DRAFT BACTERIOLOGICAL SOURCE METHOD

Prepared by:

Shawn Miller, Dustin Shull,
Pennsylvania Department of Environmental Protection
Office of Water Programs Bureau of Clean Water
11th Floor: Rachel Carson State Office Building
Harrisburg, PA 17105

And

Pamela Higgins
Pennsylvania Department of Environmental Protection
Bureau of Laboratories
2575 Interstate Drive
Harrisburg, PA 17110

2023

INTRODUCTION

Nationally, more than 30 years after the Clean Water Act was enacted, many surface waters continue to be classified as failing to meet their designated uses due to high levels of fecal bacteria (USEPA 2000b). As a consequence, protection from fecal contamination is one of the most important and difficult challenges facing environmental organizations and agencies to safeguard water used for recreation (primary and secondary contact), public water supplies, and propagation of fish and shellfish (USEPA 2005). In Pennsylvania, the number of impaired stream miles for recreational use is increasing as Pennsylvania's Department of Environmental Protection (DEP) works toward assessing 100% of Pennsylvania's surface waters. Fecal contamination can be attributed to point source discharges such as combined sewer overflows (CSO), concentrated animal feeding operation (CAFO) discharges, and even wastewater treatment plant effluents. Fecal contamination can also be attributed to sources such as urban and rural stormwater runoff, manure application, unregulated discharges, agriculture, livestock, domestic pets, and wildlife.

DEP employs the bacteriological data collection protocol and assessment method to make assessment decisions (Miller 2023, Miller and Whiteash 2021). When bacteriological results exceed the *Escherichia coli* (*E. coli*) or fecal coliform criteria found in 25 Pa. Code § 93.7, the Water Contact Sports (WC) Use is impaired. These exceedances of criteria are reflected in the Integrated Report as Recreation impairments. DEP conducts further analyses to determine the source(s) of the impairment in Pennsylvania surface waters. Bacteriological source decision may utilize *Chapter 9.1, General Source and Cause Method* (Shull and Shank 2023) found in DEP's *Water Quality Assessment Methodology for Surface Waters* (Assessment Book, Shull and Whiteash 2023); however, it is beneficial to conduct advanced source decision methods. Specific benefits of advanced source decision include more comprehensive background information and data to support TMDL development, application of targeted enforcement actions, and more appropriate implementation of best management practices for restoration activities (USEPA 2005). It is important to note that advanced source decision can be resource intensive, and multiple samples are highly recommended to confidently determine bacteria sources (USEPA 2011).

To distinguish between potential fecal contamination sources, DEP collects water samples to analyze for the presence of host-specific intestinal bacterial DNA from *Bacteroides fragilis* (*B. fragilis*). DEP Bureau of Laboratory (BOL) isolates the DNA and performs real time genotypic quantitative polymerase chain reaction (qPCR) assays with Synergy Brands Inc. (SYBR) Green dye, a fluorescent DNA-binding dye. *B. fragilis* is a common gene bacterium found in the gastrointestinal tract of humans and other warm blooded animals and has known genetic markers for many species that can be detected using specific primer couples: Human (HF183F and HF265R), Bovine (CowM3F and CowM3R), Swine (PigBac2-qBac41F and PigBac2-qPS183R), Deer (EF447F and EF990R), Horse (HoF597F), Dog (DF113F and DF472R) and Geese (CG1F and CG1R). For detecting other avian species, the BOL employs a primer couple using Gull (GFD-F and GFD-R) to detect a genetic marker from *Helicobacter sp*., which is a common bacterium found in the gastrointestinal tract of many avian species.

Results from qPCR analyses are calculated based on the amplification of target DNA sequences from test samples relative to those in calibrator samples that contain a known quantity of target organisms (Haugland et al. 2005, Wade et al. 2010, DEP 2021). The amplification of target genes in the presence of primers (short genetic sequences) that are specific to various hosts (i.e., humans, cows, pigs, horses, birds, deer, and dogs) allows for decision of an estimated concentration of each host DNA as gene copy units per 100 mL (GC/100mL) for a water sample. This provides information about which host is contributing to the elevated fecal levels in specific surface water samples.

