

ANALYTICAL CHEMISTRY · STUDY NOTES

Statistics, Calibration and Validation of Analytical Methods

From measurement to validated method: theory, criteria and worked examples

BSc in Biotechnology — 2nd year
University of Teramo
Prof. Federico Fanti · Academic Year 2026/2027

How to use these notes

These notes accompany the lectures *Applied statistics for analytical methods* and *Calibration and validation of analytical methods* of the Analytical Chemistry course. They are designed for self-study: every topic in the slides is explained in full, with the reasoning behind the formulas, step-by-step worked examples and the numerical criteria of the international guidelines. You can use them both to prepare for the exam and as a reference during laboratory practicals.

How the notes are organized

Part I (Chapters 1–8) builds the statistical tools: uncertainty and measurement errors, precision and trueness, descriptive statistics, distributions and inference, outlier tests, measurement uncertainty and linear regression. **Part II** (Chapters 9–17) applies them to the calibration and validation of methods: reference standards, calibration methods, validation parameters, quality control, matrix effect, stability, the regulatory framework and the practical aspects of routine work. The **Appendices** contain the formula sheet (A), the statistical tables (B), an Italian–English glossary (C) and the bibliographic references (D).

The two parts are linked: when a concept has already been explained, the text refers to the corresponding chapter instead of repeating it. For this reason it is best to study Part I before Part II.

The boxes

In the text you will find seven types of boxes, each always in the same color:

Definition

Formal definitions of key terms, according to the VIM, ISO 5725 or the ICH guidelines.

Formula

Important formulas, with the meaning of each symbol and the units of measurement.

Worked example

Numerical problems solved step by step. Redo them yourself with a calculator before reading the solution.

Watch out

The most common mistakes and typical pitfalls, including those that appear in exam questions.

Going further

Details for those who want to go further: they are not essential for the exam, but they help understanding.

Key points

The key points to remember, at the end of each chapter.

Test yourself

The topics of the self-assessment quiz to practice on for the chapter you have just studied.

How to study with the quiz

These notes are linked to the self-assessment quiz (file *Quiz-Statistica-Calibrazione-Validazione.html*, which opens in any browser). An effective approach is:

1. read the chapter and redo the worked examples by hand;
2. open the quiz in **Practice** mode, select the part and the topics listed in the *Test yourself* box, and read the explanations of the answers you got wrong;
3. go back to the text for the topics in which you made mistakes;
4. once you have completed both parts, try the **Exam simulation**: 30 questions in 30 minutes, with balanced difficulty.

Conventions

Decimal numbers are written with a decimal point (0.026 mg/L), and units of measurement follow the International System of Units (SI). Technical terms are those used in the international guidelines (for example *trueness*, in Italian *esattezza*); their Italian equivalents are collected in the glossary of Appendix C. The main symbols are summarized in the following table.

Table 0.1 — Main symbols used in these notes

Symbol	Meaning
$\bar{x}$	arithmetic mean of the data
s, σ	sample and population standard deviation
n, ν	number of data points, degrees of freedom
$\mathrm{RSD} = \mathrm{CV}\%$	relative standard deviation (coefficient of variation)
u, u_c, U, k	standard uncertainty, combined standard uncertainty, expanded uncertainty, coverage factor
b_0, b_1	intercept and slope of the calibration line
s_B	standard deviation of the blank signal
r, r^2	correlation coefficient and coefficient of determination
LOD, LOQ, LOL	limit of detection, quantification and linearity
LLOQ, ULOQ	lower and upper limits of quantification of the calibration curve
IS, SIL-IS	internal standard, stable isotope-labeled internal standard

The guidelines cited are those in force in the academic year 2026/2027: ICH M10 (2022) for bioanalysis, ICH Q2(R2) and Q14 (2023) for pharmaceutical methods, the Eurachem guide *The Fitness for Purpose of Analytical Methods* (3rd edition, 2025), document SANTE/11312/2021 v2026 for pesticide residues, and Regulation (EU) 2021/808 for residues of veterinary drugs.

