Chapter 3: SOME BASIC QA /QC CONCEPTS

As coordinator of a volunteer monitoring program, you are probably involved in many aspects of project planning, sample collection, laboratory analysis, data review, and data assessment. You should be considering quality assurance and quality control activities in every one of these steps.

Quality assurance (QA) refers to the overall *management system* which includes the organization, planning, data collection, quality control, documentation, evaluation, and reporting activities of your group. QA provides the information you need to ascertain the quality of your data and whether it meets the requirements of your project. QA ensures that your data will meet defined standards of quality with a stated level of confidence.

QA ensures that your data will meet defined standards of quality with a stated level of confidence.

Quality control (QC) refers to the routine *technical activities* whose purpose is, essentially, error control. Since errors can occur in either the field, the laboratory or in the office, QC must be part of each of these functions. QC should include both internal and external measures (see side box).

Together, QA and QC help you produce data of known quality, enhance the credibility of your group in reporting monitoring results, and ultimately save time and money. However, a good QA/QC program is only successful if everyone consents to follow it and if all project components are available in writing. The Quality Assurance Project Plan (QAPP) is the written record of your QA/QC program.

This chapter is designed to introduce you to the terminology of quality assurance/quality control. The key terms we will be addressing are: precision, accuracy (sometimes referred to as bias), representativeness, completeness, comparability, and sensitivity. You will

QC Measures

Internal Quality Control is a set of measures that the project undertakes *among its own samplers and within its own lab* to identify and correct analytical errors. Examples include lab analyst training and certification, proper equipment calibration and documentation, laboratory analysis of samples with known concentrations or repeated analysis of the same sample, and collection and analysis of multiple samples from the field.

External Quality Control is a set of measures that involves *laboratories and people outside of the program*. These measures include performance audits by outside personnel, collection of samples by people outside of the program from a few of the same sites at the same time as the volunteers, and splitting some of the samples for analysis at another lab.

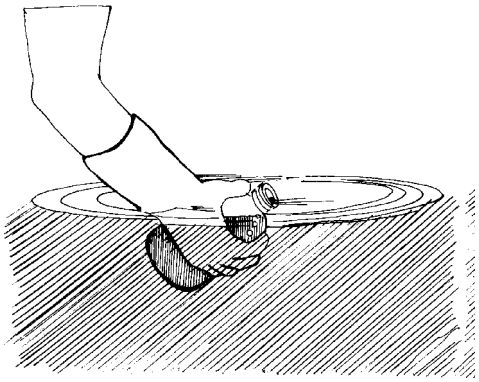
External and internal QC measures are described in more detail in the "QC Samples" box at the end of this chapter.

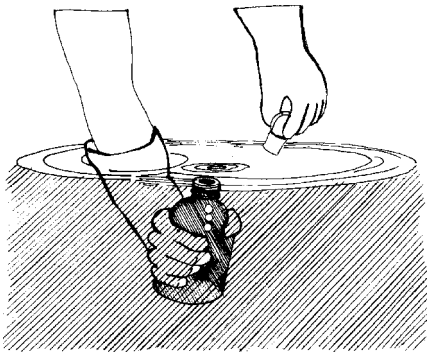
Measures of precision, accuracy, representativeness, completeness, comparability, and sensitivity help us evaluate sources of variability and error and thereby increase confidence in our data.

be seeing these terms again, so you may want to spend some time getting to know them.

In natural systems, such as streams, lakes, estuaries, and wetlands, variability is a factor of life. Changes in temperature, flow, sunlight, and many other factors affect these systems and the animals that inhabit them. Variability also occurs when we attempt to monitor such systems. Each of us reads, measures, and interprets differently; we may also apply different levels of effort in how we monitor. The equipment we use may be contaminated, broken or incorrectly calibrated. These and many other differences can lead to variability in monitoring results. Measures of precision, accuracy, representativeness, completeness, comparability, and sensitivity help us evaluate sources of variability and error and thereby increase confidence in our data.

Because all projects have different goals, data users and uses, capabilities, and methods, this document cannot tell you what levels of precision, accuracy, representativeness, completeness, comparability, and sensitivity are acceptable for your individual project. You will need to consult your advisory panel (in particular, your data users), the laboratory you deal with, and peer reviewers to determine acceptance criteria for your monitoring project.