The highest contributing host likely causes the criteria exceedance and indicates the source of impairment, but results can vary depending on the conditions (e.g., weather, season, and stream discharge), so it is critical to measure and evaluate both fecal indicator bacteria concentrations and host qPCR results that are collected at the same time to obtain relative abundance of potential sources (USEPA 2011). Tetra Tech Inc. and Herrera Environmental Consultants (USEPA 2011) provide the following example for why fecal indicator and qPCR samples, and the subsequent interpretation, need to be measured at the same location and time:

> *For example, some water samples might be dominated by human bacteria while others show bacteria from cattle. However, without the corresponding bacteria concentrations of those samples, it is unknown which water samples are exceeding water quality criteria and which sources could be contributing to those exceedances. In the example, bacteria concentration data might show that all the human-dominated samples have low levels of bacteria; however, the livestock-dominated samples have concentrations exceeding criteria. This indicates that while both human and livestock sources of bacteria exist in the watershed, it's likely that the livestock sources are causing the elevated concentrations that are leading to exceedances of water quality criteria.*

The process described below details the steps needed to make advanced source decision for dominant sources of fecal pollution using qPCR and bacteriological data.

MAKING SOURCE DECISIONS

Data Requirements

Details on bacteriological sampling design and data collection requirements are in DEP's *Water Quality Monitoring Protocols for Surface Waters* (Monitoring Book, Lookenbill and Arnold 2023), but generally, four to six qPCR samples will be collected when preliminary land cover analysis reveals high agricultural (>30%) or developed (>15%) coverages. Sampling for qPCR is also recommended if the initial collections of the fecal indicator begin to approach criteria (e.g., qPCR samples are recommended when initial *E. coli* samples are 100 colony forming units per 100 mL (CFU/100 mL) or greater; approaching the criteria of 126 CFU/100mL in 25 Pa. Code § 93.7).

Decision Process

When the data requirements have been met, focus is placed on the dominant host for source decision. Additional hosts may also be considered when qPCR samples correspond with fecal

indicator results that exceed criteria. When considering additional hosts, it is important to note that there is no direct relationship between fecal indicator concentrations and qPCR gene copy values. For example, if the E. coli criterion were exceeded and the two highest corresponding results were 200 CFU/100mL with human being the dominant host at 450 GC/100mL and 300 CFU/100mL with swine being the dominant host at 700 GC/100mL, it is not possible to infer a single host as the most dominant source leading to the *E. coli* exceedance. Therefore, consideration of qPCR results at a larger scale (i.e., order of magnitude) across all samples needs to be used and compared. The order of magnitude is defined as an exponential change of plus or minus one in the value of qPCR results. In the example above, both sources would be considered relatively equal contributors since both *E. coli* results contributed to criteria exceedance and both qPCR results were at the same magnitude.

Consistency of elevated qPCR results should also be considered when determining source. For example, if all *E. coli* samples contributed to criteria exceedance, and four out of five qPCR results showed human was the dominant source, then sources associated with humans (see Figure 1) could be justified. If the other qPCR sample suggested bovine was the dominant source, bovine related sources (see Figure 1) would only be considered if additional evidence through field reconnaissance, aerial imagery, or discharges existed to support that decision. The single bovine-dominant qPCR result alone may not be enough information to make the bovine related source decision, especially since the sample may have been collected during a time that was not representative of typical conditions (i.e., a major storm event).

Ultimately, this advanced source decision method uses the magnitude and consistency of qPCR results along with other supporting information (collected before and during the assessment) in a weight of evidence approach to make final source decisions (Figure 1). The greatest advantage of using this method is that it can select more specific sources of impairment while also providing a link between the fecal indicator and the source through DNA evidence. For example, “WATERFOWL” could be selected as a source if avian species were identified through qPCR analysis and the only likely source in the watershed was a public lake. There may be occasions when this method cannot point to a specific source. For example, if avian species were identified through qPCR analysis, but the land cover showed mostly agricultural land cover with no obvious point sources (e.g., poultry CAFO discharges), then “AGRICULTURE” may be the most appropriate source decision. Several additional examples are provided below to demonstrate this advanced source decision process.

