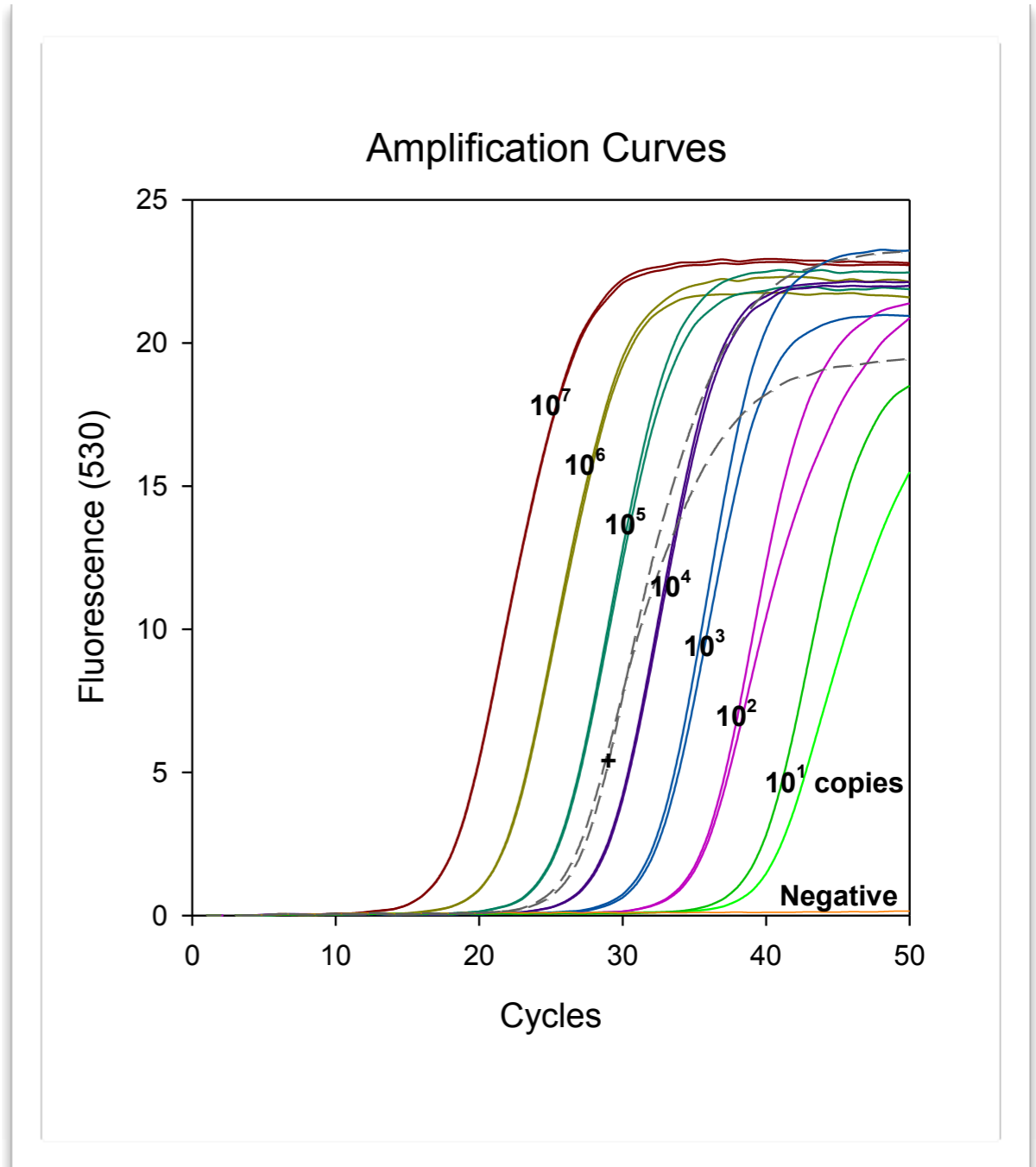


# Data Assessment

This chapter describes various methods for assessment of the results, data handling and reporting.



# Calculations, data handling and reporting

## CONTENTS

---

7.1. Data Analysis and Calculations

7.2. Data Handling

7.3. Reporting the Results

## 7.1. Data Analysis and Calculations

The complete procedures for determining target sequence quantities in water samples are described below.