Precision

Precision is the degree of agreement among repeated measurements of the same characteristic on the same sample or on separate samples collected as close as possible in time and place. It tells you how consistent and reproducible your field or laboratory methods are by showing you how close your measurements are to each other. It does not mean that the sample results actually reflect the "true" value, but rather that your sampling and analysis are giving consistent results under similar conditions.

Typically, precision is monitored through the use of replicate samples or

measurements. Replicate samples are two or more samples taken from the same place at the same time.

When you have many replicate samples, determine precision by calculating the **standard deviation(s)** of the samples. The standard deviation indicates the range of variation in the measurements you've taken. Many of today's calculators perform the standard deviation calculation.

The **relative standard deviation (RSD)**, or coefficient of variation, expresses the standard deviation as a percentage. This is generally easier for others to understand. The smaller the relative standard deviation (or standard deviation), the more precise your measurements.

When you have only two replicate samples, determine precision by calculating the **relative percent difference (RPD)** of the two samples. Again, the smaller the relative percent difference, the more precise your measurements.

STANDARD DEVIATION

The Volunteer Creek Monitoring Project wants to determine the precision of its temperature assessment procedure. They have taken 4 replicate samples:

Replicate 1 (X_1) = 21.1^0 C
Replicate 2 (X_2) = 21.1^0 C
Replicate 3 (X_3) = 20.5^0 C
Replicate 4 (X_4) = 20.0^0 C

To determine the **Standard Deviation (s)**, use the following formula:

$$s = \sqrt{\sum_{i=1}^{n} \frac{(X_1 - \overline{X})^2}{n-1}}$$

where x_i = measured value of the replicate, $\bar{x}$ = mean of replicate measurements, n = number of replicates, $\sum$ = the sum of the calculations for each measurement value--in this case, X_1 through X_4

First, figure out the mean, or average of the sample measurements. Mean = $(X_1 + X_2 + X_3 + X_4) \div 4$. In this example, the mean is equal to 20.68^0 C.

Then, for each sample measurement (X_1 through X_4), calculate the next part of the formula. For X_1 and X_2, the calculation would look like this:

$$\frac{(21.1 - 20.68)^2}{4-1} = \frac{(-0.42)^2}{3} = \frac{0.1764}{3} = 0.0588$$

For X_3 the calculation would be 0.0108; and for X_4 it would be 0.1541

Finally, add together the calculations for each measurement and find the square root of the sum: 0.0588 + 0.0588 + 0.0108 + 0.1541 = 0.2825. The square root of 0.2825 is 0.5315.

So, the standard deviation for temperature is 0.532 (rounded off).

RELATIVE STANDARD DEVIATION

If we use the same replicate measurements as above in the standard deviation example, we can determine the **Relative Standard Deviation (RSD)**, or coefficient of variation, using the following formula:

$$RSD = \frac{s}{\overline{X}} \times 100$$

where s = standard deviation and $\bar{x}$ = mean of replicate samples.

We know s = 0.5315 and that $\bar{x}$ = 20.68. So, the RSD = 2.57. This means that our measurements deviate by about 2.57%.

RELATIVE PERCENT DIFFERENCE

If the Volunteer Creek project had only two replicates (21.1^0 C and 20.5^0 C) they would use **Relative Percent Difference (RPD)** to determine precision.

$$RPD = \frac{(X_1 - X_2) \times 100}{(X_1 + X_2) \div 2}$$

where X_1 = the larger of the two values and X_2 = the smaller of the two values. In this example, $X_1 = 21.1^0$ and $X_2 = 20.5^0$.