Contents

Part I — Applied Statistics for Analytical Methods 8
1. Measurement, uncertainty and analytical data 8
1.1 Why every measurement has an uncertainty 8
1.2 How to express a result correctly 8
1.3 From data to information to knowledge 9
1.4 Types of data 9
1.5 The analytical signal and the calibration curve 10
2. Measurement errors 11
2.1 Error and uncertainty: two distinct concepts 11
2.2 Absolute error, relative error and bias 12
2.3 The three types of error 13
2.4 Systematic errors: the sources 13
2.5 Constant and proportional systematic errors 13
2.6 How to detect and eliminate systematic errors 14
2.7 Random errors 15
3. Precision, trueness and accuracy 16
3.1 Definitions according to the VIM and ISO 5725 16
3.2 The four cases: the targets 17
3.3 Repeatability, intermediate precision and reproducibility 18
4. Descriptive statistics 19
4.1 Sample and population 19
4.2 Measures of central tendency 20
4.3 Measures of dispersion 20
4.4 A complete exercise: Fe(III) in a water sample 22
5. Distributions, inference and outliers 23
5.1 The normal distribution 23
5.2 Symmetric and skewed distributions 24
5.3 The standard error of the mean 25
5.4 Student's t distribution 25
5.5 The confidence interval of the mean 26
5.6 Outliers 26
6. Measurement uncertainty 29
6.1 Error and uncertainty 29
6.2 The GUM: Type A and Type B contributions 30
6.3 Type A evaluation 30

6.4 Type B evaluation 31
6.5 Combined standard uncertainty 31
6.6 Expanded uncertainty 32
6.7 The uncertainty budget 33
6.8 How to report the result 33
7. Linear regression and the calibration curve 34
7.1 Calibration and regression 34
7.2 The simple linear regression model 35
7.3 The method of least squares 35
7.4 Interpolation, extrapolation and inverse prediction 36
7.5 The scatter plot and the choice of model 37
7.6 Decomposition of variability, r^2 and r 38
7.7 A complete example: spectrophotometric calibration of Fe(III) 40
8. Regression diagnostics and the link to validation 41
8.1 Residual analysis 41
8.2 Homoscedasticity, heteroscedasticity and weighted regression 42
8.3 Effect of systematic errors on the calibration curve 43
8.4 Link to method validation 44
8.5 Summary of the key concepts of Part I 45
Part II — Calibration and Validation of Analytical Methods 46
9. Standards and reference materials 46
9.1 The hierarchy of analytical standards 46
9.2 Metrological traceability 48
9.3 From the PS to working standards 48
10. Calibration methods 50
10.1 Absolute and relative methods 50
10.2 The analytical blank 50
10.3 Choice of signal: peak area or height 51
10.4 External standard calibration 51
10.5 The calibration curve: operational requirements 52
10.6 Internal standard calibration 53
10.7 The method of standard additions 54
10.8 Comparison of calibration methods 55
11. Validation: purpose, selectivity, sensitivity, LOD/LOQ and linear range 56
11.1 Why validate an analytical method 56
11.2 Validation parameters 57
11.3 Selectivity and specificity 57
11.4 Calibration sensitivity and analytical sensitivity 58