### *7.1.1. Calculating efficiency and amplification factor from the genomic standard curve*

Genomic DNA standards are used to generate standard curves for the determination of performance characteristics of the qPCR assays and instruments with different preparations of master mixes containing TaqMan® reagent, primers and probes. They may also be used for quantifying target sequences in calibrator sample extracts. A target sequence is the region of the gene characteristic to the organism you are trying to identify, in this case *Enterococcus*, and which differentiates it from other organisms (assay specificity). The instructions for creating a genomic DNA standard curve were previously described in Chapter 4.

To calculate the efficiency, you will need the Ct values of three replicate serial dilutions of the genomic DNA (e.g.  $4 \times 10^4$ ,  $4 \times 10^3$ ,  $4 \times 10^2$ ,  $2 \times 10^2$  and  $1 \times 10^2$  lsrRNA gene sequences per 5  $\mu\text{L}$ ) as described in Chapter 4.

Log transform the concentrations. Then plot Ct values vs these log transformed concentrations. Perform a multiple lin-

ear regression. The  $r^2$  reflects the linearity of the standard curve, therefore indicates how precisely the line fits the data.

The next step is to find out the efficiency of the reaction. The slope of the regression equation equates to the efficiency of the reaction. The efficiency should be as close to 100% as possible, which is equivalent to a slope of -3.32.

You can calculate the percentage using the following formula:

$$\text{Efficiency\%} = (10^{(-1/(-\text{slope})-1)} \times 100\%$$

For example if the slope is -3.3276 then

$$\text{Efficiency\%} = (10^{(-1/(-3.3276)-1)} \times 100\% = (1.99763845-1) \times 100\% = 0.99763845 \times 100\% = 99.763845\%$$

$$\text{Efficiency \%} = \sim 99.8$$

Once the slope is determined, the amplification factor can also be calculated. The amplification factor (AF) is a measure of the average efficiency at which the DNA target or SPC sequences are copied and detected by their respective primer and probe assays during each cycle of the qPCR reaction. It is used in comparative cycle threshold calculation methods like ddCT. AF values can range from 1 (0 % of sequences copied and detected) to 2 (100 % of sequences copied and detected) and are calculated from the standard curve.

Calculate the amplification factor using the equation below:

$$\text{AF} = 10^{(1 / (-)\text{slope value})}$$

For example, if the slope is 3.4777, the amplification factor will be:

$$\text{AF} = 10^{(1 / 3.4777)} = 1.94$$

### *7.1.2. Calculation of average target sequence in calibrator sample extracts per reaction*

The average target sequence in calibrator sample extracts is calculated from the genomic DNA standard curve and the CT value of the calibrators used during the run according to the formula below:

$$10^{((\text{average calibrator extract value} - \text{intercept of the genomic DNA standard curve}) / \text{slope of the genomic DNA standard curve})}$$

For example:

If the regression equation of the genomic DNA standard curve is:  $y = -3.477x + 38.42$

and

If the average calibrator extract Ct value obtained during the qPCR analysis of your samples is 25.21;

then the average calibrator target sequence is:

$$10^{((25.21-38.44) / -3.477)} = 6382$$

Assuming we used 1:5 dilution for samples and calibrators, the total extract volume would be: 600ul x5=3000 ul.

600 ul comes from the 590 ul Salmon DNA extraction buffer+10 ul calibrator or sample (See Chapter 6).

Now we can calculate the average calibrator target sequences per reaction by using the formula:

$$\frac{\text{Average target sequences}}{\text{Calibrator extract}} = \frac{6382 \text{ target sequences}}{5 \mu\text{L extract}} \times 3000 \mu\text{L total extract volume}$$

$$= 3,829,619$$

### 7.1.3. Calculation of CCE /100ml

For those entities interested in adopting a value for enterococci using US EPA's Enterococcus qPCR Method 1611 into their Water Quality Standards, the US EPA recommends a geometric mean (GM) criterion of 470 CCE per 100 ml; a statistical threshold value (STV) criterion of 2000 CCE/100 ml and an optional beach action value (BAV) of 1,000 CCE per 100 ml in both fresh and marine recreational waters based on its epidemiological study data.