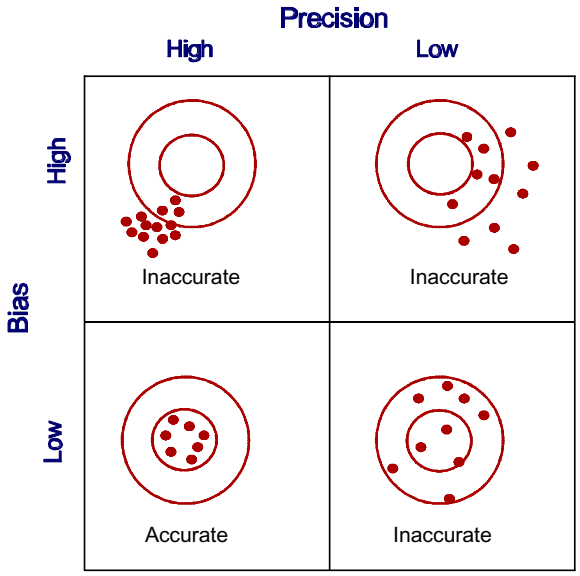
$$RPD = \frac{(21.1-20.5) \times 100}{(21.1+20.5) \div 2} = \frac{60.00}{20.8} = 2.88$$

So, in this example, the RPD between our sample measurements is 2.88%.

Accuracy

Accuracy is a measure of confidence in a measurement. The smaller the difference between the measurement of a parameter and its "true" or expected value, the more accurate the measurement. The more precise or reproducible the result, the more reliable or accurate the result.

PRECISION, BIAS, AND ACCURACY



Measurement accuracy can be determined by comparing a sample that has a known value, such as a standard reference material or a performance evaluation sample, to a volunteer's measurement of that sample (see note below). Increasingly, however, some scientists, especially those involved with statistical analysis of measurement data, have begun to use the term "bias" to reflect this error in the measurement system and to use "accuracy" as indicating both the degree of precision and bias (see "bullseye" figure at left). For the purpose of this document, the term "accuracy" will be used.

If you are concerned that other components of a sample matrix (e.g., soil or sludge) may be interfering with analysis of a parameter, one way to measure accuracy is to add a known concentration of the parameter to a portion of the sample. This is called a spiked sample. The difference between the original measurement of the parameter in the sample and the measurement of the spiked sample should equal (or be close to) the added amount. The difference indicates your ability to obtain an accurate measurement.

ACCURACY

Attendance at QC training sessions is required for Volunteer Creek monitors. In the field, monitors use a Jones Wide-Range pH Kit, which covers a full range of expected pH values. During a recent training session, the monitors recorded the following results when testing a pH standard buffer solution of 7.0 units.

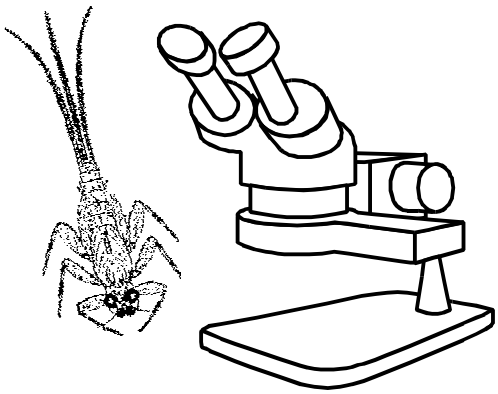
7.5	7.2	6.5	7.0
7.4	6.8	7.2	7.4
6.7	7.3	6.8	7.2

Accuracy = average value - true value

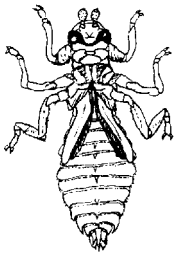
The average of these measurements is equal to 7.08 units. Since we know that the reference or "true" value is 7.0 units, the difference between the average pH value is off or biased by + 0.08 units. This level of accuracy is satisfactory for the data quality objectives of the project.

For many parameters such as secchi depth and macroinvertebrate abundance, no standard reference or performance evaluation samples exist. In these cases, the trainer's results may be considered the reference value to

which the volunteer's results are compared. This process will help evaluate if the volunteer measurements are biased as compared to the trainer's.



If you are monitoring biological conditions by collecting and identifying specimens, maintaining a voucher collection is a good way to determine if your identification procedures are accurate. The voucher collection is a preserved archive of the organisms your volunteers have collected and identified. An expert taxonomist can then provide a "true" value by checking the identification in the voucher collection.