11.5 Limit of detection and limit of quantification 59
11.6 Linear dynamic range and calibration range 60
12. Trueness, precision and quality control 62
12.1 Trueness and bias 62
12.2 Accuracy = trueness + precision 63
12.3 Precision: repeatability, intermediate precision, reproducibility 63
12.4 The Horwitz curve 63
12.5 Quality control (QC) samples 65
12.6 Performance at low and high levels 66
12.7 Consistency across the whole range 66
13. Recovery, matrix effect and process efficiency 67
13.1 Definitions and the Matuszewski scheme 67
13.2 The formulas 68
13.3 The matrix effect in LC-MS/MS 69
13.4 The IS-normalized matrix factor 69
13.5 Evaluation in the guidelines 70
13.6 Detecting and reducing the matrix effect 70
13.7 Recovery at the different concentration levels 71
14. Stability, robustness, ruggedness, carry-over and dilution integrity 72
14.1 Analyte stability 72
14.2 Robustness 74
14.3 Ruggedness 74
14.4 Carry-over 74
14.5 Dilution integrity 75
15. The regulatory framework 76
15.1 The bodies and the nature of the documents 76
15.2 Bioanalysis: ICH M10 (2022) 76
15.3 ICH Q2(R2) and ICH Q14 (2023): methods for pharmaceutical products 77
15.4 Food, veterinary drugs and laboratories in general 78
15.5 Comparison of the documents 79
16. From method to routine: practical aspects 79
16.1 Instrumental methods and main techniques 80
16.2 The analytical process 80
16.3 Uncertainty propagation in serial dilutions 81
16.4 Monitoring the curve over time 81
16.5 The intercept and forcing through the origin 82
16.6 Outliers and curve acceptance 83
16.7 The stable isotope-labeled internal standard (SIL-IS) 83

16.8 Summary of the ICH M10 criteria 84
16.9 Full validation, partial validation and cross-validation 85
16.10 Validation plan and report 85
16.11 Common mistakes in building the curve 85
16.12 The analytical sequence in LC-MS/MS 85
17. Summary: from calibration to validation 87
17.1 The big picture 87
17.2 The key messages 87
17.3 Further reading 88
Appendices 89
Appendix A — Formula sheet 89
A.1 Descriptive statistics (Chapter 4) 89
A.2 Measurement errors (Chapters 2–3) 89
A.3 Inference and outliers (Chapter 5) 90
A.4 Measurement uncertainty (Chapter 6) 90
A.5 Linear regression (Chapters 7–8) 90
A.6 Calibration (Chapters 9–10 and 16) 91
A.7 Validation parameters (Chapters 11–14) 92
A.8 ICH M10 acceptance criteria (chromatographic methods) 92
Appendix B — Statistical tables 93
B.1 Critical values of Student's t (two-tailed) 93
B.2 Critical values for Dixon's Q test 94
B.3 Critical values for the Grubbs test 94
B.4 Coverage factor and confidence level 95
Appendix C — Italian–English glossary 95
Appendix D — References 98
D.1 Guidelines and standards 98
D.2 Articles 98
D.3 Reference textbooks 99

Part I — Applied Statistics for Analytical Methods

Every chemical measurement is affected by errors. This part provides the statistical tools to describe data, estimate their uncertainty, recognize outliers and build a reliable calibration line: the foundation on which the validation of analytical methods in Part II rests.

1. Measurement, uncertainty and analytical data

Every decision based on a chemical analysis rests on a number, and that number is useful only if we know how far we can trust it. Does a blood glucose level of 7.1 mmol/L really exceed the diagnostic threshold of 7.0 mmol/L? It depends on the uncertainty. This chapter explains how a result is expressed, what types of data a laboratory produces and how the signal of an instrument becomes a concentration.

1.1 Why every measurement has an uncertainty

No chemical analysis gives results free of uncertainty. The instrument, the operator, the environment, the sample and the procedure all introduce small deviations whose size and sign are partly unknown. The true value of the **measurand** (the quantity intended to be measured) therefore remains unknown, and every result is an estimate. The analyst's task is to **minimize errors and estimate their size with acceptable accuracy**. A reliable measurement follows three steps, in this order:

1. **establish the maximum tolerable error before starting the analysis**: it depends on the purpose (*fitness for purpose*) and guides the choice of method, glassware and number of replicates;
2. **perform replicate measurements**, to estimate the variability and calculate the mean value;
3. **express the result as value ± uncertainty**, with the unit of measurement and the correct significant figures.