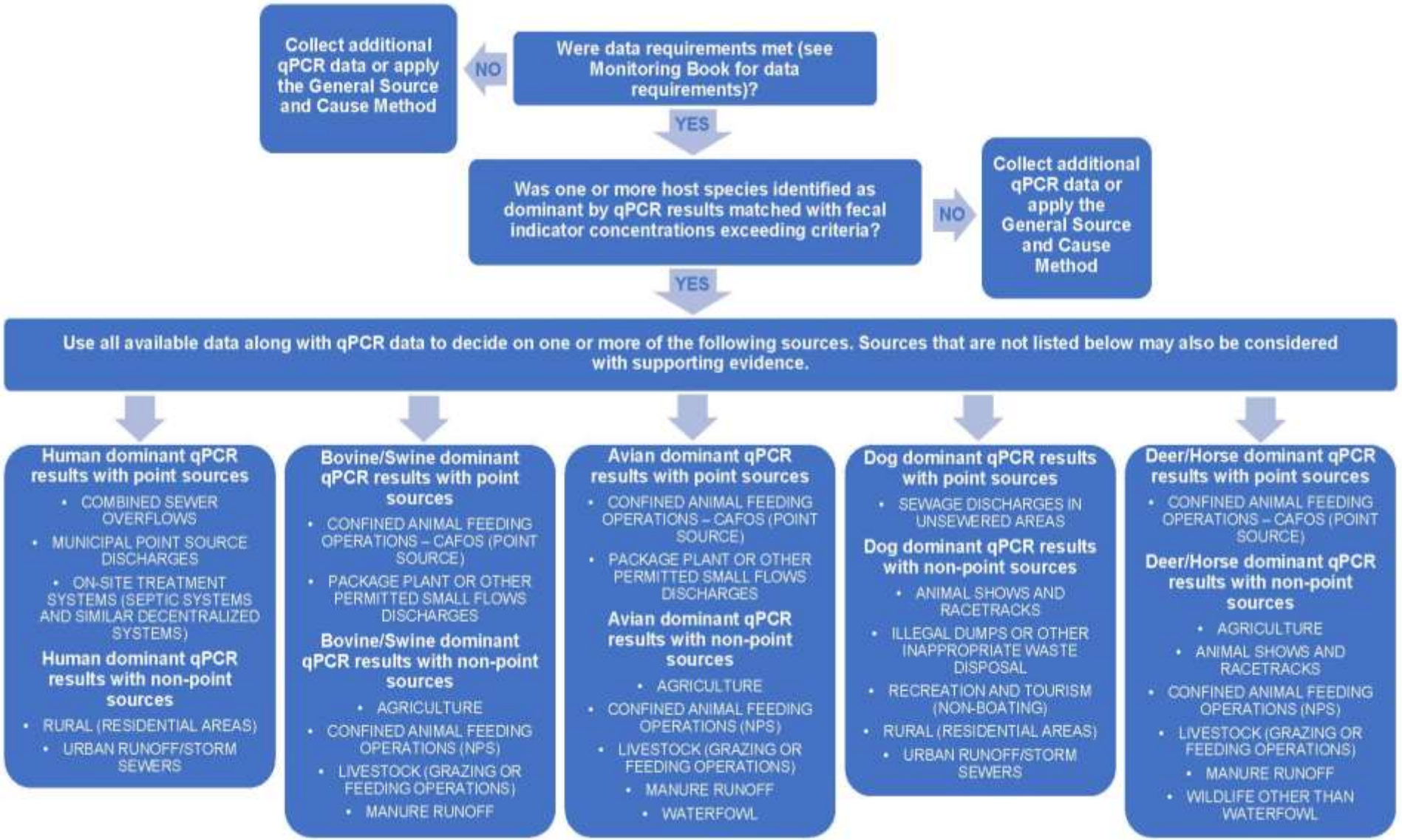

Figure 1. Bacteriological source decision process to include suggested sources that should be considered for each dominant host. Consult *Appendix A: Sources and Causes* in the Assessment Book (Shull and Whiteash 2023) for a full list of available sources.

Additional Considerations

The land application of human biosolids is often a concern when a combination of human and livestock qPCR results is discovered. Class A and Class B human biosolids have pathogen reduction requirements that are found in 25 Pa. Code § 271.932 regulations. Therefore, human DNA would exist, but the pathogen levels resulting from humans would most likely have been rendered incapacitated. When a combination of human and livestock qPCR results is discovered and human biosolids is suspected, land cover analysis and point source GIS layers should be used to make a final source decision. In this case, consideration of more general sources such as “AGRICULTURE” or “RURAL (RESIDENTIAL AREAS)” is more appropriate than a specific source, such as “DISCHARGES FROM BIOSOLIDS (SLUDGE) STORAGE, APPLICATION OR DISPOSAL”.

