

Effects of Holding Time, Storage, and the Preservation of Samples on Sample Integrity for the Detection of Fecal Indicator Bacteria by Quantitative Polymerase Chain Reaction (qPCR)-based assays.

Larry Wymer¹

Kevin Oshima¹

Jack Paar III²

Mark Doolittle³

Jennifer Lavender¹

Manju Varma¹

Rich Haugland¹

¹ US Environmental Protection Agency, National Exposure Research Laboratory, Cincinnati, OH 45268

² US Environmental Protection Agency, New England Regional Laboratory, North Chelmsford, MA, 01853

³ TechLaw, Inc., Environmental Services Assistance Team (ESAT), U.S. EPA Region 1, Lowell, MA 01853

Executive Summary

The purpose of this project was to answer questions related to storage of samples to be analyzed by the quantitative polymerase chain reaction (qPCR)-based assays for fecal indicator bacteria. The project was divided into two parts. The first part was to determine if filters that were used to collect fecal indicators could be stored frozen and analyzed at a later date and the second part was to determine if refrigerated water samples could be held for 24 to 48 hours prior to analysis by qPCR. Both of these studies answer questions that were important in the analysis of fresh and marine surface water samples for beach monitoring purposes.

I. Archived Sample Stability Study and Long-Term Holding Time Study: Evaluation of effects of freezer-storage on the preservation of filter samples for qPCR analysis.

The Beach Environmental Assessment and Coastal Health (BEACH) Act of 2000 directed the EPA to conduct studies concerning pathogen indicators in coastal recreation waters. The results of these studies are to be used by EPA to publish new or revised water quality criteria for the purpose of protecting human health in coastal recreational waters. EPA has conducted a number of studies pursuant to the BEACH Act including studies:

- To develop new rapid methods for measuring water quality
- To establish the relationship between water quality and health using the rapid methods
- To develop a system for monitoring water quality
- To provide guidance to states on the application of the new indicator criteria

The US EPA has conducted the National Epidemiological and Environmental Assessment Research (NEEAR) Water Study at four beaches on the Great Lakes and three on the Atlantic and Gulf Coasts. Water quality at each of the beaches was impacted by point sources that received combined treated sewage discharges from communities with populations of at least 15,000.

The NEEAR studies contained a health data collection component as described in Wade et al. (2006, 2008). The objective of the health portion of the study was to quantify the symptomatological observations in the swimmer vs. non-swimmer groups. The second component of the study was to collect water quality data on fecal indicator bacteria (FIB) using rapid methods such as quantitative polymerase chain reaction (qPCR) for *Enterococcus* and *Bacteroidales*.

As part of the NEEAR study design, replicate samples were filtered and the filters were retained for future study when additional methods or improvements became available. These samples could then be reanalyzed using these new or modified qPCR methods.

Relationships with health risks could then be derived for additional indicator methods based on these future analyses. The key question, however, is whether qPCR results of archived filters are equivalent to qPCR analyses that are performed on the fresh samples. In order for this to be the case, it must be shown that qPCR results from fresh and archived samples are equivalent. There are mainly two potential causes for changes in qPCR results between fresh and archived samples

(1) extraction of genetic material from intact cells may be more or less efficient from frozen, archived filters than from recently prepared filters, and (2) DNA itself may be altered during the archival process.

The key objective of the archived sample stability study was to establish if the archived filters were suitable for generating new qPCR-based assessments for modified or new qPCR methods for indicator bacteria. If the analysis indicates that the archived samples do not show a change over time with respect to these specific qPCR assays, this would lend credibility to using the archived NEEAR filters for analysis with newly-developed qPCR methods in conjunction with the health analyses from the NEEAR studies in lieu of conducting full-blown prospective epidemiological studies *de novo*. The analysis was done by comparing *Enterococcus* density estimates from both the Great Lakes and marine beach samples and *Bacteroidales* density estimates from the marine beach samples only, as determined from analyses of the original and archived filters using the same *Enterococcus* and *Bacteroidales* qPCR methods.

To further evaluate these possibilities, a long-term holding time study was conducted by the U.S. EPA, Region 1 Research Laboratory in North Chelmsford, MA. In this study, replicate filter retentates from twenty-nine freshwater and twenty-three marine water samples from the Boston, MA area were held in freezer storage for varying lengths of time for up to two years prior to analysis. In addition to qPCR analysis for *Enterococcus*, as was performed on the NEEAR samples, this study sought to provide data for other qPCR-based assays, specifically for *Bacteroidales* and *E. coli*. QPCR-based methods analysis for *Bacteroidales* (Great Lakes) and *E. coli* (Great Lakes and marine) were not available when the original samples from the NEEAR

studies were analyzed. A demonstration that the filter analysis results in this study do not show a change over time, would lend further credibility to using newly-developed qPCR assays to analyze archived filters.

Conclusions from the NEEAR archived sample stability study were, however, that significant degradation of samples had occurred. The highly significant changes and low or absent correlation between archived and original sample analyses indicate that the archived filters cannot be used in a credible manner to establish health relationships involving *Enterococcus* qPCR or, by extension, any other indicator, pathogen, or method. Any health relationship based on data derived from the archived samples is not useful because these do not reflect data that would have been obtained from the original samples in actual beach monitoring circumstances.

The results from the long-term holding time study, though of reduced importance as a result of findings from the archived sample stability study, showed significant differences in the 24-month holding time study for *Bacteroidales* and *E.coli* that further complement the findings of the archived NEEAR study samples, mainly that differences are observed between the original samples and samples that are held for extended time periods in freezer storage. It is noted that the holding temperature used in this study was -20° C.

II. Short-Term Holding Time Study: Evaluation of holding refrigerated surface water samples for up to 48 h for qPCR analysis

The short-term holding time study was conducted to determine whether holding refrigerated surface water samples for 24- and 48-hour affects density estimates of *Enterococcus*, *Escherichia coli*, and *Bacteroidales* bacteria determined by a qPCR method. The assays for these three groups of organisms in the qPCR method are designated as Entero1, EPA-EC23S, and GenBac3, respectively. These organisms are currently being considered as potential indicators of fecal pollution for use in new recreational water quality criteria. Twenty-nine freshwater and twenty-three marine water samples were collected in the Boston, Massachusetts area and analyzed by each of the assays after being held with refrigeration for 24- and 48-hours prior to the collection of target organisms by filtration. Results were compared to those from additional subsamples of the same water samples that were filtered and the filters frozen within six hours of collection. The latter filters, designated as recovery standards, were considered to be representative of freshly collected samples and were extracted and analyzed side-by-side with the Refrigerated Water samples at each holding time in an effort to minimize certain method-related sources of uncertainty. The refrigerated water sample filters were frozen and immediately thawed prior to DNA extraction in order to mimic conditions used to analyze the recovery standards.

A supplemental portion of the study, motivated by preliminary results from the primary study indicating a potential holding time effect, was also conducted using Ohio River water. A single Ohio River water sample was collected and subsamples were analyzed using the same qPCR assays after 1, 24, and 48 hours of refrigerated holding time without freezing. Twelve

subsamples were tested for each holding time. Data from these samples also provided a basis for estimating the standard deviation among pure replicate subsamples as a basis for comparison with potential holding time biases.

A summary of major findings of this study are presented below:

- Significantly lower recoveries of target sequences were realized after 24 hours of holding time for both the Enterobacter and GenBac3 assays. Observed qPCR recoveries were about 12% and 13% lower, respectively after 24 hour, with a potential range of as much as a 22% decrease.
- Recovery from the *E. coli* EPA-EC23S assay, while a net of 3% lower, was not significantly affected by a 24 hour holding time.
- In contrast there were no significant net losses in recovery after 48 hour holding time for any of the three assays.
- Supplemental data from the Ohio River water portion of the study showed significant decreases in qPCR estimated indicator densities as determined by one of the three assays after 24 hours and significant decreases by all 3 assays after 48 hours, as compared to when the sample was held for only one hour.
- Any bias due to holding time amounted to 8% (less than 17% at the 95% confidence level) of the total variation in results (actually the root mean square error, RMSE, roughly equivalent to the standard deviation) among replicate aliquots.

Conclusions

There were small, and in some cases statistically significant changes in qPCR-estimated fecal indicator densities in refrigerated water samples that were held for 24 and 48 hours. However, because these changes were neither consistent over time nor were they consistent between indicators, no conclusions on the effects of holding refrigerated water samples can be drawn from the short-term holding time study. Significant declines in qPCR signals at 24 and 48 hours were observed in the Ohio River sample portion of the study, but these observations were based on a single sample and did not consider potential influences of certain method-related sources of uncertainty. Taken together, the results of the two studies are inconclusive and, as a result, **can neither support nor refute** recommendations to hold refrigerated water samples for 24 hours or longer.

**I. Archived Sample Stability Study and Long-Term Holding Time Study: Evaluation
of effects of freezer-storage on the preservation of filter samples for qPCR analysis.**

Introduction

The Beach Environmental Assessment and Coastal Health (BEACH) Act of 2000 directed the U.S. EPA to conduct studies concerning pathogen indicators in coastal recreation waters. The results of these studies are to be used by the U.S. EPA to publish new or revised water quality criteria for the purpose of protecting human health in coastal recreational waters. The U.S. EPA has conducted a number of studies pursuant to the BEACH Act including studies:

- To develop new rapid methods for measuring water quality
- To establish the relationship between water quality and health using the rapid methods
- To develop a system for monitoring water quality
- To provide guidance to states on the application of the new indicator criteria

The U.S. EPA has conducted the National Epidemiological and Environmental Assessment Research (NEEAR) Water Study at four beaches on the Great Lakes and three on the Atlantic and Gulf Coasts. Water quality at each of the beaches was impacted by point sources that received combined treated sewage discharges from communities with populations of at least 15,000.

The NEEAR studies contained a health data collection component as described in Wade et al. (2006, 2008). The objective of the health portion of the study was to quantify the

symptomatological observations in the swimmer vs. non-swimmer groups. The second component of the study was to collect water quality data on fecal indicator bacteria (FIB) using rapid methods such as quantitative polymerase chain reaction (qPCR) for *Enterococcus* and *Bacteroidales*.

As part of the NEEAR study design, replicate samples were collected and archived at the EPA lab in Cincinnati, OH for future study when additional methods or improvements became available. Relationships with health risks could then be derived for additional indicator methods based on these future analyses. The key question, however, is whether qPCR results of archived samples are equivalent to qPCR analyses that are performed on the original samples? In order for this to be the case, it must be shown that qPCR results from original and archived samples are equivalent. There are mainly two potential causes for changes in qPCR results between original and archived samples: (1) extraction of genetic material from intact cells may be more or less efficient from frozen, archived filters than from recently prepared filters, and (2) DNA itself may be altered during the archival process.