1.2 How to express a result correctly

An analytical result is written in the following form, completed by the unit of measurement:

$$x = \bar{x} \pm U \tag{1.1}$$

where $\bar{x}$ is the best estimate (usually the mean of the replicates) and U is the **expanded uncertainty**. For example: blood glucose = (5.42 ± 0.08) mmol/L. The expanded uncertainty in Equation 1.1 is the combined standard uncertainty u_c, which gathers all the contributions (repeatability, calibration, weighings, volumes; Chapter 6), multiplied by a **coverage factor** k:

$$U = k \cdot u_c \tag{1.2}$$

In Equation 1.2, $k = 2$ gives a level of confidence of **about 95%**, and $k = 3$ about 99.7%. These correspondences hold if the distribution is approximately normal and the degrees of freedom are sufficient. U is **not** the standard deviation of the replicates, nor an error or a bias: calculating the latter would require the true value (Chapter 2). The uncertainty is often expressed in relative form, as a percentage of the result:

$$U_{\mathrm{rel}} = \frac{U}{\bar{x}} \times 100 \tag{1.3}$$

For the blood glucose example, Equation 1.3 gives $U_{\mathrm{rel}} = 0.08/5.42 \times 100 = 1.5\%$. The full width $2U$, which would give 3.0%, is not used.

1.2.1 Significant figures and rounding

Significant figures are all the digits known with certainty plus the first uncertain digit. The rules for counting them are:

- nonzero digits, and zeros between them, are significant;
- **leading zeros** are not: they only locate the decimal point;
- **trailing zeros after the decimal point** are significant, because they indicate a measured digit;
- trailing zeros in an integer are ambiguous: scientific notation is used (1.20×10^3 mL has three significant figures).

Thus 0.02050 g has **four** significant figures (2, 0, 5, 0), as does 20.50 mg. In intermediate calculations all digits are kept and **only the final result is rounded**. Four practical rules apply to a result with its uncertainty:

1. the uncertainty is rounded to **1–2 significant figures** (preferably two if the first digit is 1 or 2);
2. the central value is rounded **to the same decimal place as the uncertainty**;
3. the **unit of measurement** is always given;
4. the **coverage factor** and the level of confidence are stated (for example $k = 2$, about 95%).

Example 1.1 — Rounding a result with its uncertainty

Data: $\bar{x}$ = 12.3467 mg/L, U = 0.0832 mg/L ($k = 2$).

1. Round the uncertainty to one significant figure: 0.0832 → 0.08 mg/L, which ends at the hundredths.
2. Round the central value to the hundredths: 12.3467 → 12.35 mg/L.

Result: (12.35 ± 0.08) mg/L ($k = 2$, about 95%). With two digits in the uncertainty, (12.347 ± 0.083) mg/L is also acceptable. (12.3 ± 0.08), which loses a digit, and (12.347 ± 0.08), with different decimal places, are wrong. Also wrong are "5.42 ± 0.08" without a unit and an uncertainty with three significant figures, such as 0.00427.

1.3 From data to information to knowledge

Data are the raw measurement of a phenomenon, for example an absorbance or the area of a peak. On their own they have no meaning. **Analysis**, that is, processing and correlating the data (blank subtraction, calibration curve, calculation of the mean and uncertainty), turns them into **information**, a processed result: "glucose = (5.42 ± 0.08) mmol/L". A **model**, that is, an interpretive framework, turns information into **knowledge**, that is, scientific understanding. The composition of blood takes on meaning only through a model of the human body and its diseases: only then does 5.42 mmol/L become "normal fasting blood glucose". An error in the data propagates along the whole chain.

1.4 Types of data

Quantitative data are a **numerical value with a unit of measurement** ("the water temperature is 298.2 K"). They are the basis of the *hard sciences*, disciplines founded on rigorous data linked by mathematical models; Galilean science is founded on quantitative data. **Qualitative data** are labels, descriptors or categories ("the water is hot", "blood group A positive"): they are hard to standardize and reproduce, as in sensory analysis. Among the data about an individual, height, weight and blood glucose concentration are quantitative; skin color and blood group are qualitative.

Quantitative data can be **discrete**, i.e. integer counts (number of colonies or cells), or **continuous**, i.e. any value within an interval (mass, concentration, absorbance). Four **measurement scales** are also distinguished:

- **nominal**: categories without order (blood group);
- **ordinal**: ordered categories (turbidity +, ++, +++; sizes S, M, L);
- **interval**: conventional zero (temperature in °C);
- **ratio**: absolute zero, so that ratios are meaningful (mass, concentration, temperature in K).