A BOL in-house host source evaluation has revealed the existence of a weak cross-reaction of swine qPCR primers with human DNA. If human DNA is present in a sample, a false result indicating the detection of a small amount of swine DNA may occur. When this potential cross-reactivity occurs, the concentration of swine DNA usually amounts to approximately 10% of the human DNA concentration determined in the same sample. Human qPCR primers do not show cross-reactivity with swine DNA, so an inverse observation would not be considered a false detection. Regardless of this weak cross-reaction, the method described above would not allow a low-level swine qPCR cross-reactivity to be used to decide final sources.

Stormwater samples have been documented to have high concentrations of human sources in almost all surface waters, even when minimal anthropogenic activity exists. This is potentially due to overflow discharges at all wastewater facilities that treat sewage and exceed holding capacity. Consequently, it is important to evaluate the qPCR data in context of weather and potentially weight high stormwater flow samples lower than samples that are collected during times when water contact is more likely to occur.

Examples

The following three examples show 2017 qPCR results being used for source decisions. All data requirements were met before proceeding with the source decision process. In 2017, DEP BOL qPCR sample minimum reporting limit was <100 GC/100mL. For samples collected after 2018, the minimum reporting limit changed to <20 GC/100mL.

<u>Example 1</u>

A site on Manada Creek at Carlson Road bridge crossing in East Hanover Township was sampled for a WC assessment. The geometric mean of *E. coli* (212 CFU/100 mL) indicated impaired conditions. Land cover was 70% forest, 18% agriculture, and 12% urban. There were two municipal stormwater discharges and nine on-site septic treatment discharges within the 24 mi^2 watershed. Light rain was reported within 24 hours of sampling on August 3rd and 16th, 2017, but these rain events were not considered to be major storms as evidenced by only a 3% discharge increase at United States Geological Survey (USGS) stream gauge 01573560 on Swatara Creek near Hershey, Pennsylvania. No weather events were reported on August 10th or 30th, 2017. Three out of the four qPCR samples documented human as the dominant host. There was one sample that documented both human and

swine with the high results, but swine was an order of magnitude higher, thus indicating it was the dominant host during that sampling event. Overall, human was the more consistent host, but swine was observed to be the dominant host on one occasion (Table 1). After considering all available data and evaluating qPCR results, the sources of impairment were determined to be “RURAL (RESIDENTIAL AREAS)” and “AGRICULTURE)”. In this example, human host dominant qPCR samples were used to support ariel imagery and land cover analysis that indicated a “RURAL (RESIDENTIAL AREAS)” source decision. The swine dominant qPCR samples were used to support ariel imagery and land cover analysis that indicated an “AGRICULTURE” source decision. More specific source decisions may have been supported through additional compliance investigation and cause and effect survey information (Lookenbill and Wertz 2021).

Table 1. Manada Creek at Carlson Road bacteria (CFU/100 mL) and qPCR (GC/100 mL) results. Highest relative qPCR results for each sample are highlighted in gray.

Date	E. *coli*	qPCR Human	qPCR Swine	qPCR Bovine
8/3/2017	930	941	<100	<100
8/10/2017	140	592	2580	<100
8/16/2017	120	376	<100	<100
8/30/2017	130	499	<100	<100

Example 2

A site on Reeds Creek at Jonestown Road bridge crossing in East Hanover Township was sampled for WC use assessment. The geometric mean of *E. coli* (578 CFU/100 mL) indicated impaired conditions. Land cover was 39% forest, 36% agriculture, and 25% developed. There was one municipal stormwater discharge, two CAFOs, and two sewage pump stations within the 9 mi^2 watershed. Light rain (< 0.25-inch rain) was reported within 24 hours of sampling on August 3rd and 16th, 2017, but these rain events were not considered to be major storms as evidenced by only a 3% discharge increase at United States Geological Survey (USGS) stream gauge 01573560 on Swatara Creek near Hershey, Pennsylvania. No weather events were reported on August 10th or 30th, 2017. Two out of the four qPCR samples showed bovine as the dominant host. The other two samples had similar magnitude results for all the hosts. Overall, bovine was the dominant and most consistent host (Table 2). In this example, bovine host dominant qPCR samples were used to support ariel imagery and land cover analysis that showed a mixture of row crop production and pasture, indicating an “AGRICULTURE” source decision. Again, more specific source decision may have been supported through additional compliance investigation and cause and effect survey information.

