Ch. 5: Quality Assurance & Calibration Methods

Outline:

• 5-1 Basics of QA
• 5-2 Method Validation
• 5-3 Standard Addition
• 5-4 Internal Standards
• 5-5 Efficiency in Experimental Design
Basics of QA

Case 1: You are cooking dinner for some friends, and are making some sauce for spaghetti. You season, taste, and repeat this procedure until you have the sauce just the way you want it. Each time you taste, this is a **quality control test**.

Case 2: You run a plant that makes 1000 jars of spaghetti sauce per day. You obviously can’t test all of them, so you decide that there are three tastings per day (e.g., 11 a.m., 2 p.m. and 5 p.m.). If all the jars taste fine, then one may conclude that all 1000 jars taste fine.

**However**, this may not be true, since there is a relative risk of other jars having too much or too little seasoning (not too important, since you are easy with refunds!).

If the number of refunds per year is 100, then this is an excessive test, since you are sacrificing 365 x 4 jars per day for taste tests! This should be optimized!
Basics of QA, 2

There are three basic steps for quality assurance, all of which are intimately intertwined:

1. **Use objectives**

   Clear, concise *use objectives* for acquisition and handling of data must be decided upon.

2. **Specifications**

   We must state how good the numbers need to be and what precautions are required in the analytical procedure (there are tons of considerations here!)

3. **Assessment**

   Collection and summary of data and test to see that objectives are met.
Use objectives

1. Use objectives
In quality assurance, use objectives are a written statement of how results will be used. Use objectives are required before specifications can be written for the method. This includes (i) a decision on the acceptable uncertainty (or confidence level) of the results, (ii) the necessary precision of the measurements, and (iii) the assurance that the analysis produces results that are distinguishable from experimental error.

**e.g., sauce vs. pharmaceuticals**
The rough QA testing on the sauce cannot be done for something like pharmaceuticals, which may be near a lethal dose to be effective.

**e.g., scales vs. balances**
A bathroom scale need not measure to the nearest mg, but a pharmaceutical with an active ingredient of 2 mg cannot have an uncertainty of ±1 mg.
Use objectives, 2

e.g., drinking water and disinfection facility
Let’s say a company builds a disinfection facility with the purpose of treating water with a specially modified chlorination process, since chlorine is well known to kill most microorganisms in the water. However, there is a chance that the chlorine may react with organic matter in the water and form by-products that are harmful to humans. This modified process results in an apparent reduction of such by-products.

Analytical use objective:

(1) Analytical data and results shall be used to determine whether the modified chlorination process results in at least a 10% reduction of formation of selected disinfection by-products.

(2) Uncertainty in the analysis must be small enough that a 10% decrease in selected by-products is clearly distinguishable from experimental error. (i.e., are the tests really indicating a decrease of 10%).
Specifications

2. Specifications
We must state how good the numbers need to be and what precautions are required in the analytical procedure.

Questions (largely regarding the analytical procedure):

• How shall samples be taken and how many are needed?

• Are special precautions required to protect samples and ensure that they are not degraded?

• Within practical restraints, such as cost, time, and limited amounts of material available for analysis, what level of accuracy and precision will satisfy the use objectives?

• What rate of false positives or false negatives is acceptable?

Generally, specifications include how the sampling is done, required accuracy and precision, reagent purity, tolerances for apparatus, the use of certified reference materials, acceptable values for blanks, and the preservation of the samples/analyte after analysis.
False positives and negatives

Sampling is very important when considering the potential occurrence of false positives and false negatives in analytical chemistry. For instance, the World Health Organization and numerous governmental environmental agencies around the world place the limit for nitrates (NO$_3^-$) in potable water at 50 mg/L.

A trustworthy analysis of a good, clean source of drinking water would indicate that the nitrate levels are below this limit; however,

A **false positive** says that the concentration exceeds the legal limit when, in fact, the concentration is below the limit.

A **false negative** says that the concentration is below the limit when it is actually above the limit.

Strangely enough, in some cases, one of these “falsehoods” is actually more desirable than the other (e.g., for drinking water, it is better to have low rate of false negatives than the low rate of false positives, meaning that it is worse to certify that contaminated water is safe than to certify that safe water is contaminated).
Detection of Analyte

There are three very important considerations when deciding upon the analytical method to be used for the detection of a particular analyte.

