Chapter 18 Quality Control

It's time to go to court. The WWTP has been accused of destroying the downstream ecosystem due to continuous low pH discharge. The only defense is the lab technician data that shows the pH for the past 6 months has been between 7.1 and 7.6. Will the judge believe the lab technician or the downstream user? The judge asks the lab technician for the documentation showing his pH meter was working correctly. "Your Honor, I've been measuring pH for 15 years and I'm sure it's correct!" The judge bellows: "Where's the pH meter calibration data?" The technician responds, "The bugs ate it, your honor!" Guilty! Guilty! Guilty!

In chapter 17, the QA plan identifies what the laboratory was going to do to produce valid and defensible data. Did the lab technician follow the QA plan? No! Did the lab manager check the data? No! The laboratory failed to follow its own QA plan and has no documentation to defend its data. Quality control (QC) is the written documentation demonstrating implementation of the QA plan. The laboratory technician will spend as much as 25-50% of his/her time doing QC work. To decide what information should be recorded, think like a lawyer. The SOP details the analytical procedure, QC answers the question, "Was the action performed?" As an example, let's look at measuring pH. PH is a very quick simple measurement but the QC takes longer to perform than the sample measurement.

For example, the lab technician analyses a sample and reports a pH of 7.5. In order to measure the pH, the technician will have to be able to prove to the judge the meter was calibrated by documenting in writing:

a. What meter was calibrated?
   QC: Make/Model of pH meter
b. Who calibrated the meter?
   QC: Name of analyst
c. When was the meter calibrated?
   QC: Date/Time of standardization
d. What buffer was used?
   QC: pH buffer Lot #
   QC: Date of buffer purchase or preparation
e. Was the buffer valid?
   QC: Date of buffer expiration
f. Did the meter accept the pH values
   QC: pH buffer calibration results
   QC: Measured slope

By writing the information down in a log book or bench sheet, the laboratory will be able to provide defensible data to inspectors and courts.

Laboratories have different methods of documenting quality control but the key is to **WRITE IT DOWN IN INK**. Each SOP should identify QC parameters. Common QC information includes

- Dates/ times of collection, arrival, analysis
- Name/signature of sampler, analyst
- Calibration data
- Standardization data
- Temperatures
- Blanks
- Positive/Negative controls
- Equipment operation/maintenance
- Reagent purchase/expiration dates
- Reagent lot numbers
- Method used
- Performance evaluation sample results

Error Analysis

Even though all the QC information is recorded, the laboratory must know the result is accurate and precise. Analytical

results are always subject to errors that can affect their validity. Understanding method performance through error analysis and precision, and accuracy will improve measurement quality.

Systematic Errors
Systematic errors are errors that are repeated for every measurement that is made. For example, a reagent blank may have a slight background color due to contaminants from the reagents. This color would give a positive bias to any sample containing these reagents. This error would be present in all samples but could be determined and subtracted out from all the sample measurements.

Random Errors
Random errors are errors that occur from different sources and add either a positive or negative bias. Random errors can be divided into 2 categories:

1. Determinant errors
2. Indeterminant errors

Determinant errors are errors that can be determined. They can be method errors, instrument errors, or personal errors.

Method Errors
Method errors are errors that result because of problems in the method. Methods are developed to determine the answer for some component, for example, phosphate. The method may also cause a reaction between another components, like silica. These competing components may either make the sample answer higher (positive bias) or the answer lower (negative bias). These competing components are called interferences. Interferences create inaccurate results.

Instrument Errors
An operating instrument usually produces a signal (background noise) even when no sample is present or when a blank is being analyzed. This electronic noise causes the display to vary. The larger the variance, the more inaccurate the answer. Large variances also affect the detection limit. A common example of electronic noise is the electronic analytic balance or pH meter. The balance has been zeroed but flutters between 0.0001 and -0.0001. This flutter is the result of small electronic fluctuations. If the fluctuations change over time, the equipment may need to be serviced.

Personal Errors
Personal errors are attributable to individual mistakes which are consistently made by the analyst. These errors are the result of carelessness, lack of knowledge, or personal bias. These errors are correctable. Improvements in precision and accuracy should be found as errors are eliminated. Examples of personal errors would include

- Using the wrong glassware
- Blowing out volumetric pipets
- Poor pre-washing of TSS filters
- Poor titration endpoint determination
- Failure to calibrate instruments
- Incubators outside the temperature range
- Incorrect calculations

The laboratory manager has the most control over these errors. Prior to assigning samples to a new analyst, the lab manager should document the analyst's capability to perform the method and obtain satisfactory precision and accuracy results.

