Approaches to uncertainty estimation

'Bottom-up' vs 'Top-down'

Science
for a safer world

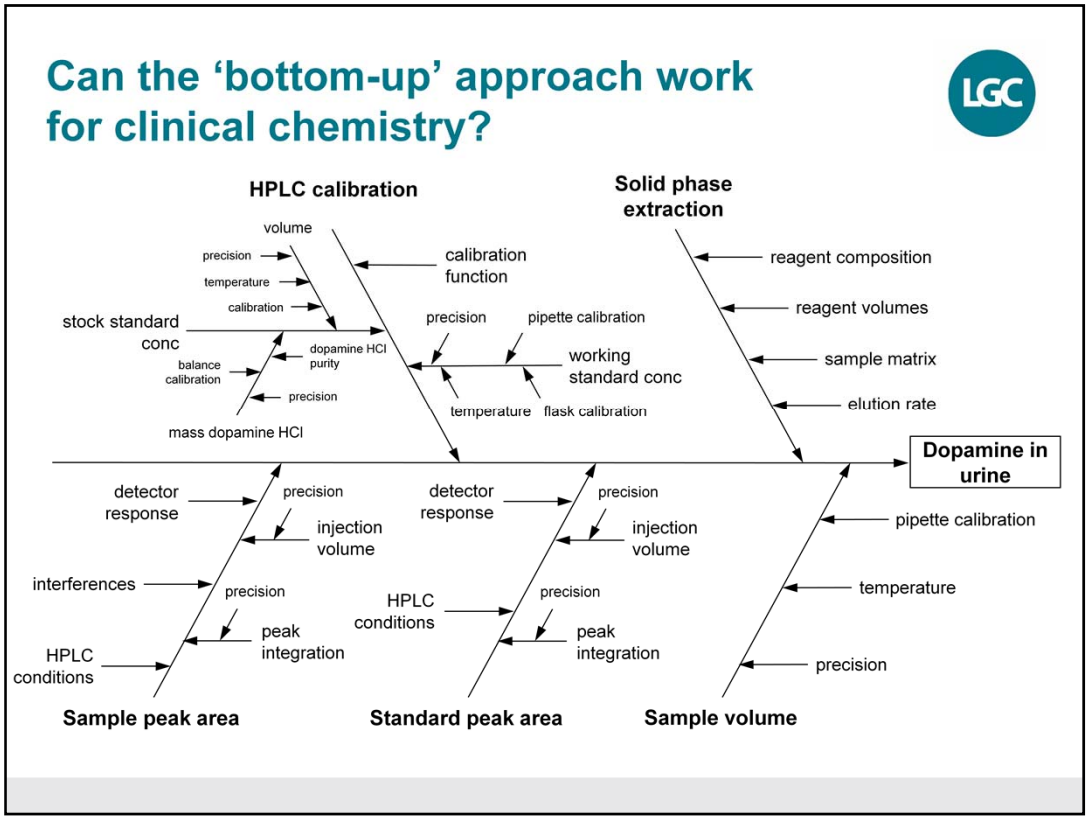
Overview

- ‘Bottom-up’ vs ‘top-down’ approach to uncertainty estimation
- Using validation and quality control data in uncertainty estimation
  - requirements for the top-down approach
- Sources of data

ISO approach – 'bottom-up'

- Write an equation that completely describes the measurement system
  - includes all parameters that could influence the measurement result
- Evaluate the uncertainties associated with all parameters in the equation
  - Type A: statistical evaluation, Type B: any other data (certificates, instrument specifications, etc)
- Express all uncertainties as standard deviations
- Combine using mathematical rules for the combination of variances
- Apply a suitable coverage factor

The slide shows the approach to uncertainty evaluation outlined in the ISO 'GUM'. There are a number of problems associated with attempting to apply this 'top-down' approach to chemical or bio-analysis. One issue is that it is generally difficult to write an equation that completely describes the measurement process.



The slide shows a 'cause-and-effect' diagram for a test method for the determination of dopamine in urine using high performance liquid chromatography. Each branch represents a source of uncertainty that could contribute to the uncertainty in the final result. Following the 'bottom-up' approach would require the evaluation of each of the individual uncertainty contributions. This would obviously be extremely difficult, if not impossible.