The key objective of this archived sample stability study was to establish if the archived filters were suitable for generating new qPCR-based assessments for modified or new qPCR methods for indicator bacteria. If the analysis indicates that the archived samples do not show a change over time with respect to these specific qPCR assays, this lends credibility to analyzing archived NEEAR filters with newly-developed qPCR methods, in conjunction with the health analyses from the NEEAR studies, in lieu of conducting full-blown prospective epidemiological studies *de novo*. Thus, the analyses in this study were performed by comparing *Enterococcus* density

estimates from Great Lakes and marine beach samples and *Bacteroidales* density estimates from marine beach samples only, as determined from analyses of the original and archived filters using the same *Enterococcus* and *Bacteroidales* qPCR methods.

To further evaluate these possibilities, a long-term holding time study was conducted by the U.S. EPA, Region 1 Research Laboratory in North Chelmsford, MA. In this study, replicate filter retentates from twenty-nine freshwater and twenty-three marine water samples from the Boston, MA area were held in freezer storage for varying lengths of time for up to two years prior to analysis. In addition to qPCR analysis for *Enterococcus*, as was performed on the NEEAR samples, this study sought to provide data for other qPCR-based assays, specifically for *Bacteroidales* and *E. coli*. QPCR methods for analysis of *Bacteroidales* (Great Lakes) and *E. coli* (Great Lakes and marine) were not available when the original samples were analyzed from the NEEAR studies. A demonstration that the results from long term holding of frozen filters do not show a change over time, would lend further credibility to using newly-developed qPCR assays to analyze archived filters.

Material and Methods

Materials and methods for archived sample stability study

Water samples. Water samples were collected over a four year period from 2003 to 2007 and held in the freezers at the U.S. EPA's National Exposure Research Laboratory (NERL), Cincinnati, OH. Sampling sites were West Beach at the Indiana Dunes National Lakeshore in Porter, Indiana on Lake Michigan in 2003; Huntington Beach in Bay Village, Ohio on Lake Erie in 2003; Silver Beach, near St. Joseph, Michigan, and Washington Park Beach in Michigan City,

Indiana on Lake Michigan in 2004; Edgewater Beach in Biloxi, MS on the Gulf of Mexico in 2005; Fairhope Municipal Beach in Fairhope, AL on the Gulf of Mexico in 2007; and Goddard State Memorial Park Beach in West Warwick, RI on Long Island Sound in 2007.

Sampling designs were similar at each of the sites. Sampling visits occurred on Saturdays, Sundays and holidays over time periods varying from approximately 10 to 12 weeks from either May through August or June through September. Sampling occurred three times daily, at 8 AM, 11 AM, and 3 PM in waist-level (1 m deep) and shin-level water (0.3 m deep) locations along three transects perpendicular to the shoreline. The sampling design at Huntington Beach included three additional shin-level locations.

Sample collection and distribution. One liter water samples were collected at each location by standard methods as recommended in Section 9060 of *Standard Methods for the Examination of Water and Wastewater*, American Public Health Association (1998). Following collection, all samples were placed in coolers and maintained on ice during transport to a local laboratory and at 1 - 4° C during the time interval before they were processed. Processing of all samples by filtration was performed within six hours of collection.

Duplicate 50 or 100 ml volumes of each water sample (either 50 or 100 ml was used consistently for all samples within a beach) were filtered through 47-mm, 0.4-µm pore size polycarbonate filters (catalog #K04CP04700, Osmonics Inc., Minnetonka, MN) and the sides of the funnels were rinsed twice with 20 ml of sterile, phosphate buffered saline. The filters were transferred to a petri dish with the sample side facing up. Using sterile forceps, each filter was folded into a

cylinder with the sample side facing inward, and then inserted into a 2 ml semiconical screw-cap microcentrifuge tube (extraction tube; catalog #506-636, PGC Scientific, Gaithersburg, MD) containing 0.3 g of acid-washed glass beads (catalog #G-1277, Sigma, St. Louis, MO). The filters were held at -20° C for no more than three days until shipment to the analytical laboratories on dry ice. One filter from each water sample was shipped to EMSL Analytical, Inc., Cinnaminson, NJ for analysis within seven days while the duplicate was shipped to the U.S. EPA NERL, Cincinnati, OH for archiving.

Sample archiving. Filters received by the U.S. EPA NERL laboratory were immediately transferred to a -40° C, 25 cubic ft. capacity, 208V upright freezer (model A25-40T, So-Low Environmental Equipment Co., Cincinnati, OH) where they were stored continuously (except as noted below) until analysis in 2009. The freezer temperature was continuously monitored by a centralized monitoring system within the Cincinnati facility. On several occasions samples were briefly removed from the freezer for reorganization and/or while defrosting of the freezer and on one occasion they were temporarily transferred to other freezers due to a facility power outage.

Sample analyses. Original filters sent to the EMSL lab and archived filters stored at the NERL laboratory were extracted to recover total DNA and the DNA extracts were subjected to qPCR analysis by the basic procedures described in Haugland et al., 2005. Briefly, cells were suspended from the filters and lysed in a bead mill for 60 seconds at maximum speed and the debris was removed by centrifugation. For all samples analyzed after 2004, including the archived samples, the published DNA extraction procedure was modified slightly by increasing the total volume of extraction buffer, containing 0.2 µg ml⁻¹ salmon DNA in AE buffer

(Qiagen, Valencia, CA), from 0.3 ml to 0.6 ml and decreasing the dilution of extracts prior to analysis from 10-fold to 5-fold. Calibrator samples (three to six replicates), consisting of clean polycarbonate filters amended with known cell quantities of *Enterococcus faecalis* (ATCC# 29212) and/or *Bacteroides thetaiotaomicron* (ATCC # 29741), and negative control samples (three to six replicates), consisting of clean filters only, were extracted in the same manner with each batch of test samples. Cells used by EMSL in the calibrator samples originated from laboratory grown cultures and were enumerated as previously described (Haugland et al., 2005; Sieftring et al., 2008). Cells used by U.S. EPA NERL were enumerated by flow cytometry and were acquired in the form of commercially available, lyophilized pellets (Bioballs™, BTF, Sydney Australia). QPCR analyses were performed using a previously described primer and TaqMan™ hybridization probe assay for *Enterococcus* target sequences (Haugland et al., 2005) on all samples (Great Lakes and marine) and a previously described primer and TaqMan™ hybridization probe assay for total *Bacteroidales* target sequences (Sieftring et al., 2008) were performed on marine samples only. QPCR amplification of water sample and calibrator sample DNA extracts, and negative control samples, was performed by using 5 µL of equally diluted extracts in a total reaction volume of 25 µL. Reagent mixes were prepared by combining 12.5 µL of TaqMan® Universal Master Mix (Applied Biosystems, Foster City, CA), 2.5 µL of 2 mg/ml bovine serum albumin, 1 µM of each primer, and 80 nM of probe for each reaction. Amplification occurred with an initial start at 50°C for 2 min followed by 95°C for 10 min, then forty PCR cycles of 95°C for 15 s and 60°C for 1 min. All analyses of the fresh samples were performed by EMSL Analytical Inc., Cherry Hill, NJ in a Cepheid SmartCycler® II (Cepheid, Sunnyvale, CA). All analyses of the archived samples were performed in the U.S. EPA NERL

laboratory in either a Cepheid SmartCycler[®] II (freshwater samples) or in an Applied Biosystems StepOnePlus[®] (marine samples).

Materials and methods for analyses of long-term holding time study samples

Water samples. Water samples analyzed in this study were collected from 29 freshwater sites and 23 marine water sites in the Boston, MA area from July through October, 2008 (Table 1). Although 25 samples from both marine and freshwater sites were planned for, some sites thought to be saltwater were later reclassified after measurements revealed low levels of salinity. Selection of freshwater and marine sampling sites was based on their diversity and their proximity to the Boston area location of the U.S. EPA Region 1 laboratory which conducted the study. A further consideration in the selection of the sampling sites was that historic data gathered by the laboratory and various collaborators indicated that the water samples could reasonably be expected to contain more than the estimated 95% confidence detection limit of the qPCR method of approximately 100 target organism cells per sample (Haugland et al., 2008). Actual samples giving lower mean results than 100 enterococci in initial 24 hour analyses were excluded from the study to eliminate the possibility of having subsequent results below the detection limit of the qPCR methods.

Sample collection and archiving. Several water samples were simultaneously collected in 1 liter bottles at each site by standard methods as recommended in Section 9060 of *Standard Methods for the Examination of Water and Wastewater*, American Public Health Association (1998). Following collection, all samples were placed in coolers and maintained on ice during

transport to the U.S. EPA Region 1 laboratory and at 1 - 4° C during the time interval before they were processed.

Upon arrival to the laboratory, and within 6 hours of collection, the contents of the individual 1 liter collection bottles for each sampling site were combined in a carboy and mixed. A total of eight 50 ml, or for some samples 100 ml volume, aliquots of each composite water sample were filtered through 47-mm, 0.4-µm pore size polycarbonate filters (catalog #K04CP04700, Osmonics Inc., Minnetonka, MN) and the sides of the funnels were rinsed twice with 20 ml of sterile, phosphate buffered saline. The filters were then transferred to a petri dish with the sample side facing up. Using sterile forceps, each filter was folded in half three times to form an umbrella and then inserted into a 2 ml semiconical screw-cap extraction tubes, containing 0.3 g of siliconized ceramic beads (Roche MagNA Lyser Green BeadsTM). All filter samples were flash frozen by placing the tubes in a cooling block, pre-chilled to -20° C for 1 hour, and then held in a -20° C freezer until they were extracted and analyzed. Duplicate filter samples from each site were extracted and analyzed after being held for 24 hour, and 6, 12 and 24 months. Multiple replicate calibrator sample filters were prepared just prior to the study by placing aliquots of a single mixed cell suspension containing pre-determined cell quantities of laboratory-grown, representative target strains of each of the assays (*E. faecalis*, *B. thetaiotaomicron* and *E. coli*) on clean filters. These calibrator filter samples were held at -20° C and replicate filters were extracted and analyzed in parallel with each batch of test samples at each time point in the holding study. Since the calibrator sample filters were held in freezer storage for the same lengths of time as the water sample filters, the holding time effects

presented in this report are specific to the water samples and do not take into account any potential holding time effects on the cultured calibrator cells.

Sample analyses. Filter samples were extracted to recover total DNA and the DNA extracts were subjected to qPCR analysis as described for the NEEAR study samples. In addition, analyses for *E. coli* 23S rRNA gene target sequences were performed in the same manner except using Gene Expression PCR Master Mix (Applied Biosystems) and an unpublished TaqMan® probe and primer set (Chern et al., manuscript submitted for publication). All of the qPCR analyses were conducted by the U.S. EPA Region 1 laboratory in a Cepheid SmartCycler® II.