Statistical Evaluation

The lab manager should perform precision and accuracy for each method used.

Accuracy is defined as the ability of the analyst to get the correct results.

Accuracy = Truth

This could be considered the bullseye on a target. Hitting the bullseye, means the analyst measured the answer correctly. If the analyst misses the bullseye, the answer was measured incorrectly. The laboratory manager must decide how far from the bullseye will be acceptable before the data can be reported. Accuracy is usually reported as % recovery. As an example, if you go to the bank to cash a $100 check, the bank teller counts back 97 one dollar bills. The bank teller is inaccurate. Is the analyst happy? What if the teller counted out 103 one dollar bills? Is the analyst happy now? How much latitude is allowed? This will be defined as accuracy or percent recovery later.

Precision is the ability of the analyst to get reproducible results. If we again use the bullseye on the target and the technician hits the bullseye he is accurate. The precision question is can he do it again and again. If the analyst consistently hits the bullseye, the analyst is both precise and accurate. (Our bank teller counts out 100 one dollar bills each time)

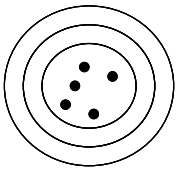


Figure: Precise and accurate

If the technician always hits the top right corner of the target, he is precise but inaccurate. (Our bank teller always counts out 97 dollars)

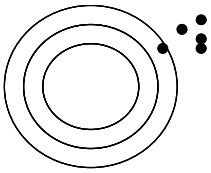


Figure: Precise but inaccurate

If the technician misses equally high, left, right, and low, the technician is accurate but very imprecise. (Our teller counts out 90 dollars one time, recounts 100 the second time, recounts 105 the third time, and finally counts 100 the fourth time. The analyst gives up and assumes 100 is correct (accurate) but the teller is very imprecise.)

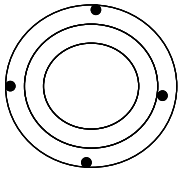


Figure: Imprecise but accurate

Finally, if the technician misses way left, a little right, a little low, and just nicks the target high, he is both imprecise and inaccurate. (Our teller doesn't know what a dollar looks like and needs additional training and he finds another bank.)

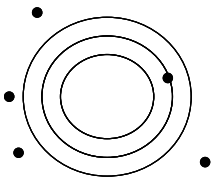


Figure: Imprecise and Inaccurate

Precision is usually reported as standard deviation or relative standard deviation.

Calculating the Mean

Precision and accuracy data are performed on normal samples. First determine the mean for a sample. To calculate the mean, 10 replicate samples are analyzed. The

average result called the mean is calculated. The mean is the sum of all the sample values divided by the number of samples run.

$$\xi = \frac{\sum X_i}{n}$$

For example: The lab analyst has run 10 TSS tests on a single homogenized influent sample with the following results.

110 mg/L	100 mg/L
105 mg/L	90 mg/L
97 mg/L	104 mg/L
103 mg/L	96 mg/L
95 mg/L	100 mg/L

To calculate the mean, all the values are added then divided by 10.

$$\xi = \frac{1000}{10} = 100 \text{ mg/L}$$

Having tested the sample 10 times, the lab analyst is pretty sure the sample answer is 100 mg/L.

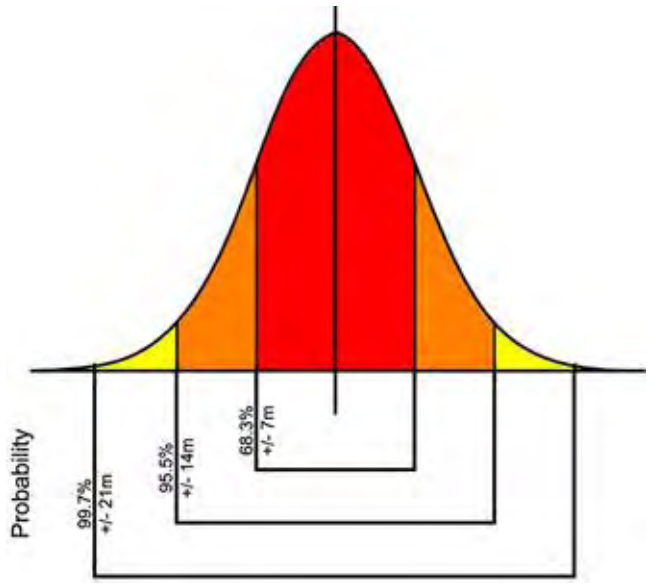
Calculating Precision

Once the mean has been determined, the analyst should find out how far his data deviates from the mean. The closer the data, the more precise the analyst. The farther the data, the poorer the precision. By calculating precision, the laboratory can check on the analyst's capability and the feasibility of using the method.

