All laboratory analyses have error associated with them, however experienced the analyst or well-established the method, and it is therefore not possible to produce an ‘exact’ result in any analysis. Error in this context does not necessarily mean that the analyst has made a mistake or that there is a mistake in the method.

Errors in laboratory analysis can be classified as systematic errors or random errors. In this topic guide you will learn about these different types of error and the effect they have on analytical results.

On successful completion of this topic you will:
• understand how the quality of an analytical result may be expressed (LO1).

To achieve a Pass in this unit you will need to show that you can:
• discuss features of the quality of analytical results (1.1)
• evaluate different types of error (1.2)
• explain how errors may be minimised (1.3)
• express the distribution of results in statistical terms (1.4).
1 Systematic errors, bias and trueness

Systematic errors are errors that are constant or vary in a predictable way. When an analysis has a systematic error, the mean, $\bar{x}$, of the results of several replicate analyses of a substance will be higher or lower than the true value, $\mu$, for that substance.

The difference between the true value and the mean of a large number of replicate analyses is called the bias, $B$. Figure 4.2.1 shows the bias in an analysis.

![Figure 4.2.1: A biased analysis – results are on average greater than the true value.](image)

Systematic errors are errors that do not change between individual repeat analyses of the same sample. They do not average to zero, and the mean value is affected by systematic errors as much as any individual result. The systematic error might be a constant value, e.g. 0.25 mg mL$^{-1}$ (a constant error), or it might be a fixed proportion of a result, e.g. 0.6% (a proportionate error).

Systematic errors may be due to a number of causes including:
- instrument miscalibration (the expected instrument response is not obtained for known analyte concentrations)
- uncorrected blank values (a sample with no analyte gives a result other than 0 concentration)
- errors in standards (solutions with known concentrations of analyte), especially when an erroneous standard is used to produce other standards
- interferences (where other species in the sample affect the measurement of the analyte)
- matrix effects (where the measurement is affected by the major species in solution, which are present as a result of the sample preparation steps)
- losses during sample pretreatment, e.g. extraction, dissolution and separation.

Once the systematic error is known for an analysis, the analytical results should be corrected for the bias, although it is better (if possible) to identify and eliminate the cause of the error.

The systematic error or bias can be calculated as:

$$B = \bar{x} - \mu$$

The percentage bias, $\%B$, can also be calculated:

$$\%B = \left( \frac{\bar{x} - \mu}{\mu} \right) \times 100\%$$

Key terms

**Systematic error:** Errors that are constant or vary predictably in the analytical process. Systematic errors do not average to zero.

**Error:** The difference between an analytical result and the true value. Errors can be systematic or random.

**Bias:** The difference between the expectation of the test results and the true value.
Bias can also be given in the form of the recovery, $R$, where recovery is the measured value as a percentage of the true value:

$$\%R = \frac{\bar{x}}{\mu} \times 100\%$$

Recovery is often determined by spiking a sample – adding a known amount of analyte (‘the spike’) to a sample that contains no analyte and going through the whole analytical procedure. Recovery can be greater than 100% as well as less than 100%.

The trueness of an analysis is the closeness of agreement between an average value and the true value. It is a measure of the lack of bias – the lower the bias, the greater the trueness of a particular analysis.

**Activity 4.2.1**

A reference sample of whole milk powder was analysed for its fat content several times and gave an average value of 29.32 g per 100 g. The true value is 26.95 g per 100 g.

Calculate the bias, percentage bias and recovery for the analysis.

2 Random errors, normal distribution and standard deviation

When an analysis is performed on the same material a number of times it will give a range of test results. This is due to random errors in the analysis, which means that there are slight variations in the analysis each time.

**Figure 4.2.2** shows a spread of results where $x$ is a value obtained in the analysis and $n(x)$ is the number of times that value is obtained. This spread of results is called a normal distribution, in which results are distributed equally either side of the mean, and in which most results are close to the mean. The results average to $\bar{x}$, the mean value.

Random errors vary from analysis to analysis but average to zero in the long run. They can be eliminated by repeating the analysis a sufficiently large number of times and taking the mean value.
Examples of random error include:

- instrument noise
- variations in technique occurring at any stage in the analysis, for example, pipetting slightly less or slightly more sample than specified, or filling a burette slightly above or slightly below the mark, etc.
- variations in the calibration of different items of volumetric ware
- variations in the composition of reagents and solutions, whether purchased or made up in the lab.

Random errors are greater with inexperienced analysts than experienced analysts. Proper training and ongoing practice will improve an analyst’s technique, and reduce random errors.