To calculate CCEs you will need:

- $\Delta\text{Ct}$  (CE)- the difference between the Ct values of your SPC and your unknown Ct value
- $\Delta\Delta\text{Ct}$  (CCE)-the sum of the sample  $\Delta\text{Ct}$  minus the calibrator dCT
- AF (from Section 7.1.1)

1. To obtain  $\Delta\text{Ct}$  values for each of these test samples, use the following formula:

- $\Delta\text{Ct}_{\text{calibrator}} = \text{Ct}_{\text{calibrator target}} - \text{Ct}_{\text{calibrator SPC}}$

2. Then calculate the average  $\Delta\text{Ct}_{\text{calibrator}}$  for calibrator replicates.

3. Calculate the average  $\Delta\text{Ct}_{\text{sample}}$  for each sample replicate using the formula:

- $\Delta\text{Ct}_{\text{sample}} = \text{Ct}_{\text{sample target}} - \text{Ct}_{\text{sample SPC}}$

4. Calculate the average  $\Delta\text{Ct}_{\text{sample}}$  for sample replicates

5. Calculate  $\Delta\Delta\text{Ct}$  by subtracting average  $\Delta\text{Ct}_{\text{calibrator}}$  (2) from average  $\Delta\text{Ct}_{\text{sample}}$ (4)

6. Calculate the ratio of target sequences in the calibrator to the sample using the following formula:

- Ratio of calibrator:sample=  $AF^{(-\Delta\Delta Ct)}$

7. Now you can calculate CCEs using the formula:

$$CCE = [AF^{(-\Delta\Delta Ct)} \times \text{Calibrator cells}/100\text{mL}] \times [100\text{mL}/\text{filtration volume (mL)}]$$

Example:

If  $AF=1.96$ ; the estimated cell count is 1200000; average Ct calibrator target is 25.71 and average Ct calibrator SPC is 34.13;

Estimated Cell Count <sup>1</sup>	Sample Type	C <sub>T,target</sub>	C <sub>T,SPC</sub>	ΔC <sub>T</sub>	ΔΔC <sub>T</sub>	Ratio calib/test $1.96^{-\Delta\Delta Ct}$	CCEs in Test Sample Extract ( $1.96^{-\Delta\Delta Ct} \times \text{Est cell count}$ )
1200000	Calibrator	25.71	34.13	-8.42	---		---
	Test	33.74	25.94	7.8	7.59	0.006	$0.006 \times 1200000 = 7264.74$

This calculation can be applied without modification to the analyses of diluted extracts if the same dilution factor was applied to both test sample and calibrator extracts (e.g. 1:5 for sample and calibrator both) and the same volume (e.g. 600 ul) of diluted extracts were analyzed.

The above calculations suppose that a 100 ml of sample was filtered through the polycarbonate filter. If sample was diluted, multiply the result by the dilution factor.

#### 7.1.4. Descriptive statistics of results

The geometric mean of the measured target sequences and associated coefficients of variation in multiple water samples can be determined from individual sample Ct values using the following procedure:

- Use ΔCt value for each individual water sample extract and the mean calibrator ΔCt value to calculate the measured target sequence numbers in each water sample extract.
- Calculate the log<sub>10</sub> of the measured target sequence numbers in each water sample (log N)
- Calculate the mean (M) and standard deviation (S) from the values of log N obtained in the previous step for all of the water sample extracts.
- Calculate the geometric mean as  $10^M$ .
- The implied coefficient of variation (CV) is calculated, based on the log normal distribution, as the square root of  $10^{V/0.434} - 1$ , where  $V = S^2$ .

## 7.2. Data Handling

### 7.2.1. Replicate pair sample agreement

As previously stated in various sections of this guidance document, it is recommended that you run each qPCR reaction in duplicate or triplicate. The suggested frequency is dependent on the type of sample you are analyzing (calibrator/standard, control or sample) and you should refer to the appropriate section for further information.

The output of your qPCR analysis will not be a concentration of the target. Rather the thermocycler generates results in units called cycle thresholds (Ct). A cycle is the unit of time for one complete PCR reaction to occur, i.e. double stranded DNA denaturation, annealing of the probe and primers to the single stranded DNA, and the extension/formation of new double stranded DNA. Method 1611 is comprised of 45 PCR cycles. The concentration of target in your calibrator, control, or unknown sample is directly proportional to the cycle at which fluorescence is first detected. This is an inverse relationship, with higher concentrations of target producing lower Ct values. Each log difference is approximately equal to a three Ct spread, e.g. a sample with a Ct of 27 would have approximately one log greater concentration of target than a sample with a Ct value of 30. A sample with a Ct value of 45 is considered a non-detection of target, i.e. no measureable fluorescence occurred within 45 PCR cycles, or during the entire duration of the assay. This does not mean that there is no target present in your sample, only that it was not detectable within

the confines of the assay's limit of detection (sensitivity). This value could represent a true negative but could also represent assay failure, e.g. inhibition, which will be discussed later in this section.