Computational method and statistical analysis for archived sample stability and long-term holding time studies

Analysis data consisted of paired observations of filters from each sample: (1) a qPCR measurement soon after the sample had been collected and (2) a qPCR measurement after archival and storage over a six month to two year period (long-term holding time study) or a two to six year period (archived sample stability study). Statistical analysis was performed on log transformed calibrator cell equivalents (CCE) estimated by qPCR using the “ $\Delta\Delta CT$ ” approach as previously described (Haugland et al., 2005) and as used in the NEEAR study analysis of swimmers’ risk-exposure relationship (Wade *et al.*, 2006, 2008).

This computational approach, which is derived from the comparative cycle threshold (CT) method (Applied Biosystems, 1997), employs an arithmetic formula to determine the ratio of target sequence quantities in DNA extracts from test sample filters relative to those in similarly-

prepared DNA extracts from calibrator sample filters containing a known quantity of target organism cells based on the difference in CT values obtained from qPCR analyses of these samples. Similar comparisons of CT values from qPCR assays for an exogenous target sequence from salmon sperm DNA, added in equal quantities to both the test and calibrator sample filters before DNA extraction, were used both as a reference to normalize results for differences in the amount of total DNA recovered from each sample (e.g., caused by test sample effects on DNA recovery) and as a sample processing control (SPC) to signal potentially non-quantifiable test sample results caused by PCR inhibition or low DNA recoveries (Haugland et al., 2005). The calculation can be expressed by the following equations:

$$\Delta\Delta C_T = \Delta C_{T,\text{target}} - \Delta C_{T,\text{reference}} \quad (1)$$

$$CCE = N_{\text{Calibrator}} \cdot A^{-\Delta\Delta C_T} \quad (2)$$

in which $\Delta C_{T,\text{target}}$ represents $C_{T,\text{sample}} - \text{Mean } C_{T,\text{calibrator}}$ for the target sequence (e.g., enterococci) and $\Delta C_{T,\text{reference}}$ represents the corresponding difference for the salmon sperm reference sequence. $N_{\text{calibrator}}$ is the known number of cells in the calibrator sample and A is the amplification factor for the assay. Ideally, $A=2$ but typically it is in the range 1.9 – 2.0 with values less than 2 resulting from less than 100% replication of the target sequence at each cycle. In practice, A is either assumed to be 2 or is estimated based on the slope of a standard curve (Applied Biosystems, 1997). For both the Frozen Storage study and Archived Preservation study, slope values were obtained from standard curves generated by each of the laboratories from pooled

results of repeated qPCR analyses of serially diluted DNA standards nominally containing target sequences in a range from 10 to 4×10^4 copies per analysis.

For both the archived sample stability and the long-term holding time studies, the analyses were conceptually the same, although differences in the designs of the respective studies necessitated slightly different treatment of the data. In the long-term holding time study, duplicate filters were analyzed for each sample and the extract from each filter was analyzed in duplicate at each time point. All of the results were well above the detection limit, this having been a criterion of sample selection in the first place. Therefore, for each water type, a nesting sampling scheme was present, (duplicate analysis nested within filter and filter nested within sample). A mixed model was used to account for these as nested random effects. Data from the Frozen Storage study were analyzed using the MIXED procedure in SAS v. 9 for Windows (SAS Institute, Cary, NC, 2009).

In the archived sample stability study, a single analysis of a single filter of each sample was performed shortly after collection and again for an archived filter two to six years later. These samples were not collected with the intention that they would be highly contaminated and, as a result, there were many non-detects. Data from the archived sample stability study could thus be analyzed for differences (archived result minus original result), but it would be necessary to consider the non-detects as censored results. If the censored result was the original analysis, the difference would be right censored, that is would represents a lower bound on the actual difference since the original result could be lower than the stated value. Similarly, if the censored result was the archived analysis, the difference would be left-censored, that is, likely to be

smaller than the stated value. Sufficient statistics comprise the set of samples for which one or both results are quantifiable so that samples can be ignored for which both analyses yield non-quantifiable results.

Maximum likelihood estimation (“Tobit analysis”) was used for the archived sample stability study analysis where the difference in log (base 10) between the qPCR results before and after archival could be censored as above. Monte Carlo Markov Chain (MCMC) estimation using WinBUGS v. 1.4 (Lunn *et al.*, 2000; Ntzoufras, 2009) with diffuse priors was employed for this purpose because of the software’s ability to capture estimates of individual differences. These were useful for evaluating the reasonableness of the normality model used and for further evaluating paired observations, particularly in terms of prediction of archived results based on the original results.

Other data analyses and summaries were performed in R v. 2.8 (2008) and Excel 2003 for Windows.

Results

Archived sample stability study

Table 1 summarizes sample characteristics from the NEEAR original and archived data analyses. A total of 2818 samples were collected at the four fresh water and three marine water beach sites. An average of 4% and 5% of these samples did not meet quality control standards or gave assay values below the detection limit for both the original and archived samples, analyzed by EMSL Analytical Inc. and EPA, respectively. These samples were excluded from further analyses.

These failure rates were similar for both fresh and marine samples (Table 1). Of particular note is the relative number of sample data that were below the detection limit (“censored”) in the original analysis vs. when they were reanalyzed after a period of archival. For the *Enterococcus* assay, 24 to 45% of the samples from the different beaches yielded at least one qPCR result that was below the limit of detection. Among the Great Lakes samples, as well as Goddard Beach, non-detects were much more frequent after archival than before. This is as expected if there is a decline in the qPCR signal as a result of archival. However, at the other two marine beaches, non-detects were about as likely to occur after as they were before being archived. The imbalance in results below the limit of detection in original vs. archived samples is one more impetus to incorporate the non-detects in the analysis. Otherwise, results would be biased by discarding more samples that had declined than those that had increased.

Non-detects among *Bacteroidales* were considerably less prevalent because of the larger number of these organisms, generally amounting to about an order-of-magnitude. Assays for *Bacteroidales* were performed at only the three marine beaches, where non-detects amounted to only about 13% of all usable samples. While the percentage of non-detects may be small in comparison to those for *Enterococcus*, a 13% non-detect rate is still substantial. Note that samples with non-detects both before and after archival do not inform the estimation of an archival effect. We may infer that where this occurred, there was increased likelihood that that particular sample was devoid of the target DNA for the respective qPCR assay.

Using one-half the detection limit for non-detect results, Table 2 shows that when qPCR-CCE are calculated for the fresh and archived samples at each beach, the mean values for archived

samples are consistently lower than those of fresh samples, although this decline was relatively small at Edgewater and Fairhope compared to the other beaches. These mean differences by beach, as well as a pooled precision parameter based on substituting one-half of the detection limit for censored data, were used as initial values in the MCMC estimation procedure.

Final analytical results from MCMC estimation that explicitly accounted for values below their respective limits of detection (Table 3) indicated strong, statistically significant declines in qPCR-CCE yields from the archived NEEAR samples from most beaches. Samples from all but Edgewater Beach in Biloxi and Fairhope Beach exhibited highly significant declines with respect to *Enterococcus* recoveries. With respect to *Bacteroidales* qPCR-CCE, Edgewater and Goddard Beach samples exhibited highly significant changes. The analysis was performed on logarithm-transformed qPCR-CCE. To put these numbers in perspective, Table 3 also shows the equivalent percent declines that corresponded to the differences in geometric means (the antilogarithms of the mean log differences). In many cases, these declines imply that only 10% or less of the original DNA remained in the archived samples.

Pearson product moment correlations based on the MCMC estimates are given in the last column of Table 3. For enterococci, the correlations were particularly low where there was any correlation whatsoever. A common measure of predictability in regression analysis is R^2 , which is equal to the square of the Pearson correlation. Multiplied by 100, an R^2 indicates the percent of variation explained by the independent variable. In this study, the independent variable would be regarded as the initial qPCR-CCE result, and knowing this value was seen to account for only

10% or less in the variation among archived sample qPCR results. Scatter plots of archived vs. original $\log_{10}(\text{qPCR-CCE})$ results for *Enterococcus* are shown by beach in Figure 1.

Bacteroidales qPCR-CCE results from the archived samples show much better correlation with their respective initial analyses compared to the *Enterococcus* analyses. These correlations still are not substantial, however, and amount to R^2 values on the order of 0.25, thus explaining only 25% of the variation among archived sample analyses for *Bacteroidales*. We show the scatter of archived vs. original $\log_{10}$ qPCR-CCE results for *Bacteroidales* in Figure 2.

Long-term holding time study

Some significant changes in qPCR-CCE results for the long-term holding time study samples between their initial analysis and their reanalysis two years later were observed (Table 4). Marine water *Enterococcus* assay results actually showed a statistically significant *increase* of 0.16 logs (44%) over this period ($p=0.002$). Mean log qPCR-CCE in this case rose between the initial analysis and the reanalysis at six months and stayed more or less constant at that level at the one and two year marks. Meanwhile, *Enterococcus* qPCR-CCE results from freshwater samples exhibited a small “marginally significant” ($p=0.063$) *decline*. Among the other assays, both *Bacteroidales* ($p<0.001$) and *E. coli* ($p=0.022$) reanalyses indicated significant declines in the PCR target in freshwater samples after two years amounting to 40% and 21% of their initial values (0.22 and 0.10 logs), respectively. The decrease in *Bacteroidales* signal was already evident at six months, but *E. coli* declined incrementally. Another decrease in qPCR results over the two year period was for *Bacteroidales* from marine water samples where the change was only marginally significant ($p=0.056$).

Pearson correlations between initial and two-year $\log_{10}(\text{qPCR-CCE})$ values were all above 0.9, being in all but one case about 0.95 (Table 4). Correspondingly, R^2 values were all 0.84 or higher. This fact is reflected in the plots of two-year vs. initial $\log_{10}(\text{qPCR-CCE})$ values in Figure 3.

Discussion

The primary purpose of the present analysis is to evaluate the feasibility of using the archived NEEAR study samples as surrogates for fresh samples in determining relationships between results of new or revised methods for potential indicator bacteria or pathogens and swimmers' health risks. For the archived data to be used in this manner, the most desirable outcome would be for there to be no change between the original analyses and the archived sample analyses two to six years later. In addition there should be a relatively high correlation between the two sets of results (the latter implying a low variance of the difference). Lacking this, at least a high correlation between the two sets of results might allow some adjustment to be made to the archived data so that they would reliably reflect the fresh sample data. Devising an adjustment factor would necessarily involve non-provable, critical assumptions, such as that other qPCR assays would follow the predictive model established by the results that were available. These assumptions may at least be more readily accepted given this information from the available data.

