Measured from the sample taken at the central sampling location, per *Standard Methods* (APHA, 1998).	mg/L
pH	Each sample measured after microbiological analysis, per *Standard Methods* (APHA, 1998).	pH units
Turbidity	Each sample measured by nephlometer after microbiological analysis processing, per *Standard Methods* (APHA, 1998).	NTU

Factors Affecting Water Quality at a Beach

Introduction

In this section, we will discuss various factors that were seen to have some association with the indicator density that was observed in our samples. This will provide answers to questions such as:

- At what locations in the water, if any, does one tend to see higher or lower levels of indicator organisms?
- Are any temporal cycles evident, such as a time of the day or a day of the week when indicator levels tend to be higher or lower than they are at other times?
- Can we expect generally higher or lower sample results under certain weather conditions? What about tidal influences at marine beaches?

These refer to *systematic* variability, as opposed to *random* variability, which will be discussed in the next section. Systematic variability implies predictable differences, albeit with a degree of uncertainty as indicated in the phrasing of these questions. The variability that is left unexplained, the uncertainty in our prediction, is random variability.

Knowledge about systematic variability is important for two reasons:

- Systematic differences that are found to exist may be exploited in the sample design, suggesting where and when to take samples.
- Because results from sampling are not available until the next day, a predictive model must be used to assess the current "state of the beach" (Of course, one possible model is to assume that nothing has changed). Even if results could be obtained instantaneously, it would still be desirable to have some lead-time for informing the public of the results.

Results from each of the five study beaches are considered individually. Any observed spatial, temporal, and environmental relationships with indicator density strictly apply only to that beach. While we have found similarities among the study beaches with regard to certain parameters, this should not be construed to mean that these relationships would apply to all beaches in general. Nor should relationships found to exist at a particular study beach be assumed to apply to a general class of which it is representative, such as an estuarine, Great Lakes, or west coast marine beach. Certainly, however, a relationship found to exist at any of these study beaches may apply to other beaches, and some of these characteristics may be true for many or most of a given class of beach. In the course of a regular beach monitoring program ample data can be generated to enable one to verify these relationships and even to discover factors that are unique in one's own environment.

Statistical methods

A distinction is made between *design variables* and *covariates*. Design variables are those factors over which we have control and were specifically accounted for in the design of the study. Samples were specifically collected in ankle-, knee-, and chest-deep water, thus, "zone" is a design variable. Likewise, three fixed transects were employed to define locations at each depth zone where samples were to be taken, so that the variable of "transect", position along the shoreline, is also designed into the study. Another design variable is time-of-day, mainly being either 9:00 a.m. or 2:00 p.m. on a given day, but also hourly from 9:00 a.m. to 6:00 p.m. on selected days. Depth below the surface from

which the sample was collected is also a design variable, since we occasionally sampled from depths below the standard 0.3 meters from the surface.

"Covariates" include all factors that were outside of our control which were not a part of our sampling plan, such as rainfall, cloud cover, tidal movements, currents, and bather load. In general, these constitute a "messier" set of observations than do design variables. We may not be able to attribute an effect due to one of these variables at a given beach simply because there was too little variability in that factor at that beach. Such was the case at Imperial Beach, for example, where rainfall occurred only twice during the course of our study. In addition, many of these covariates may be interrelated ("co-linear" or "confounded" variables), a potential example being temperature, cloud cover, prior rainfall, and bather load. Are the increased counts due to the greater number of bathers at the beach or are they due to the high temperatures that bring these bathers out? Do we fail to see an increase in indicator density with greater numbers of bathers simply because there are more bathers on sunny days when, at the same time, the sunlight increases the rate of bacteria die-off?

Criterion variable

Log_{10} indicator density per 100 mL is used as the criterion variable (the "dependent" or "response" variable) in the analyses that follow, thus having a direct correspondence with models that relate mean log_{10} indicator density to swimming-related health effects [Cabelli *et al.*, 1982; Cabelli, 1983; Dufour, 1984]. Existing U.S. EPA guidelines [1986], based on these models, specify limits for the geometric mean - the antilog of the mean log_{10} density. The implications of these health effect models with respect to a sampling will be discussed later, in the section **Designing a Beach Monitoring Program**.

Single sample limits

Another criterion recommended by the EPA [U.S. EPA, 1986] is the single-sample limit. Over 1500 samples were collected at each of the study beaches (see Table 1), giving us a multitude of "case studies" for evaluating the performance of the single-sample limit *vis-à-vis* any given factor. These case studies will be presented in parallel to the analyses on log means in a boxed format such as this.

Design variables

Relationships between log_{10} indicator densities per 100 mL and design variables were investigated via an analysis of variance (ANOVA) model. Factors used in the ANOVA were zone (ankle-, knee-, and chest-deep), transect (labeled 1-3, as in Figure 2), time-of-day (9:00 a.m. through 6:00 p.m.), and weekend (Friday through Sunday).

Covariates

The initial analysis of the dependencies of log_{10} indicator density per 100 mL on the ancillary variables recorded in this study (Table 3) was conducted via a stepwise regression. Because these variables, such as air temperature, cloud cover, etc., apply to the entire sampling visit, the dependent variable is the average log density for the respective visit. This is calculated as a weighted average, by first averaging replicate observations, if any, within a location, then averaging over transects within a zone, and finally averaging over zones. Only samples taken from the usual sampling depth of 0.3 M below the surface are used in this average. So that whatever conditions happen to exist on hourly sampling

days do not disproportionately influence the results, only data from the 9:00 a.m. and 2:00 p.m. visits are used.

Some ancillary data were transformed before being used as candidate covariates. In some cases, a single variable was expressed in more than one form in the search for best predictor. Many were expressed as *dummy* variables, that is, variables that can take on only two values, zero and one, where one indicates the presence of some condition and zero, the absence of that condition. For example, cloud cover determined the value of a dummy variable, "Sunny?", which took on the value 1 when cloud cover was coded as "Sunny" or "Mostly Sunny", and 0 when cloud cover was coded as "Cloudy", "Mostly Cloudy", or "Overcast." Some of the variations on covariate that were considered are:

- Temperature (air and water) was used in °C, as originally measured. In addition, this was expressed as a high-temperature dummy variable (greater than the median for the respective time-of-day) and as 1st quartile and 4th quartile dummy variables (among lowest/highest 25% for the respective time-of-day).
- Rainfall was used in inches, as originally measured, and in the form of dummy variables indicating whether it had rained in the past 24 hours or in the past 3 days.
- Wind speed and wind direction were combined to produce a wind vector variable (the component of wind velocity that is onshore, which is negative if the wind is blowing away from the shore). Also, dummy variables indicate onshore wind and offshore wind.
- Wave height was used in meters, as originally measured, and as a "high wave" dummy variable.
- Bather densities (on the beach/in the water) were considered as dummy variables (1=less than 20, 0=20 or more). They were also combined in order to consider cases where there were more than 20 bathers both on the beach and in the water. (Note: Categories of "100-200" and "over 200" were too infrequently observed to be considered.)

One covariate that was included in all models was the previous day's $\log_{10}$ indicator density at the same time-of-day. This was either considered as an independent variable or subtracted from the current day's $\log_{10}$ indicator density to create a dependent variable equal to the change in log density.

Several regression model-building techniques were attempted, including stepwise and best R^2 searches, along with considerable trial-and-error. In the end, however, it was a matter of judgment among several competing models, each of which was as good as the other in terms of degree of complexity and the percentage of variation explained (R^2). A model for a given beach that is presented here is only one of several possible models, any of which may be a reasonable alternative. In general, however, these models do tend to be similar in that some form of a major effect is common to all models with only minor contributions being made by disparate effects. For example, rainfall and tidal stage, in one form or another, may be common to all the models, while some models may include a bather density effect, others, a temperature effect.

Spatial factors

The physical significance of spatial factors

We defined the physical location from which a sample is collected in terms of transect, zone, and sampling depth.

Transect – an imaginary line extending at a right angle to the beach's orientation from a given point along the shoreline, and the water column below it. Specifying transect fixes a location with respect to its distance along the shoreline.

Zone – the water column above a contour line on the floor of the bathing area, corresponding to constant water depth, that is, ankle-deep, knee-deep, or chest-deep water. Where the water depth is constant or decreases as one moves away from shore, as when the floor is level or undulates, our zones refer to the *first* occurrence of the specified depth. Roughly, "zone" means "distance from shore", at least in the sense that going from ankle-deep to chest-deep water takes one farther from shore. The actual distance from shore, however, may vary between two locations that are in the same zone.

Sampling depth – the depth below the surface of the water from which a sample is collected.

For this study, samples were collected along parallel transects located 20 meters apart, each transect being defined by two points on the shore that were fixed for the duration of the study. Samples were collected in zones described as "ankle-deep", "knee-deep", and "chest-deep" water, but in actuality fixed by measurement to determine where the mean water depth was 0.15, 0.5, and 1.3 M, respectively. Samples were collected from approximately 0.3 M below the surface (0.075 M in ankle-deep water), and, on certain days, additional samples were collected at greater depths.

Barring some rather unusual beach topography, there are several common factors that ought to differentially affect water quality among depth zones in qualitatively the same way at practically any beach:

- Contamination on the shore or in the swash zone ought to have, if anything, a greater impact on water quality in ankle-deep water than in chest-deep water.
- The shallower the water, the more the water quality at or near the surface will be affected by any disturbance of the sediment, such as might be caused by bathers or wave action.
- Indicator organisms that come directly from bathers, not from sewage contamination, should be more common in shallower water due to children and a higher bather density in general.
- Deeper water, on the other hand, may be exposed to any offshore point source of contamination to a greater degree than shallower water.

In contrast, there are few, if any, such generalities one can make with regard to transects-different locations along a shoreline. Currents along or near the shoreline may have negative or positive effects on contamination levels downstream, and may have significant impact at some beaches and be negligible at others. Their impact may vary even at the same beach. A physical barrier in or near the bathing area, such as a sandbar, reef, or pier, may effect a gradient in the level of contamination parallel to the shoreline at some beaches, but such effects will be absent at others. Lacking such generalities, systematic variability along the shoreline must be evaluated with respect to a specific beach.

Findings from the EMPACT Beaches study with regard to spatial factors

Major findings on spatial variation are:

- In every case, the zone from which the sample was collected was found to have the greatest predictable impact on microbial indicator densities of all factors investigated in this study, spatial or temporal. Bacterial densities become progressively lower as one moves from ankle-deep to knee-deep to chest-deep water.
- Two of the study beaches, Belle Isle and Miami Beach Park, exhibited some form of systematic spatial variation that was not adequately accounted for by zones alone. It may or may not be a coincidence that both of these beaches are associated with river systems.

Variation in target density associated with zones

For each of the five study beaches, the greatest single determinant of microbial indicator level is the zone, or, roughly, distance from the shoreline at which the sample was collected. Bacterial densities become substantially lower as one moves from ankle-deep to knee-deep to chest-deep water. Figure 5 and Table 4 illustrate the relative magnitudes of measurements taken 0.3 meter (0.075 M for ankle-deep water) from the surface in these respective depth zones during the 9:00 a.m. and 2:00 p.m. visit for each beach. Although the effect appears to be relatively small at Imperial Beach, even here this was the single most important influence on sample indicator density.

Table 4. Comparisons of point-in-time geometric mean indicator density by zone.

Location	9:00 a.m.			2:00 p.m.		
	Ankle	**Knee**	**Chest**	**Ankle**	**Knee**	**Chest**
West Beach	34	19	14	41	19	9
Belle Isle	2113	487	11	754	224	4
Wollaston	43	24	12	36	12	7
Imperial Beach	6	5	4	6	4	3
Miami Beach Park	426	221	15	231	67	4

The relationship between distance from shore and indicator density is important in terms of its implication for sample design. First, if the bather population exposed to contamination in one zone essentially differs from those exposed to the contamination in another zone, then an overall estimate that combines indicator densities from both zones may not adequately reflect exposure for either population. Given that the usual route of exposure for gastrointestinal effects is through ingestion, an example of two different bather populations would be small children, who are likely to be exposed to contamination only in shallow water, and adults and older children, who are most likely to be exposed only in deeper waters. Second, if combining indicator densities from different depth zones is appropriate for a given bather population, then the sampling scheme should take these known systematic differences between zones into account.

Variation in target density associated with distance along the shoreline

In contrast to indicator densities from samples taken at different distances out from the shoreline, samples that were collected in water of the same depth but at different locations relative to the beach front (that is, on different transects) generally are not that different

from one another. This is evident in Figure 6, which shows, by depth zone for each of the study beaches, the geometric mean densities along each transect at 9:00 a.m. and 2:00 p.m. Any systematic differences that may exist along the parallel-to-shore dimension are minor, particularly in comparison to the obvious differences between depth zones, as seen previously in Figure 5.

Figure 5. Point-in-time geometric mean indicator density per 100 mL by zone.
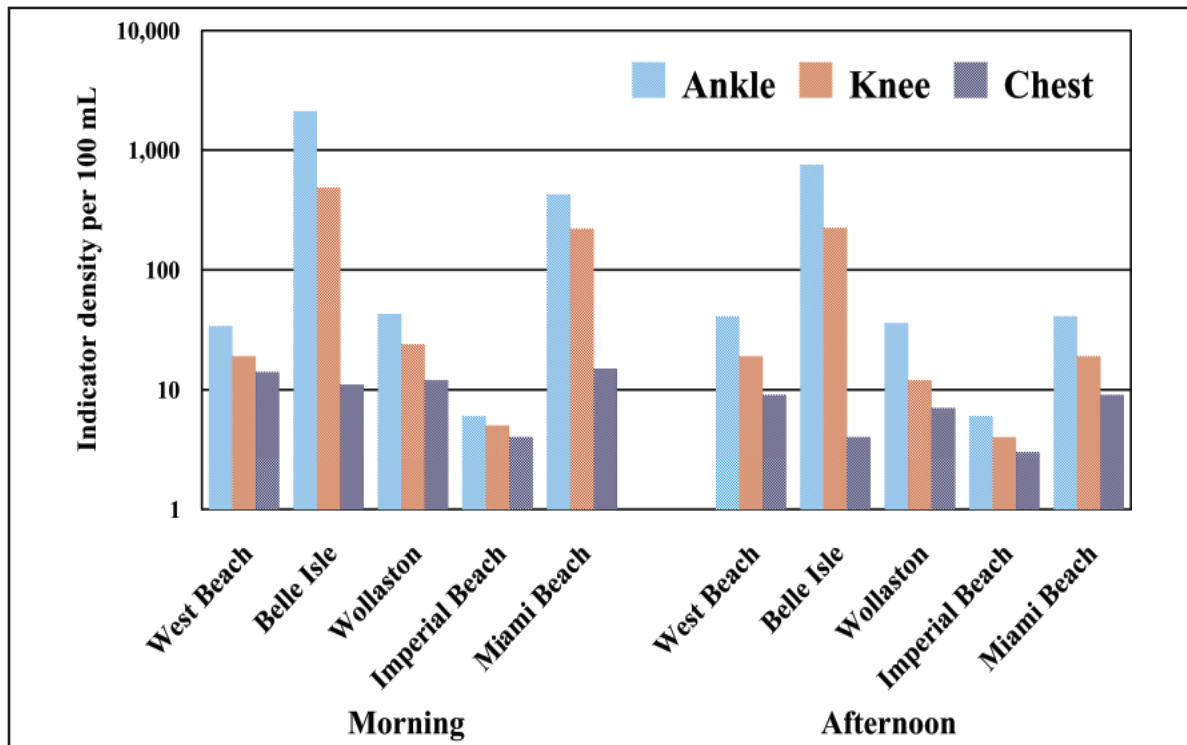
There are, however, two instances in which we note a statistically significant "transect effect" – a persistent difference in microbial density along one transect relative to another. The more pronounced of these is at Belle Isle, where *E. coli* densities in knee-deep water are consistently lower along the leftmost transect (Transect 1 in Figure 6), than in the other two knee-deep sampling locations. At Miami Beach Park, where the other transect-related difference is indicated, the situation is not so clear and the relationship is comparatively weak. Furthermore, analysis of variance reveals an interaction between sampling location and time-of-day, indicating that any such difference varies between the 9:00 a.m. and 2:00 p.m. visits. As seen in Figure 6, values in ankle-deep water were somewhat higher along one transect, the leftmost transect in the morning, but the rightmost transect in the afternoon. We can also discern an increasing trend, particularly in the morning, as one moves left to right (*i.e.*, from Transect 3 to Transect 1) in chest-deep water.

Whether these observed differences in indicator levels among transects have any practical significance can be determined only by a detailed sanitary survey of these respective beaches. It may be noted, however, that at Belle Isle, where we see the only truly substantial transect effect, a drainage ditch empties very near the vicinity of the rightmost transect from an area adjacent to a water park and bathhouse. This might be hypothesized as a point source capable of affecting

---

Single sample limits and depth zones

Zones ("distance from shore") from which to collect samples will be a critical consideration in designing a monitoring program. How far from shore should one go? Two important points:

- If zones are not accounted for, it may be impossible to compare monitoring results from two different days, or even to interpret the results of a single sample.
- Our current knowledge of human health effects is based on studies that have measured indicator levels in approximately waist-deep water.

Among the five study beaches, those samples taken in ankle-depth water were the most likely to exceed single-sample criteria [U.S. EPA, 1986], and, except at Imperial Beach where only one sample failed the criterion, samples taken in chest-deep water were least likely to do so.

Table 5. **Percentage of total samples collected that exceeded the single sample limit as recommended in EPA guidelines [U.S. EPA, 1986]**

Location	% of samples exceeding the single sample limit: Number of exceedences / total number of samples (total number of samples in parentheses)					
	Ankle		**Knee**		**Chest**	
West Beach	21.9	(183)	12.0	(183)	9.3	(172)
Belle Isle	97.0	(167)	88.5	(174)	2.3	(174)
Wollaston Beach	24.0	(183)	14.2	(183)	6.6	(183)
Imperial Beach	1.6	(183)	0.0	(183)	0.5	(183)
Miami Beach Park	80.1	(186)	61.2	(185)	12.9	(186)

---

contamination levels in a manner consistent with observations. Mainly, we might see higher levels in the knee depth zone at the rightmost and middle transects, which are upstream in this riverine environment, simply because the plume is more concentrated at these points. Other features of the observed pattern at Belle Isle are consistent with this hypothesis - particularly the elevated microbial levels in chest-deep water at the leftmost transect (downstream from the ditch), and in ankle-deep water near the outlet of the ditch at transects 2 and 3 (although this is clear only in the afternoon).