It is important to note that the relationship between a voucher collection and accurate identification cannot be expressed numerically in your QAPP. Rather, the QAPP document should indicate that you have a voucher collection and describe how it is used to evaluate consistent accurate identification in your program.

Note: Standard reference material (in the form of solids or solutions with a certified known concentration of pollutant) can be obtained from a variety of companies, including the National Institute of Standard and Technologies, that sell quality control, proficiency, or scientific reference materials.

Representativeness

Representativeness is the extent to which measurements actually depict the true environmental condition or population you are evaluating. A number of factors may affect the representativeness of your data. For instance, are your sampling locations indicative of the waterbody? Data collected just below a pipe outfall is not representative of an entire stream. Minimizing the effects of variation is critical in the development of your sampling design.

Completeness

Completeness is a measure of the number of samples you must take to be able to use the information, as compared to the number of samples you originally planned to take. Since there are many reasons why your volunteers may not collect as many samples as planned, as a general rule you should try to take more samples than you determine you actually need. This issue should be discussed within your QAPP team and by peer reviewers before field activities begin.

COMPLETENESS

The Volunteer Creek Monitoring project planned to collect 20 samples, but because of volunteer illness and a severe storm, only 17 samples were actually collected. Furthermore, of these, two samples were judged invalid because too much time elapsed between sample collection and lab analysis. Thus, of the 20 samples planned, only 15 were judged valid.

The following formula is used to determine **Percent Completeness (%C).**

$$\%C = \frac{v}{T} x 100$$

where v = the number of planned measurements judged valid and T = the total number of measurements.

In this example, v = 15 and T = 20. In this case, percent completeness would be 75 percent. Is this enough information to be useful?

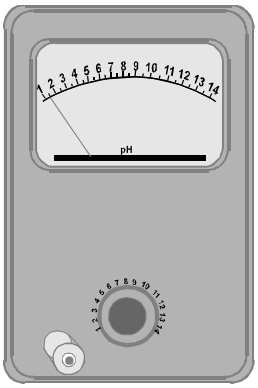
To calculate percent completeness, divide the number of measurements that have been judged valid by the total number of measurements you originally planned to take and then multiply by 100.

Remember, completeness requirements can be lowered if extra samples are factored into the project. The extra samples in turn, increase the likelihood of more representative data.

Comparability

Comparability is the extent to which data from one study can be compared directly to either past data from the current project or data from another study. For example, you may wish to compare two seasons of summer data from your project or compare your summer data set to one collected 10 years ago by state biologists.

Using standardized sampling and analytical methods, units of reporting, and site selection procedures helps ensure comparability. However, it is important to keep in mind that some types of monitoring rely heavily on best professional judgement and that standard methods may not always exist.



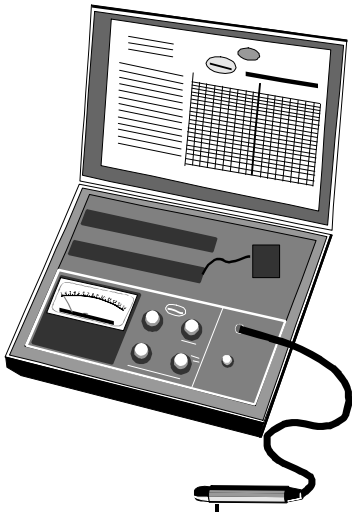
Detection Limit

The term *detection limit* can apply to monitoring and analytical instruments as well as to methods. In general, detection limit is defined as the lowest concentration of a given pollutant your methods or equipment can detect and report as greater than zero. Readings that fall below the detection limit are too unreliable to use in your data set. Furthermore, as readings approach the detection limit (that is, as they go from higher, easier-to-detect concentrations to lower, harder-to-detect concentrations) they become less and less reliable. Manufacturers generally provide detection limit information with high-grade monitoring equipment such as meters.

Measurement Range

The *measurement range* is the range of reliable measurements of an instrument or measuring device. Preassembled kits usually come with information indicating

the measurement range that applies. For example, you might purchase a kit that is capable of detecting pH falling between 6.1 and 8.1. However, pH can theoretically range from 0.0 to 14.00. If acidic conditions (below 6) are a problem in the waters you are monitoring, you will need to use a kit or meter that is sensitive to the lower pH ranges.