The analysis of the previous section shows, however, that suitability of the archived samples as surrogates for original samples is contraindicated by comparisons between archived and original

sample results with respect to the available data, i.e. qPCR results for *Enterococcus* and marine *Bacteroidales*. Not only are large, highly significant declines in qPCR-CCE recoveries observed, but the differences are largely unpredictable from sample to sample as evidenced by the low correlation between results of archived and original sample analyses. This precludes the possibility of using any sort of “adjustment factor” to rectify this change.

Given the negative results for the archived sample stability study data, results from long-term holding time study have a greatly reduced relevance. The value of the long-term holding time information would have been in supporting the inference that the equivalency or predictability of the NEEAR archived sample data could be extrapolated to other qPCR targets (most notably *E. coli* and fresh water *Bacteroidales*), but neither equivalency nor predictability were observed for any of the NEEAR study results that could be directly compared. Even if one or both of these conditions had been met, there would have been additional challenges in making such extrapolations. Results from the long-term holding time study, particularly the high correlations between held and initial sample qPCR-CCE values, indicate that archived samples may be capable of serving as surrogates for fresh samples under some circumstances, at least with a statistical correction. However, the predictive relationship observed in the long-term holding time study would need to have been extrapolated from two years out to as long as six years in order to be applicable to the archived sample stability study samples. The length of time that samples are held would logically seem to be an important factor, as is borne out by the archived data, except for enterococci results at Goddard Beach. That data showed a two-year decline in mean $\log_{10}$ (qPCR-CCE) that was more in line with the five year declines seen at Silver Beach and Washington Park Beach (Table 3). However, the long-term holding time study data

themselves, which were designed to track change over time, were not consistent among the different indicator and water matrix combinations with respect to showing progressive declines, if any at all, over time, nor were they consistent with the changes seen in the archived sample stability study Fairhope, and particularly Goddard marine beach, samples that were held for the same total amount of time of two years.

A short-coming of the archived sample data from the archived sample stability study, with respect to the interpretation of their change, is that the initial analysis of original samples and analysis of archived samples two to six years later were performed by different laboratories. Thus, the effects of change over time are confounded with effects of different analysts and equipment. These data were not specifically designed to evaluate archival effects – samples were simply archived for later use. However, even for the long-term holding time study data, time effects are confounded with any potential changes that may have occurred within the single laboratory that performed all of the qPCR analyses. An experiment designed to properly capture all of the between lab and between time variation would have to involve several labs performing both the original and archived sample qPCR analysis. The cost of doing so would be prohibitive and even then, one would have to assume that there were no systematic changes in labs and/or qPCR quality over time. By way of a rough comparison of the changes observed among the archived samples over time and the magnitude of changes that may be expected from lab-to-lab variation, some preliminary data on qPCR inter-laboratory variance are available (U.S. EPA Office of Water 2008, unpublished data; *Ad Hoc* multi-laboratory study, 2010, unpublished data). These data indicate that a two standard deviation (*i.e.*, 95% confidence level) difference on the order of about 0.5-0.7 logs might be expected between two labs analyzing the identical

sample. Most of the significant changes observed in the archived samples are outside this range. Therefore, it does not seem likely that inter-laboratory differences themselves could have accounted for all of the observed changes.

As a final point, we note that *Enterococcus* and *Bacteroides* target organism cells used for the preparation of calibrator sample filters by the two laboratories came from different sources. Cells used by EMSL originated from laboratory grown cultures and were enumerated as previously described (Haugland et al., 2005; Siefring et al., 2008). Cells used by U.S. EPA NERL were enumerated by flow cytometry and were acquired in the form of commercially available, lyophilized pellets (Bioballs™, BTF, Sydney Australia). To evaluate the comparability of target organism CCE estimates in the test samples using calibrator samples prepared from these two cell sources, target sequence recoveries from calibrator sample DNA extracts prepared by each of the laboratories were examined. Quantitative estimates of target sequence recoveries per calibrator cell were obtained by interpolating qPCR CT values from analyses of the calibrator extracts on master standard curves generated by each of the laboratories from pooled results of repeated qPCR analyses of DNA standards nominally containing from 10^1 to 4×10^4 target sequence copies per analysis. Results from these analyses indicated that the mean target sequence recoveries from the laboratory grown *Enterococcus* cells used by EMSL were approximately 2-fold higher than those from the Bioball™ *Enterococcus* cells used by U.S. EPA NERL. This apparent difference may have been related to uncertainty in the accuracy of the standard curves, particularly those used for the EMSL analyses. If real, this difference could result in a systematic bias towards relatively high CCE estimates for the archived U.S. EPA NERL test samples compared to those obtained for the EMSL samples implying that, if anything,

the actual difference between initial and archived data may be even greater than the differences in enterococci results of Table 3. In contrast, the mean target sequence recoveries from the laboratory grown *Bacteriodes* calibrator cells used by EMSL were nearly identical to those from the Bioball™ cells used by U.S. EPA NEHL as determined from the master standard curves generated by the respective laboratories, a condition that does not lead to change in Table 3 for *Bacteroidales*.

Conclusions

The archived sample stability study results from Great Lakes and marine beaches indicate that significant degradation of samples had occurred. The highly significant changes and low or absent correlation between archived and original sample analyses indicate that the archived filters cannot be used in a credible manner to establish health relationships involving *Enterococcus* qPCR or, by extension, any other indicator, pathogen, or method. Any health relationship based on data derived from the archived samples is not useful because these do not reflect data that would have been obtained from fresh samples in actual beach monitoring circumstances.

The results from the long-term holding time study, though minor in importance compared to the analysis of the archived sample stability study samples, showed small but significant differences in the 24-month holding time study for *Bacteroidales* and *E. coli* that further complement the findings of the archived sample stability study samples that differences are observed between the original and archived samples.

630

Table 1. NEEAR study archived sample characteristics

	West Beach	Huntington Beach	Silver Beach	Washington Park Beach	Edgewater Beach	Fairhope Beach	Goddard Beach
<i>Enterococcus</i> qPCR-CCE							
Total Samples	294	420	423	421	396	438	426
- Failed QC ¹	11	5	16	19	29	36	8
- Both times < DL ²	8	10	18	7	9	49	35
Samples Used	275	405	389	395	358	353	383
# samples							
- Original < DL	5	7	38	23	38	74	24
- Archived < DL	102	111	105	113	46	85	109
- Neither < DL	168	287	246	259	274	194	250
% < DL before or after archival ³	38.9	29.1	36.8	34.4	23.5	45.0	34.7
<i>Bacteroidales</i> qPCR-CCE							
Total Samples	- ⁴	-	-	-	396	438	426
- Failed QC	-	-	-	-	16	5	12
- Both times < DL	-	-	-	-	26	36	5
Samples Used	-	-	-	-	354	397	409
# samples	-	-	-	-			
- Original < DL	-	-	-	-	14	36	29
- Archived < DL	-	-	-	-	33	14	20
- Never < DL	-	-	-	-	307	347	360
% < DL before or after archival	-	-	-	-	13.3	12.6	12.0

¹ Sample Processing Control out of range of ± 3 cycle thresholds from mean. ² Detection limit (DL) of 40 cycles reached without a positive signal in analyses for both the initial and archive sample. ³ Among samples used. ⁴ *Bacteroidales* qPCR was not performed at the 4 freshwater beaches.

631

632

Table 2. NEEAR study mean $\log_{10}$ (qPCR calibrator cell equivalents) based on qPCR calibrator cell equivalents $\div 2$ for non-detects

	West Beach	Huntington Beach	Silver Beach	Washington Park Beach	Edgewater Beach	Fairhope Beach	Goddard Beach
<i>Enterococcus</i> qPCR-CCE							
- Original	2.201	2.244	1.696	1.746	2.062	1.847	2.239
- Archived	0.812	1.210	1.038	0.872	1.935	1.670	1.451
- Change	-1.389	-1.034	-0.658	-0.873	-0.128	-0.178	-0.788
<i>Bacteroidales</i> qPCR-CCE							
- Original	¹	-	-	-	2.999	3.050	3.120
- Archived	-	-	-	-	2.494	2.862	2.807
- Change	-	-	-	-	-0.505	-0.187	-0.313

¹ *Bacteroidales* qPCR was not performed at the 4 freshwater beaches.

Table 3. Maximum likelihood estimates of difference between archived and original samples

	Log10 change	P-value ¹	Equivalent % change	Pearson correlation
<i>Enterococci</i> qPCR-CCE				
West Beach	-1.68	< 0.001	-98%	0.34
Huntington Beach	-1.23	< 0.001	-94%	0.37
Silver Beach	-0.73	< 0.001	-81%	-0.04
Washington Park Beach	-1.02	< 0.001	-90%	-0.03
Edgewater Beach	-0.14	0.195	-28%	0.11
Fairhope Beach	-0.21	0.113	-38%	-0.04
Goddard Beach	-1.00	< 0.001	-90%	0.11
<i>Bacteroidales</i> qPCR-CCE				
Edgewater Beach	-0.51	< 0.001	-69%	0.52
Fairhope Beach	-0.06	0.255	-12%	0.46
Goddard Beach	-0.28	< 0.001	-47%	0.48

¹ P-values in **bold-face** indicate statistically significant differences.

Table 4. Long-term holding time study samples mean log₁₀(qPCR calibrator cell equivalents) and change from initial results (24 hr) for samples held for two years

	Mean log ₁₀ (qPCR-CCE)				2 yr net change ¹	P-value ²	Equivalent % change	Pearson correlation
	24 h	6 mo	1 yr	2 yr				
	<i>Enterococcus</i> qPCR-CCE							
Freshwater	3.82	3.93	3.77	3.73	-0.08	0.063	-17	0.94
Marine	3.34	3.61	3.45	3.50	0.16	0.002	44	0.92
	<i>Bacteroidales</i> qPCR-CCE							
Freshwater	6.49	6.23	6.32	6.27	-0.22	< 0.001	-40	0.96
Marine	6.10	6.30	5.98	6.00	-0.10	0.056	-20	0.96
	<i>E. coli</i> qPCR-CCE							
Freshwater	4.65	4.64	4.58	4.55	-0.10	0.022	-21	0.95
Marine	3.99	4.02	3.98	4.01	0.02	0.658	5	0.95

¹ Difference from 24 h to 2 yr mean log₁₀(qPCR-CCE). ² P-values in **bold-face** indicate statistically significant differences.