Figure 6. Indicator density per 100 mL by zone and transect.
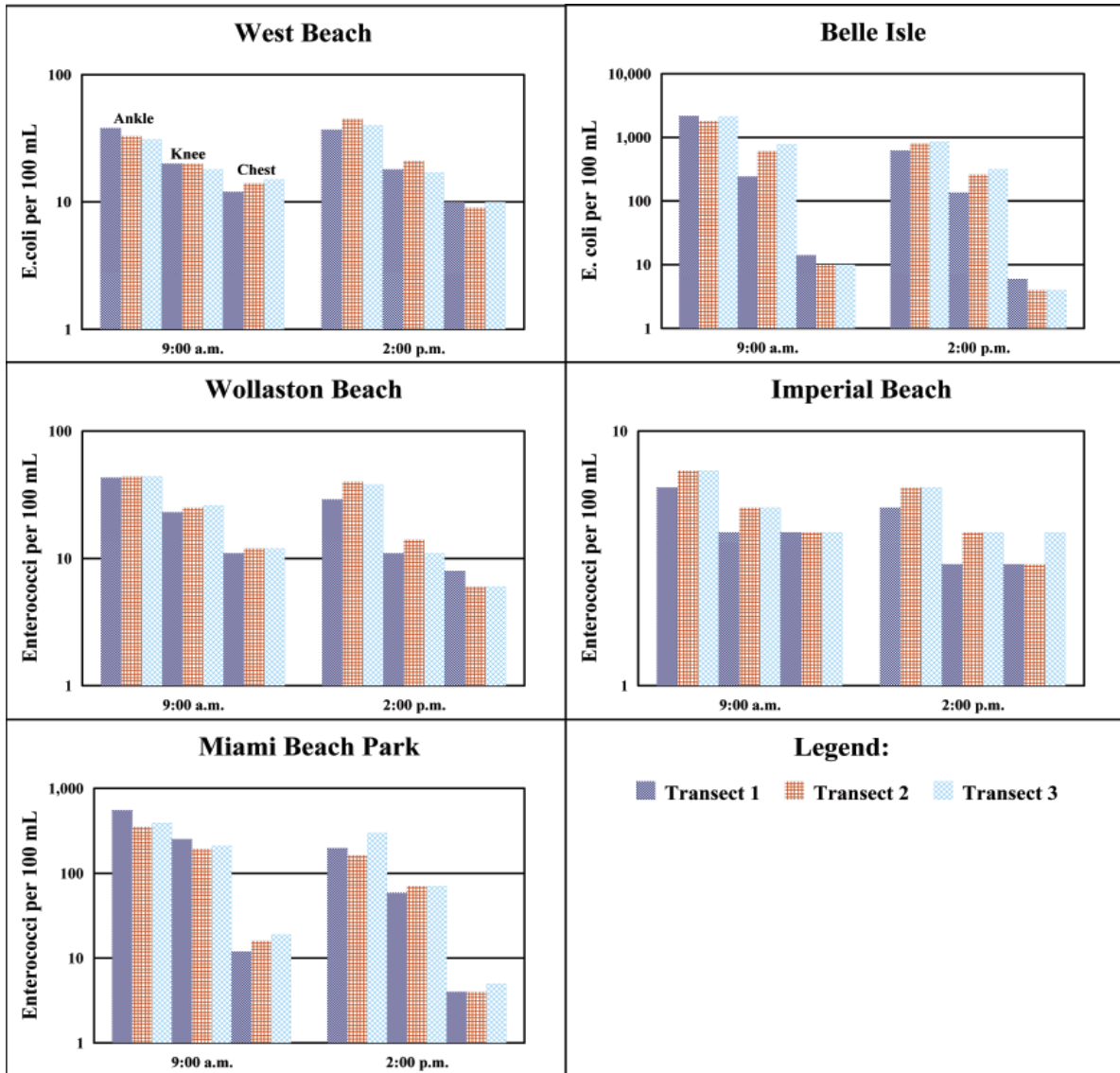
The Public Health Laboratory Service (PHLS) Water Surveillance team (PHLS, 1995) examined multiple parameters in eight inland recreational waters. Water samples were collected at 10 predetermined sites on each water body on four consecutive weeks. Unlike the results in the current study, the PHLS study results showed significant differences between results from the ten sites even when the sites were very near each other. The sampling sites in the PHLS studies were, in general, further apart than the 20 meter distances in our study and, in addition, some of the sampling sites were at the inlet and outlet points of the lakes. However, on two of the lakes, six of the sampling points were in a grid one meter apart and the results from these points showed differences greater than random variation. The differences between our data and the PHLS data probably indicate that transect parameter cannot be generalized from one location to another.

Single sample limits and shoreline location

We offer two suggestions with regard to where to sample along the beach. The first is obvious. The other may entail extra effort (and funding) to implement.

- Naturally, we want to sample in areas where bathers tend to congregate.
- Other than this, "hot spots" that deserve special consideration may exist. These may only be determined by knowledge of peculiarities at a particular beach or by exploratory sampling.

Consider the exceedences found in the five study beaches among samples taken in knee-deep water, again based on the 1986 single-sample criteria [U.S. EPA, 1986]. We chose the knee-depth zone since this is likely to be an area that is sampled and yields a fair number of exceedences among our study beaches (with the exception of Imperial Beach, where few exceedences were seen among any of the samples).

Table 6. Percentage of samples collected in knee-deep water that exceeded the single sample limit as recommended in EPA guidelines [U.S. EPA, 1986] (9:00 a.m. samples)

Location	% of samples exceeding the single sample limit: Total # of exceedences/total # of samples (total number of samples in parentheses)					
	Transect 1		**Transect 2**		**Transect 3**	
Belle Isle	81.0	(58)	93.1	(58)	91.4	(58)
West Beach	11.5	(61)	13.1	(61)	11.5	(61)
Wollaston Beach	14.8	(61)	11.5	(61)	16.4	(61)
Imperial Beach	0.0	(61)	0.0	(61)	0.0	(61)
Miami Beach Pk.	61.3	(62)	64.5	(62)	57.4	(61)
Belle Isle p.m. samples	54.4	(57)	73.7	(57)	82.5	(57)

For the most part, there are no meaningful differences in the rate of exceedence among the three transects. A possible exception is Belle Isle, where fewer exceedences were observed at the 1st (leftmost) transect. While the p-values for the observed differences among transects for the other three beaches that had exceedences were all greater than 0.70, indicating that the magnitude of the differences shown in Table 6 could very likely result from simple random variation, the p-value for Belle Isle's inter-transect differences in exceedence was 0.08, indicating a somewhat unlikely result unless there are truly systematic differences among these locations. As further evidence of some transect effect at this beach, afternoon results for Belle Isle only are shown (afternoon results for the other beaches do not indicate any different variability than the morning results). The difference in the afternoon is even more dramatic for Belle Isle, with a p-value of less than 0.01. Managers at Belle Isle would be well advised to pay particularly close attention to the area defined by transects 2 and 3, and to investigate the cause for this phenomenon with an eye toward possible remediation.

Sampling depth

One of the more interesting outcomes of these beach studies was the lack of a clear gradient in the indicator density with respect to the depth from the surface at which the sample was collected (Table 7). Such a difference with depth might be expected from the fact that water near the bottom is more likely to contain suspended particles from the sediment, which is presumably relatively rich in bacteria. Others have observed depth effects [Public Heath Laboratory Service Water Surveillance Group, 1995].

Bottom samples (0.075 m from the bottom) were collected in knee-deep and chest-deep water and mid-depth samples (0.65 m from the bottom) were collected in chest-deep water. Geometric mean densities over the duration of the study and over all transects for each type of depth sample are compared to corresponding surface densities (from only those samples collected at the same time as depth samples) in Figure 7.

The effect of depth on indicator density results was examined in one other study [PHLS, 1995]. The findings of that study indicated that paired comparisons between surface, 30 cm and 100 cm depth samples, showed a significant tendency for higher fecal streptococci counts at the surface. A similar trend was shown for *E. coli*, but the differences were not statistically significant. Although the procedure for taking surface samples was not described, an accumulation of microbes at the water-air interface could have accounted for the differences observed in the PHLS studies.

Table 7. Comparison of geometric mean density[1] of indicator densities by depth of water from which the sample was collected. "P_0" is the "p-value" for the difference ($\log_{10}$ means) between the surface and respective depth means.

Location	Knee deep water			Chest deep water					
	Surface[2]	Bottom[3]	P_0[4]	Surface[5]	Bottom	P_0	Surface[5]	Mid-depth[6]	P_0
West Beach	32	36	0.237	19	22	0.229	20	21	0.954
Belle Isle	154	142	0.324	3	4	0.082	3	4	0.331
Wollaston Beach	31	47	0.040	18	16	0.732	18	14	0.315
Imperial Beach	3	3	0.716	2	2	0.673	2	3	0.292
Miami Beach Park	79	87	0.499	2	3	0.478	3	4	0.045

[1] Geometric means are over all transects and sampling visits. For surface samples, only those samples for which a corresponding depth sample was taken at the same time are considered. Numbers of samples range from 18 to 30. [2] Samples collected 0.3 m below the surface. [3] Samples collected 0.075 m from the bottom. [4] "P-value" for difference in mean $\log_{10}$ values between surface and the respective depth (paired samples). Generally, a value less than 0.05 is considered "significant" [5] The two values for surface samples from chest-deep water may differ slightly due to missing data among the depth samples. [5] Samples collected midway between surface and bottom.

Figure 7. Geometric mean density of surface samples vs. depth samples.
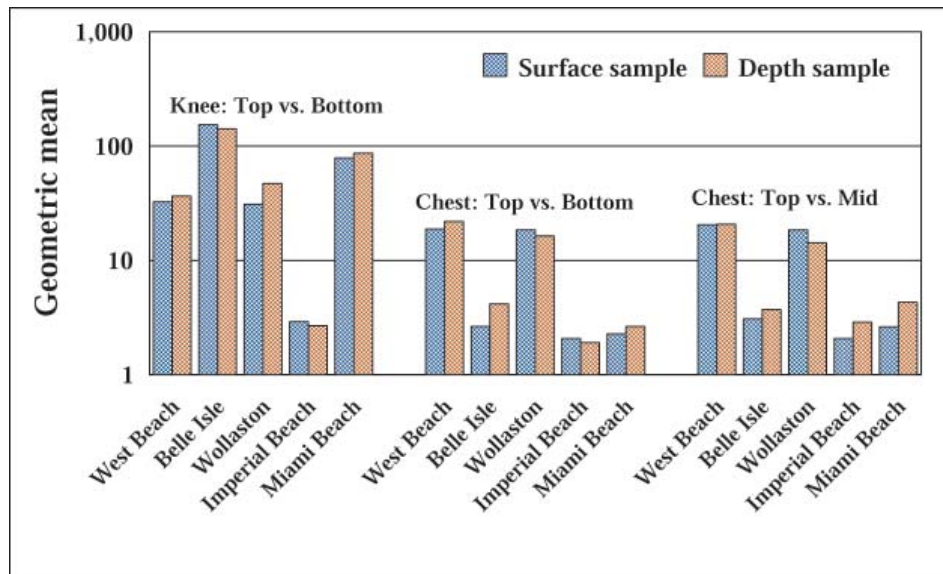
Single sample limits and sampling depth

We see little difference between samples collected near the surface and those collected near the bottom with respect to the rate at which they exceeded the single-sample criteria [U.S. EPA, 1986]. Lest Table 8 lead one to believe that the closeness of the exceedence rates indicates agreement between samples taken from different depths, we note that, overall, the results were in disagreement 44% of the time when an exceedence was recorded in at least one of the two samples from different depths.

Table 8. **Percentage of total samples collected that exceeded the single sample limit as recommended in EPA guidelines [U.S. EPA, 1986]**

Location	% of samples exceeding the single sample limit in knee-deep water, top vs. bottom samples		
	Top	Bottom	Total #
Belle Isle	42.9	33.3	21
West Beach	11.1	11.1	27
Wollaston Beach	41.7	41.7	24
Imperial Beach	0.0	0.0	24
Miami Beach Park	20.8	25.0	24

Temporal factors

The significance of temporal factors in sample design

Current EPA guidelines refer to limits on indicator organisms based on the geometric mean of "not less than 5 samples equally spaced over a 30-day period" and further specify that, for designated bathing beaches, no single sample exceed the 75th percentile of the corresponding log-normal distribution [U.S. EPA, 1986]. Based on health studies conducted by the EPA in the 1970's and early 1980's [Dufour, 1984; Cabelli, 1983], the time-averaged recommended maximum geometric means of 126 *E. coli* per 100 mL in freshwater and 35 enterococci per 100 mL in marine water correspond to marginal illness rates of 8 and 19 cases of gastroenteritis per 1000 swimmers, respectively.

In current practice, a reasonable scenario for a beach monitoring authority might be to collect a sample every, say, Tuesday morning. This would allow adequate time to obtain the results from that sample and, if necessary, take corrective action and resample before the weekend, or give sufficiently advanced notice to the public that the beach will be closed the following weekend. Closure, or action to avoid closure, would be based on the above 30-day geometric means or single sample limits.

Bearing this in mind, the daily, and sometimes hourly, collection of samples in both the morning and afternoon in this analysis will give insight into the following questions on sampling strategy:

- How well does a 30-day mean characterize recreational water quality on a given day?
- How does an observed exceedence (or non-exceedence) in today's sample correspond to water quality four (or three, two, one) days in the future?
- When is a good time of day to sample?

Findings from the EMPACT Beaches study with regard to temporal factors

Major findings on temporal variation are:

- From one day to the next, the geometric mean indicator density over all sampling locations changed by a factor of 2 (doubling or halving) or more about half the time at each studied beach.
- There was a limited statistical relationship observed between levels on the sampling day and the following day at three out of the five beaches studied.
- Indicator densities tended to decline from morning to afternoon at three of the beaches. This was true for both freshwater and marine indicators. Early morning results were fairly indicative of afternoon water quality, realizing that the density may be lower as the day progresses.

Day-to-day variability

Time series plots of geometric mean indicator densities over all sampling locations in knee-deep and chest-deep water at 9:00 a.m. are shown in Figure 8 for each of the five beaches. From the analysis of spatial factors, depth zone was seen to have a substantial impact on indicator densities. These plots are representative of samples from zones in which adult swimmers would be exposed.

The time series plots of Figure 8 afford us a rather accurate daily "snapshot" of each beach. At the same time, we may simulate a realistic beach monitoring scenario. For illustration, we will use scenario described in the previous section – a single Tuesday morning sample. Having examined

other scenarios, however, we can offer assurance that the conclusions apply to a single-sample scheme for any day of the week or time of day, although specific details in the behavior of the simulated monitoring results will, of course, vary.

Based on this scenario, the four-week moving geometric mean is superimposed on these graphs for a sample collected on Tuesday morning, approximating a monitoring result based on 5 samples collected over a 30-day period (actually a 29-day period) as specified in U.S. EPA guidelines. Naturally, a four week moving average cannot be calculated until August 1, four weeks from the first Tuesday of the study. From then on, the running geometric mean is based on an idealized Tuesday sample result – a single sample that matches the geometric mean of six knee- and chest-deep samples.

In the first of these graphs, which is for West Beach, note that the only occasion during August when the point-in-time geometric mean for the morning exceeded 126 was on Sunday the 27th. On this day, the geometric mean reached 446, well above the single-sample limit of 235 *E. coli* per 100 mL in fresh water. Not only does the 29-day geometric mean fail to predict this event, it is, in fact, at its lowest level of the study. There appears to be very little possibility that any single sample that had been collected on the preceding Tuesday would have itself exceeded the single-sample limit (235), given that the highest single value actually obtained on that Tuesday morning of August 22 was 11 *E. coli* per 100 mL. Even the day before August 27, the highest observed single density was 23. In fact, there is nothing from these data that would indicate this high indicator levels on Sunday, August 27 at West Beach, other than the monitoring results from that very day. The last rainfall prior to this day had been observed about four days earlier, and was fairly light (0.3"). There was a brisk on-shore wind on this particular Sunday, a factor which will be shown to have an impact at this beach, but stronger winds that were observed nine to eleven days prior to this failed to exhibit such a strong effect.

This is just one contradiction of many that can be gleaned from these plots. For example, at Wollaston Beach, the only beach at which there was not a preponderance of results on one side or the other of the recommended limit, only five out of eleven exceedences were captured by the 29-day Tuesday moving geometric mean.

Aside from the geometric mean over 29 days, it is obvious that variations in the geometric mean at a point in time can be quite large in themselves from one day to the next. In fact, for each of these beaches and considering the 9:00 a.m. results only, on at least half of the days during this study the change in geometric mean indicator level from the previous day was by a factor of 2 or more, that is, it more than doubled or halved from the previous day (Table 9). Day-to-day changes in indicator density in this study were large enough that the geometric mean over five Tuesdays, or, for that matter, any geometric mean of five observations over a 30-day period, will often not reflect true conditions at a beach at a given time.

Figure 8. Time series plots of geometric mean morning (9:00 a.m.) indicator levels. Heavy line indicates the 4-week geometric mean for samples collected on Tuesday morning of each week.
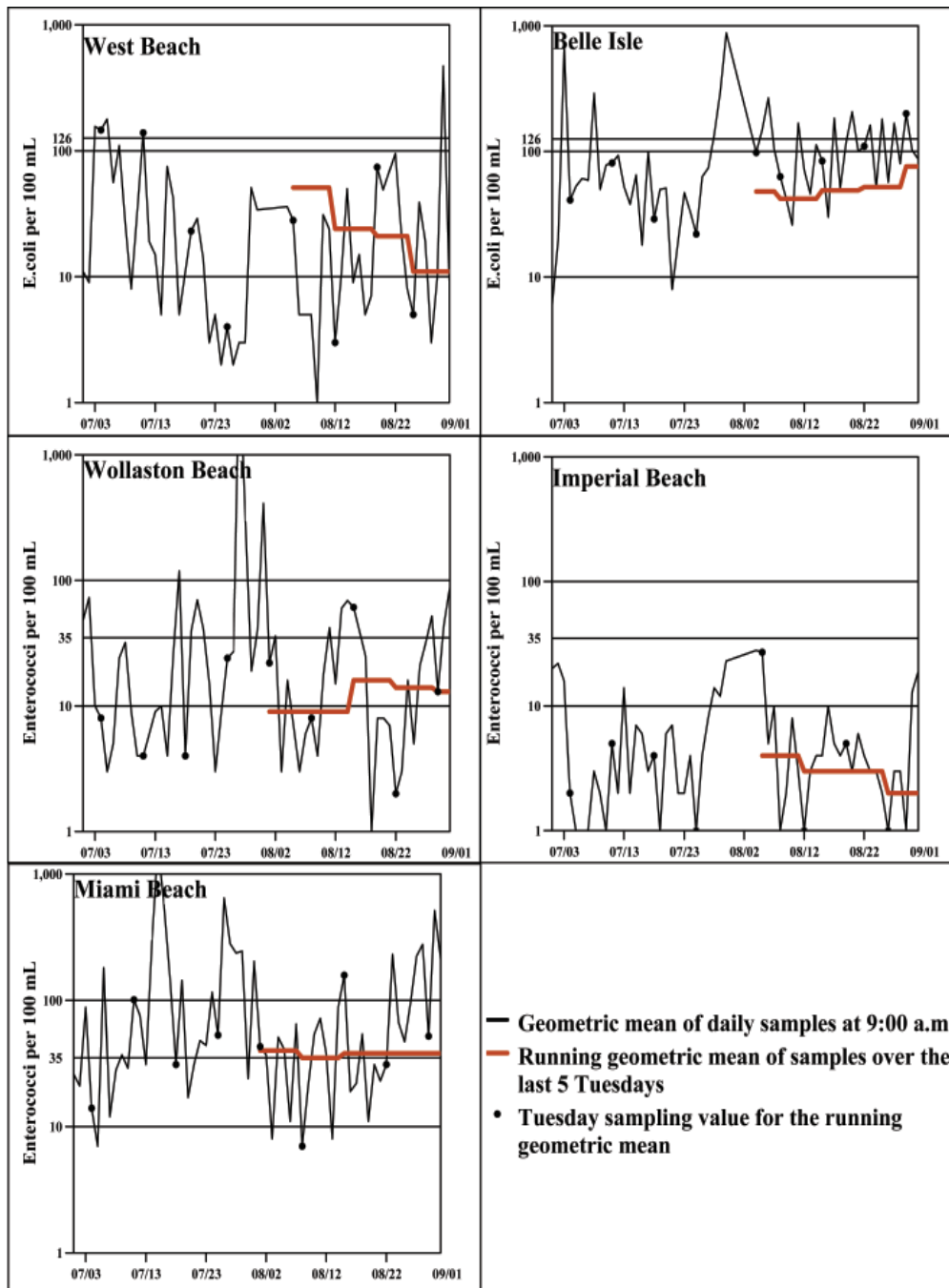
Table 9. Ratio to the previous day of the geometric mean (GM) of all knee- and chest-depth samples taken at the 9:00 a.m. visit. Interquartile range (25th and 75th percentiles) and frequency with which the geometric mean changes by a factor of 2 or more (ratio<0.5 or ratio>2) and 5 or more (ratio<0.2 or ratio>5).

Location	Number	Interquartile range for change in GM		Percent of days that the ratio to previous GM is:	
		25th %-ile	75th%-ile	<0.5 or >2	<0.2 or >5
West Beach	54	0.32	1.99	67%	20%
Belle Isle	56	0.55	2.34	43	21
Wollaston Beach	59	0.39	2.50	66	25
Imperial Beach	60	0.50	2.00	55	13
Miami Beach	61	0.33	2.64	69	28

Table 10 compares the serial correlation between indicator densities at each beach using a 1-day, 2-day, and 4-day time lag. This is simply the correlation between two time-related variables, the log mean indicator density over all samples on a given day at 9:00 a.m. and the corresponding log mean at 9:00 a.m. 1, 2, or 4 days later. The higher this correlation, a perfect correlation being +1 (or -1), the better will be our ability to predict results 1, 2, or 4 days in the future based on "today's" results. While none of these correlation values are particularly impressive, at least a positive relationship is indicated at most of the beaches between one day and the very next day. Correlation becomes markedly weaker when the lag is two days, and negligible at four days. These results do not support a scenario whereby water is sampled on Tuesday for a decision on whether it will be acceptable to swim next Saturday.