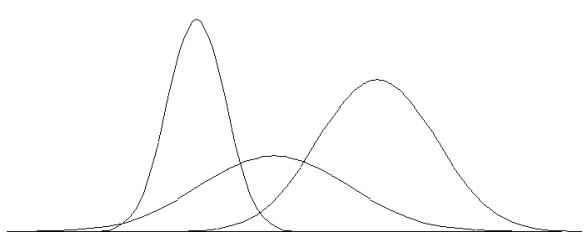
To calculate precision, the standard deviation is calculated from the data used to calculate the mean.

$$S^{n-1} = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

A scientific calculator with a statistics function can easily perform this calculation. In this example, the standard deviation is ±6 mg/L. Two standard deviations would be ±12 mgL and three standard deviations would be ±18mg/L. The normal distribution of data follows the Gaussian or normal distribution curve.



In a normal distribution of data, most points fall close to the mean. For instance, if the analyst performs 10 replicate titrations, most of the titration volumes should be the same. There will be a few titrations that take a little bit more and few that take a little bit less. These deviations are due to error. The fewer mistakes the analyst makes, the sharper the curve will be. If the analyst makes lots of big mistakes, the curve will be very broad.



The lab manager's goal is to produce as sharp a peak as possible.

A fixed percentage of data points should fall within each standard deviation. Sixty-eight percent of the sample values should fall within 1 standard deviation. In this example, 7 of the 10 values should fall between 94 and 106 mg/L (100 $\pm$ 6 mg/L). Ninety-five percent of the values should fall within 2 standard deviations (100 $\pm$ 12 mg/L) and Ninety-nine percent of the values should fall within 3 standard deviations (100 $\pm$ 18 mg/L).

This data could have been generated for 3 different purposes. First, if this is a new method being used in the laboratory, the standard deviation could be compared with the reference standard deviation expected for the method. If the references coincide, the lab may decide to use the procedure. For example, the TSS procedure might indicate a sample with a concentration of 100 mg/L might have a standard deviation of 5 mg/L. Since the laboratory standard deviation is $\pm$ 6 mg/L, the laboratory may not be performing the method satisfactorily.

Likewise, the laboratory could be using this data to define its capabilities. The precision data is basically telling the lab manager or customer that the lab is 95% confident a sample with a TSS of 100 mg/L will be within 88 - 112 mg/L.

Finally, if the laboratory is using the standard deviation in Standard Methods ($\pm$ 5 mg/L) as its guideline, and the data represents a new analyst's capabilities, the new analyst is not yet ready to perform samples to the laboratories QC levels.

Laboratories use the 95% values to establish control charts.

The process can be repeated at lower levels to establish the standard deviation for low level samples such as effluent.

Example 2: Effluent TSS data

5.4 mg/L	5.5 mg/L
6.0 mg/L	6.0 mg/L
5.5 mg/L	5.8 mg/L
6.0 mg/L	5.6 mg/L
4.8 mg/L	5.9 mg/L

The calculated mean and standard deviation for this effluent data is found to be

$$\xi = \frac{56.5}{10} = 5.6 \text{ mg/L}$$

SD = 0.4 mg/L

The two different standard deviations are difficult to compare numerically but can be compared by percentage. The relative standard deviation (RSD) can be calculated by dividing the standard deviation by the mean and then multiplying the result by 100.

$$\text{RSD} = \frac{\text{SD} \times 100}{\xi}$$

The RSD for the influent data is :

$$\text{RSD (influent)} = \frac{6 \times 100}{100} = 6\%$$

$$\text{RSD (effluent)} = \frac{0.4 \times 100}{5.6} = 7\%$$

Although the SD for the effluent (0.4 mg/L) is lower than the SD for the influent (6.0 mg/L), the RSD is higher. The higher RSD indicates the effluent data is actually less precise than the influent data. Remember precision is the ability to get the same answer, so the higher the RSD, the less precise the answer.

The precision statement should read: In a single laboratory, using wastewater samples at concentrations of 6 and 100 mg/L, the relative standard deviation was between 6% and 7%.