The spread of results is measured by the standard deviation. If the number of results is small (a ‘sample’ of the whole set of possible results), the standard deviation is given the symbol \( s \). If the number of results is large, the standard deviation is given by the symbol \( \sigma \).

When there is a normal distribution:

- 68.3% of results will lie within one standard deviation of the mean, i.e. between \( \bar{x} - s \) and \( \bar{x} + s \) (shown in Figure 4.2.3)
- 95.5% of results will lie within two standard deviations of the mean, i.e. between \( \bar{x} - 2s \) and \( \bar{x} + 2s \) (shown in Figure 4.2.4)
- 99.7% of results will lie within three standard deviations of the mean, i.e. between \( \bar{x} - 3s \) and \( \bar{x} + 3s \).
Activity 4.2.2
The mean of a large number of analyses of aspirin in tablets is 302 mg, and the standard deviation is 8 mg.
Within how many standard deviations of the mean does each of the following measurements lie?
• 281 mg
• 307 mg
• 315 mg

Standard deviation of a sample, \( s \), can be calculated using the equation:

\[
s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}
\]

However, in a laboratory it is more likely that standard deviation is calculated using a computer spreadsheet package such as Microsoft Excel®.

In the example shown in Figure 4.2.5, the average and standard deviation of the mass of aspirin in eight tablets was determined in Excel® by selecting the ‘Formulas’ tab, then selecting ‘More functions’ then ‘statistical’ (or selecting ‘Insert function’ then selecting ‘Statistical’ in the category box) then selecting ‘AVERAGE’ and ‘STDEV.S’ (standard deviation of a sample).

**Precision**

The precision of an analysis is the closeness of agreement of a set of test results. A method that has low random errors, i.e. a narrow spread of results, is described as being precise, and the standard deviation will also be a small value. Precision is affected by many factors and therefore there are different levels of precision.

Repeatability is the level of precision when the test results are obtained with the same method in the same laboratory by the same operator within a short space of time.

Reproducibility is the level of precision when the test results are obtained with the same method but in different laboratories by different operators using different equipment.

**Accuracy**

The bias and precision combine to determine the accuracy of an analysis. Accuracy is defined as the closeness of an individual test result to the true value, and therefore it is affected by both systematic and random errors.

Figure 4.2.6 shows how trueness and precision both contribute to accuracy, and that the most accurate analysis has both trueness (lack of bias) and good precision. Accuracy is therefore really an absence of error, both systematic and random.

While it is quite common to hear or read that an analytical method is ‘accurate’, this is not actually the correct terminology. An individual test result is accurate if it is close to the true value. A method that gives accurate results shows good trueness and precision.
4.2: The quality of analytical data

Improving precision, decreasing bias, improving accuracy

Portfolio activity (1.1, 1.2, 1.4)
The Analytical Methods Committee (AMC) of the Royal Society of Chemistry has produced a number of ‘technical briefs’ on different aspects of analysis. See the link below:
http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/TechnicalBriefs.asp
Read AMC Technical Brief number 13, ‘Terminology – the key to understanding analytical science. Part 1: Accuracy, precision and uncertainty’. This technical brief summarises and discusses the concepts and terms introduced above.
Sign the technical brief to show you have read it and include it in your portfolio.

Measurement uncertainty

Since all analytical methods have inherent random and systematic errors there is always some uncertainty in any measurement. This means that it is highly unlikely that the result of an analysis will give the true value. What is important is that the results of the analysis are within acceptable tolerance limits. Measurement uncertainty is a quantitative estimate of the limits within which the true value of a measurand (such as the fat content of whole milk powder) is expected to lie, to a specified level of confidence. This is usually given as twice the standard deviation, and is referred to as the 95% confidence level (since 95% of results lie within two standard deviations of the mean value).

The sources of uncertainty fall into four general categories:
- short-term variability, i.e. repeatability
- longer-term variability covering uncertainty in calibration of equipment and inter-laboratory variability, i.e. reproducibility
- uncertainty in reference materials
- sampling effects (e.g. sample inhomogeneity), matrix effects and assumptions about the procedure (e.g. has a reaction gone to completion?).

The overall measurement uncertainty for an analysis is calculated by estimating and combining all the individual uncertainties (given the symbol u) that contribute to the analysis. The overall uncertainty is then multiplied by a factor.
of 2, called the coverage factor, to take into account the fact that the measurement uncertainty is quoted to two standard deviations (the 95% confidence level). The resulting value is called the expanded uncertainty, and is given the symbol U.

Measurement uncertainty for individual components is obtained from standard deviations from repeat measurements in an analysis, certified uncertainties in reference materials, and certified tolerances on equipment such as pipettes and burettes (where a value is given by a tolerance, this is usually divided by $\sqrt{3}$).