Significant day-to-day, week-to-week, or month-to-month variation has been observed in several studies (Seyfried, 1973, Cheung, 1990, Brenniman, 1981, PHLS, 1995). The differences were as great as two orders of magnitude between days, weeks, or months. Similar differences were observed in the current study.

Table 10. Serial correlations between log means separated by 1, 2, or 4 days

Location	Time lag between samples		
	1-day lag	2-day lag	4-day lag
Correlation coefficients[1]:			
West Beach	0.27	0.04	-0.03
Belle Isle	0.18	0.06	0.17
Wollaston Beach	0.44	0.18	0.13
Imperial Beach	0.45	0.28	-0.09
Miami Beach Pk.	0.40	0.30	0.18
P-values[2]:			
West Beach	0.039	0.789	0.805
Belle Isle	0.191	0.644	0.233
Wollaston Beach	0.001	0.176	0.329
Imperial Beach	<0.001	0.031	0.514
Miami Beach Pk.	0.001	0.021	0.180

[1]Correlations greater than about 0.24 are statistically significant ($\alpha=0.05$). Sample sizes are in the range of N=50-60 (days).
[2]Probability associated with the sample correlation under the null hypothesis that the true correlation is zero.

Variation between morning and afternoon results

A phenomenon that was observed to occur in three of the five beaches studied was a tendency for indicator levels to decrease between the morning (9:00 a.m.) and afternoon (2:00 p.m.) sampling events (Figure 9). We use $\log_{10}$ of the indicator density per 100 mL as the y-axis in these plots, since there are problems with scale in attempting to show the changes in geometric mean or percentage changes. Note that the predominant effect seen at Miami Beach Park, Wollaston Beach, and Belle Isle, the three beaches where indicator levels were relatively high, was a decline in the 2:00 p.m. reading as compared to the 9:00 a.m. value.

Consider the scenario of taking one Tuesday morning sample for a decision whether to close the beach on Saturday and those instances when a Tuesday sample from knee-deep water exceeds the EPA recommended limit for a single sample of 235 *E. coli* or 104 enterococci per 100 mL [1986]. Had that sample been collected on Saturday instead, would the decision to close the beach or not, have been the same? How well does the Tuesday sample discriminate between acceptable and unacceptable Saturday results? Does an exceedence on Tuesday indicate a higher risk of exceedence the following Saturday?

The first part of Table 11 examines each single sample from knee-deep water in terms of whether it exceeds the EPA limit and whether that sample taken from the very same location four days earlier had exceeded the limit. In effect, this gives the probability of exceedence four days from now when (in column 1) today's sample does not exceed the limit or (in column 2) today's sample does exceed the limit. One would hope that the probability of exceedence four days hence is substantially higher when today's sample exceeds the limit than when it does not. That is, column 2 should be substantially higher than column 1.

This is not the case. In fact, for the four-day time lag, the percentages in both columns are remarkably similar except at Wollaston Beach. It may well be that the validity of this sampling scenario varies by beach, and has merit at Wollaston. However, note that the difference at Wollaston is not statistically significant (p-value = 0.31).

What about sampling on Friday morning, the day before a decision must be made? We can see from Table 11 that a sample was generally more likely to exceed the limit if that same sample had exceeded the limit the previous day ("1 day earlier"). Sampling two days earlier may be somewhat, but not much, better than sampling four days earlier. In fact, note that all exceedences at Wollaston, the only beach where the four-day lag appeared to hold promise, are missed by the two-day sampling lag.

Table 11. Percentage of total samples collected that exceeded the single sample limit [U.S. EPA, 1986], classified by whether the sample from the same location was above or below the limit 1, 2, or 4 days earlier.

Location	Percent of samples that exceeded the EPA-recommended single sample limit (total number of samples in parentheses)[1]					
	4 days earlier the sample was:			1 day earlier the sample was:		
	< Limit		> Limit		< Limit	> Limit
West Beach	9.2%	(152)	9.1%	(22)	9.3% (161)	31.8% (22)
Belle Isle	90.6	(32)	91.5	(130)	50.0 (32)	93.2 (148)
Wollaston Beach	13.8	(152)	18.2	(22)	12.7 (158)	25.0 (24)
Imperial Beach[2]	0.0	(174)	-	(0)	0.0 (183)	- (0)
Miami Beach Park	62.5	(72)	62.4	(101)	52.8 (72)	67.3 (110)
	2 days earlier the sample was:					
	< Limit		> Limit			
West Beach	11.4%	(158)	19.0%	(21)		
Belle Isle	71.9	(32)	92.3	(142)		
Wollaston Beach	16.0	(156)	0.0	(22)		
Imperial Beach[2]	0.0	(180)	-	(0)		
Miami Beach Park	52.8	(72)	68.2	(107)		

[1] For example, of 152 samples found to be less than the single-sample limit (235 *E. coli* per 100 mL) in West Beach, 9.2% exceeded the limit 4 days later, while of the 22 samples that did exceed the limit, 9.1% were in exceedence 4 days later.

[2] There were no single sample exceedences at Imperial Beach.

Figure 9. Change in mean log₁₀ indicator density between the morning (9:00 a.m.) sample and afternoon (2:00 p.m.) sample.
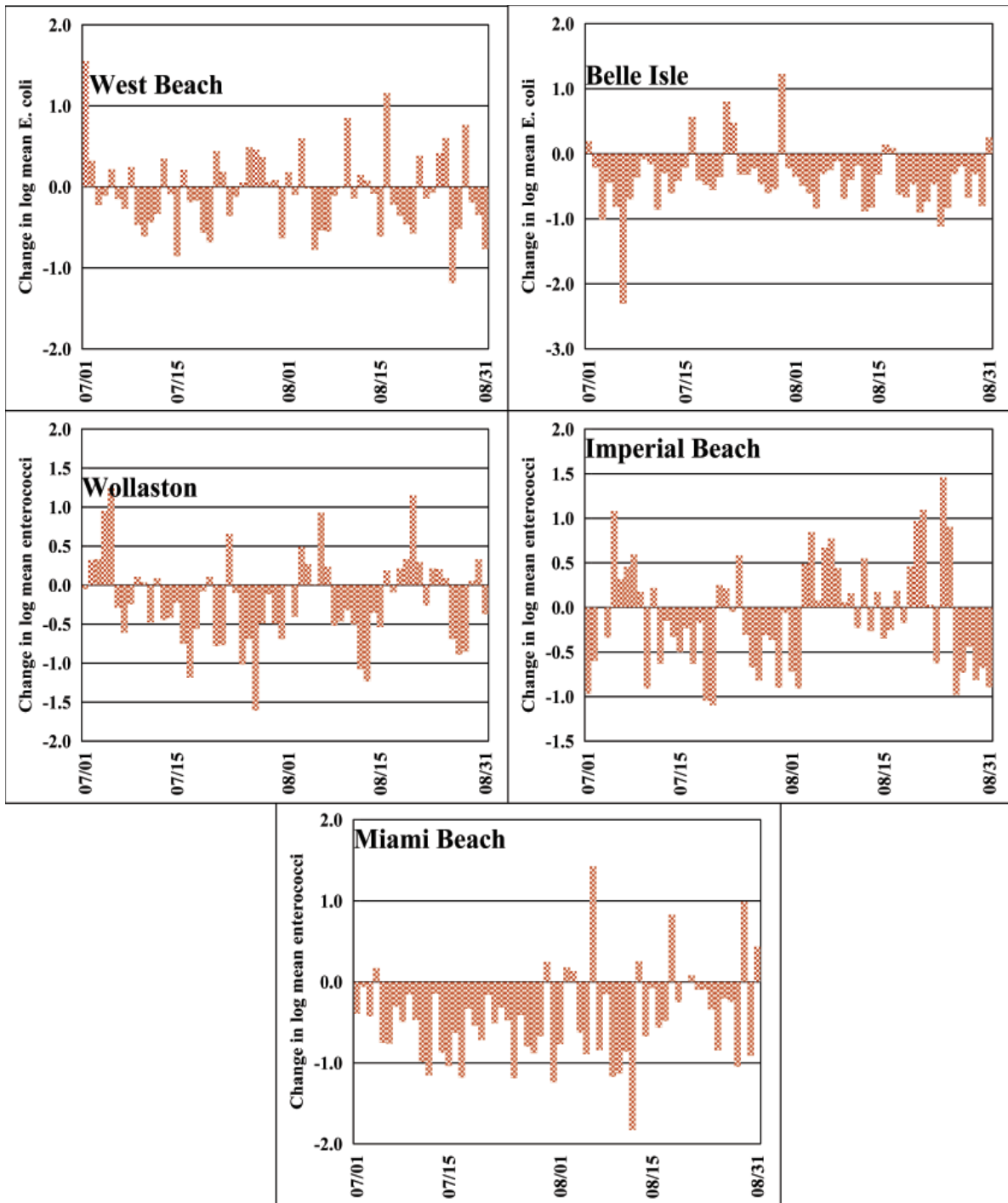
Previous studies [Sieracki, 1978] have demonstrated the die-off of *E. coli* after exposure to sunlight, an effect supported by these data when cloud cover at the time of the afternoon sample is considered. As seen in Table 12, and illustrated by Figure 10, indicator levels tended to be lower in the afternoon relative to the same morning on sunny afternoons (*i.e.* less than 50% cloud cover) at both freshwater beaches. At West Beach, lower *E. coli* densities were observed in sunny or mostly sunny afternoons, but higher levels were seen in the afternoon on cloudy days. *E. coli* at Belle Isle generally dropped in the afternoon even on overcast days, but decreased even more on those occasions when cloud cover was less than 50%.

Table 12. Comparison of geometric mean density of respective indicator organism at each study beach between morning and afternoon visits – sunny vs. cloudy days.

Location	Cloudy days (Greater than 50% cloud cover)			Sunny days (Less than 50% cloud cover)		
	Number	9:00 a.m.	2:00 p.m.	Number	9:00 a.m.	2:00 p.m.
West Beach	19	24	31	41	22	17
Belle Isle	19	202	125	41	235	60
Wollaston Beach	32	27	14	25	21	15
Imperial Beach	30	7	4	31	4	4
Miami Beach	32	149	80	30	84	20

Although enterococci have been shown to be relatively more resistant to solar radiation than are *E. coli* [Sieracki 1978], significantly lower densities of this indicator were noted at Miami Beach Park in the afternoon relative to corresponding morning levels on days when there was less than 50% cloud cover at the time of the afternoon visit. Neither of the other two saltwater beaches exhibited this effect to a significant degree. In our evaluation of changes between morning and afternoon, other effects, such as tidal stage, rainfall, air and water temperature, and bather load were included as co-variable with cloud cover, but failed to account for the apparent effect of sunlight at the Miami Beach Park location. Of course, this research only indicates association between indicator levels and cloud cover, not causation of die-off of either *E. coli* or enterococci because of exposure to sunlight. There may be other, unmeasured associations with cloud cover that influence indicator levels at any of the beaches.

Some studies in the literature have reported significant differences between the indicator densities obtained in the morning when compared with results obtained in the afternoon [Seranno *et al.*, 1998; Brenniman, 1981; Seyfried, 1973]. Indicator densities were generally highest in the early morning and lowest in the mid-afternoon. These authors point out that this time interval is closely related to the period of greatest solar radiation intensity. What we may be seeing is a confirmation of this phenomenon.

Figure 10. **Geometric mean indicator density for morning (9:00 a.m.) samples and afternoon (2:00 p.m.) samples, by whether afternoon was sunny (< 50% cloud cover) or not sunny.**
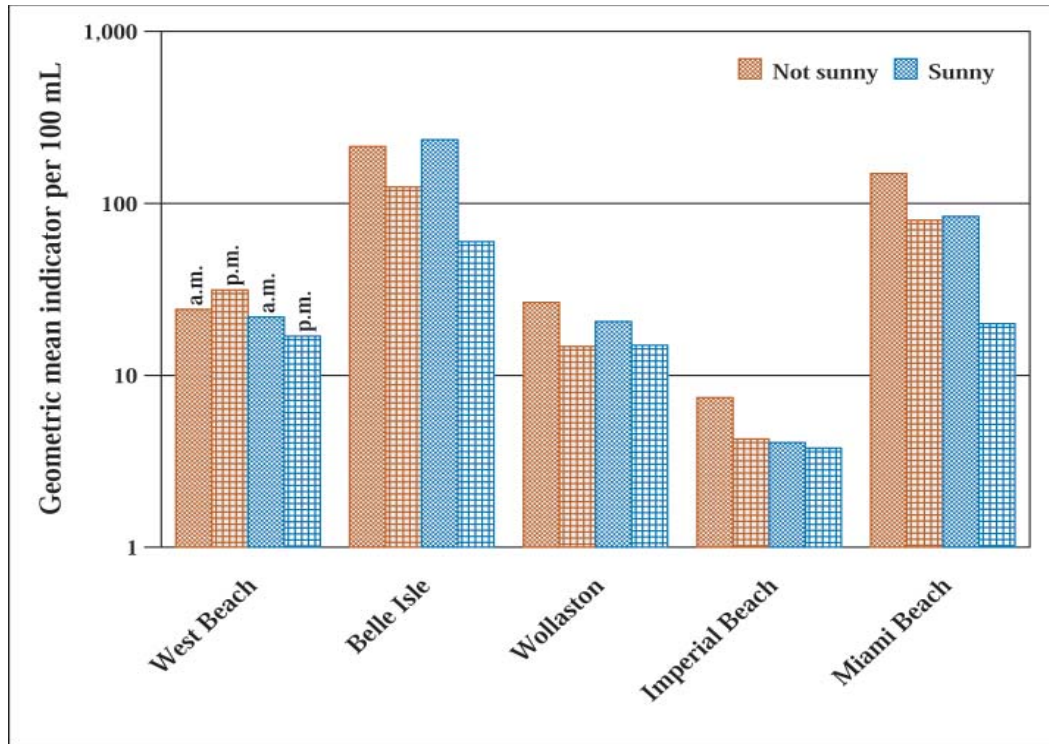
Results from the hourly sampling that was performed on several days at each study beach are also available for ascertaining whether there were any predominant diurnal patterns in indicator density. We can make at least a qualitative assessment of this by examining Figure 11, in which indicator levels observed on the hour (9:00 a.m. through 6:00 p.m.) relative to the overall geometric mean indicator level for that day are shown. The heavy lines in these graphs indicate the average (geometric mean) of this ratio for each hour.

With exception of West Beach, the tendency is for elevated indicator levels during the first sampling event of the day, at 9:00 a.m. This is particularly evident at Miami Beach Park, where average values for the first two hours were substantially higher than the overall daily mean, after which values tended to be much lower. As can be seen from the actual hourly results, deviations from this overall trend can be substantial in any case. These data are quite noisy; the best indication of overall trend still comes from the normal 9:00 a.m. and 2:00 p.m. daily sample results, which cover the entire two-month period.

Single sample limits and time of day

Again, consider the single-sample exceedence rate for those samples taken at 9:00 a.m. from knee-deep water, using the EPA limits of 235 *E. coli* or 104 enterococci per 100 mL. On those occasions when the morning sample exceeded the limit, at three of the study beaches a sample taken from the same location in that same afternoon was much more likely also to be in exceedence, when compared to those occasions when the morning sample did not exceed the limit. This is illustrated in Table 13. At West Beach, for example, of the 22 knee-depth samples that were over the 235 per 100 mL standard for *E. coli* at 9:00 a.m., 40.9% were also higher than this limit at 2:00 p.m., as compared to 7.3% of the "acceptable" 9:00 a.m. samples that exceeded 235 by 2:00 p.m.

Table 13. **Percentage of knee-deep samples collected at 2:00 p.m. that exceeded the single sample limit as recommended in EPA guidelines [U.S. EPA,1986], by whether the sample from the same location was above or below the limit at 9:00 a.m.**

	Percent of samples above the EPA recommended single sample limit at 2:00 p.m. (total number of samples in parentheses)	
Location	**< Limit at 9:00 am**	**> Limit at 9:00 am**
West Beach	7.3 (164)	40.9 (22)
Belle Isle	21.9 (32)	74.8 (151)
Wollaston Beach	7.5 (160)	11.5 (26)
Imperial Beach	1.1 (186)	- (0)
Miami Beach Park	19.2 (73)	52.2 (113)
	% above the limit:	
	At 9:00 a.m.	**At 2:00 p.m.**
West Beach	11.8 (186)	11.3 (186)
Belle Isle	82.5 (183)	65.6 (183)
Wollaston Beach	14.0 (186)	8.1 (186)
Imperial Beach	0.0 (186)	1.1 (186)
Miami Beach Park	60.8 (186)	39.2 (186)

In total, single sample exceedences were much more likely to occur in the morning at both Belle Isle and Miami Beach Park. At West Beach, the relative number of exceedences was about the same at the morning and afternoon visits. Considering both parts of Table 13, it can be seen that a morning sample from a given location in the water that exceeds the recommended limit is nearly four times as likely also to exceed that limit in the afternoon compared to what may be expected by chance (40.9% compared to 11.3%).

Figure 11. Geometric mean (GM) indicator density by time-of-day (9:00 a.m. – 6:00 p.m.) for samples collected on hourly sampling days - ratio of GM on the hour to overall GM for the day. Average ratio for each hour is shown by the heavy line.
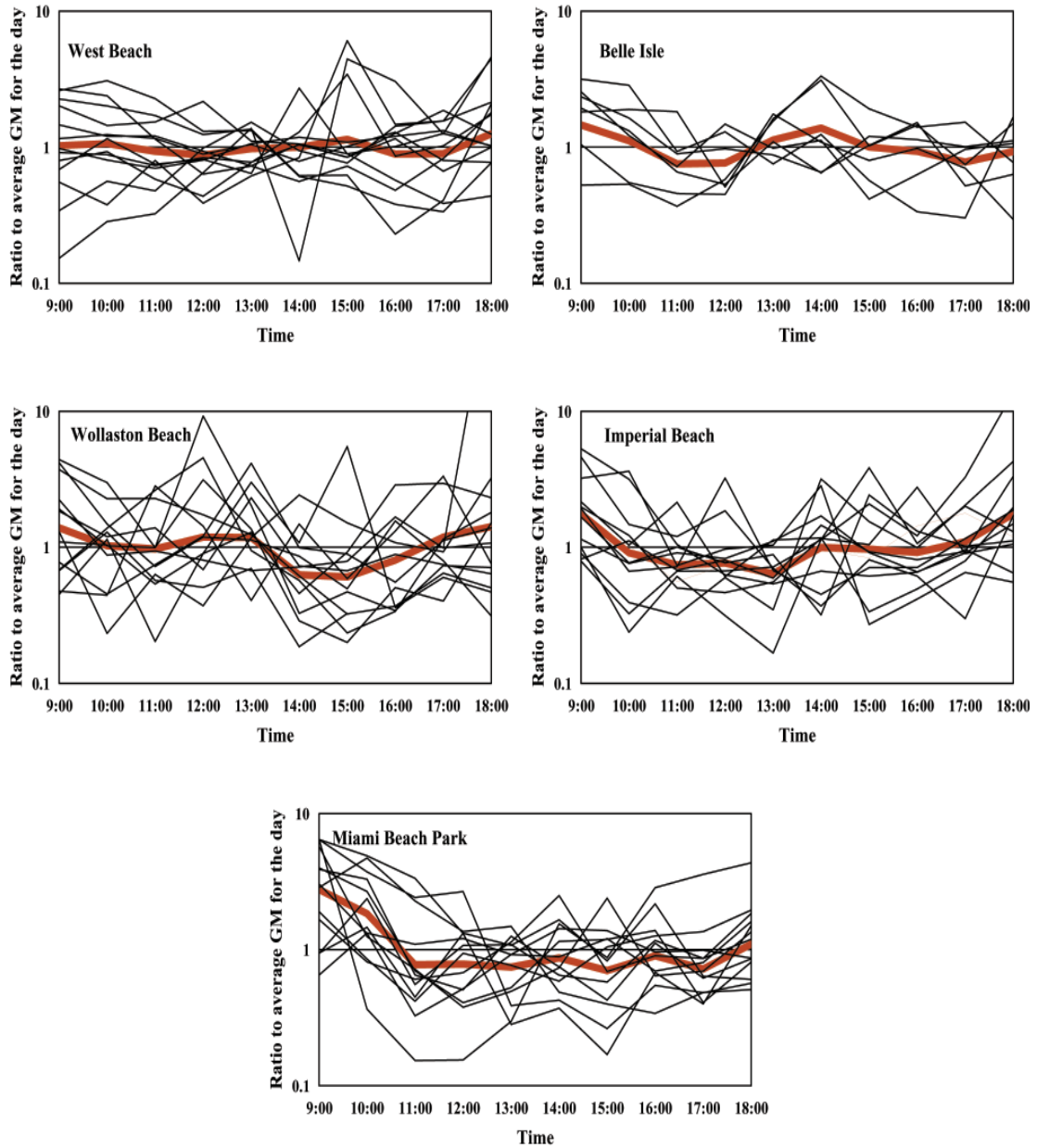
Hourly single-sample exceedence rates among knee-deep water samples are shown in Table 14. Exceedences were lower by mid-afternoon at Wollaston and Miami, confirming what we saw earlier when comparing only 9:00 a.m. and 2:00 p.m. results. Belle Isle, which showed a tendency to decline between 9:00 a.m. and 2:00 p.m. when all data were considered, does not show this tendency when hourly results alone are considered. Then too, because of logistic problems, only eight days of hourly sampling were completed at this beach, compared with a full two weeks at the other sites.