**Selectivity** (or *specificity*): the ability of the technique to distinguish the analyte from other species in the sample (i.e., avoiding interference which can corrupt the measurement).

**Sensitivity**: the capability of the technique to respond reliably and measurably to changes in analyte concentration.

**Detection limit**: the lower limit where any signal from the analyte can be detected by the analytical method - this limit must be significantly lower than the signal or response arising from the concentrations to be detected!
Blanks

Blanks account for interference by other species in the sample and for traces of analyte found in reagents used for sample preservation, preparation, and analysis.

*Method blank*: a sample containing all components except analyte, and it is taken through all steps of the analytical procedure. We subtract the response of the method blank from the response of a real sample prior to calculating the quantity of analyte in the sample.

*Reagent blank*: a sample similar to a method blank, but it has not been subjected to all sample preparation procedures.

*Field blank*: a sample similar to a *method blank*, but it has been exposed to the site of sampling.

*e.g.*, To analyze particulates in air, a certain volume of air could be sucked through a filter, which is then dissolved and analyzed. A field blank would be a filter carried to the collection site in the same package with the collection filters. The filter for the blank would be taken out of its package in the field and placed in the same kind of sealed container used for collection filters. The difference between the blank and the collection filters is that air was not sucked through the blank filter (i.e., so volatile compounds encountered during transportation or in the field are potential contaminants of a field blank).
Spikes and the Matrix

**Spike (or fortification):** A known quantity of analyte added to a sample to test whether the response to a sample is the same as that expected from a calibration curve.

In this way, the response of the analyte can be directly correlated to the known amount of the “spike” which is added - this is useful in cases where elements of the *matrix* (i.e., the remainder of the sample) may influence the response of the experiment to the analyte.

**e.g., drinking water and nitrate levels**
Let’s say your analytical method has lead to the prediction that the concentration of nitrates is 10 μg/L. One could add a spike of 5 μg/L (i.e., spike the solution), and the analytical method, if all is well, should provide a result of 15 μg/L. If this is not the case, then we known that the matrix is interfering with the analytical method.
Spike recovery

If $C$ is the concentration of analyte in a sample, then the *spike recovery* is defined as

$$\% \text{ recovery} = \frac{C_{\text{spiked sample}} - C_{\text{unspiked sample}}}{C_{\text{added}}} \times 100\%$$

**Example:**
An unknown was found to contain 10.0 $\mu$g of analyte per litre. A spike of 5.0 $\mu$g/L was added to a replicate portion of unknown. Analysis of the spiked sample gave a concentration of 14.6 $\mu$g/L. Find the percent recovery of the spike. The acceptable recovery limit is ±4%.

**Solution:**
The percent of the spike found by analysis is:

$$\% \text{ recovery} = \frac{14.6 \ \mu\text{g/L} - 10.0 \ \mu\text{g/L}}{5.0 \ \mu\text{g/L}} \times 100\% = 92\%$$

Since the acceptable recovery limit range is 96 to 104%, this result is unacceptable, and something must be adjusted in terms of analytical method, technique, sample prep, etc.
**Control charts**

*Control chart*: A visual representation of confidence intervals for a Gaussian distribution as a function of time. Recall that 95.5% of observations are within 2 SD of the mean, and 99.7% are within 3 SD.

**e.g., vitamin C tablets**

Consider a manufacturer making tablets with a mean value of $\mu$ g per tablet. The SD is monitored over time. If the following conditions occur, production should shut down:

- 1 observation outside the action lines
- 2 out of 3 consecutive measurements between the warning and action lines
- 7 consecutive measurements all above or all below the centre line
- 6 consecutive measurements all increasing or all decreasing, wherever they are located
- 14 consecutive points alternating up and down, regardless of where they are located
- an obvious nonrandom pattern

**Note**: these are not strict rules, but rather, specifications decided upon by the manufacturer.


**Checks and balances**

**Calibration checks**: Samples are routinely analyzed during the course of analysis (i.e., concentration) to ensure that the instrument is working properly and the methodology is being practiced properly. For example, a specification may that every 10 sample, a calibration check is done with a standard sample to test the instrument.