Activity 4.2.3
1 In an analysis, the pipette used to dispense aliquots of the sample has a manufacturer’s specification of 10.00±0.02 mL. Calculate the uncertainty in the pipette.
2 A reference material for heavy metals in rye grass has a certified value of 82±4 mg kg$^{-1}$ for manganese, with the uncertainty given as an expanded uncertainty. What is the uncertainty in the manganese value?

Take it further
To learn more about measurement uncertainty, look at the National Measurement System (NMS) website (www.nmschembio.org.uk), which is full of useful information.
The NMS has produced a useful leaflet which describes sources of uncertainty:

Outliers
Occasionally there will be an analytical result which is very different from the others – an ‘outlier’. The natural instinct is to reject this result on the assumption that ‘something went wrong’ with that particular test, i.e. a specific mistake or error occurred. You should bear in mind, however, that in a normal distribution, values can theoretically occur from $-\infty$ to $+\infty$ due to random error. There is a very small but distinct probability that a result may genuinely lie more than two or even three standard deviations away from the mean.

Take it further
Read AMC Technical Brief number 39, ‘Rogues and suspects: How to tackle outliers’. (Use the link below.)
http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/TechnicalBriefs.asp

Key terms
Aliquot: A portion of a whole, for example a sample of solution that is used for analysis.
Outlier: A member of a set of values which is inconsistent with the rest of the set.
3 Minimising errors in analysis

Analyses are made more accurate by reducing the systematic and random errors that may occur. The accuracy of analyses can be improved in a number of ways.

Systematic errors often occur in sample extraction and dissolution steps, so the sample preparation should be optimised. To avoid samples decomposing before analysis, sample storage has to be appropriate, for example, storing samples in fridges or freezers or in the dark, as appropriate. The analytical methods must be appropriate for the types of sample and they should be validated (see Topic guide 4.3, section 1).

All types of quantitative analysis (measuring how much sample is present) will require some form of calibration that may range from the calibration of balances and glassware to the production of calibration charts. If there are interfering species present, these should be identified, and if there is no suitable alternative analytical method, a means of overcoming the interferences must be used. This may include, for example, adding suitable reagents that combine with or remove the interferent, or selecting an alternative wavelength at which to measure absorbance.

Random errors can be reduced by ensuring that staff are trained to perform the analyses correctly and consistently. Errors that give rise to test results that are very different from other test results may be due to ‘sporadic blunders’. Sporadic blunders often arise from a lapse of concentration leading to sample contamination, sample mislabelling or data transcription errors, or from a temporary instrumental problem.

Contamination can be avoided by, for example, ensuring lab coats, benches and equipment are clean and ensuring that disposable equipment (such as gloves and pipette tips) are used only once.

Samples should be double-checked to ensure that they are correctly labelled and that results are assigned to the correct sample.

Data handling errors can be reduced by double-checking values when writing data into lab note books or when transferring data, and performing calculations on computers using preset programmes.

A laboratory information management system (LIMS) can significantly reduce sample misidentification and transcription errors by barcoding samples and having data transferred electronically from instrument to lab reports.

**Link**
The material in this topic guide is relevant to Unit 3: Analysis of Scientific Data and Information. In Unit 3, you will learn more about presenting data, processing data using statistics, classifying and handling errors and understanding the limitations of analytical results.

**Portfolio activity (1.3)**
Read AMC Technical Brief number 49 ‘Sporadic blunders’ (use the link: [http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/TechnicalBriefs.asp](http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/TechnicalBriefs.asp)).

This technical brief discusses the causes of sporadic blunders, and how they can be detected and eliminated.

Sign the technical brief to show you have read it and include it in your portfolio.
4.2: The quality of analytical data

Checklist
At the end of this topic guide you should:
✓ understand what is meant by trueness, precision and accuracy
✓ be able to express the accuracy of an analytical result in terms of bias, recovery and standard deviation
✓ understand that every analysis is subject to measurement uncertainty
✓ be aware of ways in which systematic and random errors can be reduced.

Further reading
Many textbooks cover statistics relevant to their subject area, for example:
Analytical Chemistry (S.J. Higson, 2003), Oxford

Acknowledgements
The publisher would like to thank the following for their kind permission to reproduce their photographs:
Shutterstock.com: Olivier
All other images © Pearson Education
Every effort has been made to trace the copyright holders and we apologise in advance for any unintentional omissions. We would be pleased to insert the appropriate acknowledgement in any subsequent edition of this publication.

References
ISO 5725-1:1994 Accuracy (trueness and precision) of measurement methods and results – Part 1 General principles and definitions