Therefore, the initial action following the completion of the assay is to examine the returned Ct values of your replicate analyses (remember that it is recommended that you run each reaction in duplicate or triplicate). Replicate analyses, from a single DNA extract, should agree within 1.0 Ct. If there is more than a 1.0 Ct spread, the assay should be repeated. Replicate reaction agreement is a good measure of precision and frequently a direct result of the analyst's ability to pipet in a manner that generates reproducible results. If an analyst has consistent problems with replicate pair mismatches, pipetting drills may assist in improving method performance. If your lab does not have an SOP for this procedure, there are several examples available online; e.g.

<http://www.johnmorris.com.au/files/files/PDFs/Gilson/Verification%20Procedure%20for%20Accuracy%20and%20Precision.pdf>.

If your replicate pairs are in agreement the next step in data review is to examine your positive and negative control values.

### 7.2.2. Negative controls

The filter blank, method blank, and no template control are considered negative controls. As such the CT value of these

reactions should ideally be 45, e.g. a non-detection of target (enterococci) DNA. There are some exceptions however, please refer to Table 2.1. If your values do not conform to the acceptable parameters delineated in this table you likely have sample, reagent, or instrument/equipment contamination. You should repeat the reaction. If it comes back negative you can rerun your samples without additional actions. If the contamination persists you will have to summarily rule or rule out sources through stepwise substitution of reagents or consumables prior to repeating your sample analysis. A good first place to start is to thoroughly clean your workstations and equipment.

### 7.2.3. *Positive controls*

Your positive controls are the calibrator(s) and specimen processing controls (SPC). These samples have known quantities of DNA which allows you to compare the anticipated to the actual result as a measure of assay performance. The returned values of your positive controls will directly influence the quality of your analytical (reportable) results as they are used in the  $\Delta\Delta C_t$  calculations which translate  $C_t$  values into units of measure called calibrator cell equivalents (CCE). This is the value that will be reported as a measure of target per volume of surface water analyzed (typically 100 ml) in the same way as culture-based data are reported. The concentration of qPCR target, i.e. enterococci, is directly tied to regulatory compliance via epidemiological studies which linked threshold fe-

cal indicator bacteria (FIB) levels to human health outcomes (US EPA 2012).

The calibrator(s) employed as positive controls in your test runs should represent point(s) on your master standard curve (Section 4.3.6). As a general guideline, the  $C_t$  value of your daily control should not differ from the master standard curve value by more than two standard deviations (see Table 2.1). If the calibrator value does not meet acceptability criteria you should rerun the reaction. If results are acceptable you may repeat your run. If it still fails to meet criteria for acceptance, a new calibrator should be prepared. In either case all of your unknown samples in that run will need to be repeated.

The same acceptance criteria also applies to the SPC. You will remember that the SPC is a known quantity of non-target (salmon) DNA that you added to all samples (positive and negative controls as well as unknown samples) at the DNA extraction phase of Method 1611. Because you know the amount of DNA added, the SPC functions as a recovery control, i.e. you should get out what you put in. However, it also serves a second purpose as a sentinel for assay performance with respect to incomplete or frank failure to detect a DNA target.

### 7.2.4. *Incomplete target detection*

Incomplete target detection means that you are quantifying some, but not all, DNA present in your sample. This could be due to a variety of factors including: malfunction of the ther-

mal cycler, incorrect qPCR reaction mixture, poor DNA polymerase activity, or the presence of interfering substances in the sample (Råström et al. 2003). Incomplete quantification can result from competition for the oligonucleotides (primers) and Taqman™ DNA polymerase (target competition). This scenario can sometimes be seen in methods containing an internal amplification control (IAC). In contrast to a (external) positive control such as a calibrator, an IAC is a non-target DNA sequence, which is co-amplified simultaneously with the target sequence, conceptually similar to the SPC. However, unlike the SPC, an IAC will produce a signal even if there is no target sequence present. While the US EPA considered the addition of an IAC to Method 1611, it is currently not a component of this assay.

Experimental factors such as the length, secondary structure, and GC (guanine and cytosine) content of the amplicon or the dynamics of the reaction itself, the use of non-optimal reagent concentrations, and enzyme quality can result in efficiencies below 90%. These conditions may lead to inaccurate target quantification, referred to as underestimation. If this happens, samples should be rerun with a new set of reagents. If this situation arises frequently, e.g. in surface water samples containing high amounts of algal biomass, consider analyzing samples by an alternative method (not qPCR).