**Performance test samples** (also quality control samples or blind samples) are a quality control measure to help eliminate bias introduced by an analyst who knows the concentration of the calibration check sample:

- These samples of known composition are provided to an experienced analyst as unknowns.

- Results are then compared with the known values, usually by a quality assurance manager.
Standard Operating Procedures

Standard operating procedures state what steps will be taken and how they will be carried out are the bulwark of quality assurance. Adhering to these procedures guards against the normal human desire to take shortcuts based on assumptions that could be false, or to unconsciously bias or alter results. If SOPs are not followed rigidly (and enforced) within an analytical chemistry unit, much doubt can be cast on the analysis, which is potentially costly and/or dangerous.

Sample handling:

• Samples must be stored in containers and under conditions that do not allow relevant chemical characteristics to change (e.g., oxidation, photodecomposition, introduction of foreign substances, growth of invasive organisms, etc.

• The *chain of custody* is the trail followed by a sample from the time it is collected to the time it is analyzed and archived (e.g., drug testing in athletes).
### Chain of Custody

- Documents are signed each time the material changes hands to indicate who is responsible for the sample.
- Each person in the chain of custody follows a written procedure telling how the sample is to be handled and stored.
- Each person receiving a sample should inspect it to see that it is in the expected condition in an appropriate container.

Check out the flowchart for chain of custody and procedure for a laboratory analysis of items or materials confiscated by the **Canada Border Services Agency** - this is for quality assurance!
Assessment

Assessment is the process of (1) collecting data to show that analytical procedures are operating within specified limits and (2) verifying that final results meet use objectives. The progress towards a successful assessment (integration and summary of data) is continuously monitored (e.g., control charts, QA managers, etc.).

<table>
<thead>
<tr>
<th>Question</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Use Objectives</strong></td>
<td>• Write use objectives</td>
</tr>
<tr>
<td>Why do you want the</td>
<td>• Write specifications</td>
</tr>
<tr>
<td>data and results</td>
<td>• Pick methods to meet specifications</td>
</tr>
<tr>
<td>and how will you use</td>
<td>• Consider sampling, precision, accuracy, selectivity, sensitivity,</td>
</tr>
<tr>
<td>the results?</td>
<td>detection limit, robustness, rate of false results</td>
</tr>
<tr>
<td><strong>Specifications</strong></td>
<td>• Employ blanks, fortification, calibration checks, quality control</td>
</tr>
<tr>
<td>How good do the</td>
<td>samples, and control charts to monitor performance</td>
</tr>
<tr>
<td>numbers have to be?</td>
<td>• Write and follow standard operating procedures</td>
</tr>
<tr>
<td><strong>Assessment</strong></td>
<td>• Compare data and results with specifications</td>
</tr>
<tr>
<td>Were the specifications achieved?</td>
<td>• Document procedures and keep records suitable to meet use objectives</td>
</tr>
<tr>
<td></td>
<td>• Verify that use objectives were met</td>
</tr>
</tbody>
</table>
Method Validation

Method validation is the process of proving that an analytical method is acceptable for its intended purpose. For instance, in pharmaceutical chemistry, method validation requirements for regulatory submission include studies of method specificity, linearity, accuracy, precision, range, limit of detection, limit of quantitation, and robustness:

Specificity

Linearity

Accuracy

Precision

Range

Limit of Detection

Robustness

or “SLAPRLR”
Specificity

Specificity is the ability of an analytical method to distinguish analyte from everything else that might be in the sample.

An electropherogram is a graph of detector response versus time in an electrophoretic separation (electrophoresis). Below is an electropherogram of the drug cefotaxime (peak 4) spiked with 0.2 wt% of known impurities normally present from the synthesis.

Impurity peak 3 is not completely resolved from the cefotaxime. In this case, a reasonable criterion for specificity might be that unresolved impurities at their maximum expected concentration do not affect the assay of cefotaxime by more than 0.5%.

In the analysis of a drug formulation, we would want to compare the pure drug with one containing additions of all possible synthetic by-products and intermediates, degradation products, and excipients (substances added to give desirable form or consistency).

Degradation products might be introduced by exposing pure material to heat, light, humidity, acid, base, and oxidants to decompose ~20% of the original material.
Linearity

Linearity measures how well a calibration curve follows a straight line, showing that response is proportional to the quantity of analyte.