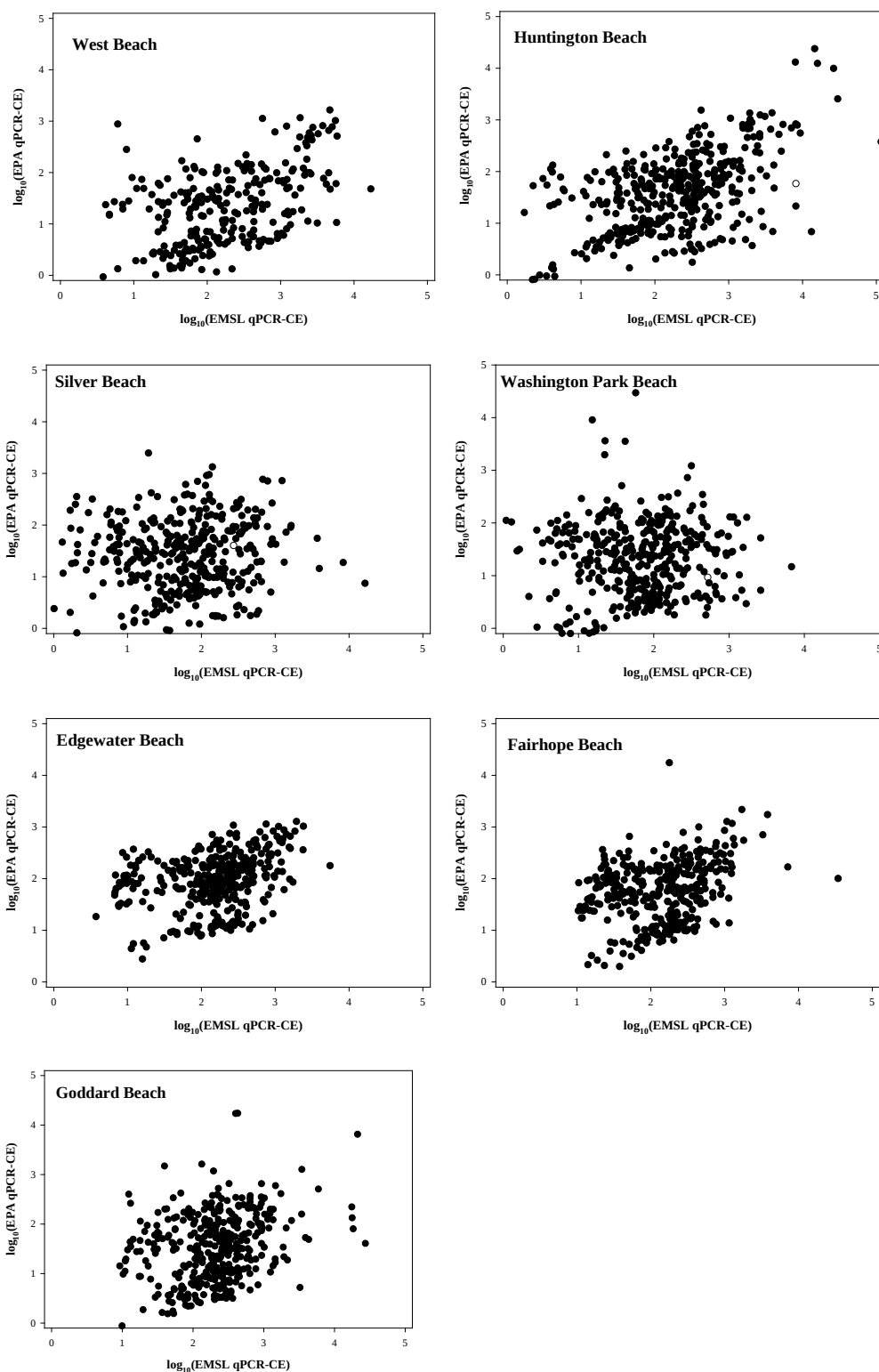


Figure 1. Scatter plots of archived sample stability study samples, *Enterococcus* qPCR archived vs. initial values.

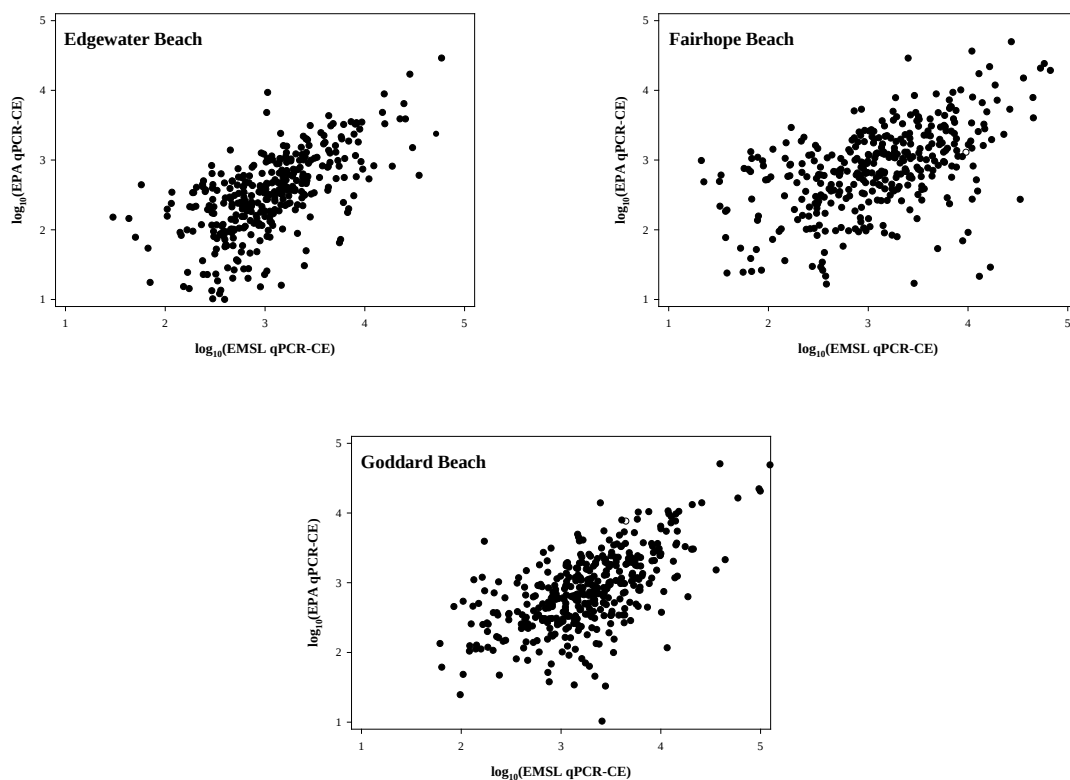


Figure 2. Scatter plots of archived sample stability study samples, *Bacteroidales* qPCR archived vs. original values.

References

Applied Biosystems. 1997. User Bulletin #2. ABI PRISM 7700 Sequence Detection System.

http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_040980.pdf.

Haugland RA, Sieftring SC, Wymer LJ, Brenner KP, Dufour AP (2005) “Comparison of *Enterococcus* measurements in freshwater at two recreational beaches by quantitative polymerase chain reaction and membrane filter culture analysis,” Water Research 39(4):559-568.

Haugland, R. A., S. Sieftring, M. Varma and L. Wymer. 2008. Laboratory evaluations of the *Enterococcus* QPCR method for recreational water quality testing: Method performance and sources of uncertainty in quantitative measurements. In Abstracts of the 108th General Meeting of the American Society for Microbiology, 2008. American Society for Microbiology, Washington, D.C.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000) WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing, 10:325--337.

Ntzoufras I. (2009) Bayesian Modeling Using WinBUGS, John Wiley & Sons, Hoboken, NJ

R Foundation for Statistical Computing (2008) R: A Language and Environment for Statistical Computing, Vienna, Austria

680

681 Siefring SC, Varma M, Atikovic E, Wymer LJ, Haugland RA. (2008) Improved real-time PCR
682 assays for the detection of fecal indicator bacteria in surface waters with different instrument and
683 reagent systems. J. Water Health 6:225-237

684

685 Wade TJ, Calderon RL, Sams E, Beach M, Brenner KP, Williams AH, Dufour AP (2006)
686 Rapidly measured indicators of recreational water quality are predictive of swimming-associated
687 gastrointestinal illness. Environ Health Perspect. Jan;114(1):24-8.

688

689 Wade TJ, Calderon RL, Brenner KP, Sams E, Beach M, Haugland R, Wymer L, Dufour AP
690 (2008) “High sensitivity of children to swimming-associated gastrointestinal illness: Results
691 using a rapid assay of recreational water quality,” Epidemiology, 19(3):375-383

692

II. Short-Term Holding Study: Evaluation of holding refrigerated water samples for up to 48 hours for qPCR analysis.

Introduction

This study was conducted to determine whether holding surface water samples under refrigeration for 24- and 48-hours affects recoveries of DNA target sequences from *Enterococcus*, *Escherichia coli*, and *Bacteroidales* bacteria as determined by a quantitative polymerase chain reaction (qPCR) method. While the qPCR method is normally intended to provide rapid determinations of fecal indicator bacteria densities in surface or recreational waters, which is contrary to the holding of water samples, instances could arise where temporary holding of water samples would be beneficial or necessary. Such instances might include when water samples are collected at remote locations where it is not possible to transport them to an analytical laboratory immediately after collection. Another example might be where it is desired to transport samples to several different laboratories for round-robin testing. The current recommended limit on water sample holding time is 6 hours for microbiological culture methods (U.S. EPA, 2000). This time limit imposes constraints on the applicability of the culture methods for reasons such as those mentioned above. Previous studies have suggested that the DNA target sequences of the qPCR method may be more stable than cultivability of the respective target organisms (Dupray et al., 1997, Walters, et al., 2009, Wery et al., 2008). Thus water samples might be held longer than the 6 hour microbiological culture method time limit without significant effects on quantitative density estimates of these organisms as determined by qPCR. The objective of the current study was to provide additional data to support this hypothesis.

Sample size (the number of different fresh water and marine water samples) used in this study was based on results from a U.S. EPA study in which filter retentates of replicate water samples

with greater than 100 enterococci cells were distributed to multiple laboratories (U.S. EPA Office of Water, 2008, unpublished data). A design of approximately 25 samples per water matrix (fresh or marine water), with 2 subsamples (filters) per sample was estimated as the required number to enable detection of a decline in target recovery by qPCR of 36% at the 0.05 critical level ($\alpha=0.05$).

Methodology

Water samples from 29 freshwater and 23 marine locations in the Boston, MA area were collected from July through October, 2008 (Table 1). Selection of freshwater and marine sampling sites was based on their ambient levels of fecal indicator bacteria (FIB) and their proximity to the laboratory that conducted the study. All locations had to be close enough to the laboratory to allow collection, transport to the laboratory and filtration of the samples within the currently recommended time limit of 6 hours for microbiological culture methods. A further consideration in the selection of the sampling sites was that historic data gathered by the laboratory and various collaborators indicated that the water samples could reasonably be expected to contain more than the estimated 95% confidence detection limit of the qPCR method of approximately 100 target organism cells per sample (Haugland et al., 2008). Actual samples giving lower mean results than 100 enterococci in initial 24 hour analyses were excluded from the study to eliminate the possibility of having subsequent results below the detection limit of the qPCR methods.