Problems

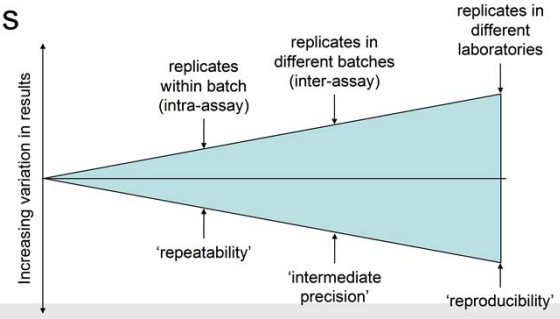
- Difficult to write an equation that includes all influence factors
  - what about sample clean-up conditions, recovery of analyte from matrix, instrument conditions, interferences….
- Difficult (impossible?) to evaluate individual uncertainty components
- Process is too time consuming and unworkable in routine testing laboratories
  - a ‘reasonable estimation’ is required

While there will be an equation available to calculate the final result of an analysis (even if it is hidden in instrument software!), it is unlikely to include terms representing all of the uncertainty components. As mentioned previously, it is generally difficult to produce a mathematical model that completely describes the measurement process.

The full ‘bottom-up’ approach has been found to be inappropriate for testing laboratories carrying out chemical or biochemical analyses. An alternative approach, which is still compliant with the principles of the GUM, has therefore been developed. This approach makes use of method performance data.

Key terminology: (im)precision

- A measure of the variability in results from one measurement to the next
  - expressed as a standard deviation, relative standard deviation (or %coefficient of variation, %CV)
  - distribution of random errors determines the precision
- Precision estimate is obtained under specified measurement conditions



This workshop focuses on the use of method performance data in uncertainty estimation. It is therefore important to have a clear understanding of the key parameters used to describe method performance.

Precision is defined in ISO 3534-2 as, 'The closeness of agreement between independent test/measurement results obtained under stipulated conditions'.

Precision is therefore a measure of the spread of repeated measurement results and depends only on the distribution of random errors – it gives no indication of how close those results are to the true value.

Precision is estimated by making repeated independent measurements on identical samples. The precision is usually expressed as the standard deviation (or relative standard deviation) of the results. 'Independent measurements' means that the whole analytical process must be replicated, including any sample preparation.

The conditions under which the independent replicate measurements are made will determine the type of precision estimate obtained.

Repeatability (within-batch precision) refers to a precision estimate obtained when measurement results are produced in one laboratory, by a single analyst using the same equipment over a short timescale. Repeatability gives an indication of the short-term variation in measurement results.

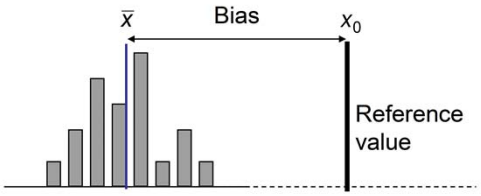
Reproducibility refers to a precision estimate obtained when measurement results are produced by different laboratories (and therefore by different analysts using different pieces of equipment).

If a single laboratory is studying the performance of a method for its own use, a study of repeatability is likely to underestimate the real variation in results when the method is used routinely. Laboratories should therefore consider evaluating the intermediate precision (also known as within-laboratory reproducibility). This involves making replicate measurements on different days, under conditions which mirror, as far as possible, the conditions of routine use of the method (e.g. measurements made by different analysts).

[ISO 3534-2:2006 Statistics - Vocabulary and symbols - Part 2: Applied statistics]

Key terminology: Bias

- Difference between observed mean value and reference value
  - bias is a measure of *trueness*



- Bias expressed as:
  - absolute value $B = \bar{x} - x_0$
  - percentage $\%B = \frac{(\bar{x} - x_0)}{x_0} \times 100$
  - %recovery $\%R = \frac{\bar{x}}{x_0} \times 100$

- Reference values from:
  - certified reference material
  - material prepared in-house (e.g. spiked sample)
  - reference method
  - External Quality Assessment (EQA) samples

Trueness is defined in ISO 3534-2 as, 'The difference between the expectation of a test result or measurement result and a true value', with a note that the measure of trueness is usually expressed in terms of bias. In practice, the true value is replaced by an accepted reference value (e.g. the concentration of the analyte in a certified reference material). Bias represents the total systematic error.