To prepare a calibration curve for an impurity that might be present at, say, 0.1 to 1 wt%, you might prepare a calibration curve with five standards spanning the range 0.05 to 2 wt%. The whole analysis would entail $3 \times 5 = 15$ standards plus three blanks.

A relatively superficial, but common, measure of linearity is the $R^2$ coefficient (which is always output on Excel when you select linear trendlines (or with the LINEST function).

$$R^2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})^2}{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}$$

$R^2$ must be very close to 1 to represent a linear fit. For a major component of an unknown, a value of $R^2$ above 0.995 or, perhaps, 0.999, is deemed a good fit for many purposes.

Another criterion: $y$-intercept of the calibration curve (after the response of the blank has been subtracted from each standard) should be close to 0.
Accuracy

*Accuracy* is “nearness to the truth.” To demonstrate accuracy:

1. Analyze a *certified reference material* in a matrix similar to that of your unknown.

2. Compare results from *two or more different analytical methods*. They should agree within their expected precision.

3. Analyze a *blank sample spiked with a known addition of analyte*. The matrix must be the same as your unknown. When assaying a major component, three replicate samples at each of three levels ranging from 0.5 to 1.5 times the expected sample concentration are customary. For impurities, spikes could cover three levels spanning an expected range of concentrations, such as 0.1 to 2 wt%.

4. If you cannot prepare a blank with the same matrix as the unknown, then it is appropriate to make standard additions of analyte to the unknown (next section...).
Precision

*Precision* is how well replicate measurements agree with one another, usually expressed as a standard deviation.

- *Instrument precision* (or *injection precision*), is the reproducibility observed when the same quantity of one sample is repeatedly introduced (≥10 times) into an instrument.

- *Intra-assay precision* is evaluated by analyzing aliquots of a homogeneous material several times by one person on one day with the same equipment.

- *Intermediate precision* (or *ruggedness*), is the variation observed when an assay is performed by different people on different instruments on different days in the same lab.

- *Interlaboratory precision* (or *reproducibility*), is the most general measure of reproducibility observed when aliquots of the same sample are analyzed by different people in different laboratories.
Range of detection

*Range* is the concentration interval over which *linearity*, *accuracy*, and *precision* are all acceptable.

An example of a specification for range for a **major component** of a mixture is the concentration interval providing:

- a correlation coefficient of $R^2 \geq 0.995$ (a measure of *linearity*)

- spike recovery of $100 \pm 2\%$ (a measure of *accuracy*), and

- interlaboratory *precision* of $\pm 3\%$.

(For an *impurity*, an acceptable range might provide a correlation coefficient of $R^2 \geq 0.98$, spike recovery of $100 \pm 10\%$, and interlaboratory precision of $\pm 15\%$)

Range has a lot of different meanings, and we hope to avoid confusing the use of range as a specification, with the other uses of range covered in lecture 4 (see below).

**Linear range**: concentration range over which calibration curve is linear.

**Dynamic range**: concentration range over which there is measurable response.
Limit of Detection

The *detection limit* (or *lower limit of detection*) is the smallest quantity of analyte that is “significantly different” from the blank. Curves are Student’s t distributions for 6 degrees of freedom and are broader than the corresponding Gaussian distributions.

Distribution of measurements for a blank and a sample whose concentration is at the detection limit. Only ~1% of measurements for a blank are expected to exceed the detection limit. However, 50% of measurements for a sample containing analyte at the detection limit will be below the detection limit. There is a 1% chance of concluding that a blank has analyte above the detection limit (false positive). If a sample contains analyte at the detection limit, there is a 50% chance of concluding that analyte is absent because its signal is below the detection limit (false negative).

Curves are Student’s t distributions for 6 degrees of freedom and are broader than the corresponding Gaussian distributions.
Setting the detection limit

1. After estimating the detection limit from previous experience with the method, prepare a sample whose concentration is ca. 1 to 5 times the detection limit.