In most studies of this nature, samples would be analyzed as soon as possible (within 6 hours) to establish the initial concentration against which samples that have been held for 24 or 48 hours

would be compared. Complete analysis of the samples within six hours was not possible in the present study because of the amount of time necessary to collect samples from diverse locations during the day and transport them to the laboratory as well as the amount of time necessary to filter multiple aliquots of each water sample to support both this study and the parallel Long-Term Holding Time Study. Instead, within 6 hours of collection aliquots of each sample was filtered in the lab and the filters were stored in a freezer at -20°C to serve as recovery standards as described below.

A complete description of the treatments to which each sample was subjected follows.

1. Recovery Standard filters: Multiple 50 or 100 ml aliquots of each water sample were filtered within six hours of collection. Replicate filters from each water sample were held at -20°C for 24 and 48 hours and then extracted for DNA and analyzed by qPCR.
2. Refrigerated Water: The remainder of each of the water samples was refrigerated and then equivalent aliquot volumes to those used to prepare each of the respective Recovery Standard filters were filtered at 24 and 48 hours. After filtration, the filters were flash frozen in a -20°C cooling block for 1 hour and then immediately thawed, extracted for DNA and analyzed by qPCR. A freeze/thaw cycle was important in order to be consistent with the also frozen and thawed Recovery Standard filters.

The purpose of the Recovery Standard filters was to establish standards representing the initial densities of indicator organisms that could be recovered by the qPCR method when the water samples were filtered within six hours of collection. This study design assumed that results from

764 samples held for 6 hour or less prior to filtration could be considered to be representative of
765 freshly collected samples. This assumption was based on previous holding time studies that have
766 shown no significant losses of culturable bacteria counts over this time period (The Public Health
767 Laboratory Service Water Sub-Committee, 1953). The Recovery Standard filters were frozen
768 and stored instead of processing and analyzing the water samples on the same day that they were
769 collected (e.g. within 6 hours) in part because of the logistical and time challenges associated
770 with collecting, transporting and filtering multiple water samples, as indicated above. A second
771 important consideration in the decision to use frozen Recovery Standard filters was that this
772 procedure enabled side-by-side extraction and analysis of these filters under the same conditions
773 as the Refrigerated water samples at each holding time. Freezing the Recovery Standard filters
774 assumed that the DNA targets of the qPCR method would be preserved in this manner and thus
775 allowed these filters to be used as a basis for comparing target organism density estimates in the
776 original samples with those in the samples that had been refrigerated. By using this approach,
777 run to run variation (batch effects) can be eliminated from the analysis.

778
779 Each water sample was subjected to both of the treatments (Recovery Standard and Refrigerated
780 water) described above and each treatment was analyzed at 24 and 48 hours. Each of the 4
781 combinations of treatment and holding time (Recovery Standard and Refrigerated Water at 24
782 and 48 hours) was performed on duplicate filters and the DNA extracts from each filter were
783 analyzed via qPCR in duplicate, giving a total of 12 subsamples and 24 qPCR analyses for each
784 sample. QPCR assays for *Enterococcus* (Enter01), general *Bacteroidales* (GenBac3), *E. coli*
785 (EPA-EC23S), and spiked salmon sperm as sample processing controls (Sketa2) were performed
786 on each subsample. Quantitative estimates of target organism calibrator cell equivalents (CCE)

in the test samples were obtained by a previously described approach (Haugland et al., 2005) that is based on the comparative cycle threshold (C_T) method (Applied Biosystems, 1997). This approach employs an arithmetic formula to determine the ratio of target sequence quantities in DNA extracts from test sample filters relative to those in similarly-prepared DNA extracts from calibrator sample filters containing a known quantity of target organism cells based on the difference in C_T values obtained from qPCR analyses of these samples. Multiple replicate calibrator sample filters were prepared from a single mixed suspension containing pre-determined cell quantities of representative target strains of each of the assays (*E. faecalis*, *B. thetaiotaomicron* and *E. coli*) just prior to the study. These calibrator filter samples were held at -20° C and three filters were extracted and analyzed each week of the study. Similar comparisons of C_T values from qPCR assays for an exogenous target sequence from salmon sperm DNA, added in equal quantities to both the test and calibrator sample filters before DNA extraction, were used either as a reference to normalize results for differences in the amount of total DNA recovered from each sample (*e.g.*, caused by test sample effects on DNA recovery) or as a sample processing control (SPC) to signal potentially non-quantifiable test sample results caused by PCR inhibition or low DNA recoveries. The calculation can be expressed by the following equations:

$$[1] \Delta\Delta C_T = \Delta C_{T,\text{target}} - \Delta C_{T,\text{reference}}$$

$$[2] \text{CCE} = N_{\text{calibrator}} \cdot A^{-\Delta\Delta C_T}$$

in which $\Delta C_{T,target}$ represents $C_{T,sample} - \text{Mean } C_{T,calibrator}$ for the target sequence (*e.g.*, Enterococci); $\Delta C_{T,reference}$ represents the corresponding difference for the salmon sperm target sequence; $N_{calibrator}$ is the known number of cells in the calibrator sample; and A is the amplification factor for the assay. Ideally $A=2$ but typically it is in the range 1.9 – 2.0 with values less than 2 resulting from less than 100% replication of the target sequence at each cycle. In practice, A is either assumed to be 2 or is estimated based on the slope of a standard curve (Applied Biosystems, 1997). In this study slope values were obtained from standard curves generated from pooled results of repeated qPCR analyses of serially diluted DNA standards nominally containing target sequences in a range from 10 to 4×10^4 copies per analysis. It is noted that while quantitative estimates of target organism densities in the samples were calculated by the comparative cycle threshold method and are reported as CCE as described above, the role of the calibrator sample CT measurements had no influence on the comparisons between Recovery Standard and refrigerated water samples in this study. Common calibrator sample CT measurements were used in the calculations for the two sets of samples in all instances.

Statistical analysis was performed using a linear mixed model on logarithm (base 10) of the number of CCE as calculated by $\Delta\Delta CT$. Samples were treated as random. Filters, which were performed in duplicate for each sample, were treated as another random factor. Holding time (24 and 48 hours), water type (freshwater and marine water), and treatment (Recovery Standard, Refrigerated Water) comprise fixed, controllable, effects. Of particular interest are comparisons between Recovery Standard and Refrigerated water at each holding time as the estimator of the respective holding time effect.

Given that it was not possible to complete analyses of samples described above within 6 hours, a separate portion of the study was conducted for the purpose of obtaining data from a more traditional design in which a sample was analyzed immediately after holding times of 1, 24 and 48 hours without freezing. In February, 2009, a sample of Ohio River water was collected in Cincinnati. The sample was refrigerated and subsequently 12 aliquots were filtered each at 1 hour, 24 hours and 48 hours and the extracts from each filter were analyzed in duplicate. Identical assays, methods and calculations to those described above were performed and used in evaluating the results. This portion of the study also provided data on variability among replicate aliquots from the same sample that was used in evaluating the importance of any variability introduced by sample holding time, but does not account for run to run or “batch” variation in the analysis.

Results

The samples described in Table 1 were analyzed by the three qPCR assays and the results are given in Figures 1 and 2 and Table 2. Figures 1 and 2 show box and whisker plots of the *difference* in $\log_{10}$ (qPCR calibrator cell equivalents per 100 mL) between refrigerated water and their respective Recovery Standards at 24 and 48 hours for freshwater and marine samples, respectively. The box indicates the interquartile range, wherein 50% of the samples lie, and the horizontal line indicates the median value for the respective difference. “Whiskers” indicate the range of the 90th percentile. Data lying outside the 90th percentile range are individually plotted. Table 2 shows that there are small but significant mean differences between Refrigerated water

and respective Recovery Standards for enterococci and *Bacteroidales* assays in freshwater but not in marine water after a 24, but not after a 48 hour holding time.

These data showed a lack of significant “interaction effects” between water type (freshwater/marine water) and treatment (Recovery Standard/Refrigerated water) regardless of organism or holding time as shown in Table 3. This suggests that the effect of holding a sample on qPCR results may be the same regardless of whether it is a freshwater or marine water sample. Comparing data combined from both water types continued to show small but significantly lower recoveries in Refrigerated water samples held for 24 hours for the Enterol and GenBac3 assays (Table 2). After 48 hours holding time, on the other hand, none of Refrigerated Water samples, using combined freshwater and marine results, were seen to be significantly different from the frozen and presumably fixed, Recovery Standard samples. As Table 2 shows, Enterol and GenBac3 recoveries among Recovery Standards were slightly lower at 48 hours compared to the 24 hour analyses, while Refrigerated Water samples remained fairly constant.

Table 4 shows the data from the Ohio River sample. These data show holding time effects at 48 hours for all assays and at 24 hours for the GenBac3 assay (Table 4).

Discussion

The results from the samples described in Table 1 indicated small but statistically significant lower recoveries of qPCR targets from two of the three target organism groups among

Refrigerated Water samples held for 24 hours based on comparison with their respective Recovery Standards. Ninety five percent confidence intervals for these differences between the held Refrigerated Water and their Recovery Standards among both freshwater and marine water samples combined are shown in Table 5. These are further interpreted in terms of equivalent percentage differences simply by taking the antilogarithms of the differences in $\log_{10}(\text{cell equivalents per 100 ml})$. This is the corresponding percentage change in average (*geometric mean*) recovery by qPCR over 24 or 48 hours of holding time.

At 24 hours, geometric mean recoveries for both the Enterol and GenBac3 assays were 12-13% lower in the Refrigerated Water samples than in the Recovery Standards, with a potential range of up to 28% lower. For EPA-EC23S the difference amounted to 3%, possibly as high as 13%. Near parity between Refrigerated Water and their Recovery Standards for EPA-EC23S was indicated by the low percent differences and the fact that their respective 95% confidence intervals bracket zero.

The use of Recovery Standards for these samples was an approach to evaluating holding time effects on microorganisms in water samples that would not be available for culture based methods because samples for culture can not be preserved through freezing like DNA. By controlling predation, chemical reactions, or other factors that might degrade DNA in a water sample, the target DNA in organisms deposited on a filter and subsequently frozen were assumed to persist without any losses for at least the short holding times involved in this investigation. Thus, the Recovery Standards acted as a control group against which the Refrigerated Water analyses were compared in a side-by-side manner for the 24 and 48 hour time points. Differences

in the qPCR results of the two sets of samples using this side-by-side analysis approach was anticipated to minimize any potential influences of method-related “batch” effects, i.e. variability between the results of two sets of samples associated with their being extracted and analyzed at different times. A potential illustration of this benefit can be seen in comparisons of the mean CCE densities estimated in the Recovery Standards at 24 and 48 hours in Table 2. Although the overall mean CCE estimates obtained by the *E. coli* EPA-EC23S assay remained constant, the overall mean CCE estimates obtained by the Enterobac 1 and GenBac3 assays differed appreciably between 24 and 48 hours. While not found to be statistically significant, the latter differences may be indicative of the aforementioned “batch” effects. A similar difference was observed in the 24 and 48 hour Refrigerated Water results obtained by the *E. coli* EPA-EC23S assay. In this instance an unexpected increase in the 48 hour CCE densities was observed, although the possibility of growth by indicator organisms in stored water samples can not be completely ruled out.