Table 14. Percentage of knee-deep samples that exceeded the single sample limit [U.S. EPA, 1986], by time-of-day. Only hourly sampling results are used. For all but Belle Isle, there were 14 hourly sampling days and 39-42 valid knee-deep observations (for Belle Isle there were 23 or 24 observations over 8 days).

| Time of day | Percent of knee-deep samples that were in exceedence | | | | |
	West Beach	Belle Isle	Wollaston Beach	Imperial Beach	Miami Beach Park
09:00	2.4	79.2	23.8	0	78.6
10:00	4.8	52.2	16.7	0	63.4
11:00	4.8	39.1	14.3	0	52.4
12:00	7.1	54.2	33.3	0	45.2
13:00	0	58.3	26.2	0	40.5
14:00	2.6	75.0	7.1	4.8	35.7
15:00	4.8	58.3	14.3	0	26.2
16:00	0	75.0	5.1	0	51.2
17:00	0	66.7	20.5	0	38.1
18:00	2.4	79.2	19.0	2.4	50.0

Environmental and bather effects

A summary of environmental and bather density data collected reveals some striking differences among the study beaches (Table 16). West Beach was by far the most heavily used, followed by Imperial Beach. The water at Wollaston Beach was particularly turbid in comparison to the other beaches with an average of 56 NTU. Note that turbidity at any of the study beaches was highly variable, with the standard deviation of the turbidity values approximately equal to its mean in every case (data not shown).

Particularly noteworthy is that weather conditions at Imperial Beach were fairly uniform throughout the period of the study. Most of the afternoons were sunny or mostly sunny, only trace amounts of rainfall were recorded, and the onshore wind was ever present. This constancy obviously limits any conclusions that may be made regarding weather related effects at Imperial.

Table 15. Environmental characteristics and bather densities of EMPACT study beaches.

	West Beach	Belle Isle	Wollaston Beach	Imperial Beach	Miami Beach Park
Sunny/mostly sunny days [1]	67%	44%	50%	71%	48%
Avg. air temperature (ºC) [1]	25.0	24.7	26.6	22.5	27.5
Avg. water temperature(ºC) [1]	23.6	24.1	22.2	19.8	27.4
Total rainfall (in.)	9.9	13.2	14.2	< 0.1	13.7
# days with rain	14	17	14	2	22
Offshore winds [1,2]					
- Frequency	31%	47%	26%	0%	24%
- avg. wind speed (mph)	4.1	4.5	3.0	-	3.3
Onshore winds [1,2]					
- Frequency	59%	25%	39%	81%	56%
- avg. wind speed (mph)	8.3	6.3	5.8	4.9	4.6
Avg. wave height (M)	0.30	0.02	0.11	0.36	0.13
Avg. turbidity (NTU) [3]	6	4	56	1	7
Avg. tide range (ft.) [4]	-	-	9.9	4.2	0.6
# bathers on the beach [1]					
- < 20	6	34	52	19	50
- 20-100	21	21	10	37	11
- 101-200	20	1	0	2	1
- > 200	14	1	0	1	0
# bathers in the water [1]					
- < 20	19	37	62	30	52
- 20-100	31	18	0	29	10
- 101-200	9	2	0	0	0
- > 200	2	0	0	0	0

[1] at 2:00 p.m.   [2] Wind at a direction of at least 45º with respect to shoreline
[3] as measured from the central sampling location in the water (9:00 a.m. & 2:00 p.m.)
[4] average difference in water level between succeeding low and high tides

Statistical evaluation

In evaluating environmental and bather effects, the goal was to develop a model for each beach that might be used in predicting log mean indicator density 24 hours later based on its current value, since with the culture methods used, "current" values are not known until the next day. Thus, the previous day's $\log_{10}$ density is present in all of these models, even though it proves to be non-significant in one case. A stepwise regression procedure [Draper and Smith, 1981] was utilized on the potential predictor variables, with the significance level ("p-value") for entry or removal from the equation set at 0.10. Trial and error was involved in selecting the "best" regression as well. In some cases, forcing one or two variables into the equation, in addition to the previous day's density, resulted in

some improvement in its explanatory power with little or no increase in its complexity (number of independent variables in the model).

These analyses are exploratory, indicating variables which *might be* useful predictors of log microbial densities. There is a tendency for any technique designed to select a "best" regression model to overstate the fit, with regard to significance levels, partial or total R^2, or any other measure. To the extent that the same variables are suggested as being good predictors in a number of different beaches, this suggests that these variables really are important. Several variables show up consistently in these models: rainfall, bathers on the beach, etc. There is good theoretical rationale to believe that these variables make a difference in water quality. Consequently further examination of these variables in more detail will be presented.

Potential explanatory factors considered in this section are not "design variables", in contrast to the time-of-day, day-of-week, depth zone, and transect of the previous section. Rather, we take environmental and human factors as they occur. Often, these variables may be correlated with one another, and, thus, we do not observe an effect that is purely a result of a single factor. For example, it is probable that bather density at a beach, the occurrence of rain in the past 24 hours, whether it is sunny or not, and the temperature are all interrelated. We cannot isolate their effects, because we did not, for example, prevent bathers from going to the beach on warm, sunny days. Therefore, some variables may be surrogates for others or may mask the effects of other variables.

In addition, if some condition never occurred, or happened only infrequently, in the study, we cannot infer an effect from that condition. Such was the case for offshore winds and rainfall at Imperial Beach. Rainfall was observed on only two occasions at this location. It may well be that rainfall has an overwhelming effect on conditions at Imperial Beach, but, if so, we were not afforded the opportunity to observe this effect.

Relationships found at the study beaches are specific to the respective location during the time period of the study; blindly generalizing these results to other beaches should be avoided. Such correlates with indicator levels as found in this section, however, may serve as a general guideline for related research at other sites.

Table 16 summarizes the results of modeling for each of the study beaches. The regression models shown are not the only models possible, but are representative of a group of models, any of which might be appropriate.

The corresponding log density of indicator organisms 24 hours earlier is included as a covariate in all models in Table 16. In one case, that of Miami Beach Park, the relationship is seen to be non-significant. In this case, there is a highly significant relationship between the 1-day lagged log indicator densities by themselves (Table 10 shows this), but the inclusion of other variables is sufficient to account for this relationship.

The last column of Table 16 gives the *coefficient of determination*, R^2, for each model. This is the proportional of total variation (sum of squared deviations from predicted value) of the dependent variable that is accounted for by the regression. A "partial R^2" is also listed, which indicates the extra proportion of variation accounted for by the other covariates over that which is accounted for by knowledge of the previous day's log density alone. The partial R^2 is a critical consideration in improving our prediction of water quality based on the previous day's microbial levels, as will be discussed in the next section, <u>Sources of Sampling Variance</u>.

Table 16.   Results of regression analysis on $\log_{10}$ density of indicator organism.

Factor	Coefficient	P-value[1]	R^2
West Beach (N=120) [2]	Int.=1.032[2]		
Sunny?[3]	-0.329	< 0.001	R^2 = 0.34[5]
Wind vector[4]	0.029	< 0.001	Partial R^2 = 0.24 [6]
Rain in past 3 days? [3]	0.229	0.024	
PM & rain in past 24 hrs.?[3]	-0.326	0.032	
Previous day $\log_{10}$ density	0.280	< 0.001	
Belle Isle (N=113)	Int.=1.769		
Sunny?[3]	-0.180	0.005	R^2 = 0.42
2:00 p.m.?[3]	-0.281	< 0.002	Partial R^2 = 0.28
Rain in past 24 hrs.? [3]	0.173	0.016	
<20 bathers on beach?[3]	0.148	0.076	
Previous day $\log_{10}$ density	0.212	0.010	
Wollaston Beach (N=121)	Int.=0.928		
2:00 p.m.?[3]	-0.259	0.007	
Rain in past 24 hrs.? [3]	0.396	0.001	R^2 = 0.37
Low air temperature?[3,7]	0.239	0.029	Partial R^2 = 0.24
Onshore wind?[3]	0.195	0.049	
<20 bathers on beach?[3]	-0.334	0.051	
Water level[8]	0.043	0.003	
Previous day $\log_{10}$ density	0.234	0.005	
Imperial Beach (N=118)	Int.=0.236		
<20 bathers in water?[3]	-0.246	0.005	
Water level			R^2 = 0.31
- if 9:00 a.m. sample[9]	0.158	< 0.001	Partial R^2 = 0.21
- if 2:00 p.m. sample[9]	0.102	0.007[10]	
Previous day $\log_{10}$ density	0.221	0.010	
Miami Beach Park(N=122)	Int.=1.326		
Rain in past 24 hrs.? [3]	0.258	0.009	
Wind vector	0.041	0.004	R^2 = 0.42
Water level[8]	0.290	< 0.001	Partial R^2 = 0.37
PM & sunny?[3]	-0.111	0.031	
Water temp. > median?[3]			
- if 2:00 p.m. sample[11]	-0.387	0.004	
Previous day $\log_{10}$ density	0.071	0.371	

[1] "Significance level" of the variable – required to be < 0.10 for stepwise selection.
[2] N=# of observations, Int.=intercept [3] Dummy variable: value = 1 if the condition is true, 0 if false.   [4] Wind speed x direction vector, where the direction vector is +1 for a wind blowing straight onshore, -1 for offshore. [5] Proportion of variation explained by the regression. [6] Additional proportion of variation explained by environmental and bather factors over that explained by previous day's log density alone. [7] Air temperature less than the 25th percentile for the respective time-of-day.   [8] Water level in feet above mean lower-low tide mark. [9] Effect differs between morning and afternoon. [10] P-value for the difference between morning and afternoon effects.   [11] Water temperature higher than the median (for that time of day) is a significant effect only at 2:00 p.m.

Tides

Among the three beaches in this study subject to tidal movement, tide stage is seen to be a significant determinant of enterococci density within the swimming area; in fact, in terms of the p-values of Table 16, tide stage has the strongest association with microbial contamination of all the environmental factors examined at these beaches. We note that the water level, measured in feet above the mean lower-low tide mark (the lower of the two low tides in a 24-hour tidal cycle), was found to have a stronger association with bacteria densities than the actual tide stage itself, which we defined as high or low when samples were collected within 1.5 hours of the actual event. Neither did it seem to matter whether the tide was incoming or outgoing. Figure 12, data for which are shown in Table 17, illustrates the observed water level effect, comparing geometric mean enterococci levels for times when the water level was in the lower and upper quartiles of its range at each particular beach for both the morning and afternoon visits. Ranges of water level, at least over the times when sample collection took place at that beach, are also indicated in Table 17.

Table 17. **Comparisons of geometric mean enterococci density per 100 mL at low, mid, and high water levels for marine beaches (based on 25th and 75th percentiles of water level).**

Location	9:00 a.m.			2:00 p.m.		
	Water level:			Water level:		
	Low	Mid	High	Low	Mid	High
Wollaston Beach	12	28	37	11	12	22
Imperial Beach	2	5	11	4	3	6
Miami Beach Park	85	101	140	20	49	228

In each case, enterococci densities are seen to be higher when the water level was in its upper quartile compared to densities obtained during lower quartile water stages; this is so even at the estuarine beach, Miami Beach Park, where tidal variations were relatively low. The fact that water level shows a stronger association with microbial levels than does tidal stage itself or whether the tide is incoming or outgoing suggests that suspension of sediment or debris from the beach may contribute to contamination, perhaps in tandem with waste washing in from deeper waters. Shoreline sediment, of course, would constantly mix with bathing waters, and may ultimately affect water of swimming depths whether the tide is coming in or going out, but sediment higher on the beach is more exposed to humans and animals as well as to air and may be richer in microbial content.

Weather

We had previously discussed the apparent effect of sunlight on afternoon readings. Other weather factors are also seen to play important roles in determining microbial contamination. These vary among the beaches of this study, and include wind, particularly whether the wind is blowing onshore or offshore, the occurrence of rain in the hours prior to sampling, and air or water temperature. Rainfall exhibited a strong association with water quality in three of the study beaches, as did wind velocity and direction, while association with air or water temperature was found in two of the beaches.

Figure 12.   Geometric mean enterococci density by tide stage at marine beaches.
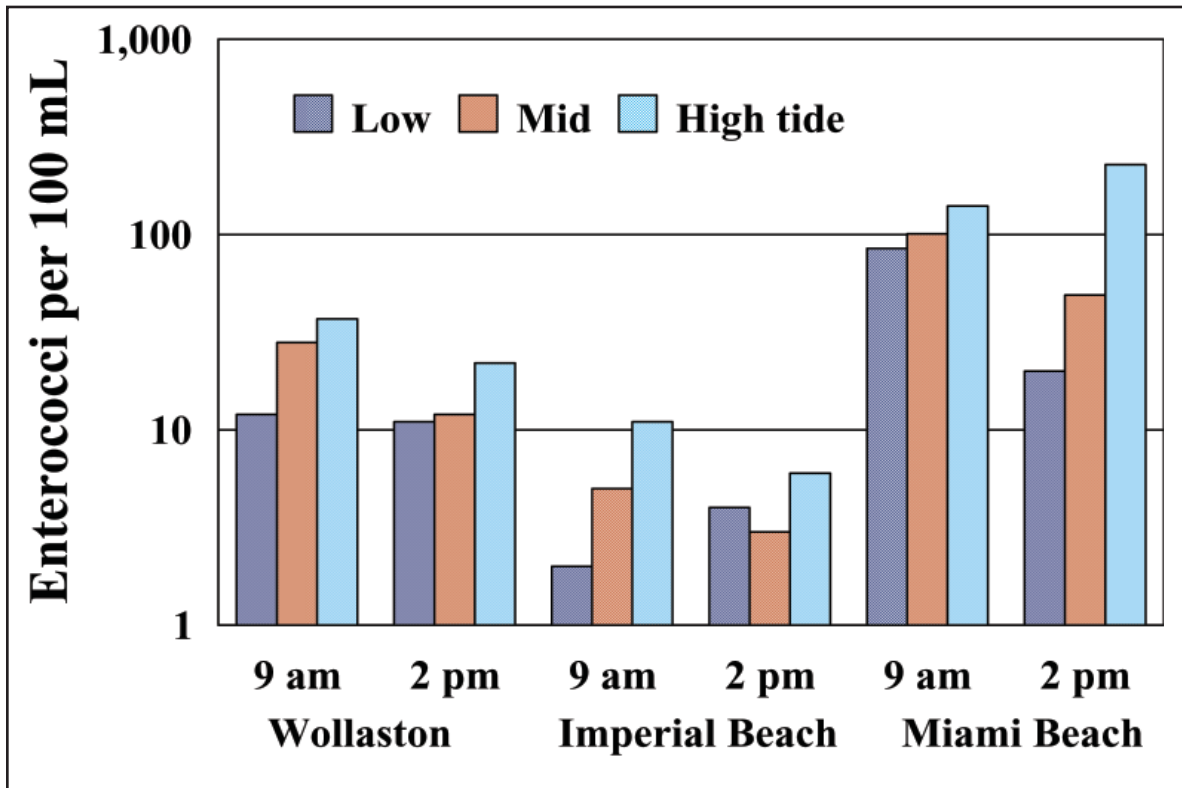
These various weather effects are illustrated in Figure 13 and summarized in Table 18. Significant rain effects were observed at Miami Beach Park, Wollaston Beach, and Belle Isle. In all three cases, as might be expected, the occurrence of rainfall within the past 24 hours is associated with higher indicator levels at the beach. Note that measurable rainfall occurred on only two occasions at Imperial Beach; thus, we have no real opportunity to discern rain effects due to rain at this location. While previous 24-hour rainfall fails to be a significant factor at West Beach, there is an evident rainfall effect when we consider rain in the past three days, as indicated in Table 18. In essence, then, rainfall was observed to have an effect on water quality in all cases where we had the opportunity to observe this effect. A 24-hour window is arbitrary; in truth, rain effects on recreational water quality may likely vary from beach to beach, depending on such factors as the size and gradient of the drainage area.

Wind vector at the beach, which takes into account the wind direction as well as wind speed, is a significant factor at Miami Beach Park, Wollaston, and West Beach, in each case resulting in higher geometric mean indicator densities with a relatively strong onshore wind component and lower densities with an offshore component. Once again, we do not have the opportunity to gauge an effect at Imperial Beach, because offshore winds never occurred at this site; its graph in Figure 13 compares light or calm onshore wind conditions with moderate onshore winds. At Belle Isle, *E. coli* levels were actually somewhat lower, but not significantly so, when winds were onshore in the morning, but virtually zero difference with wind vector is seen in the afternoon.

Figure 13.   Geometric mean indicator density for various weather conditions.
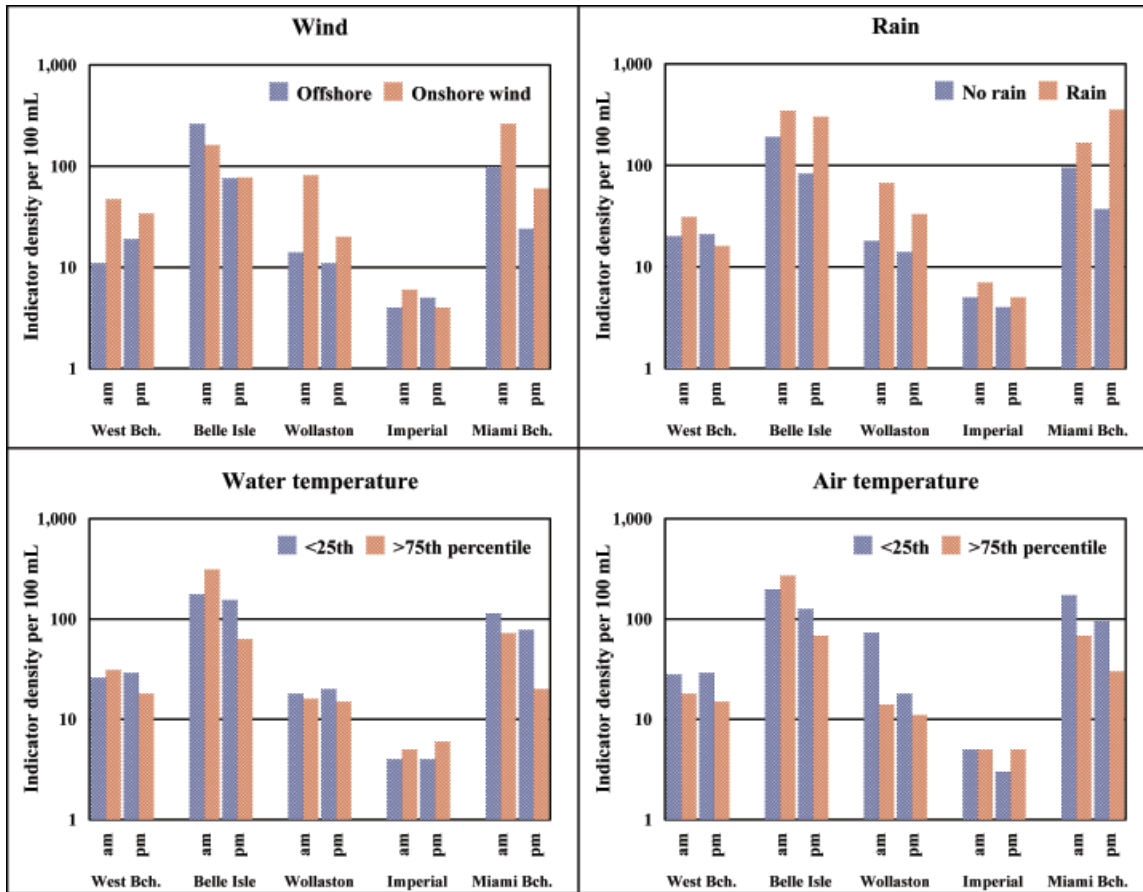
Elevated water or air temperatures might be predicted to result in an increase in microbial counts at a beach. However, either may also be associated with increased sunshine. Even when cloud cover at the time of sampling is accounted for, there may have been occasions on which cloud cover was more than 50%, at the time of sampling, yet water temperature was relatively high because of earlier sunshine. This is one possible explanation for both cases where either water or air temperature was associated with decreased densities of indicators organisms, at Miami Beach Park (water temperature) and Wollaston (air temperature), as indicated by Figure 13. Note that *E. coli* levels were generally higher at Belle Isle when water temperature was relatively high compared to those occasions when water temperature was on the low side, but only on the afternoon visits. The opposite was true for this beach on morning visits, and this is a statistically significant difference between morning and afternoon effects of temperature at this particular beach. Whether changes in water temperature have an effect on growth of these organisms may also be supposed to depend on the range of temperatures involved; water temperatures at Imperial Beach were generally the lowest among these study beaches, reaching a high of only 23º C, possibly still too low to promote bacterial growth.