Evaluating bias generally involves carrying out repeat analysis of a suitable material containing a known amount of the analyte (this is the reference value). One of the problems facing an analyst when planning a study of method bias is the selection of a suitable reference value. There are a number of options. In chemical analysis the most common are:

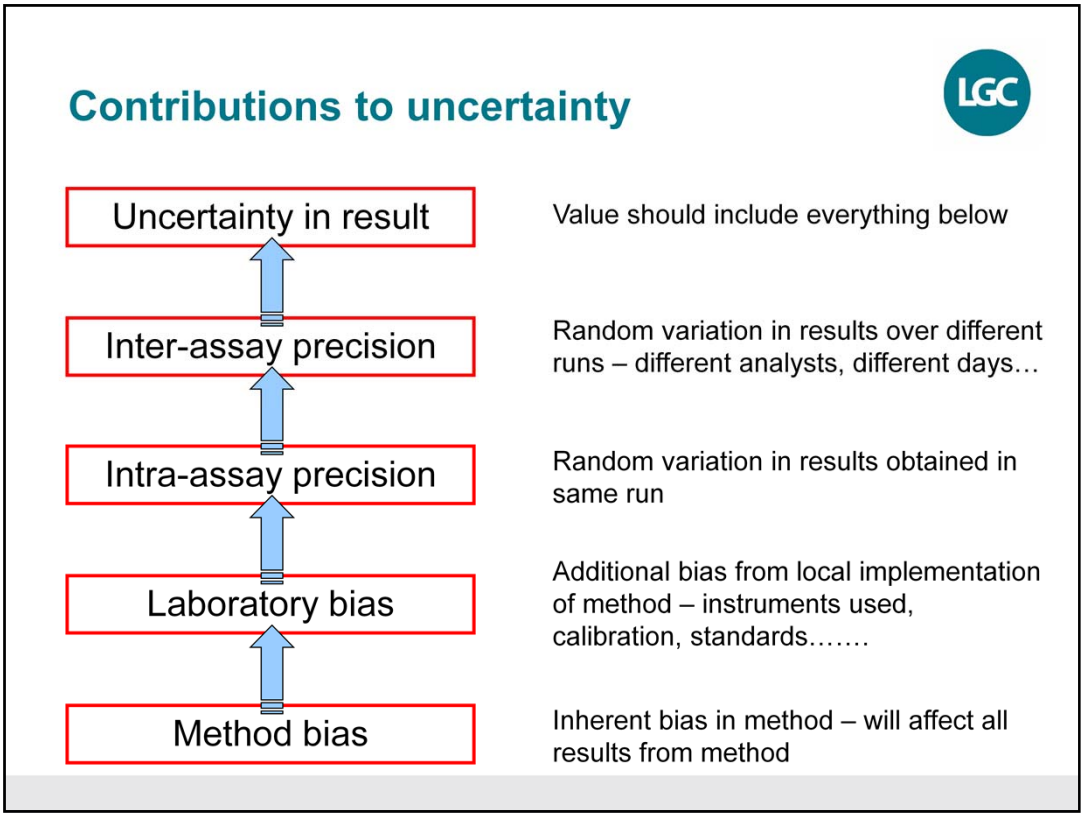
- certified reference material (CRM)
- spiked test samples
- reference method.

A certified reference material is a material that has been produced and characterised to high standards and that is accompanied by a certificate stating the value of the property of interest (e.g. the concentration of a particular analyte) and the uncertainty associated with the value. Unfortunately, compared to the very large number of possible analyte/matrix combinations, the number of CRMs available is relatively limited so a suitable material may not be available. An alternative is to prepare a reference sample in the laboratory by spiking a previously analysed sample with an appropriate amount of the analyte of interest.