Accuracy (Percent Recovery)

The laboratory should calculate accuracy to determine whether a method is compatible with a specific sample by using standard additions (spiking). A standard spike is a method used to determine if the sample matrix contains materials that will interfere with the analytical method. If the method either adds or detracts analyte, an accurate answer will not be obtained.

To perform a standard spike, first analyze the unspiked sample to determine the mean. For example, the analyst runs seven replicate total phosphate samples and finds the mean to be 5.0 mg/L.

Next, add a known amount of phosphate standard to a second set of replicates of the same sample and repeat the analysis. Check to see if the additional standard added has been recovered. For instance, the sample above was spiked with 1 ml of a 100 mg/L phosphate standard and analyzed. If the percent recovery of the spiked sample is 90-110%, the method may be deemed acceptable. If the spike result is outside the range, a substance may be present in the sample matrix which interferes with the analytical method.

Calculating Percent Recovery (Standard Addition)

A lab technician analyzes a 50 ml sample for total phosphate and finds the answer to be 4.0 mg/L. The sample is spiked with 1.0 ml of a 100.0 mg/L phosphate standard. The theoretical concentration of the spiked sample is

$$\text{Theoretical conc} = \frac{(C_u V_u)+(C_s V_s)}{V_u + V_s}$$

Where;
C_u = Original measured concentration of unspiked sample
V_u = Volume of sample to which the spike is added
C_s = Concentration of the standard used to spike
V_s = Volume of standard used to spike

$$\text{Theoretical Conc.} = \frac{(50 \times 4)+(100 \times 1)}{(50 + 1)} = \frac{200 + 100}{51} = 5.9$$

The spiked sample is now analyzed and found to be 5.6 mg/L. The percent recovery is calculated by dividing the original sample answer by the spiked sample answer then multiplying by 100.

$$\text{\% accuracy} = \frac{\text{actual spike}}{\text{theoretical spike}} = \frac{5.6 \times 100}{5.9} = 95\%$$

Spiked samples should be determined at both high and low sample concentrations to show the method is acceptable for the entire method range. A second sample at the high end of the analytical range (15 mg/L) was also spiked with 1 ml of 100 mg/L phosphate standard.

$$\text{Theoretical Conc.} = \frac{(50 \times 15)+(100 \times 1)}{(50 + 1)} = \frac{750 + 100}{51} = 16.7$$

The analyst measures 16.9 mg/L. The second percent recovery is:

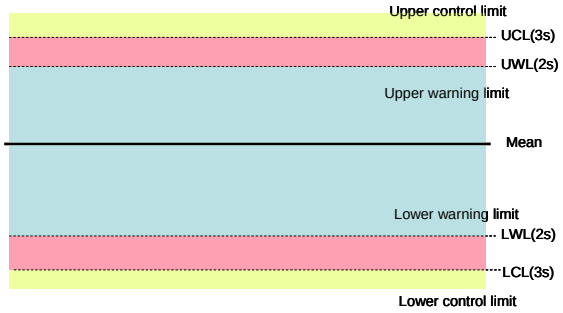
$$\text{\% accuracy} = \frac{\text{actual spike}}{\text{theoretical spike}} = \frac{16.9 \times 100}{16.7} = 101\%$$

Accuracy is reported as:
In a single laboratory, using wastewater at concentrations of 4 mg/L and 15 mg/L, recoveries of 95% and 101% were found. The laboratory manager will have to decide

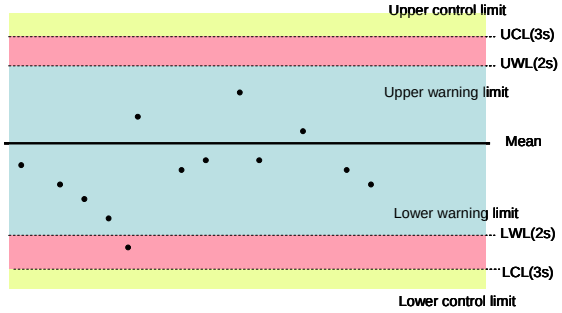
if the % recovery is satisfactory for the method chosen.

Control Charts

Control charts are visual tools used by the laboratory to track performance. Each time a standard is run, ie GG standard for BOD, the results are plotted on the control chart. Control charts can be used to track both accuracy and precision over a period of time. Accuracy charts can track QC samples such reagent blanks, standards, spikes, etc. The control chart will have several horizontal lines which act as performance limits. WL is the warning limit and is defined as two times the standard deviation (95% confidence level). CL is the critical limit or control limits and defines the outer boundary for the test. It is defined as three times the standard deviation.