Table 18. Comparisons of geometric mean indicator density per 100 mL for various weather-related variables.

	9:00 a.m.		2:00 p.m.	
	Rained in the past 24 hours?			
Location	**No**	**Yes**	**No**	**Yes**
West Beach	20	31	21	16
Belle Isle	191	346	83	301
Wollaston Beach	18	67	14	33
Imperial Beach	5	7	4	5
Miami Beach Park	95	167	37	354
	Wind vector[1]			
	Strong offshore	**Strong onshore**	**Strong offshore**	**Strong onshore**
West Beach	11	47	19	34
Belle Isle	261	161	76	77
Wollaston Beach	14	81	11	20
Imperial Beach[2]	4	6	5	4
Miami Beach Park	99	262	24	60
	Water temperature extremes[3]			
	Cold	**Warm**	**Cold**	**Warm**
West Beach	26	31	29	18
Belle Isle	176	311	154	63
Wollaston Beach	18	16	20	15
Imperial Beach	4	5	4	6
Miami Beach Park	114	72	78	20
	Air temperature extremes[3]			
	Cold	**Warm**	**Cold**	**Warm**
West Beach	28	18	29	15
Belle Isle	198	270	126	68
Wollaston Beach	73	14	18	11
Imperial Beach	5	5	3	5
Miami Beach Park	172	68	95	30

[1] Upper 50th percentiles of offshore & onshore winds. [2] No offshore winds; comparison is between lower & upper 25th percentile of (onshore) winds. [3] Lower & upper 25th percentile of recorded temperatures correspond to "Cold" & "Warm", respectively.

Bather density

Bather density at a beach was recorded in order to ascertain whether the presence of bathers itself influences microbial contamination. Both bathers on the beach and bathers in the water at the time of sample collection were categorized as fewer than 20, 20-100, 101-200, and more than 200. These are further condensed for our purposes into two categories, 0 to 19 and 20 or more, since occasions on which over 100 bathers were recorded at any of the beaches are rare. Figure 14 compares indicator densities at the 2:00 p.m. sample collection for each of the beaches with regard to whether there were fewer than 20 bathers either on the beach or in the water, data for this figure being shown in Table 19.

Bather effects were found to be significant at Wollaston and Imperial Beaches, where an increase in numbers of bathers was found to result in a corresponding increase in enterococci levels. At Wollaston Beach, few bathers were ever observed to be in the water at any given moment, but the presence of substantial numbers of bathers on the beach was adequate to trigger the increase in microbial levels in the water. People on the beach in greater numbers implies the possibility that there are more actual bathers in the water, whether more than 20 or not, or that there had been more bathers in the water prior to the time of sampling.

Three studies looked at the potential for bather load to affect indicator density [Sekla, 1987; Cheung *et al.*, Seranno *et al.*, 1998]. Sekla indicated that in spite of bather numbers varying between 0 and 3000, there was no related variation in indicator densities. Seranno *et al.*, on the other hand, clearly showed that as the number of bathers increased the density of indicator bacteria decreased. Cheung *et al.* were able to observe a correlation between indicator density and number of bathers at only one beach out of nine beaches and only on one day. These differences in observed results may be related to the location of the study sites and to the type of dispersion of contamination at any one site. Locations with very little water movement may result in data where indicator densities increase with number of bathers. At locations where there is significant wave action, indicators shed into the water may be dispersed very rapidly and the relationship between indicator and bather density cannot be developed. It appears that the effect of bather load on water quality measurement will have to be determined on a case-by-case basis.

Figure 14. **Geometric mean indicator organism density by number of bathers on the beach and in the water.**
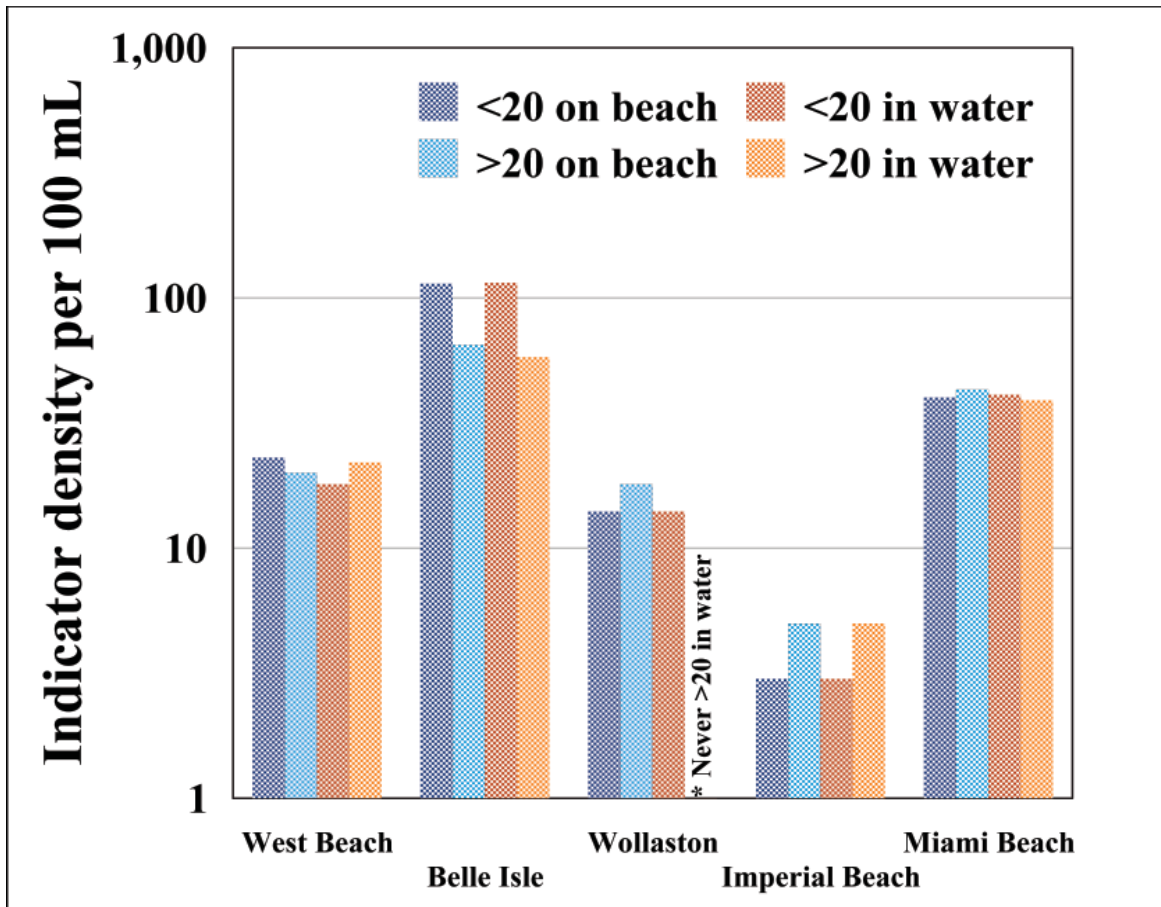
Table 19. **Comparisons of geometric mean indicator density per 100 mL. vs. number of bathers in the water and on the beach, for the afternoon (2:00 p.m.) sampling visits.**

Location	Bathers on the beach		Bathers in the water	
	0-19	20 or more	0-19	20 or more
West Beach	23	20	18	22
Belle Isle	114	65	115	58
Wollaston Beach	14	18	14	--
Imperial Beach	3	5	3	5
Miami Beach Park	40	43	41	39

Sampling Variance

Sources of variation

Populations, sampling, variance, and components of variation

Sampling variance is a measure of the variability of a measurement, and is calculated from a sample as

$$V = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}{n-1}, \qquad (1)$$

in which n is the sample size, $\bar{x}$, the sample mean, and x_i is the ith sample value for the measurement (i=1 to n). The x_i are assumed to be results that have been realized from a random sampling of some *population* encompassing all of the possible individual x values that might be obtained.

By the term "population," we mean the aggregate from which samples are collected and to which the variance given by Formula 1 applies. It is the population to which any inferences derived from sampling apply. Later, we will discuss in detail the Data Quality Objectives (DQO) Process [U.S. EPA 2000b], which is a formal framework for considering study objectives (the inferences and decisions to be made), population, and sampling. For now, we wish simply to emphasis the importance of the concept of a population to sampling and appropriate variance components for various population concepts.

In the context of water sampling, the population might be, for example, daily values of the indicator density in waist-deep water midway along the designated beachfront at 9:00 a.m. In this case, a random sample of days would enable one to make an inference such as "the geometric mean density was X over the last thirty days." Another example of a population could be the entire stretch of waist-deep water within the "bather area" at noon on a specified date. The inference becomes "the geometric mean density was Y in waist-deep water at noon on the 23rd of July."

If day-to-day variability (at the same point in the water and time-of-day) is greater than between-locations variability (at the same point in time and water depth), then Equation 1 is expected to give a higher value for V for samples that have been collected among different days than for those collected among different locations. "Among days" and "among locations" are two sources of variation, or variance components, that can be thought of as arising from two different, independent populations, as in this example. Sources of variation relevant to the EMPACT Beaches Study are discussed next. For each source, the variance component due to that source is calculated with all other sources, spatial and temporal, held fixed.

Sampling distribution

Estimates of variance components are used to assess the accuracy of a mean derived from sampling based on a model of the *sampling distribution*, which gives the underlying probabilities of obtaining a sample value within a given range. We hope to be able to utilize fairly small sample sizes in beach monitoring – current EPA guidelines [1986] specify a single sample - a worthy goal of any sample design, and, therefore, cannot rely upon some large sample theory. Some assurance is needed that normality applies, or else we need to determine the sampling distribution that does apply.

Previous research has indicated that $\log_{10}$ indicator densities from environmental samples tend to be approximately normally distributed [Fleisher and McFadden, 1980; Gannon and Busse, 1989; Mujeriego *et al.*, 1980]; that is, valid inferences regarding the means of such data can be made in using normal theory. Sufficient data have been collected in this study to enable an evaluation of the consequences of using a normal model. This evaluation indicates that this is a reasonable model of the $\log_{10}$ densities, resulting in probability statements that are in error by only +/-1% (*i.e.*, a 95% confidence interval may actually be a 96% confidence interval). Given that a normality assumption is appropriate for indicator sampling, this allows us to construct confidence limits for the log mean density based on the sample log mean, log variance, and sample size using "z" (standard normal distribution) values and to perform z tests in comparing the sample log mean to some critical value.

Components of spatial variation

On eight days at each of the five beaches that participated in this study, replicate samples were collected at each location in the water during both the 9:00 a.m. and 2:00 p.m. sampling visits. On these occasions, ten samples were collected from the central location (Location 5, knee-deep water at Transect 2) and two samples, at the remaining eight locations. These samples yield information on the small-scale variability inherent in measuring recreational water quality. Only if the water of the bathing area were perfectly mixed, by definition, would the variability of water quality between different locations in the water equal this small-scale value. Otherwise, between-locations variability will always be higher.

The study design used explicitly accounted for all three physical dimensions. Locations in the water at which samples were collected varied in directions parallel to the beachfront, at right angles to the beachfront, and vertically from different depths within the water column.

We have seen that the zone from which the sample was collected has the greatest predictable impact on microbial indicator densities of any factor considered in this study. In each of the study beaches, bacterial densities become progressively lower as one moves from ankle-deep to knee-deep to chest-deep water.

On the other hand, as long as the sample came from the same depth zone, there were only minor consistent differences among transects at two of the beaches. Therefore, given a sample of water, it makes a great deal of difference in our ability to assess overall water quality to know how deep the water was where that sample was collected, but makes little or no difference to know what where along the beachfront it was collected.

Likewise, we found no consistent differences among the various sampling depths. Recall that, on four days, samples were taken from two depths in knee-deep water (0.3 and 0.425 M from the surface) and from three depths in chest-deep water (0.3, 0.75, and 1.425 M from the surface), whereas a standard sampling depth of 0.3 M was used on all other days.

In evaluating their respective variances, the different zones, transects, and sampling depths from which sample are taken are regarded as having been randomly selected. In reality, however, these dimensions were fixed at the start of the study. Thus, the sampling plan is similar to a *systematic* sample (with the exception that a random starting point is not determined each time). Given the observed differences among zones, it is apparent that the value for the variance among zones will be dependent on exactly how these zones are defined. Had waist-deep been used instead of knee-deep, the resulting variance may have been much different. Likewise, for different transect spacing or sampling depths. We shall

see, however, that this is not so great a concern – zone and sampling depth variance are not relevant, and a way of accounting for transect separation is shown in the next section.

How is the population defined spatially?

- If we consider the population to consist of water 0.3 M below the surface within one zone, then variation in sampling depth or between zones are not considerations in our sample design.
- Variance between zones is important when we expand the population to include all of a given area of water. This may be beyond limits of practicality, and is possibly not even desirable since health studies themselves were based on approximately waist-deep water results.
- We may not have to consider them in sample design, but knowledge of the magnitude of inter-zone and depth variances is useful in describing the indicator levels that may be experienced along these dimensions.

Components of temporal variation

We estimate the variance of indicator density among days for samples collected at the same location in the water and at the same time of day. This estimates the variance among, for example, samples collected in knee-deep water along the leftmost transect at 9:00 in the morning, where a single sample is taken on each of several days. Because the duration of this study was 62 days, from July 1 to August 31, the variance thus estimated is appropriate for a set of random dates chosen out of a sixty-two day period of time.

Another temporal component is the intra-day variability, the variation in indicator levels throughout a day. The eight to fourteen days of sampling at each beach during which samples were collected on the hour from 9:00 a.m. through 6:00 p.m. will enable us to determine an hourly variance.

How is the population defined temporally?

- EPA criteria refer to a geometric mean over 30 days. The population implied by this consists of microbial water quality during those 30 days. Are samples collected at the same time in the morning? Then the population is further limited to water quality at that time in the morning over 30 days.
- We have shown that a mean over 30 days does not reflect water quality at any given point in time except by chance. An estimate of water quality at a point-in-time is needed. Our population is, then, the water quality at that point in time, and temporal variation ceases to be a concern.
- Unfortunately, given the state of the art of recreational water assessment, even though we may base a sample design on a population that consists of the water quality indicator at 9:00 a.m. on Friday, the population of concern is the water quality at noon on the following day.
  - Temporal variation is again a concern, but now affects our ability to *predict*.
  - No amount of sampling will reduce this type of uncertainty. This takes us into the realm of modeling in an attempt to reduce the *unexplained* part of temporal variability.

Estimates of variance components for the EMPACT beaches

The various components of spatial and temporal variance, and standard deviations (simply the square root of the variance), as estimated for each of the five study beaches are presented in Table 20. Two kinds of variance are given for each component. One is a "pure" factor variance for each component, that is, the variance that would be seen if we had "perfect" knowledge of the indicator density at each sampling location. This is added to the variance within sampling location (that is, the variance among the replicate samples taken on occasion at each location), to derive the variance that would actually be observed when a single sample is collected at each location and many such samples are taken from different transects, zones, days, etc. The pure and observed variances for a given spatial or temporal factor apply to the situation in which all of the other factors are held fixed; for example, the variance among zones is applicable to samples that are taken along the same transect at the same time.

Statistical tests would reveal significant differences among the different variance components for any individual beach, and, likewise, significant differences among the different beaches for any individual variance component for most comparisons. There is, however, no reason to suspect that they should be the same.

Note that the incremental factor variance components, except for replicates variance, are the lowest, or among the lowest, at Imperial Beach and among the highest at Belle Isle. Variance among replicates, the "pure error" component, however, follows the opposite trend, being lowest at Belle Isle and highest at Imperial Beach. This implies that, on a large scale, the distribution of indicator organisms is the most uniform at Imperial Beach, the least so at Belle Isle.

Spatial components of variation

Small-scale (replicate) variance

This variance is a result of the variation among different samples taken at the same time from the same location in the water from a depth of 0.3 M from the surface. There are two components that may contribute to the small-scale variation:

1. Method variance is likely negligible for both methods. Precision (standard deviation) for method 1600 is on the order of +/-0.02 [U.S. EPA, 1997b], implying a $\log_{10}$ standard deviation of about 0.01, affecting only the least significant digit of small-scale standard deviation displayed in Table 20.
2. This leaves true variation among replicates as the only real contributor to small-scale variance. Although we describe replicate samples as being collected from the same location at the same time, they, of course, are not. They are as close to one another, temporally and spatially, as we can get.

Table 20. Standard deviation and variance component estimates for the EMPACT beaches.

Location	Source of variation					
	Replicate Samples[1]	Sampling Depth[2]	Depth Zones	Location Within Zone[3]	Hourly 9:00am - 6:00pm)	Among Days
Pure factor variances[4]						
West Beach	0.055	- [5]	0.078	0.031	0.091	0.234
Belle Isle	0.037	0.021	1.480	0.103	0.182	0.168
Wollaston Beach	0.096	0.052	0.203	0.036	0.190	0.339
Imperial Beach	0.138	0.010	0.017	0.032	0.118	0.114
Miami Beach Pk.	0.039	0.047	0.911	0.169	0.329	0.432
Total variances[6]						
West Beach	0.055	0.045	0.133	0.086	0.147	0.289
Belle Isle	0.037	0.058	1.517	0.140	0.219	0.205
Wollaston Beach	0.096	0.147	0.299	0.132	0.285	0.434
Imperial Beach	0.138	0.147	0.155	0.169	0.256	0.252
Miami Beach Pk.	0.039	0.086	0.950	0.208	0.368	0.471
Standard deviations						
West Beach	0.235	0.212	0.365	0.294	0.383	0.538
Belle Isle	0.193	0.241	1.232	0.375	0.468	0.453
Wollaston Beach	0.309	0.384	0.547	0.363	0.534	0.659
Imperial Beach	0.371	0.384	0.394	0.412	0.506	0.502
Miami Beach Pk.	0.198	0.293	0.975	0.456	0.607	0.686

[1] Samples collected at the same point in the water 0.3 M from the surface. [2] Samples collected at the same position in the water, but at different depths below the surface. [3] Samples collected from water of the same depth & from the same depth below the surface, but at different points in water. [4] Variance attributable solely to its respective factor. [5] Estimate of this variance component was negative, so is assumed to be zero. [6] Total variance = pure factor variance + replicate sample variance (except in the case of replicate samples themselves, where it is identical to pure factor variance).

Replicate variance represents the least variability that we may expect among any set of water samples, even if the water comprising the bathing area were "well-mixed" prior to our collection of the samples. This small-scale variability is a component of every other variance shown in Table 20.

Variance among sampling depths

In addition to samples that were collected at a depth of 0.3 M from the surface, on occasion samples also were collected at 0.075 M from the bottom in both knee- and chest-deep water, and at 0.75 M from surface (mid-depth) in chest-deep water. Variance among sampling depths represents the variation among these samples (including the 0.3 M sample). In most cases, we note that the contribution to replicate variance is relatively small, being at most on the same order as the magnitude of the replicate variance itself.

In one case (West Beach), the estimate of pure variance due to different sample depths is negative, leading to a total variance for this factor that is smaller than the corresponding replicate variance. That this should be true is not believable, but that this might occur for one of these estimates is not particularly surprising. This simply implies that the variance due to sampling depth is sufficiently small that such a result is likely to occur.