### 7.2.5. *Inhibition*

Inhibition equates to assay failure resulting in underestimation or complete absence of target detection. qPCR inhibition is defined as a delay in target DNA amplification and can be caused by various environmental compounds present in the sample matrix. qPCR inhibitors may include humic substances, polysaccharides, salts, acids, proteins/enzymes, ions, and alcohols. These inhibiting agents can interfere with qPCR by binding to and degrading the target DNA, interacting with Taq Polymerase and inhibiting its enzymatic activity, and/or preventing annealing of the probe and primers to target DNA template. The presence of PCR inhibitors in one or more of the reagents can also produce amplification efficiencies (AF) of greater than 110%. If inhibition is present, the qPCR results are considered invalid and no management decision on beach closure can be made based on the qPCR analysis.

The specimen processing control (SPC) is used to test for inhibition. The SPC is a known quantity of Salmon testes DNA that is added to all samples and calibrators before DNA extraction (see Section 6.2.2). To identify inhibition, compare Ct values of samples to those of the calibrator. Because the same quantity of SPC was added to all the samples and calibrator, the Ct values should be relatively close. If the sample Ct value is significantly higher than that of the calibrator, the sample is considered inhibited and qPCR results for that sample are considered invalid. Inhibition can be quantified by taking the difference between sample SPC Ct and calibrator SPC Ct. This difference is referred to as an inhibition score. Ideally, the inhibi-



tion score should be less than 2.0 Ct. By US EPA definition, a sample is considered inhibited if the inhibition score is  $\geq 3.0$  Ct.

If inhibition is found, you can try to reduce it by using one of the following techniques: serial dilution of the DNA extracts, a DNA extraction or clean/concentrate kits, and/or adding of compounds that stabilize qPCR (for example, bovine serum albumin or BSA). Serial dilution of the sample DNA extract is the fastest and cheapest resolution for qPCR inhibition but may not be useful in samples where the target concentration is projected to be low because you will effectively decrease the target to below the limit of quantification. DNA extraction kits can remove inhibitors and concentrate target, an alternative for samples with low target copy numbers, but they will increase the cost of, and time required, to perform the assay. Beach sanitary survey data may be used to identify ambient environmental conditions which are more likely to be associated with qPCR inhibition, e.g. high turbidity or antecedent precipitation. Foreknowledge will allow you appropriately pre-treat your sample prior to the initial analysis, saving time and money. Not all inhibition is resolvable. If your site suffers from frequent, irresolvable inhibition it may not be a good candidate for qPCR implementation (see Chapter 2).

If all quality control criteria are met and there is no evidence of target underestimation or inhibition of the qPCR assay you may report out your sample results (as CCE/100 ml using the  $\Delta\Delta$ Ct calculation method as described in Section 7.1.3).

## 7.3. Reporting the Results

Replicate sample analyses are recommended and should be performed for each unknown surface water sample. Calculate the results as Enterococcus (large subunit ribosomal RNA gene) CCE/100 ml.

Please consider the following prior to reporting out your results. While you may have good sample replicate agreement with respect to CT values, i.e. less than 1.0 Ct spread, the calculated CCE/100 ml may be far apart. Remember that in calculating CCE/100 ml you are looking at the relative difference between the calibrator/SPC paired analyses and your unknown sample/SPC paired analyses as a function of your standard curve. The  $\Delta\Delta$ Ct calculation used to determine this relationship, and your final results, may result in numerical disagreement of the CCEs from replicates 1 and 2. If both fall above or below your regulatory action threshold there is no operational difference even though the individual numbers may be relatively far apart, e.g. 140 and 900 CCE/100 ml. However, if the CCE values are variable and fall on either side of the criteria thresholds, you will need to rerun the analysis until you get replicate pair agreement (at minimum with respect to regulatory action decision agreement) or default to an alternative method such as a culture or predictive model estimation.

It may be desirable to report out a single CCE value, rather than the range of values returned from the analysis of the replicate pairs. If both values fall above or below the regulatory

threshold and they are within one order of magnitude for values less than or equal to 10<sup>2</sup> and two orders of magnitude for values greater than 10<sup>2</sup> they can be arithmetically averaged. In other situations, you may use the geometric mean calculation as described in Method 1611. When the qPCR Ct values are 45, non-detection, report your CCE/100ml as one half of the reciprocal of the dilution factor employed, i.e. a non-detect on a sample processed using a 1:10 dilution would be reported as 5 CCE/100 ml. Reporting data in this manner will provide consistency and the ability to conduct future statistical analysis.