2. Measure the signal from \( n \) replicate samples (\( n \geq 7 \)).

3. Compute the standard deviation (s) of the \( n \) measurements.

4. Measure the signal from \( n \) blanks and find the mean value, \( y_{\text{blank}} \).

5. The minimum detectable signal, \( y_{\text{dl}} \), is defined as:

\[
\text{signal detection limit: } \quad y_{\text{dl}} = y_{\text{blank}} + 3s
\]

6. The corrected signal, \( y_{\text{sample}} - y_{\text{blank}} \), is proportional to sample concentration:

\[
\text{calibration line: } \quad y_{\text{sample}} - y_{\text{blank}} = m \times \text{conc.}
\]

where \( y_{\text{sample}} \) is the signal observed for the sample and \( m \) is the slope of the linear calibration curve. The minimum detectable concentration (detection limit) is obtained by substituting \( y_{\text{dl}} \) for \( y_{\text{sample}} \)

\[
\text{detection limit } = \frac{3s}{m}
\]
Take it to the limit(s)

**Lower limit of detection**: when the signal is 3 times greater than the noise (3s/m), it is detectable, but still too small for accurate measurement.

**Lower limit of quantitation**: A signal that is 10 times greater than the noise or the smallest amount that can be measured with reasonable accuracy.

**Instrument detection limit**: obtained by replicate measurements \((n \geq 7)\) of aliquots from one sample.

**Method detection limit**: obtained by preparing \(n \geq 7\) individual samples and analyzing each one once (always greater than the instrument detection limit).

**Reporting limit**: the concentration below which regulations say that a given analyte is reported as “not detected,” which does not mean that analyte is not observed. It means that analyte is below a prescribed level. Reporting limits are set at least 5 to 10 times higher than the detection limit, so that detecting analyte at the reporting limit is not ambiguous.
Robustness is the ability of an analytical method to be unaffected by small, deliberate changes in operating parameters (i.e., how well the technique stands up!)

**e.g., chromatography**
A chromatographic method is robust if it gives acceptable results when small changes are made in solvent composition, pH, buffer concentration, temperature, injection volume, and detector wavelength.

In tests for robustness, the organic solvent content in the mobile phase could be varied by, say, ±2%, the eluent pH varied by ±0.1, and column temperature varied by ±5°C.
Standard Addition

In **standard addition**, known quantities of analyte are added to the unknown. From the increase in signal, we deduce how much analyte was in the original unknown. This method requires a linear response to analyte.

This analysis is particularly useful when the sample composition is unknown or complex and affects the analytical signal (i.e., there is potentially a **matrix effect**, an influence on the analytical response by something else in the sample other than the analyte).

**ClO\textsubscript{4}^- in water, and the matrix effect:**

This figure shows a strong matrix effect in the analysis of perchlorate by mass spectrometry. Perchlorate at a level above 18 \( \mu \text{g/L} \) in drinking water is of concern because it can reduce thyroid hormone production. Standard solutions of ClO\textsubscript{4}^- in pure water and ground water gave the upper and lower calibration curves, respectively. Reduction of the signal in the latter is a matrix effect attributed to other anions present in the groundwater.
Consider a standard addition in which a sample with unknown initial concentration of analyte $[X]_i$ gives a signal intensity $I_X$. Then a known concentration of standard, $S$, is added to an aliquot of the sample and a signal $I_{S+X}$ is observed for this second solution. Addition of standard to the unknown changes the concentration of the original analyte because of dilution. Let’s call the diluted concentration of analyte $[X]_f$, where “f” stands for “final.” We designate the concentration of standard in the final solution as $[S]_f$.

\[
\frac{[X]_i}{[S]_f + [X]_f} = \frac{I_X}{I_{S+X}}
\]

For an initial volume $V_0$ of unknown and added volume $V_S$ of standard with concentration $[S]_i$, the total volume is $V = V_0 + V_S$ and the concentrations above are:

\[
[X]_f = [X]_i \left( \frac{V_0}{V} \right) \quad [S]_f = [S]_i \left( \frac{V_S}{V} \right)
\]

The quotient (initial/final volume) is known as the *dilution factor*. Since we know $[X]_i$, $V_0$ and $V$, we can rearrange and easily solve for $[X]_i$. 
Standard Addition, 3

**Example**  Standard Addition

Serum containing Na⁺ gave a signal of 4.27 mV in an atomic emission analysis. Then 5.00 mL of 2.08 M NaCl were added to 95.0 mL of serum. This spiked serum gave a signal of 7.98 mV. Find the original concentration of Na⁺ in the serum.