We had previously shown that there were no consistent differences among the various sampling depths. Additionally, the human health studies that are the basis for recreational water criteria (Dufour, 1984; Cabelli, 1983) are based on water samples taken at a depth of 0.3 M from the surface. For these reasons, there is little reason for collecting samples at any other than the standard depth of 0.3 M. Thus, the variation in sampling depth is not a relevant consideration in recreational water sampling.

Variance among zones

We have divided the remaining spatial variation into two components, variation among zones (outward from the beach) and variance among locations within depth zone (parallel to the beachfront) for reasons mentioned earlier, mainly, that variation among zones was shown to be substantial and, to a large degree, predictable. The consistent differences in indicator levels among zones can be exploited by the technique of stratification – sampling within each zone and then combining the individual zone means using a weighting scheme to account for their relative importance.

Variance among depth zones as given in Table 20, however, represents variation that would be encountered if one were to randomly select sampling locations (*i.e.*, ankle-, knee-, chest-deep and locations) along a fixed transect. Variance among zones is seen to have the largest component of spatial variation at all but Imperial Beach, a fact that is particularly evident when unique, pure factor variances are examined. At Imperial Beach, all components of spatial variation seem to result in virtually the same total variance, whether we look down through the water column, across the beach, or outwards from the beach.

This variance, though, is avoidable because of the consistent ordering of indicator levels, decreasing as one goes out from the shoreline from ankle-deep to chest-deep water. Besides the technique of sample stratification, it seems reasonable that water of a particular depth be a separate population of study in itself, given that, as previously mentioned, epidemiological models for swimmers' illnesses have relied on indicator values from waist- to chest-deep water. Additionally, contamination in shallow water is likely to affect different bather sub-populations.

Variance among locations within zone (among transects)

This leaves the variation within depth zone, or "transect" variance, to contend with. Variance among locations within the same depth zone in Table 20 represents an inescapable source of sampling variation. This source of variation must be considered in any sampling plan because rarely, if ever, would we be interested in the indicator level at a single point in the water. Even if we limit our sampling to one such location, the question remains, "How 'representative' is this of the water quality in general?"

We can use between locations variance to estimate how much we might expect a single sample taken in, say, knee-deep water to differ from the true mean for knee-deep water along that stretch of beach. Using a normal distribution model for log indicator density the half-width of the 95% confidence interval is calculated as two times the standard deviation among locations by 2. At West Beach, for

example, this gives a 95% confidence interval of 2 x 0.294=0.588 logs. The interpretation is that 95% of the time the value from a single sample will be within 0.588 logs of the actual geometric mean density in knee-deep water. This is equivalent to a factor of nearly 4 ($10^{0.588}$=3.9), so that our 95% confidence interval implies that the density of a single sample may range from about ¼ of the true mean to 4 times the true mean value.

It may be more interesting to consider what kind of difference might be expected between single samples, such as two samples taken in knee-deep water but at somewhat different locations. At West Beach, we would expect the two samples to differ from one another by more than 0.28 logs 50% of the time, based on the normal distribution. This means that if one relies on a single sample, half of the time a different sample would have given a result more than about twice ($10^{0.28}$=1.9) or less than half (1/1.9) the value.

Replication, sampling from several different locations, is the means of reducing the impact of this variance on the uncertainty of the sample estimate. If V represents the variance of single samples among locations, or transects, the variance of the mean of n samples taken from different locations within the same zone is V/n.

A valid concern is whether variance among sampling locations itself varies from one zone to another. In fact, this question may be asked of all sources of variation. On theoretical grounds, if these data are log-normally distributed, and this does appear to be at least a reasonable model, this implies proportionality between mean and standard deviation. Their logs (base 10 or any other base), then, will follow a normal distribution with constant variance. In particular, the variance, or, equivalently, standard deviation, within the zone will be the same value for each zone. When separate variances are computed for ankle, knee, and chest depth zones, as they were for Table 21, there is no evidence to refute that this is the case. This is found to be true for other factors of variation as well, even in the case of day-to-day variation where one may suppose that whatever environmental factors may affect this sort of variation, the magnitude of their effects close to shore may differ from that in deeper water.

Table 21. Standard deviation among locations (transects) within each separate depth zone.

Location	Depth zone		
	Ankle	**Knee**	**Chest**
West Beach	0.329	0.275	0.270
Belle Isle	0.389	0.361	0.369
Wollaston Beach	0.360	0.336	0.393
Imperial Beach	0.410	0.423	0.402
Miami Beach Pk.	0.458	0.428	0.470

Temporal components of variation

Hourly variation

At least for the period from 9:00 a.m. to 6:00 p.m., we can determine the variance that takes place from hour to hour within a day. Variation on this time scale might be of interest if one were to collect samples at random times during the day in order to estimate some daily average.

In our case, however, it is interesting to compare the variance hour-by-hour with that on the larger day-to-day time scale. Despite this difference in time scale, the hourly variation is equal to the daily variation, or approximately so, at three of these beaches. At two of these beaches, Miami and Belle Isle, we had previously noted a pronounced tendency for indicator levels to drop between the early morning and mid-afternoon visits (specifically, between the normal visit times of 9:00 a.m. and 2:00 p.m.). This may explain some of the apparent "excess variance" among hourly results at these two beaches, just as the drop off in indicator levels from shallow to deep water accounts for much of the inter-zone variance we saw earlier. While we could not discern a substantial decline between morning and afternoon at Imperial Beach, its hourly graph in Figure 11 does show a generally U-shaped curve, which may account for its relatively high hourly variance as well. We do note, also, that all of the variance components there, spatial and temporal, appear to be much more like one another than at any of the other beaches studied.

Variance among days

This source of variation refers to single samples that are collected at the same time of day at the same relative location in the water (*e.g.*, along transect 1 in knee-deep water) at a depth of 0.3 M from the surface. This, in fact, is the variance referred to in the EPA criteria document [Dufour and Ballantine, 1986] in relation to single sample limits in recommending "a one-sided confidence limit ... based on a *site specific standard deviation*" (emphasis added). In the absence of such information, recommended log standard deviations are given – 0.4 for *E. coli* in freshwater and 0.7 for enterococci in marine water – which appear to be entirely reasonable based on the standard deviation results among days given in Table 20.

The total variance among days would also be used in assessing the precision of the log mean of 5 samples over a thirty-day period, as recommended in the criteria document. Letting V represent the variance among days, the corresponding variance of the log mean is V/5, and we could construct confidence intervals or upper bounds for the true mean. Here, the "true mean" refers to the actual log mean value at the same (relative) location in the water at the same time of day over the entire 30-day period.

Variance of the change over a 24-hour period

Current technology for monitoring recreational water quality results in a one-day lag between the time that the samples are collected and the time that results from these samples are known. Therefore, another source of temporal variability that we must be concerned about is that of the change in indicator levels that occur over approximately a 24-hour period. This variance component reflects how good a *prediction* of today's water quality based on yesterday's samples will be.

That a prediction is being made implies that a model exists for making that prediction. The simplest, and probably most often used, such model is that of *status quo* - that there will be no change over the 24 hours. Lacking any other information, this would be as good a model as any. In this case, the variance of the error in our predictions is that of the raw change in log mean density, tomorrow's log mean density minus today's.

Earlier, in examining environmental and bather effects on indicator levels, we developed a linear regression for each EMPACT study beach that predicted what the log mean density will be 24 hours later based on various weather, tide, and bather density information along with the current day's sample value (see Table 16). If these models are used in place of the *status quo*, the error variance for our predictions becomes the difference between the true log mean density and our predicted value.

These two variances are given in Table 22 for each of the EMPACT Study beaches. With remarkable consistency, the variance of predictions based on the regression models of Table 16 is about one-half that of the raw change (the "*status quo*" model). Variances for predictions based on the regression models are likely understated in Table 22, simply because of the amount of "data mining" that went into the development of these models, leaving plenty of opportunity for over-fitting the regressions. Nevertheless, the benefit of basing the assessment on information in addition to yesterday's samples is apparent.

Table 22. **Variance and standard deviation of 24-hour change in $\log_{10}$ indicator density and of 24-hour change in $\log_{10}$ indicator density when modeled as in Table 16.**

Location	Raw change in mean $\log_{10}$ density		Modeled change in mean $\log_{10}$ density	
	Variance	Standard Deviation	Variance	Standard Deviation
West Beach	0.397	0.631	0.201	0.448
Belle Isle	0.205	0.452	0.100	0.316
Wollaston Beach	0.423	0.651	0.213	0.462
Imperial Beach	0.195	0.442	0.108	0.329
Miami Beach Park	0.540	0.735	0.225	0.474

Designing a Beach Monitoring Program

The preceding sections have led us through an examination of the spatial, temporal, and environmental factors that influence recreational water quality to the description of potential sources of spatial and temporal variance and estimation of their respective magnitudes. The concept of a population as it relates to sampling, how it affects which variance components come into play, and the importance of the sampling distribution have been emphasized. These, together with what we know about human health effects, are all factors to be considered in developing an optimal sampling strategy.

A standard for systematically planning for the collection of environmental data have been established by the U.S. Environmental Protection Agency by its Data Quality Objectives (DQO) Process [U.S. EPA 2000b]. Implications from the EMPACT Beaches study with respect to monitoring recreational waters will be given in the context of this standard.

The DQO approach

The DQO Process [U.S. EPA 2000b] addresses the requirements for systematic planning to "ensure that data collected for the characterization of environmental processes and conditions are of the appropriate type and quality for their intended use …" enumerating seven steps to be taken in planning for data collection projects that support environmental decision-making. These seven steps are shown in Figure 15. The reader is encouraged to refer to the U.S. EPA DQO Guidance document, which presents in a readable fashion a formal framework for planning any environmental sampling program, and can be downloaded from the Internet at *http://www.epa.gov/quality1/qs-docs/g4-final.pdf*.

Figure 15. The Data Quality Objectives Process (from "Guidance for the Data Quality Objectives Process", U.S.EPA [2000b])

Step 1. State the Problem
Define the problem; identify the planning team; examine budget, schedule.

Step 2. Identify the Decision
State decision; identify study question; define alternative actions.

Step 3. Identify the Inputs to the Decision
Identify information needed for the decision (information sources, basis for Action Level, sampling/analysis method).

Step 4. Define the Boundaries of the Study
Specify sample characteristics; define spatial/temporal limits and units of decision-making.

Step 5. Develop a Decision Rule
Define statistical parameters (mean, median); specify Action Level; develop logic for action

Step 6. Specify Tolerable Limits on Decision Errors
Set acceptable limits for decision errors relative to consequences (health effects, costs).

Step 7. Optimize the Design for Obtaining Data
Select resource-effective sampling and analysis plan that meets the performance criteria.

Results from the EMPACT Beaches Study will be related primarily to steps 4 and 7 of the DQO Process. Step 6 will be discussed in some detail in order to present concepts that are useful in understanding the rationale behind sample size determination. Some brief comments on the remaining steps are:

- State the problem. Fecal contamination at a recreational beach poses a threat to human health and because of this, water quality must be monitored. Questions to be answered at a local level are: Who is to be involved in the monitoring program? What is our budget? What is the timeframe for developing the program?
- Identify the decision. At a minimum, results from monitoring will lead to the decision to close the beach or not, depending on whether contamination is found to be at an unacceptable level. Additional actions may be identified, for example, public posting of microbial counts or posting a warning when contamination levels are in a questionable range.
- Identify the inputs to the decision. The very purpose for conducting the EMPACT Beaches Study itself was to provide this input. Other studies on sampling recreational beaches have been published [Brenniman *et al.*, 1981; Cheung *et al.*, 1990; Churchland and Kan, 1982; Noble *et al.*, 2000; Pande *et al.*, 1983; Van Ess and Harding, 1999]. The World Health Organization (WHO) has published a series of guidelines for beach monitoring [1975, 1977, 1998]. A good starting point for Internet resources is the U.S. EPA's Beach Watch website at http://www.epa.gov/OST/beaches/. Other resources may be available locally.
- Develop a decision rule. Examples of decision rules are given in current EPA guidelines [Dufour and Ballentine, 1986; U.S. EPA, 1986] and are based on limits of acceptable illness rates among swimmers of 8 per 1000 in freshwater and 19 per 1000 in marine waters. Beach monitoring programs using federal grants authorized by the BEACH Act must use the water quality standards (decision rules) adopted by the state [U.S. EPA, 2004].

Defining the boundaries of the study: the population

The concept of a population in its statistical sense was discussed earlier in the section on sources of variation. We defined the term "population" as "the aggregate from which samples are collected and to which inferences and, ultimately, decisions are made." The phrase "define the boundaries of the study" tells us to identify this population in detail. What should the target population, the body of water, be in terms of sampling depth, distance from shore, distance along the beachfront, and time?

Sampling depth

We have used 0.3 meters from the surface as the standard depth at which to collect water samples in water that is of swimming depth [APHA, 1998]. Sampling at this depth, at which a swimmer is likely to be exposed to pollution, is recommended in the WHO approach to monitoring [1998] and was the depth used for microbial sampling in the U.S. EPA epidemiological studies [Cabelli *et al.*, 1982; Dufour, 1984], upon which United States recreational water quality standards have been based.

While this study found little in the way of systematic differences among samples collected at different depths, this does not mean that there may not be meaningful differences at other beaches. Regardless, sampling from different depths below the surface would seem to introduce an additional component of variance, unnecessary in light of the lack of interpretability in regard to human health for sample results that come from depths other than 0.3 meters from the surface.

55

Distance from shore (Depth zone)

This study has demonstrated that the depth zone from which samples are to be collected will likely be a critical decision to be made in a beach sampling plan. At all of the study beaches, indicator levels were seen to fall off substantially as one goes from ankle-deep to chest-deep water, to such a degree that the likelihood that a beach met current U.S. EPA guidelines (1986) at any point in time was highly dependent upon the depth zone from which samples were taken.

Depth zones in a bathing area can be associated with particular usage patterns and segments of the human population that are likely to be affected by the water quality in that zone. Infants and toddlers, given their activities at the beach, are most likely to ingest water in the "ankle zone" as defined by this study. For adults, exposure to contamination from immersing one's head in water seems most likely to occur in waist-deep or deeper water, given that substantial swimming activity probably does not occur until one reaches this depth. It is advisable that a sampling plan account for these differences in the affected population, and be capable of yielding separate estimates of indicator density for shallow water and water of swimming depth.

Epidemiological studies have concentrated on adults, and have measured indicator densities in water of swimming depth. U.S. EPA guidelines for beach water quality (1986) are based on these studies, and, thus, on fecal contamination levels in deeper waters. Unfortunately, epidemiological data that would enable quantitative evaluations of the risks inherent to infants and toddlers with respect to exposure to sewage-contaminated water are not available.
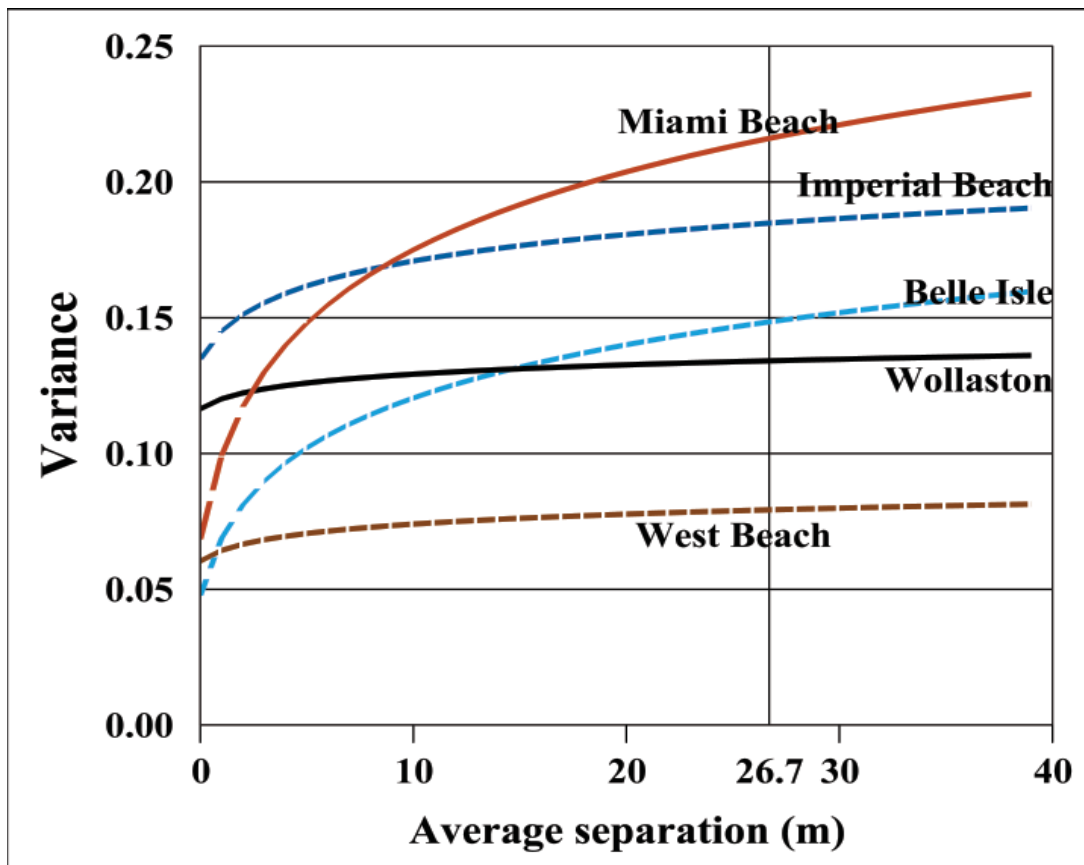
Distance along the beachfront

We define the beach itself in setting the boundaries for this final spatial dimension, the shoreline to be encompassed by a monitoring program. This expanse comprises the beach with regard to which decisions will be made. One will need to determine areas that may be exposed to pollution, as determined by the locations of point sources on the shore or in the water. Conversely, other areas may be shielded from contamination. Usage is another consideration. Where do bathers tend to congregate and when? What areas are used primarily by surfers or by swimmers? Answers to these questions can help to define unique recreational areas of the beach. In some cases, political boundaries or the authority responsible for managing the beach may come into play in determining what constitutes a beach. Public perception of the beach area will need to be considered, regardless of the officially recognized boundaries of the beach. Those responsible for communicating the quality of recreational waters to the public, along with reporting the results of the day's sample, must unambiguously state the area of the beach to which the results apply, that area decided upon in this phase of the design process.

The World Health Organization, in its recommendations for beach monitoring [1998], addresses the issue of defining a beachfront for sampling purposes through its "primary microbiological categorization protocol". This protocol entails a sanitary survey to determine sources of contamination and a preliminary sampling phase to determine if "significant variation" along the beach is indicated. The term "significant variation", as used in the WHO report, connotes systematic variation in which a section of the beach is consistently more, or less, contaminated than other sections of the same beach. We have found "significant variation" to exist at each of the five study beaches in the statistical meaning of this term, mainly that there is a between-locations variance component, over and above the variance we see from replicate sampling at the same location in the water. This variation, however,

is not systematic, as seen from the lack of any apparent consistent differences among transects (with possible exceptions at Belle Isle and Miami Beach Park).

The length of shoreline that one decides upon to define the monitoring area may itself affect the magnitude of the variance encountered in the process of sampling. Sampling variances shown in Table 20 apply to the forty-meter, centrally located expanses of beach that comprised the sampling area at each of the study beaches. If sampling had encompassed larger sections of the beaches, these variances would likely be higher. This is so because values obtained from two points separated by a relatively large distance are likely to be less correlated than values obtained from two points that are closer together. The effect can be illustrated by considering only those points along transects 1 and 3, the two extreme transects, in our variance calculation, yielding a variance between sampling locations that are separated by 40 meters. Variances computed from all the data, as they are in Table 20, reflect an average separation of 26.7 meters (40 m separation for one pair of points, and 20 m for each of the two other pairs). In addition, "pure error" variance is available, reflecting zero separation between sampling points (replicate samples from the same location). Respective variances calculated for average separations of 0, 26.7, and 40 meters are plotted in Figure 16, and joined by a smooth curve to show hypothetical relationships

Figure 16.   **Variance of log10 indicator density per 100 mL vs. average separation among sampling locations.**

between average distance separating the sampling locations and the corresponding sampling variance. Stratifying the beach along shoreline segments is a potential counter to this trend, as discussed later, in the section "Optimize the design for obtaining data."