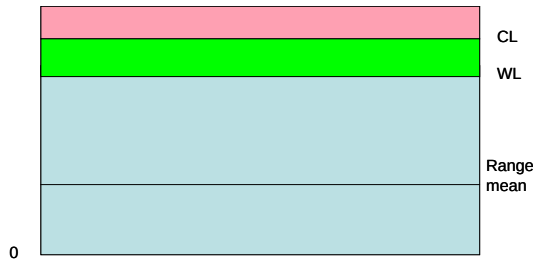
As an example, the GG standard for the BOD test could be plotted. The upper and lower control limits are defined in the procedure as being 198 $\pm$ 30.5. So the lower control limit is 167.5 mg/L and the upper control limit is 228.5 mg/L. Since 30.5 is 3 times the standard deviation, the warning limit would be about 20 mg/L. Each time the BOD analysis is run, the standard result would be plotted on the graph. Any standard result falling below the upper or lower warning limit would be an indication to the laboratory that the test is in danger of becoming invalid and investigation and correction is needed. By using the warning limits, the laboratory can take preventative measures to correct problems before the process is out of control and sample data becomes invalid.



Control charts should be constructed for each instrument and analytical method used.

For procedures where the control limits are not established by the method, the laboratory can establish there own control limits by running 15-20 standards and calculating the mean and standard deviation.

The second type of control chart is a precision control chart. Precision charts commonly track differences between duplicates. As before, if the warning limit or control limit are exceeded, the process is out of control and corrections must be made before continuing with analysis of samples. A minimum of 20 sets of duplicates should be run to determine the warning limit and control limit.



For simplicity, we'll use an example with 10 values instead of 20.
10 sets of a COD standard data were run over 10 weeks.

		Range
Standard 1:	187, 194 mg/L	7
Standard 2:	195, 196 mg/L	1
Standard 3:	198, 198 mg/L	0
Standard 4:	200, 205 mg/L	5
Standard 5:	189, 190 mg/L	1
Standard 6:	198, 193 mg/L	5
Standard 7:	199, 189 mg/L	10
Standard 8:	207, 205 mg/L	2
Standard 9:	201, 205 mg/L	4
Standard 10:	187, 189 mg/L	2

Calculate the standard deviation of the range values. In this example, s = 3.0 mg/L
To determine the mean range, the table in standard methods requires using the factor 1.128 for duplicates. The mean range is:

Mean range (R) = 1.128(s) = 1.128 (3.0) = 3.4 mg/L

The warning limit is:

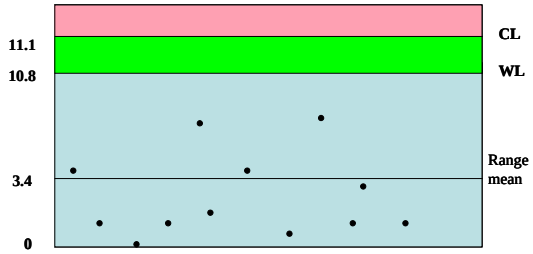
WL = mean range + 2/3 (D_4)(mean range)

D_4 again comes from the table in standard methods for duplicates.

WL = 3.4 + 2/3 (3.267)(3.4) = 3.4 + 7.4 = 10.8
The control limit is:

CL = (D_4)(mean range)

CL = 3.267(3.4) = 11.1 mg/L



Once the control chart is prepared, the range between duplicates can be compared.

Control limit exceeded:
If any range exceeds the CL, the measurement must be repeated. (duplicates are too far apart). If the repeat duplicate is still beyond the CL, the analysis must be stopped and the problem corrected. If the repeat duplicate is within the CL, continue analysis. For example: Standard 21 has a range of 15 which exceeds the CL of 11.1. Run another set of duplicates. If the next set has a range of 12, problems must be identified and corrected before more samples can be analyzed.

Warning limit exceeded
If 2 out of 3 consecutive range values are greater than the WL, analyze another standard. If the next point does not exceed the WL, continue analysis. If the next point exceeds the WL, problems must be identified and corrected before samples can be analyzed. For example, standards 22, 23, 24 have ranges of 11, 11, 10 and exceed the WL of 10.8. Analyze another standard. If this standard is 11 mg/L, problems must be identified and corrected before samples can be analyzed.

While establishing control charts seems like a lot of work, once established they provide the analyst, lab manager, and inspectors a quick review to show the laboratory data is under control.