**Solution**  From Equation 5-8, the final concentration of Na⁺ after dilution with the standard is \([X]_f = [X]_i\left(\frac{V_o}{V}\right) = [X]_i\left(\frac{95.0 \text{ mL}}{100.0 \text{ mL}}\right)\). The final concentration of added standard is \([S]_f = [S]_i\left(\frac{V_S}{V}\right) = (2.08 \text{ M})(5.00 \text{ mL/100.0 mL}) = 0.104 \text{ M}\).

Equation 5-7 becomes

\[
\frac{[\text{Na}^+]_i}{[0.104 \text{ M}]} + 0.950[\text{Na}^+]_i = \frac{4.27 \text{ mV}}{7.98 \text{ mV}} \Rightarrow [\text{Na}^+]_i = 0.113 \text{ M}
\]

**Test Yourself**  If spiked serum gave a signal of 6.50 mV, what was the original concentration of Na⁺? (Answer: 0.182 M)
Standard Addition: One solution

If the analysis does not consume solution, we begin with an unknown solution and measure the analytical signal. Then we add a small volume of concentrated standard and measure the signal again. We add several more small volumes of standard and measure the signal after each addition. Standards should be concentrated so that only small volumes are added and the sample matrix is not appreciably altered.

The concentration of ascorbic acid (vitamin C) was measured in orange juice by an electrochemical method (current proportional to concentration). Eight standard additions increased current from 1.78 to 5.82 μA, which is at the upper end of the desired range of 1.5- to 3-fold increase in analytical signal. The original concentration of the unknown can be found on the graph, and the theoretical response is:

\[ I_{S+X} \left( \frac{V}{V_0} \right) = I_X + \frac{I_X}{X_i} \left[ S_i \right] \left( \frac{V_S}{V_0} \right) \]

Graphical treatment of standard additions to a single solution with variable total volume. Standard additions should increase the analytical signal to between 1.5 and 3 times its original value (that is, B = 0.5A to 2A).
Standard Addition: Multiple solutions

If the analysis consumes solution, equal volumes of unknown are pipetted into several volumetric flasks. Increasing volumes of standard are added to each flask and each is diluted to the same final volume (i.e., each flask contains the same concentration of unknown and differing concentrations of standard). For each flask, a measurement of analytical signal, $I_{S+X}$, is made.

Plot $I_{S+X}$ versus $[S]_f$, and the $x$-intercept is $[X]_f$. The initial concentration of unknown, $[X]_i$, is calculated from the dilution applied to make the final sample.
Internal Standards

An *internal standard* is a known amount of a compound, different from analyte, that is added to the unknown. Signal from analyte is compared with signal from the internal standard to find out how much analyte is present. Internal standards are useful in cases where the quantity of sample or the instrument response varies slightly from run to run.

*e.g.*, GC or HPLC experiments, variations in gas/liquid flow may alter the response of the instrument, and change the magnitude of the observed signal.

**Calibration curves**: accurate under conditions where they were obtained

**Internal standards**: the *relative response* of the detector to both the analyte and standards is constant over a range of conditions

If the concentration of the standard is precisely known, then the concentration of the analyte (unknown) can be calculated.

**Another useful fact about internal standards**: Internal standards are desirable when *sample loss* can occur during sample preparation steps prior to analysis. If a known quantity of standard is added to the unknown prior to any manipulations, the ratio of standard to analyte remains constant because the same fraction of each is lost in any operation.
Internal Standards, 2

Prepare a known mixture of standard and analyte to measure the relative response of the detector to the two species.

In the chromatogram on the left, the area underneath peak is proportional to the concentration of the species; however, the response of the detector to each species is different in this case.

**e.g.,** [X] and [S] = 10 mM; however, the area under the S curve, $A_S$ is 2.3 times that of the X curve, $A_X$ (i.e., the response factor, $F$, is 2.3 times greater for S).

$$\frac{A_X}{[X]} = F \left( \frac{A_S}{[S]} \right)$$

where [X] and [S] are the concentrations of the analyte and standard after they have been mixed together (of course, this only works if the detector response is linear).