Time

A geometric mean calculated from five values over a four-week period did not accurately represent actual conditions on any given day at any of our five study beaches. Figure 8 in the section on temporal variation clearly illustrates this fact. A point-in-time estimate of indicator levels is needed.

The water quality measurement from which decisions will be made should be the contaminant level in the swimming area at a time when beachgoers will be exposed, and, unfortunately, current membrane filtration and culture methods do not allow us to obtain the sample results any earlier than about 24 hours after the samples have been collected. A scenario was introduced earlier, in which the beach manager sampled on a Tuesday in order to make decisions for the next weekend. Results from a Tuesday sample were shown to have little bearing on conditions that exist the following Saturday. Correlations between indicator levels on one day and those four days later (see Table 10) were seen to be negligible, while serial correlations for a one-day lag were, if not impressive, at least positive.

Collecting samples the day before one must make a decision with regard to public safety at a beach may not be convenient, but it appears that data obtained earlier than this may have little or no relevance to such a decision.

The time of day when sampling is performed is also likely to be governed by practical considerations - one would prefer to collect samples in the morning in order to assure ample time for transporting the samples to the lab and for the labs to process the samples and start incubation. In addition, the earlier the sampling time, the more timely the results from the analysis. This study has demonstrated that, for the five beaches studied, contamination levels in the afternoon were, if anything, likely to be lower than in the morning. Thus, morning sampling would tend to err on the safe side.

To summarize, this study indicates appropriate spatial and temporal boundaries for beach sampling to be:
- 0.3 M from the surface
- in water of swimming depth
- along a pre-determined beachfront
- in the morning
- of the day before

Monitoring in shallow (ankle-deep) water may be performed as an adjunct to, and viewed separately from, monitoring in deeper waters, although results with respect to children's health effects may be difficult to interpret without further health studies

---

Specifying tolerable limits on decision errors

The U.S. EPA's DQO guidance document [EPA, 2000b] uses the concept of a power curve to define tolerable limits on decision errors. In our case, a power curve gives the probability of rejecting a beach as unfit for bathing as a function of the true (log) indicator level at that beach. This probability will be the complement of what is commonly known as the "Type II" or "β" (beta) error in those cases in which the true indicator level exceeds a given standard but a sample fails to detect this exceedence,

and equal to the "Type I" or "α" (alpha) error when the true indicator level falls below this standard. We may also view this in terms of acceptance sampling(see box), in which the beach is considered a product that we wish to accept or reject as suitable for use.

For either the power function or the operating characteristics curve, it is necessary and sufficient to specify two points on that curve, the shape of the curve being given by the normal (usually) probability density function. The DQO guidance document illustrates this by showing a true value equal to the "action level" for the decision variable and another true value equal to an upper limit specified by the user. Figure 18 emphasizes two true density levels that determine the producer and consumer points.

Meaningful producer and consumer points or action levels can be determined based on the relationship between implied risk and indicator level. Health studies conducted by the EPA give these relationships in terms of extra cases of "highly credible gastrointestinal illness" (HCGI) per 1000 swimmers as

$$-11.9 + 9.5 \cdot \log_{10}(E.coli \text{ per } 100 \text{ mL}) \tag{2}$$

in freshwater [Dufour, 1984], and

$$0.1 + 12.3 \cdot \log_{10}(\text{enterococci per } 100 \text{ mL}) \tag{3}$$

in marine water [Cabelli, 1983]. At EPA recommended limits for the 30-day geometric mean, 126 per 100 mL for *E. coli* and 35 per 100 mL for enterococci in marine water, the implied illness rates are about 8 and 19 per thousand swimmers, respectively.

As an example to be used for sample size estimation in the following section, "Optimizing the design for obtaining data", consider the freshwater case and assume that we would desire a low probability of accepting a beach as safe for swimming if the swimming-related illness (HCGI) rate is 10 per 1000 or higher. Somewhat arbitrarily, based on custom more than anything else, we set this probability at 0.05 (5%), and this represents our consumer's point, as in Figure 18. Conversely, should the swimming-related illness rate be 6 per 1000 or less, we consider this to be an acceptable risk, and want a high probability of accepting this beach, *i.e.*, declaring it to be suitable for swimming. We set this probability at 0.95 (again, based on a commonly accepted value for what we mean by a "high probability"), to obtain the producer's point in Figure 18.

Equation 2 can be used to determine the values of the $\log_{10}$ *E. coli* density per 100 mL that correspond to each of these decision points. An incidence of 6 cases of HCGI per 1000 is seen to correspond to about 1.9 logs, and of 10 cases per 100, to 2.3 logs. Whenever the beta risk at the consumer's point is equal to the complement of the alpha risk at the producer's point, the action level will lie exactly halfway between the two corresponding illness rates, or log densities for the indicator organisms. These decision points are centered about an illness rate of 8 per 1000, corresponding to 2.1 log density per 100 mL, the current EPA-recommended action level for a geometric mean over time ($\log_{10} 126 = 2.1$).

This is not to be construed to advocate the retention of current EPA recommendations, which, after all, apply to a geometric mean over 30 days (or to single samples). The EPA recommended action level simply offers a convenience reference point for this presentation of a rationale for determining an appropriate action level based on the acceptance sampling concepts of consumer and producer's points and an operating characteristics curve. Indeed, as far as sample size requirements are concerned,

what is important is the *width* of the interval between the $\log_{10}$ indicator values at the producer and consumer's points (the tolerance interval). In our case, this width is 0.4 logs, or +/- 0.2 logs with respect to the action level (the half-width of the tolerance interval). A different action level may be selected, but as long as our decision points remain +/- 0.2 logs from this value (corresponding to +/- 2 illnesses per 1000 swimmers in fresh water), the operating characteristics of the sampling plan, as exemplified by the curve in Figure 18, do not change.

Figure 17.   Correspondence between health effects criteria and indicator density.
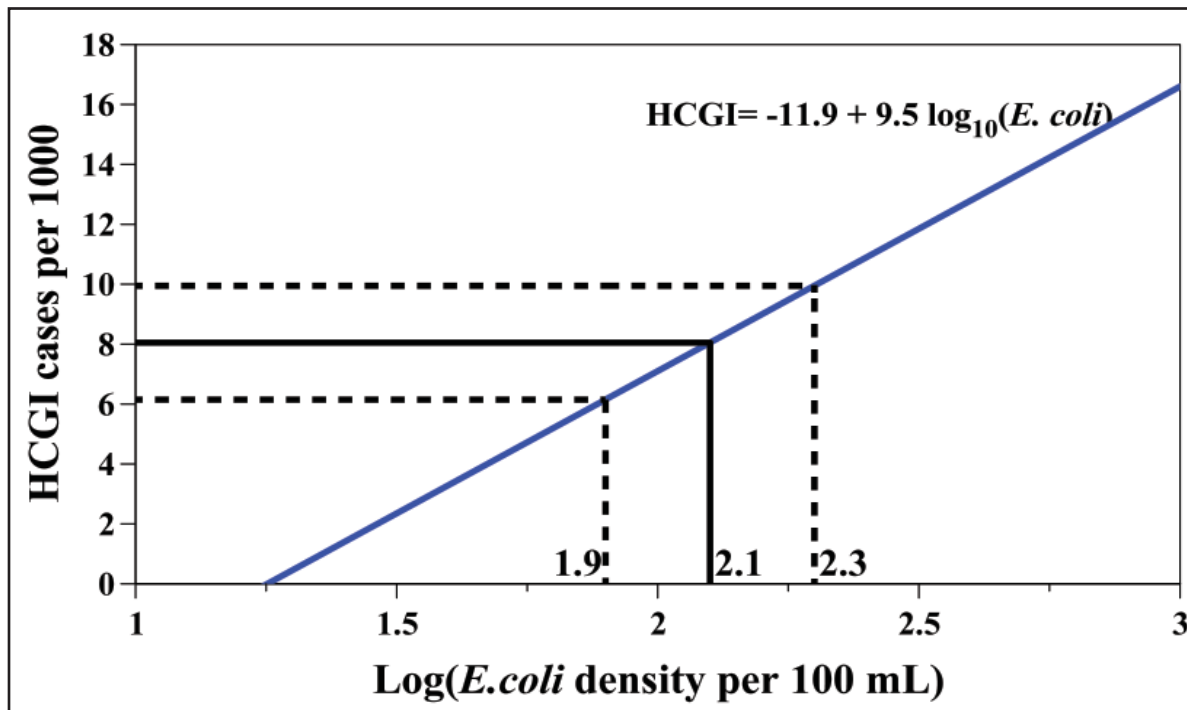
A rationale for setting tolerance intervals

We may choose a tolerance interval to be as narrow as we wish. The narrower this interval is, that is, the closer the producer and consumer points are to each other, the more precise our sampling plan becomes, and the better we are able to "pinpoint" the condition of the beach.

Is there a point at which increased precision, and the associated expense, is wasted? The ultimate purpose of sampling is to determine if a beach is safe for recreational use. Measurements of indicator organisms help to make this determination through what we know about their relationship to human health, as given in Equations 2 and 3. Thus, uncertainty in water quality assessment should be viewed not only in terms of sampling precision, but also in terms of the precision of our knowledge of health effects.

The +/-0.2 log tolerance interval that we have described corresponds to a 90% confidence interval for the estimated mean log density, because of the 5% tail regions at each end of the interval. This spread for log densities was shown to be equivalent to +/-2 HCGI cases per 1000 swimmers. Available data from the freshwater health studies [Dufour, 1984] indicate a 90% confidence interval for the expected health effect of about +/-4 HCGI cases per 1000 swimmers at 126 *E. coli* per 100 mL (this will be

dependent to a degree on the actual indicator density). Thus, the uncertainty due the estimated density is substantially less than the uncertainty in its health effect, at least based on the EPA studies alone. In fact, the square of the total uncertainty will be equal to the sum of the squares of the individual components, leading to the conclusion that the total uncertainty will increase by only 12%, from 4 to 4.5, when the sampling error alone is equivalent to +/-2 illnesses. For marine waters [Cabelli, 1983], the 90% confidence interval for the expected illness rate is about +/- 3 HCGI cases per 1000 at 35 enterococci per 100 mL. In this case, a sampling error equivalent to +/-2 illnesses per 1000 contributes about 20% over the uncertainty in the health effect itself.

Another component of uncertainty in recreational water quality assessment unfortunately comes into play when we consider the 24 hour lag in information from a membrane filtration assay. Using ancillary data collected at the study beaches, we were able to achieve a meaningful reduction in the standard deviation of the change in indicator density (Table 22). However, the residual standard deviation, which represents the uncertainty in our ability to project today's values based on what they were yesterday, is still on the order 0.4 logs. This is equivalent to +/- 6 illnesses per 1000 swimmers for a 90% confidence interval, larger than even the uncertainty in the health effect itself, and underscores the need for refined predictive models and rapid methods for assessing water quality.

Optimizing the design for obtaining data

We have described appropriate spatial and temporal boundaries to encompass sampling from within a single depth zone (say knee- to chest-deep water), at a point in time (the morning before), from the same depth below the surface (0.3 M). The only difference among samples, then, is the locations (transects) along the beachfront from which they have been collected. The appropriate sampling variance for such a design is given by "total variance between locations" in Table 20 as experienced in each of the individual EMPACT study beaches.

One possible specification for tolerance requirements of a sampling was given in the previous section, mainly, that the sampling plan should be capable of correctly classifying the beach 95% of the time when the true mean indicator density at the beach is 0.2 logs removed in either direction from our action level. This was derived from a consideration of the implied health effects at these indicator densities specifically for a freshwater beach, where +/-0.2 logs is equivalent to +/-2 illnesses (HCGI) per 1000 swimmers. In the case of marine beaches, a 0.2 log increase in indicator (enterococci) density has a somewhat greater predicted effect on illness due to exposure, being equivalent to an increase of 2.5 illnesses per 1000 swimmers per Equation 3. However, rather than basing freshwater and marine designs on equivalent illness rates, we will use the +/-0.2 log range for each so that each has the same precision with respect to the decision variable itself. In practice, one might choose to consider the same marginal illness rates for each type of beach, or, perhaps, choose a wider interval for marine beaches, given that their expected illness rate is more than twice that of freshwater beaches at existing recommended indicator levels.

Sample sizes in Table 23, the number of samples to be collected in a single visit, are based on the respective sampling variances within depth zone at each of the study beaches using the tolerance requirements of a 95% probability for correctly identifying +/-0.2 log deviations from the action level. In addition, we show sample size requirements for +/-0.3 log deviations (equivalent to a health effect of about 3 illnesses per 1000 swimmers in freshwater and 3.7 illnesses per 1000 swimmers at marine beaches). Note that, because the tolerance interval is expressed in terms of a log scale, the effect is that, in terms of the raw indicator densities, the lower and upper limits of this interval will

be a constant factor of the action level. For instance, 0.2 logs over the action level is about 1.6 times the action level and the log action level -0.2 logs is about 1/1.6, or about 0.6, times the action level. Likewise, +0.3 logs and -0.3 logs from the log of the action level are equivalent to twice and one-half the action level, respectively.

Any given beach, after a routine monitoring program has been in place for some time, will have data approaching the quality we have from this study upon which to base a sampling program. At the outset, one necessarily will have to make assumptions about sampling variances. Results from the EMPACT studies may serve as a guideline to some provisional estimate of the appropriate sampling variances, but these provisional estimates will soon be superseded by experience. Because monitoring will be an ongoing process, once data is obtained this serves as a feedback to further data obtainment. The feedback addresses earlier steps in the DQO process as well, for example, defining the boundaries of the monitoring program, which may change as heretofore-unknown "hot spots" are discovered.

Table 23. Sample size requirements based on estimated overall sampling variance for the geometric mean within a single depth zone in the water.

Location	Sample size required for a 95% tolerance limits of:	
	+/-0.2 logs[1]	+/-0.3 logs[1]
West Beach	6	3
Belle Isle	9	4
Wollaston Beach	9	4
Imperial Beach	11	5
Miami Beach Park	14	6

[1] Specifies a range centered on the action level such that there is a 95% probability of accepting the beach as safe for swimming when the true log mean indicator density per 100 mL is at the lower end of the range and a 95% probability of rejecting the beach when the true log mean indicator density is at the upper end of the range.

Because estimates of spatial variance are critical to evaluating precision and, subsequently, determining sample size requirements, samples should be analyzed individually during the initial phases of a monitoring program so that microbial density can be established for each location in the water from which a sample was collected. This procedure should be followed at least until enough samples have been collected over a number of days to enable a reliable estimate of the underlying variances. After that, one may consider compositing samples prior to analysis (see the box on composite sampling), with the benefit of potential cost savings. Even so, whole samples should be assayed periodically in

order to monitor and update sampling variances.

Sample size requirements, when individual samples are assayed, may be calculated using the same formula as was used to create Table 23. The sample size, n, required is given by

$$n = V \left(\frac{z_{0.05}}{\Delta x} \right)^2. \tag{4}$$

Here, V is the variance within the zone, Δx is the half-width of the desired tolerance interval for $\log_{10}$ indicator density per 100 mL, and $z_{0.05}$ represents the upper 5th percentile of the standard normal distribution (=1.645). Again, the value of V will be a rough estimate at first, perhaps based on these EMPACT beach study results. As more data are collected, a more appropriate value for V will be obtained.

Given the common practice of collecting a single water sample at the beach, we are interested in knowing the sampling properties of a sample size of one. In this case, the only estimate of spatial variance would necessarily come from prior knowledge, since at least two samples are required in order to derive a variance estimate. Corresponding tolerance intervals for a sample size of one are shown for each of the study beaches are shown in Table 24, based on their respective variances within depth zone. The effects of more intensive sampling are illustrated by considering the width of this tolerance interval as the sample size increases. Log indicator density tolerance intervals are translated into equivalent health effect tolerances in the second part of Table 24 based on the relationships between incidence of HCGI among swimmers and indicator organism density given in Equations 2 and 3.

Table 24. Width of tolerance intervals for various sample sizes based on sample variances within depth zone at the study beaches.

Location	Width of tolerance interval in $\log_{10}$ indicator density per 100 mL for sample size =				
	1	**2**	**4**	**8**	**16**
West Beach	+/- 0.45	+/- 0.32	+/- 0.22	+/- 0.16	+/- 0.11
Belle Isle	0.60	0.42	0.30	0.21	0.15
Wollaston Beach	0.60	0.42	0.30	0.21	0.15
Imperial Beach	0.68	0.48	0.34	0.24	0.17
Miami Beach Park	0.74	0.52	0.37	0.26	0.18
Location	Width of tolerance interval in terms of equivalent health effect (cases of HCGI per 1000 swimmers) for sample size =				
	1	**2**	**4**	**8**	**16**
West Beach	+/- 4.3	+/- 3.0	+/- 2.1	+/- 1.5	+/- 1.1
Belle Isle	5.7	4.0	2.9	2.0	1.4
Wollaston Beach	7.4	5.2	3.7	2.6	1.8
Imperial Beach	8.3	5.9	4.2	3.0	2.1
Miami Beach Park	9.1	6.4	4.5	3.2	2.3

Composite sampling, while reducing analysis costs, will necessitate collecting additional samples beyond the requirements of Equation 4 in order to achieve the same precision in terms of a fixed tolerance interval. There also must be consideration of the fact that a composite sample yields an estimate of an *arithmetic* mean rather than a *geometric* mean (the geometric mean being the antilog of the log mean which we have used thus far in describing sampling strategy). See the boxed discussion on composite sampling and its effects on sample size requirements and equivalent action levels.

Stratification along the shoreline

Study areas limited to a beachfront of forty meters resulted in variances among locations within depth zone as shown in Table 20. In practice, the shoreline comprising a beach may be much longer than this, and division of the beach into "shoreline strata" may be indicated in order to maintain variance within each such stratum at levels comparable to those given in Table 20. Stratification along the shoreline may be based on a priori considerations, for example, the locations of point sources of pollution into the beach area or of physical barriers that may serve to isolate an area of the beach from contamination. Stratification also may be decided upon as a result of continued sampling that reveals systematic differences between different sections of beachfront. In any event, strata need not be contiguous - an area affected by drainage may comprise one stratum while the expanses of beach on each side together comprise another single stratum.

Table 21 shows there to be no substantial differences among standard deviations within different depth zones. Because these standard deviations pertain to log densities, this tells us that the coefficient of variation, or "percent standard deviation," is roughly the same for each zone. Assuming that the same holds true for shoreline strata, one would substitute a common (log) variance for V in Equation 4 in calculating the total sample size requirement. The total sample size would then be allocated among the various shoreline strata in proportion to the total length of each respective shoreline. If composite sampling is to be used, only samples collected from the same stratum (*i.e.*, segment of shoreline) should be composited, and the results from different strata combined via a weighted geometric mean (antilog of the weighted mean of log densities), the weights being the proportion of beach represented by each stratum.

Large systematic differences may warrant special consideration. Depending on the magnitude of the difference, one may want to limit sampling to known "problem areas." For example, in the knee-deep zone at Belle Isle, densities between transects 2 and 3 were observed to be over twice that of transects 1. If this holds true for waist-deep water as well, we would limit sampling to the area within transects 2 and 3. Cut this difference by one-half, and we would be inclined to combine all three transects.

Acceptance sampling

EPA's guidance document on the DQO process discusses the decision process, and decision errors in particular, in terms of a "decision performance curve", which describes the probability of deciding that the decision parameter exceeds some allowable level based on a sample given the true value of that parameter. Similar in concept is the *operating characteristics* (OC) *curve*, an example of which is shown in Figure 18.