Bias can also be evaluated by comparing results obtained from a reference method with those obtained using the method under study. If this approach is taken, the evaluation can be carried out using test samples rather than a CRM or spiked sample.

It may also be possible to evaluate bias using results obtained from the analysis of EQA samples. This is discussed in a later session.

[ISO 3534-2:2006 Statistics - Vocabulary and symbols - Part 2: Applied statistics]



An uncertainty estimate should take into account everything that is known about the measurement process.

In any test method there will be a number of sources of error, both random and systematic, which will influence the measurement result. The uncertainty estimate should include both random and systematic effects.

The slide shows the general sources of uncertainty in a measurement process. There will be a large number of factors that will contribute to the parameters identified on the slide. Note that for empirical methods* the method bias is zero. However, there may still be a laboratory bias due to the way a particular laboratory has implemented an empirical method.

* An ‘empirical method’ is a standardised method agreed on for comparative measurement within a particular field of application where the measurand depends upon the method in use. The empirical method used accordingly defines the measurand.

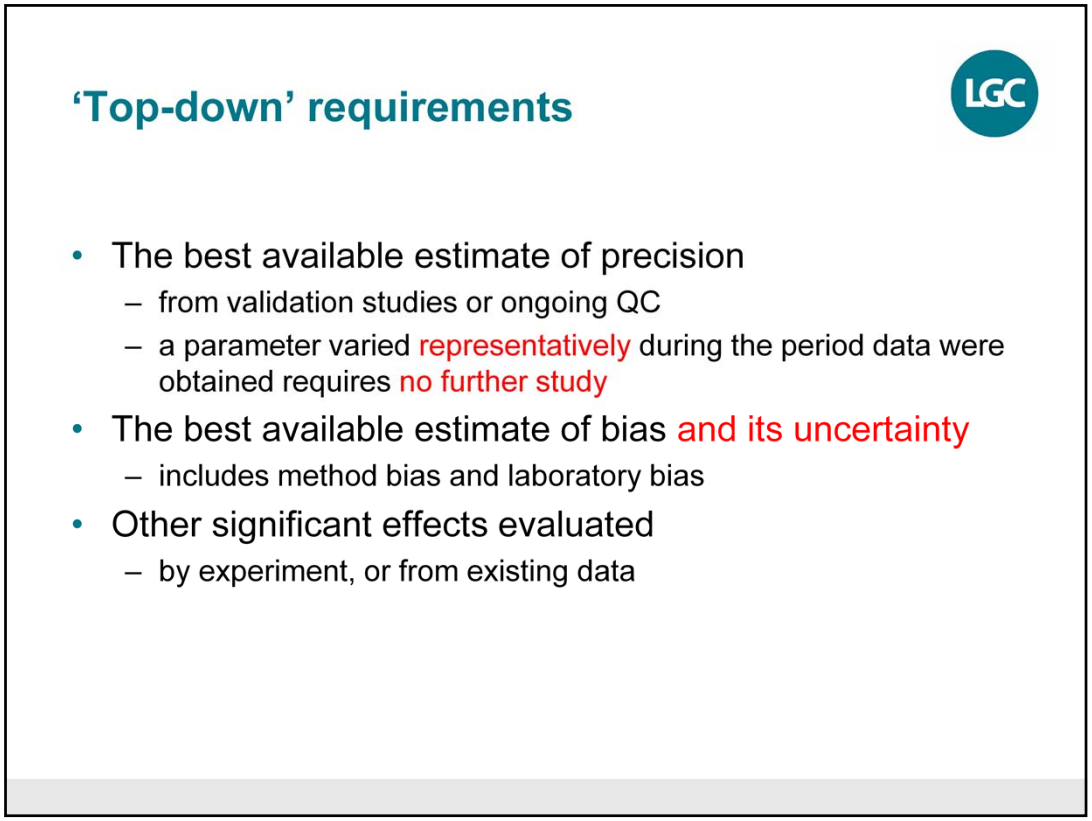
‘Top-down’ approach

- Use method performance data
  - validation data on precision and bias
  - ongoing internal quality control (IQC) data
  - EQA data
- Capture the effect of a number of sources of uncertainty
- Look at the variation in method *outputs* (i.e. results) rather than method *inputs*
- Cover method scope
  - matrix, analyte concentration

The ‘top-down’ approach makes use of method performance data which evaluate the variation in method outputs (i.e. the results obtained) rather than considering the effect on the results of individual inputs (i.e. the individual steps of the method).