Quality Control (QC) Samples

Contamination is a common source of error in both sampling and analytical procedures. QC samples help you identify when and how contamination might occur. For most projects, there is no set number of field or laboratory QC samples which must be taken. The general rule is that 10% of samples should be QC samples. This means that if 20 samples are collected, at least one additional sample must be added as a QC sample. The laboratory must also run its own QC samples. For a new monitoring project or for a new analytical procedure, it is a good idea to increase the number of QC samples (up to 20%) until you have full confidence in the procedures you are using.

The general rule is that 10% of samples should be quality control (QC) samples.

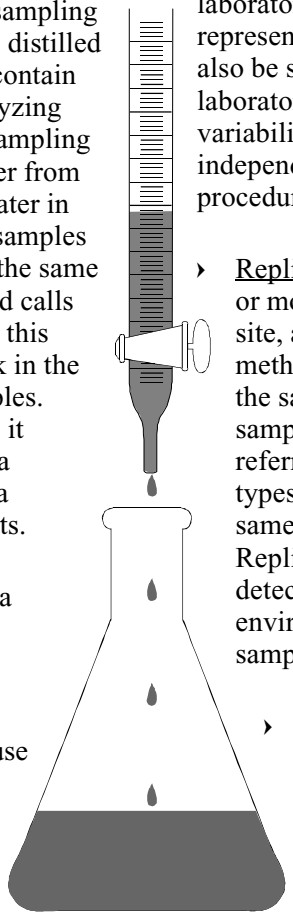
When the project is over, determine data quality by evaluating the results of all the QC samples and determining precision and accuracy. For QC samples that are not blind to the lab, require the lab to calculate and report precision and accuracy results. Lab reported precision and accuracy results can then be checked during data validation.

The decision to accept data, reject it, or accept only a portion of it is should be made after analysis of all QC data. Various types of QC samples are described in the box on the next page.

QC SAMPLES

- A field blank is a "clean" sample, produced in the field, used to detect analytical problems during the whole process (sampling, transport, and lab analysis). To create a field blank, take a clean sampling container with "clean" water (i.e., distilled or deionized water that does not contain any of the substance you are analyzing for) to the sampling site. Other sampling containers will be filled with water from the site. Except for the type of water in them, the field blank and all site samples should be handled and treated in the same way. For example, if your method calls for the addition of a preservative, this should be added to the field blank in the same manner as in the other samples. When the field blank is analyzed, it should read as analyte-free or, at a minimum, the reading should be a factor of 5 below all sample results.

- An equipment or rinsate blank is a "clean" sample used to check the cleanliness of sample collection equipment. This type of blank is used to evaluate if there is carryover contamination from reuse of the same sampling equipment. A sample of distilled water is collected in a sample container using regular collection equipment and analyzed as a sample.



- A split sample is one sample that is divided equally into two or more sample containers and then analyzed by different analysts or labs. Split samples are used to measure precision. Samples should be thoroughly mixed before they are divided. Large errors can occur if the analyte is not equally distributed into the two containers. A sample can be split in the field, called a field split, or in the laboratory, a lab split. The lab split measures analytical precision while the field split measures both analytical and field sampling precision. In addition, a sample split in the field and submitted to the laboratory without informing the laboratory represents a blind sample. Split samples can also be submitted to two different laboratories for analysis to measure the variability in results between laboratories independently using the same analytical procedures.

- Replicate samples are obtained when two or more samples are taken from the same site, at the same time, using the same method, and independently analyzed in the same manner. When only two samples are taken, they are sometimes referred to as duplicate samples. These types of samples are representative of the same environmental condition. Replicates (or duplicates) can be used to detect both the natural variability in the environment and that caused by field sampling methods.

- Spiked samples are samples to which a known concentration of the analyte of interest has been added. Spiked samples are used to measure accuracy. If this is done in the field, the results reflect the effects of preservation, shipping, laboratory preparation, and analysis. If done in the laboratory, they reflect the effects of the analysis from the point when the compound is added, e.g. just prior to the measurement step. Percent recovery of the spike material is used to calculate analytical accuracy.