It is important to note the difference between this example and the first example. In this example, percent developed land cover was higher than what was observed in the first example, but qPCR results did not support human-related sources using this method. On August 30th, human host was the highest qPCR result, but this result was at the same magnitude as the swine qPCR result. Thus, the August 30th qPCR results were determined to be inconclusive. Additionally, although all *E. coli*

concentrations were higher than 25 Pa. Code § 93.7 criteria, the highest *E. coli* concentrations coincided with bovine host qPCR results, further supporting an “AGRICULTURE” source decision.

Table 2. Reeds Creek at Jonestown Road bacteria (CFU/100 mL) and qPCR (GC/100 mL) results. Highest relative qPCR results for each sample are highlighted in gray.

Date	*E. coli*	qPCR Human	qPCR Swine	qPCR Bovine
8/3/2017	720	<100	102	7570
8/10/2017	140	<100	174	<100
8/16/2017	4100	<100	175	2350
8/30/2017	270	329	113	<100

Example 3

A site on Quittapahilla Creek at Bellegrove Road bridge crossing in North Annville Township was sampled for WC use assessment. The geometric mean of *E. coli* (173 CFU/100 mL) indicated impaired conditions. Land cover was 16% forest, 51% agriculture, and 35% developed. There were 37 permitted discharges for municipal and industrial stormwater and waste, one on-site septic treatment system, and three sewage treatment facilities within the 73 mi^2 watershed. Light rain was reported within 24 hours of sampling on August 3rd and 16th, 2017, but these rain events were not considered to be major storms as evidenced by only a 3% flow increase at United States Geological Survey (USGS) stream gauge 01573560 Swatara Creek near Hershey, Pennsylvania. No weather events were reported on August 10th or 30th, 2017. All four qPCR samples documented less-than-detect results (Table 3). After considering all available data and consulting the *General Source and Cause Method* (Shull and Shank 2023), aerial imagery and land cover analysis indicated “RURAL (RESIDENTIAL AREAS)” and “AGRICULTURE” as source decisions.

Table 3. Quittapahilla Creek at Bellegrove Road bacteria (CFU/100 mL) and qPCR (GC/100 mL) results.

Date	*E. coli*	qPCR Human	qPCR Swine	qPCR Bovine
8/3/2017	470	<100	<100	<100
8/10/2017	90	<100	<100	<100
8/16/2017	150	<100	<100	<100
8/30/2017	140	<100	<100	<100

LITERATURE CITED

DEP. 2021. Detection and relative quantification of gene targets in Bacteroides using real-time PCR (qPCR) with SYBR Green dye. Pennsylvania Department of Environmental Protection, Bureau of Laboratories, Harrisburg, Pennsylvania.

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

Lookenbill, M. J., and E. Arnold (editors). 2023. Water quality monitoring protocols for surface waters. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania.

Lookenbill, M. J., and T. A. Wertz. 2021. Cause and effect surveys. Chapter 2.1 *in* M. J. Lookenbill and E. Arnold (editors). Water quality monitoring protocols for surface waters. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania.

Miller, S. (editor). 2023. Bacteriological data collection protocol. Chapter 3.11 *in* M. J. Lookenbill and E. Arnold (editors). Water quality monitoring protocols for surface waters. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania.

Miller, S., and R. Whiteash (editors). 2021. Bacteriological assessment method. Chapter 5.1 *in* D. R. Shull and R. Whiteash (editors). Water quality assessment methodology for surface waters. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania.

Shull, D. R., and M. K. Shank (editors). 2023. General source and cause method. Chapter 9.1 *in* D. R. Shull and R. Whiteash (editors). Water quality assessment methodology for surface waters. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania.

Shull, D. R., and R. Whiteash (editors). 2023. Water quality assessment methodology for surface waters. Pennsylvania Department of Environmental Protection, Harrisburg, Pennsylvania.

USEPA. 2000b. Atlas of America’s polluted waters. EPA 840-B00-002. U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

USEPA. 2005. Microbial source tracking guide document. EPA-600/R-05/064. U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C. p. 131.

USEPA. 2011. Using microbial source tracking to support TMDL development and implementation. Prepared for U.S. Environmental Protection Agency, Region 10, Watersheds Unit, by Tetra Tech Inc. and Herrera Environmental Consultants, Seattle, Washington.

Wade, T. J., E. Sams, K. P. Brenner, R. Haugland, E. Chern, M. Beach, L. Wymer, C. C. Rankin, D. Love, Q. Li, R. Noble, and A. P. Dufour. 2010. Rapidly measured indicators of recreational water quality and swimming-associated illness at marine beaches: A prospective cohort study. Environmental Health. 9:66.