The aim is to obtain method performance data that capture the effect of variations in as many of the inputs as possible. Data of this type is available for many methods from method validation studies and ongoing quality control.

It is important to remember that the uncertainty study should cover the method scope. A representative range of sample types, taking into account different sample matrixes and analyte concentrations should therefore be studied as required by the method scope.



The most important inputs to a 'top-down' uncertainty estimate are described below.

• The best available estimate of precision

The precision contribution should be estimated as far as possible over an extended time period, and chosen to allow natural variation of all factors affecting the result. For established methods, data from the ongoing analysis of QC materials is a useful source of information. For new methods being validated in-house, an estimate of within-laboratory reproducibility (intermediate precision) is appropriate.

• The best available estimate of bias and its uncertainty

An estimate of overall bias (including the effects of method and laboratory bias) requires a reliable and traceable reference value. The issues associated with finding a suitable reference are discussed in a later session.

• Other significant effects evaluated

Where an effect is not clearly covered by the available precision or bias data, additional data are needed. Evaluating some additional effects may involve experimental work; others may be resolved simply by inspection of known uncertainties in similar systems. For example, weighing uncertainties are typically very small and need not normally be studied explicitly once established.

Evaluating precision

- Aim to cover as many sources of variation as possible
  - extended time period, different analysts, different calibration standards, environmental conditions
- A parameter varied representatively during a precision study requires no further evaluation
- Types of data
  - method validation study (intermediate precision)
  - quality control data – repeated analysis of QC materials
  - data form interlaboratory studies (method validation or EQA)

The key to the top-down approach is a sound estimate of precision. A repeatability estimate is unlikely to be sufficient. When validating a method, the aim is to plan a precision study that captures as many of the sources of variation in a method as possible.

For established methods, data from the analysis of quality control materials can also provide a useful estimate of long term precision.

The key issue to remember is that a parameter varied representatively during the generation of the precision data requires no additional study.

The precision estimate should hopefully include contributions from the majority of the random effects associated with the method. If the data were obtained over a sufficient time period, effects that would normally be considered to be 'systematic' (e.g. calibration of equipment) may also be covered if, for example, the equipment has been recalibrated a number of times.

Evaluating bias

- A reasonable estimate of the bias can be obtained from
  - validation data (using CRMs or spiked samples)
  - EQA data (depending on the nature of the scheme/samples)
- Is the bias significant?
  - statistically significant?
  - significant compared to the method precision?
- Bias and its uncertainty should be considered as part of the uncertainty evaluation process

Obtaining an estimate of the precision is generally relatively straight forward. The true value for the material analysed to generate precision data doesn't need to be known - it is only the variation in results that is of interest. In addition, precision data generally provide 'Type A' uncertainty estimates (i.e. a standard deviation is calculated directly from experimental observations).

Evaluating bias is somewhat more complex. First, a suitable reference value is required. Second, we need to consider whether the observed bias is significant. Most bias estimates will be non-zero. However, when the precision of the measurement results and the uncertainty in the reference value are taken into account the observed bias may not be significantly different from zero. Finally we need to decide how to include any bias in the uncertainty estimate (e.g. how might we combine an estimate of precision with an estimate of bias?). It is this last point which is still open to debate. The ISO guide requires that results are corrected for all known significant biases. However, in many chemical analyses, results are not routinely corrected. Without the correction, a known bias is being ignored. If the uncorrected result is reported with its uncertainty, the range will not include the best estimate of the 'true value' – a simple report of the result and its uncertainty is therefore misleading.

Summary

- The ‘bottom-up’ approach is impractical for many test methods
- The ‘top-down’ approach utilises method performance data
  - requires a reliable estimate of method precision and information on bias
  - available from method validation studies, IQC and EQA