Despite the efforts taken to eliminate as many method-related variables as possible that might confound the results of this water sample holding time study, the observation of going from a significant difference between Refrigerated Water and Recovery Standard results at 24 hours to no significant difference at 48 hours was unexpected. It is noted that the net change in Recovery Standards between 24 and 48 hours was not significant, even though the differences themselves go from being significant at 24 hours to non-significance at 48 hours. As always, lack of statistical significance is not convincing evidence for no difference, but only lack of convincing evidence for a difference.

Nevertheless, it is possible that there are variables still unaccounted for in this study. An effect associated with the freezing of the Recovery Standard samples could be such a variable. While an attempt was made to control for the potential influence of sample freezing on target DNA recovery in the qPCR method by also flash-freezing the Refrigerated Water filters prior to extraction, it could only be assumed that these two freezing methods had the same net effects on DNA recovery. It was also necessary to assume that there were no differential effects of holding frozen Recovery Standard samples for 24 vs. 48 hours. This uncertainty associated with the potential effects of sample freezing in this portion of the study was part of the rationale for also conducting the supplemental Ohio River water portion of the study where no freezing of the samples was involved. It is noted, however, that while the Ohio River water portion eliminated freezing effects as a variable, it reintroduced the method-related variables indicated above.

The experimental design used in the Ohio River sample allowed for the comparison of recoveries for 24 and 48 hours based on recoveries at 1 hour which may be a more standard approach for comparing recoveries over time. However this analysis approach created the potential to introduce greater uncertainties in comparing time point results than with the samples from Massachusetts. The uncertainties are associated with extracting and analyzing the different holding time samples in different batches. The results from the Massachusetts samples indicate that there was run to run or batch effects but they were accounted for in the study design. These effects were not considered in the Ohio River sample analysis. The Ohio River results indicate that different conclusions may be reached if run to run or batch effects are not a component of the analysis.

To put the results of this study in further perspective, any bias as a result of holding a sample should be compared to the difference that might be expected among different aliquots from that sample. In practice only a single aliquot most likely would be drawn from the sample for analysis. A 12% difference in recovery, such as shown in Table 5 for the Enterol and Genbac3 assays at 24 hours, might be important if the range of results that could be reasonably expected from the “luck of the draw” is, for example, $\pm 10\%$, but of little importance if the range is more like $\pm 50\%$. Representative data that could be used for determining such differences in the analyses of a single aliquot were available from the Ohio River study, where twelve aliquots were taken for analysis at each time point. The variances in the indicator density estimates from the twelve aliquots that were taken just one hour after collection were used as a basis for comparisons with the potential biases that could be attributable to holding time effects.

To use the Ohio river data for evaluating a holding-time bias, the concept of mean square error (MSE) can be used (Cochran, 1963). MSE is the average of the squared deviations of “all possible” results from the “true” value of what is being measured (the overall mean log qPCR-CE of the water sample at or near the time of collection). If there is no bias (*i.e.*, the sample is analyzed immediately), the MSE is obviously the same as the variance of $\log_{10}(\text{qPCR-CE})$. If there is a bias (*e.g.*, due to holding time degradation) the MSE is equal to this variance plus the square of the bias. The square root of the MSE is the root mean square error (RMSE), the same as the standard deviation for the unbiased result and larger than this, influenced by the size of the bias, for a biased result. Dividing MSE by variance gives the MSE relative to pure variance, and the square root of this (minus 1), the relative increase in RMSE due to bias. Based on the observed sample aliquots variances for the respective targets from the 1-hour $\log_{10}$ variances

pooled over subsamples of Ohio River water, the last section of Table 5 shows the relative increase in RMSE corresponding to biases represented by the Refrigerated Water minus respective Recovery Standard $\log_{10}$ differences after 24 and 48 hours of holding time. These range from nil up to an 8% increase in RMSE. At the 95% confidence level, the increase in RMSE amounts to less than 17%.

This indicates that the potential difference in mean log recovery of enterococci and *Bacteroidales* in held samples may not be an important factor in relation to normal sampling variation. Note, however, that while sample variability can be compensated for by the collection and analysis of additional sample replicates, bias cannot.

Summary

- Holding time effects did not appear to depend on whether samples were from freshwater or marine sources.
- Significantly lower recoveries of CCE were realized after 24 hours of holding time for both the Enterol and GenBac3 assays in Refrigerated Water samples compared to the corresponding Recovery Standards. Observed qPCR recoveries determined by the two assays were about 12% and 13% lower, respectively, in the Refrigerated Water samples compared to the Recovery Standards after 24 hours, with a potential range of up to 22%.
- Recovery for the *E. coli* (EPA-EC23S) assay, while a net of 3% lower, was not significantly different in the Refrigerated Water samples at 24 hours compared to the Recovery Standards.

- There were no significant net losses in recovery after 48 hours of holding time in the Refrigerated Water samples compared to the corresponding Recovery Standards for any of the three bacterial indicator assays. This reversal of observed effect compared to the 24 hour results was a result of a small decline in recovery among the two sets of Recovery Standards, which simply may be a method-related “batch” effect.
- Supplemental data from the Ohio River water portion of the study showed significant decreases in qPCR estimated indicator densities as determined by two of the three assays (GenBac3, and EPA-EC23S) after an Ohio River sample was held for 24 hours and significant decreases as determined by all 3 assays (Enterol, GenBac3, and EPA-EC23S) in the sample held for 48 hours, as compared to when the sample was held for only one hour.
- Any bias due to holding time amounted to 8% (less than 17% at the 95% confidence level) of the total variation in results (actually the root mean square error, RMSE, roughly equivalent to the standard deviation) among replicate aliquots.

Conclusions

There were small, and in some cases statistically significant changes in qPCR-estimated fecal indicator densities in Refrigerated Water samples that were held for 24 and 48 hours. However, because these changes were neither consistent over time nor were they consistent between indicators, no conclusions on the effects of holding refrigerated water samples can be drawn from the Short-Term Holding Time Study. Significant declines in qPCR signals at 24 and 48 hours were observed in the Ohio River sample portion of the study, but these observations were

1016 based on a single sample and did not consider potential influences of certain method-related
1017 sources of uncertainty. Taken together, the results of the two studies are inconclusive and, as a
1018 result, **can neither support nor refute** recommendations to hold refrigerated water samples for
1019 24 hours or longer.

1020

Table 1. Sample locations, description

ID	Date	
	Collected	Source
Freshwater samples		
1-5	08/05/08	Charles River, Stony Brook Tributary
6-8	08/11/08	Charles River, Boston Water & Sewer Outfalls
9-11	08/13/08	Charles River, Muddy River Outfalls
22, 24-26	08/25/08	Winns Brook
30	08/27/08	Newtown
36	09/03/08	Salem Sound
37	09/08/08	Ell Pond, Melrose
38	09/08/08	Mill Brook, Melrose
39, 40	09/08/08	Lower Mystic Lake, Arlington
41	09/08/08	Spy Pond, Arlington
52, 53	09/22/08	York Beach, ME, River Rd, Sewage Spiked
58	09/25/08	Concord, NH WWTF, pre-UV
59	09/25/08	Concord, NH WWTF, post-UV
61	09/25/08	Outfall at Wollaston Beach, Sewage Spiked
62	09/25/08	Furnace Brook, Sewage Spiked
63	10/08/08	Lowell WWTF (chlorinated)
Marine samples		
14-15	08/18/08	Mystic River, Marine Side of Dam
17-21	08/20/08	Mill Creek, Chelsea
27, 28	08/27/08	E. Boston
29	08/27/08	Revere
32-36	09/03/08	Salem Sound
43, 44	09/15/08	Kings Beach, Lynn
45	09/15/08	Fisheries Beach, Lynn
54-57	09/22/08	York Beach, ME, Clark Rd
60	09/25/08	Wollaston Beach, Sewage Spiked

1021

1022

1023

Table 2	Holding time effects: qPCR recovery comparisons between Refrigerated Water (RW) and the Recovery Standards (RS)				
Fresh/ Marine	Holding time (h)	Mean log Calibrator Cell Equivalents			P-value
		RS	RW	Difference	
		Enterol			
Combined	24	3.58	3.52	-0.06	0.042
	48	3.51	3.53	0.02	0.457
F	24	3.82	3.74	-0.08	0.028
	48	3.75	3.71	-0.03	0.356
M	24	3.34	3.31	-0.03	0.460
	48	3.28	3.35	0.07	0.132
		GenBac3			
Combined	24	6.29	6.23	-0.06	0.023
	48	6.23	6.26	0.02	0.380
F	24	6.49	6.38	-0.11	0.001
	48	6.41	6.41	0.00	0.892
M	24	6.10	6.09	-0.01	0.887
	48	6.06	6.10	0.04	0.289
		EPA-EC23S			
Combined	24	4.32	4.31	-0.01	0.577
	48	4.32	4.36	0.03	0.159
F	24	4.65	4.62	-0.03	0.276
	48	4.60	4.65	0.05	0.111
M	24	3.98	3.99	0.01	0.823
	48	4.04	4.06	0.02	0.641

1024

Table 3 Evaluation of difference in holding time effects between fresh and marine waters

	"Interaction" P-value ¹	
	24 hours	48 hours
Enterol	0.53	0.08
GenBac3	0.06	0.65
EPA-EC23S	0.57	0.25

¹ Test of difference between fresh and marine waters with respect to holding time effect on recovery.