Figure 18. Operating characteristics curves
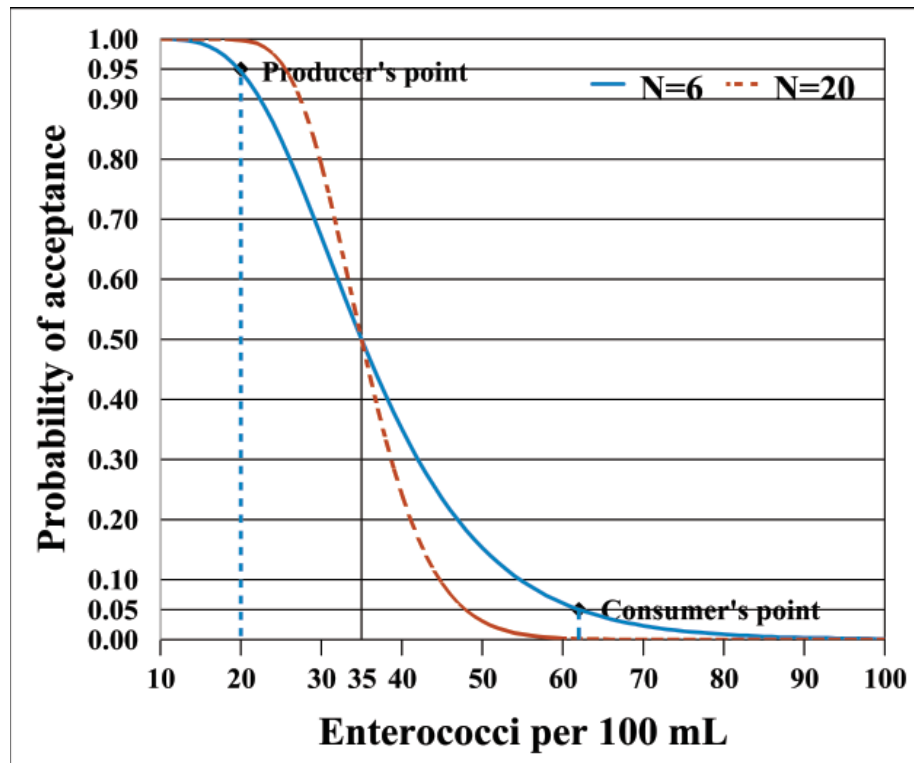
For every possible value that the true geometric mean indicator density at a beach might be (on the horizontal axis), the OC curve indicates the probability of *not* declaring beach to be unacceptable (the vertical axis) based on a sample result. The *producer's point* and the *consumer's point*, as labeled on the OC curve, are concepts of *acceptance sampling*, a statistical quality control technique used in industry to accept or reject a lot (batch) of some product based on a sample. The consumer's point specifies the probability of accepting a "bad" lot, and the producer's point, the probability of accepting a "good" lot. Usually, "good" and "bad" are defined in terms of percent defective. When this percent is truly some specific, acceptably low value, we want the probability of acceptance to be high (commonly, 95%), thus specifying the producer's point. Conversely, for some unacceptably high value for percent defectives, we want the probability of acceptance to be low (*e.g.*, 5%), giving the consumer's point. The *producer's risk*, corresponding to the Type I (α) error, is 100% minus the probability of acceptance at the producer's point, and the *consumer's risk*, the Type II or β error, is the probability of acceptance at the consumer's point.

Composite Sampling, Part I

Composite sampling is a commonly used technique in environmental monitoring [Lancaster and Keller-McNulty, 1998] and may be a cost-effective practice for monitoring recreational water quality. For our purposes, we will define composite sampling as follows.

Composite sampling involves thoroughly mixing together equal volumes from various samples collected from different locations in the water and subsequently assaying a subsample of the resulting mixture. For example, ten 100 mL samples are combined and mixed, and a single 100 mL aliquot of this 1 L mixture is selected for analysis. Two or more composites may be formed from a like number of distinct sets of samples, such as those drawn from knee-deep and those drawn from chest-deep waters and combined to create knee-deep and chest-deep composites.

However, an indicator density obtained from a composite sample differs in a fundamental way from the geometric mean of the individual samples; namely, a composite sample result will almost always be higher. This is purely a mathematical artifact arising from the fact that, for any series of non-negative numbers, the geometric mean is always less than the arithmetic mean (unless all the values are identical). Consider the following hypothetical counts obtained from five 100 mL samples:

$$11, 24, 68, 56, 16$$

The geometric mean of this series is 27.6. If, instead, these samples had been combined and a single 100 mL of the resulting composite assayed, the *expected* count would the arithmetic mean of this series, 35. Of course, because of additional sampling variation introduced by subsampling, the actual count may be greater or less than 35; however, there is a 90% probability that the count will be greater than the geometric mean, 27.6 (we know this from the *Poisson* distribution, which describes the distribution of possible counts from "well-mixed" samples).

The use of the geometric mean is based on well-founded principles, as stated elsewhere in this report. There is, however, a correspondence between geometric and arithmetic means that results if densities among the original samples follow a lognormal distribution, and we have shown that our data is reasonably described by such a distribution. Given lognormal data, a geometric mean estimates the *median* of those data (actually, the geometric mean is biased on the high side, but this bias becomes negligible as the sample size increases). The median of a lognormal distribution is related to its mean by:

$$Median = Mean \times 10^{-1.15 \cdot V} \quad (4)$$

Here, V is the variance of the $\log_{10}$ densities, as in Table 20 for example. Multiplying the count per 100 mL obtained from a composite sample by the factor $10^{-1.15V}$ yields an estimate that is approximately equivalent to that of the geometric mean of the individual samples. Alternatively, one could use the inverse relationship to express a mean in terms of the corresponding median, and inflate the decision level for the indicator density accordingly. In that case, the equivalent decision level for a composite sample mean becomes 35 x $10^{1.15V}$ enterococci or 126 x $10^{1.15V}$ *E. coli* per 100 mL.

Note that the preceding assumes we know the value of V. We also need to know this variance in order to properly evaluate precision of the composite sample estimate, as discussed in Composite Sampling Part II. The appropriate value for V can be known only from sufficient historical results based on assays of individual samples. We recommend that this variance estimate be based on individual analyses of at least 50 samples (from each depth zone when multiple zones are sampled); this precludes composite sampling during an initial start-up phase of a monitoring program. More samples may be required for an adequate estimate of V if the sampling variance is greater than about 0.3 for the $\log_{10}$ densities. Additionally, assays of individual samples should be performed periodically, on the order of every 2-4 weeks, thereafter in order to verify and update this variance estimate.

Composite Sampling, Part II

Having addressed the problem of non-equivalence between geometric means and arithmetic means, we must face another problem inherent in composite sampling. As briefly mentioned in Part I, subsampling from a composite adds another layer of estimation; a single aliquot is used to estimate the indicator density over all samples, which, in turn, are used to estimate the indicator level at the entire beach. The extra step in the estimation process, naturally, adds additional sampling error to the final estimate. This may be illustrated using the example in Part I; the expected value of the composite count is 35, which is the average of the individual counts, but, 95% of the time, will range between 23 and 47.

However, what about collecting more samples than we otherwise would have and then compositing these samples? Say, in the above example, we collect ten samples instead of five and take a 100 mL composite from these ten samples. Certainly, ten samples are superior to five, and this may offset the additional uncertainty of subsampling from a composite.

In order to evaluate sampling errors resulting from compositing samples compared to that of complete analysis, simulation studies were performed under the assumptions that (1) individual 100 mL samples follow a lognormal distribution with known variance of log densities, and (2) samples are composited and well-mixed so that the number of CFU captured in a 100 mL aliquot follows a Poisson distribution. Conclusions from this, in terms of sample size requirements with and without composite sampling, are shown below. Note that the additional sampling required when compositing is utilized increases rapidly when variance exceeds about 0.20 or so – another reason for stratifying the beach front into relatively homogeneous sections.

Table 25. Comparison of sample size requirements for equivalent precision with and without compositing.

Sampling variance[1]	Number of samples to collect for equivalent precision	
	Individually analyzed	Composited
0.10	3	5
0.15	5	8
0.20	6	9
0.25	8	14
0.30	9	17
0.40	12	32
0.50	15	43

[1] Variance of $\log_{10}$(indicator density per 100 mL)

<u>Summary</u>

Several factors have been identified in this study as correlates of microbial indicators in recreational water. These include:

- Spatial factors
  - Zone – Indicator levels decreased the farther one goes from shore into deeper water (not to be confused with the depth from which a sample is taken, which was *not* found to be an important factor in this study)
  - Distance along the shoreline ("transect") – There were only minor systematic effects at two of the five beaches that were monitored for this study.
- Temporal factors
  - Morning to afternoon (9:00 a.m. to 2:00 p.m.) - Indicator levels generally decreased by the afternoon at three of the beaches (2 marine and one freshwater). Afternoon levels at the other freshwater beach tended to be lower on sunny days, but higher on overcast days. Only Imperial Beach in Southern California, where indicator densities were very low at all times, failed to show this effect.
  - Fecal indicator levels were found to vary significantly from day to day. There was a limited statistical relationship observed between levels on the sampling day and the following day at three out of the five beaches studied.
- Environmental & bather factors
  - Rain – Resulted in increased indicator levels at the four beaches where substantial rainfall occurred (there was very little rain at Imperial Beach). The relationship may be complicated; at West Beach, there was no apparent effect unless rainfall in the past days was considered, in contrast to the others, where rainfall in the past 24 hours was sufficient.
  - Wind – Associated with increased indicator levels when blowing onshore at three of the beaches. No effects were observed at Belle Isle, and Imperial Beach experienced too little variation, as its winds were constantly onshore.
  - Cloud cover – Under mostly sunny conditions, both freshwater beaches tended to have lower indicator densities. This effect is in addition to its association with the morning to afternoon change at West Beach, mentioned above, since levels were lower on sunny mornings also. At Miami Beach Park, the estuarine beach, sunshine tended to enhance the decrease that was observed in the afternoon.
  - Tides – Tidal effects on water quality were observed. Absolute water level, rather than the tide stage *per se*, was found to be a predictor of water quality, high water being associated with increased microbial counts at the two East Coast beaches (very much so at Miami Beach Park), and with lower counts at the West Coast beach.
  - Temperature – Only Miami Beach Park yielded a relationship with air or water temperature. The tendency here was for counts to be *lower* with increased water temperature. While this may seem counterintuitive, this relationship does not necessarily imply causation, and may indicate the common influence of some unmeasured factor, such as deep-water currents, on both water temperature and indicator levels.
  - Bather density – Lower bather density was associated with lower indicator density at two of the beaches, but *higher* density at another. However, relationship of bather density to microbial counts may be influenced by conditions that bring bathers to the beach (or keep them away) in the first place.

Spatial and temporal components of variation have been identified and estimated. These are:

- Spatial sources of variation
    - Replicate variance ("Pure error") – A factor at any level of analysis that represents the smallest attainable variation in sampling (standard deviation=0.235).
    - Sampling depth – Adds to sampling variability, but is not a constraining factor, since all samples should be taken from the same depth below the surface.
    - Variance between zones – Would add considerably to sampling variance if not accounted for in the design. This variability can be eliminated by restricting sampling to water of the same depth.
    - Variance within zone (variance between transects) – The key component of variance for a point-in-time estimate. May be able to reduce this variance if the beach can be stratified along its length into "hot" and "clean" regions.
- Temporal sources of variation
    - Variance among days
        - Equivalent to the standard deviation that is referred to in the EPA criteria document [Dufour and Ballentine, 1986] in its recommendations for single-sample limits.
        - Would be the principal source of variability in a sample estimate over time for the same location at the beach each time, *e.g.* a 30-day geometric mean.
        - Not a source of sampling variability for a point-in-time estimate, but must be considered when using such an estimate to project results to future days.
    - Hourly variation – based on hourly sampling that was performed over several days at each beach, hour-to-hour variability was not much less than day-to-day variability.

The following points would be included in a protocol for sampling recreational waters for indicators of fecal contamination:

- Time and location of sampling must be carefully considered.
    - Depth zones from which samples are collected are likely to have great effect on the resulting estimate of indicator density. Sampling in knee- to waist-deep water would seem to offer a reasonable, but still conservative, approach to estimating water quality, particularly given that health effects are based on quality of water at waist-depth or deeper.
    - Sampling at 0.3 M below the surface is justified based on exposure considerations. This study failed to discern differences at lower collection depths.
    - Sampling in the morning will likely be a conservative practice, in addition to perhaps being a convenient time to sample.
    - Sampling should be performed as close as practical to the day on which a decision is to be made regarding beach closure or advisement. Preferably this should be the day before, given current conditions of a one day turnaround for the results.
- A number of samples should be collected from different points in knee- to waist-deep water.
    - Results from the EMPACT Beaches Study can be used as a guideline for the initial determination of sample size requirements. For the beaches in this study, sample sizes of from 3 to 6 would be adequate to allow for 95% certainty of detecting a 0.3 log exceedence from the action level for the geometric mean, equivalent to a health risk of 3 to 3.5 cases of HCGI per 1000 swimmers.
    - As data are gathered, the sampling plan should be refined and ultimately based on a variance estimate that is uniquely associated with the subject beach.

- "Hot spots" within the bathing area, which may be known *a priori* or discovered as a result of sampling, should be considered as separate strata for sampling purposes.
  - These should be sampled independently and weighted appropriately in the final result.
  - If indicator density within the problem area is very different from that elsewhere, sampling should be limited to this area. A proposed rule-of-thumb is a two-fold difference.
  - Lesser differences would warrant a combined, stratified estimate, weighting the "hot spot" in proportion to its extent relative to the rest of the beachfront.
- Composite sampling may be used as a cost-efficient technique, enabling better sample coverage at minimally increased cost.
  - In the initial stages of a monitoring program, composite sampling should *not* be used in order to develop data that is necessary in estimating the appropriate sampling variance for a particular beach.
  - A composite sample estimates an arithmetic mean, which would require adjustment in order to equate this to standards based on a geometric mean.
- If resources are available to collect the data necessary for developing a predictive model for the change in indicator density over a 24 hour period, this will likely result in much-improved assessment of water quality, given the 1-day lag in obtaining results from membrane filtration assays.

References

American Public Health Association, American Water Works Association, Water   Environment Federation. 1998. Standard methods for the examination of water and   wastewater, 20th edition. American Public Health Association, Washington, DC.

Bartram, J. 1999. Health-based monitoring of recreational waters: The feasibility of a new approach (the 'Annapolis Protocol'), WHO/SDE/WSH/99.1, World Health Organization,Geneva.

Bordner, R., Winter, J. and Scarpino, P. 1978. Microbiological methods for monitoring the environment: water and wastes. EPA-600/8-78/017, Environmental Monitoring and Support Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, Ohio.

Brenniman, G.R., Rosenberg, S.H. and Northdrop, R.L. 1981. "Microbiological sampling variables and recreational water quality standards," Am. J. Public Health, 71(3):283-289.

Cabelli, V.J. 1983. Health effects criteria for marine recreational waters. EPA-600/1-80-031, U.S. Environmental Protection Agency, Research Triangle Park, NC.

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

Cabelli, V., Levin, M., Dufour, A. and McCabe, L. 1974. "The development of criteria for recreational waters," Proc. International Symposium on Discharge of Sewage from Sea Outfalls, Gameson, A. (editor), Pergamon Press, London/New York, p. 63.

Cheung, W.H.S., Chang, K.C.K., and Hung, R.P.S. 1991. "Variations in microbial indicator densities in beach waters and health-related assessment of bathing water quality," Epidemiol. Infect. 106(2):329-44

Churchland, L.M., Kan, G. and Ages, A. 1982. "Variation in fecal pollution indicators through tidal cycles in the Fraser River estuary," Can J Microbiol. 28(2):239-47

Cochran, W.G. 1977. Sampling Techniques, 3rd Ed.. Wiley, NY.

Code of Federal Regulations. 2000. Protection of environment. Code of Federal Regulations, Title 40, Part 136.3. pp. 27-29.

Dufour, A.P. 1984. Health effects criteria for fresh recreational waters, EPA-600/1-84-004, Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, OH.

Dufour, A.P. and Ballentine, P. 1986. Ambient water quality criteria for bacteria – 1986 (Bacteriological ambient water quality criteria for marine and fresh recreational waters). EPA A440/5-84-002, Washington, DC.

Fatal, B., Vasl, R., Katzenelson, E. and Shuval, H. 1983. "Survival of indicator bacteria organisms and enteric viruses," Water Research,17:379.

Fleisher, J. M. and McFadden, R.T. 1980. "Obtaining precise estimates in coliform enumeration," Water Research, 14:477-483.

Fujioka, R., and Narikawa, O. 1982. "Effect of Sunlight on Enumeration of Indicator Bacteria Under Field Conditions," Appl. Environ. Microbiol., 44:395.

Gannon, J.J. and Busse, M.K. 1989. "*E. coli* and enterococci levels in urban stormwater, river water and chlorinated treatment plant effluent," Water Research, 23(9):1167-1176.

Lancaster, V.A. and Keller-McNulty, S. 1998. "A review of composite sampling methods," J. Am. Statistical Assoc., 93(443): 1216-1230.

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

Mujeriego, R., Bravo, J. M., Pinas, M. and Sanchez Murias, B. 1980. "Statistical variation of microbiological quality of coastal waters: regulatory implications," V[es] Journées Étud. Pollutions, 1980:49-60.

National Technical Advisory Committee. 1968. Water Quality Criteria. Report on the National Technical Advisory Committee, Washington DC, Federal Water Pollution Control Administration, p. 7.

Noble, R.T., Leecaster, M.K., McGee, C.D., Moore, D.F., Orozco-Borbon, V., Schiff, K., Vainik, P.M. and Weisberg, S.B. 2000. Southern California Bight Region Monitoring Program: Storm Event Shoreline Microbiology. Southern California Coastal Research Project., Westminster, CA.

Olivieri, V., Kruse, C., Kazuyoshi, K. and Smith, J. 1977. Microorganisms in urban stormwater. EPA-600/2-77-087, U.S. Environmental Protection Agency, Cincinnati, OH.

Pande, J., Das, S. M. and Misra, D. S. 1983. "Studies on bacteriological quality of Naintal Lake water in different seasons," Water, Air, and Soil Pollution, 19:149-154.

Public Health Laboratory Service Water Surveillance Group. 1995. "Preliminary study of microbiological parameters in eight inland recreational waters," Letters in Appl. Microbiol., 21:267-271.

Sekla, L., Williamson, D., Greensmith, C., Balacko, G., Brown, D. and Stackiw, W. 1987. "Bacteriological characteristics of 15 freshwater beaches in Manitoba," Canadian J. Public Health, 78(3):181-184

Serrano, E., Moreno, B., Solaun, M., Aurrekoetxea, J.J. and Ibarluzea, J. 1998. "The influence of environmental factors on microbiological indicators of coastal water pollution," Water Sci. Tech., 38(12):195-199.

Seyfried, P.L. 1973. "Sampling bacteria in Lake Ontario and the Toronto Harbour," Proc. 16th Conference on Great Lakes Research, Internat. Assoc. Great Lakes Res., 163-182.

Sieracki, N. 1980. "The effects of short exposures of natural sunlight on the decay rates of enteric bacteria and a colipahge in a simulated sewage outfall microcosm," Master of Science Dissertation, Univ. of Rhode Island.

Smith, B., and Dufour, A. 1993 "Effects of swimmers on the microbiological quality of recreational waters: a simulation study," Abst. Annual Meeting of ASM, p. 406.

Stevenson, A.H. 1953. "Studies of bathing water quality and health," Am. J. Public Health, 43:529.

U.S. Environmental Protection Agency. 1986. Bacteriological ambient water quality criteria availability. Federal Register, 51(45): 8012.

U.S. Environmental Protection Agency. 1997a. Manual for the certification of laboratories analyzing drinking water: criteria and procedures, quality assurance, 4th edition. EPA-815-B-97-001, Office of Ground Water and Drinking Water, U.S. Environmental Protection Agency, Cincinnati, OH.

U.S. Environmental Protection Agency. 1997b. Method 1600: membrane filter test methods for enterococci in water. EPA-821/R-97/004a, Office of Water, Washington, DC.

U.S. Environmental Protection Agency. 2000a. Improved enumeration methods for the recreational water quality indicators: enterococci and *Escherichia coli*. EPA/821/R-97/004, Office of Science and Technology, Washington, DC.

U.S. Environmental Protection Agency. 2000b. Guidance for the data quality objectives process, EPA/600/R-96/055, Office of Environmental Information, U.S. Environmental Protection Agency, Washington, DC.

U.S. Environmental Protection Agency. 2002. National beach guidance and required performance criteria for grants, EPA-823-B-02-004, Office of Water, U.S. Environmental Protection Agency, Washington, DC.

Van Ess, E. and Harding, A. K. 1999. "A bacteriological study of swimming areas in Western Oregon," Environ. Health, 14-19.

World Health Organization. 1975. Guide and criteria for recreational quality of beaches and coastal waters. Bilthoven.

World Health Organization. 1977. Health criteria and epidemiological studies related to coastal water pollution,. Athens.

World Health Organization. 1998. Guidelines for safe recreational-waters environments: coastal and fresh waters. Consultation draft. World Health Organization, Geneva.

**United States
Environmental Protection
Agency**

Office of Water and Office of Research and Development
National Exposure Research Laboratory
Cincinnati, OH 45268

Official Business
Penalty for Private Use
$300

EPA/600/R-04/023
August 2005