Table 4 Ohio River holding time effects: qPCR recovery comparisons

Holding time (h)	Mean log CCE	Difference from 1 hour	P-value
Enterol			
1	2.98	-	
24	2.79	-0.19	0.073
48	2.74	-0.24	0.029
GenBac3			
1	4.59	-	
24	4.32	-0.27	0.024
48	4.26	-0.33	0.007
EPA-EC23S			
1	3.12	-	
24	2.87	-0.25	0.080
48	2.69	-0.43	0.004

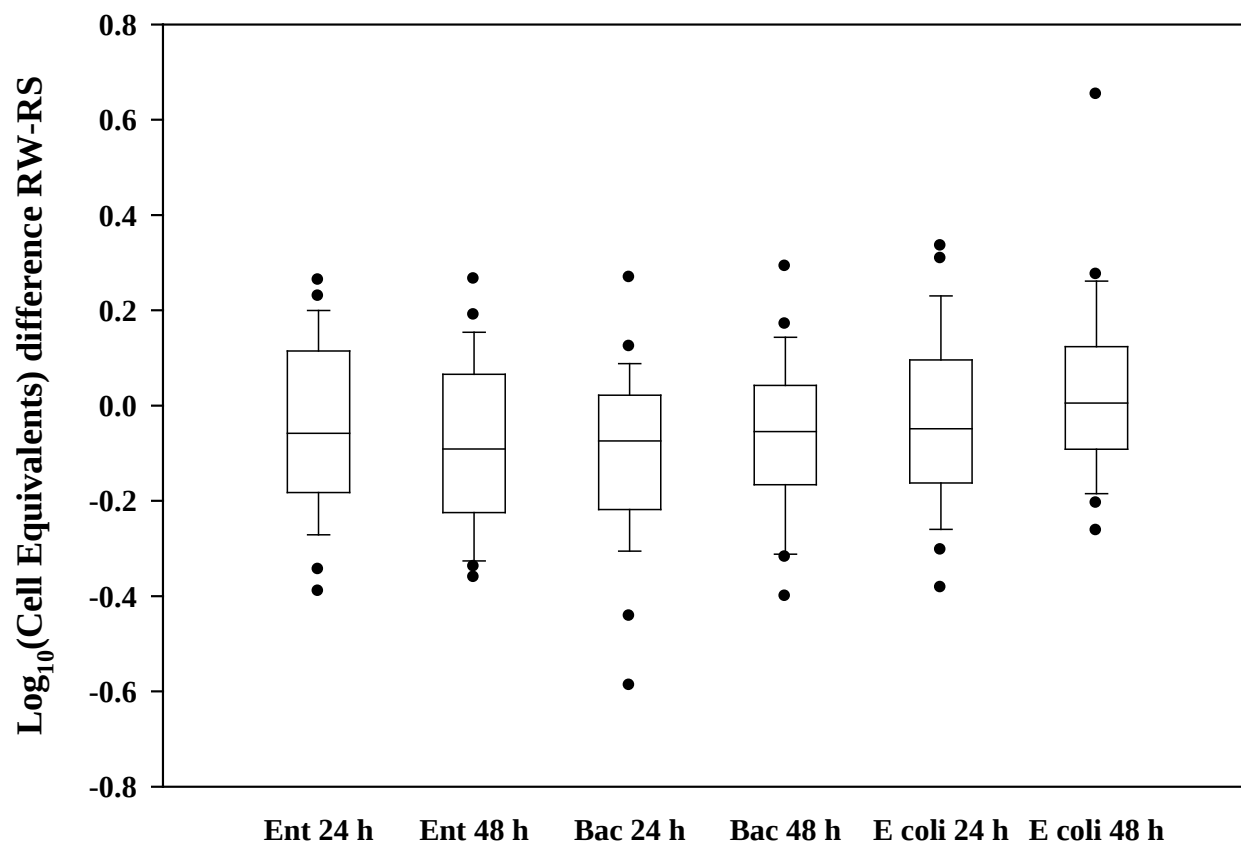
Table 5 Holding time effects: qPCR recovery comparisons
 between Refrigerated Water (RW) held for 24 or 48 hours
 and respective Recovery Standards (RS)

	24 hours		48 hours	
	RW-RS	95% CI	RW-RS	95% CI
Difference in log ₁₀ (CCE)				
Enterol	-0.06	(-0.11, -0.00)	0.02	(-0.03, +0.07)
GenBac3	-0.06	(-0.11, -0.01)	0.02	(-0.03, +0.07)
EPA-EC23S	-0.01	(-0.06, +0.03)	0.03	(-0.01, +0.08)
Difference as a percent of geometric means				
Enterol	-12%	(-22%, -0%)	+5%	(-7%, +18%)
GenBac3	-13%	(-22%, -2%)	+5%	(-6%, +18%)
EPA-EC23S	-3%	(-13%, +8%)	+8%	(-3%, +21%)
Relative increase in root mean square error				
Enterol	1%	(< 5%)	0.2%	(< 2%)
GenBac3	3%	(< 10%)	0.4%	(< 5%)
EPA-EC23S	0%	(< 2%)	0.6%	(< 4%)

RW-RS: Refrigerated Water minus Recovery Standard (log₁₀ difference)

1038

Figure 1. Freshwater box and whisker plot¹ of differences in $\log_{10}$ (Calibrator Cell Equivalents per 100 ml): Refrigerated Water – Recovery Standards

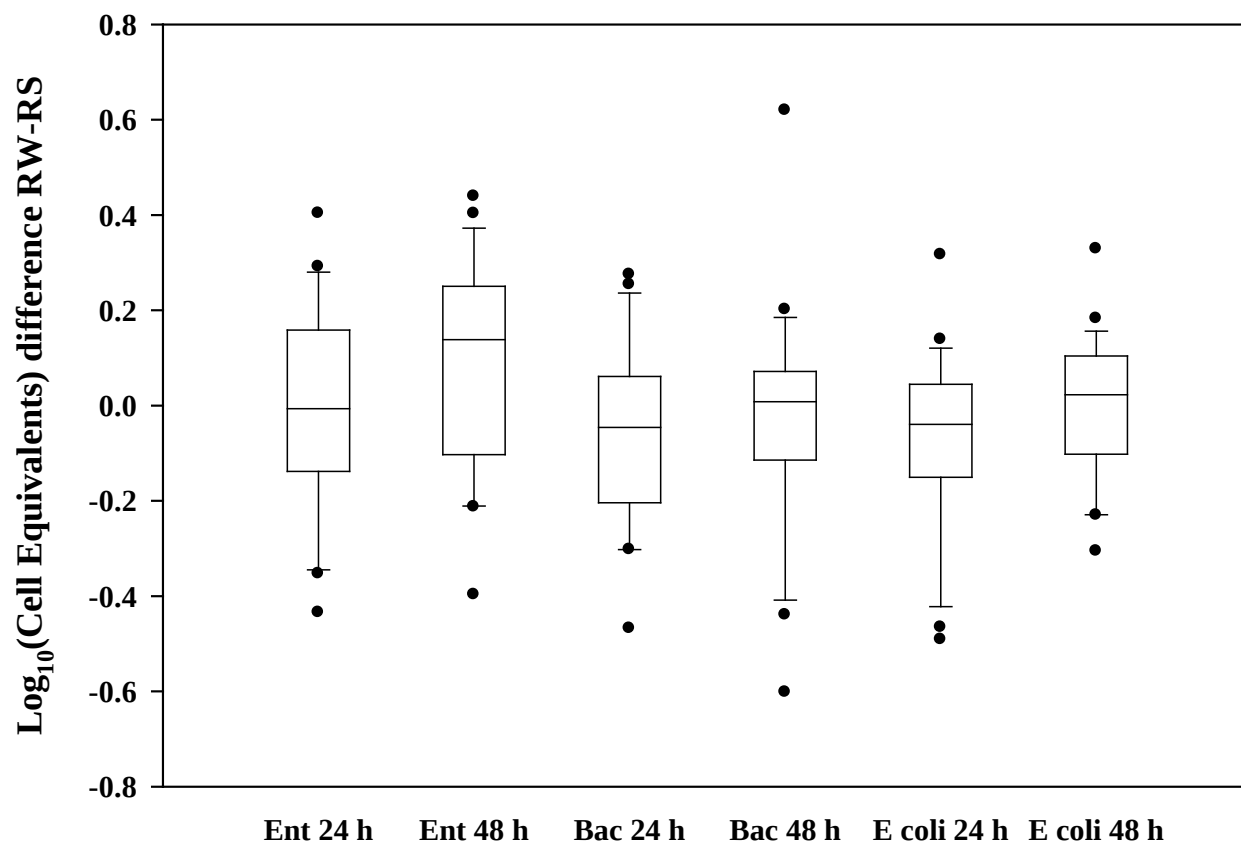


¹ Box shows 50% range, “whiskers” show 90% range, individual points shown outside this range. Median value is indicated within each box.

1039
 1040
 1041
 1042
 1043

1044

Figure 2. Freshwater box and whisker plot¹ of differences in $\log_{10}$ (Calibrator Cell Equivalents per 100 ml): Refrigerated Water – Recovery Standards



¹ Box shows 50% range, “whiskers” show 90% range, individual points shown outside this range. Median value is indicated within each box.

1045
 1046
 1047
 1048
 1049

References

Applied Biosystems. 1997. User Bulletin #2. ABI PRISM 7700 Sequence Detection System.

http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_040980.pdf..

Cochran WG (1963) Sampling Techniques, John Wiley & Sons, New York

Duprey, E., Caprais, M.P., Derrien, A, Fach, P. 1997. *Salmonella* DNA persistence in natural seawaters using PCR analysis. *Journal Applied Microbiology* **82**, 507-510.

Haugland RA, Sieftring SC, Wymer LJ, Brenner KP, Dufour AP. (2005) “Comparison of *Enterococcus* measurements in freshwater at two recreational beaches by quantitative polymerase chain reaction and membrane filter culture analysis,” *Water Research*, 39(4):559–568.

Haugland, R. A., S. Sieftring, M. Varma and L. Wymer. 2008. Laboratory evaluations of the *Enterococcus* QPCR method for recreational water quality testing: Method performance and sources of uncertainty in quantitative measurements. *In* Abstracts of the 108th General Meeting of the American Society for Microbiology, 2008. American Society for Microbiology, Washington, D.C.

1073 The Public Health Laboratory Service Water Sub-Committee (1953) “The effect of storage on
1074 the coliform and bacterium coli counts of water samples. Storage for six hours at room and
1075 refrigerator temperatures,” *The Journal of Hygiene*, 51(4): 559-571
1076
1077 U.S. EPA. 2000. Improved Enumeration Methods for the Recreational Water Quality Indicators:
1078 Enterococci and *Escherichia coli*. Office of Science and Technology, Washington, DC. EPA-
1079 821/R-97/004.
1080
1081 Walters, S,P,, Yamahara, K.M., Boehm, A,B. (2009) Persistence of nucleic acid markers of
1082 health-relevant organisms in seawater microcosms: implications for their use in assessing risk in
1083 recreational waters. *Water Res.* 43, 4929-4939.
1084
1085 Wery, N., Lhoutellier, C., Ducray, F., Delgenes, J-P., Godon, J-J. 2008. Behaviour of pathogenic
1086 and indicator bacteria during urban wastewater treatment and sludge composting, as revealed by
1087 quantitative PCR. *Water Res.* 42, 53-62.
1088
1089