The mean, the standard deviation and the CV% (Chapter 4) are fully meaningful only on the last two scales; the CV% only on the ratio scale.

1.5 The analytical signal and the calibration curve

The **analytical signal** is the quantity measured by the instrument that depends on the amount of analyte: absorbance, emission intensity, chromatographic peak area, ion counts. A necessary condition is a **known functional relationship** between the amount of analyte and the source of the signal, such as the Beer–Lambert law in spectrophotometry. The signal is **proportional** to concentration only within the linear range and if the blank signal is zero or subtracted. In general, within the linear range:

$$y = b_0 + b_1 x \tag{1.4}$$

with intercept b_0 (blank signal) and slope b_1. If $b_0 \neq 0$ the relationship is linear but not proportional. To build the **calibration curve**, standard solutions of known concentration are prepared, their signal is measured and b_0 and b_1 are estimated (Chapter 7). The concentration of a sample is then obtained from its signal y_0 by inverting Equation 1.4:

$$x_0 = \frac{y_0 - b_0}{b_1} \tag{1.5}$$

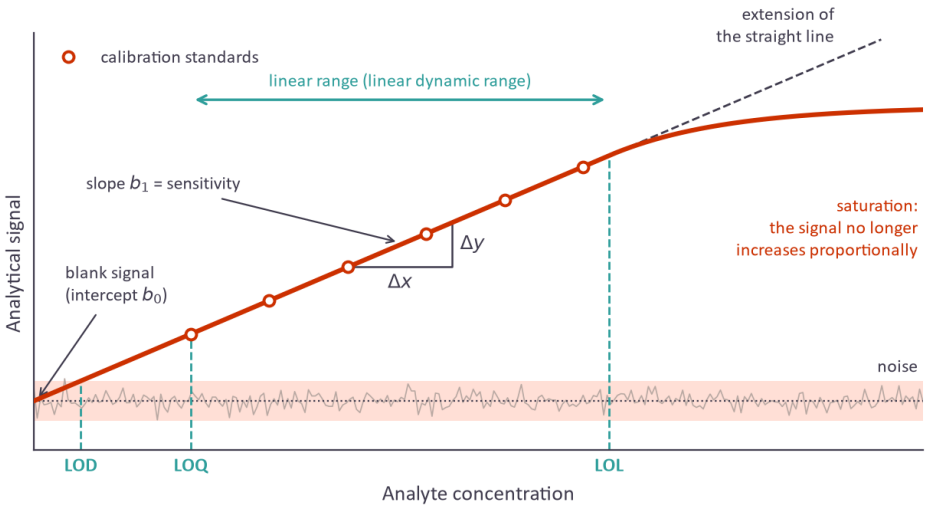


Figure 1.1 — Typical behavior of the analytical signal as a function of concentration: blank signal with noise, LOD and LOQ, linear range up to the LOL, saturation beyond the LOL.

Figure 1.1 shows the main quantities:

- the **slope** $b_1 = \Delta y / \Delta x$ is the **sensitivity**: the steeper the line, the better the method distinguishes small differences in concentration. It must not be confused with the detection limit;
- **LOD** and **LOQ** are the lowest concentrations that can be detected and quantified, respectively (Chapter 11);
- the **linear range** extends from the LOQ to the **LOL** (limit of linearity);

- beyond the LOL the detector **saturates** and the signal levels off to a plateau: proportionality no longer holds.

A sample above the calibration range is not quantified by extrapolation: it is diluted and reanalyzed (Chapter 14). The plateau does not depend on the blank, which would shift the line parallel to itself (constant error, Chapter 2), nor on random errors.

Example 1.2 — From signal to concentration, and when to dilute

Data: blank-corrected line through the origin, b_1 = 0.0760 L/mg, calibration range 0.50–10.0 mg/L.

1. A sample with A = 0.380 gives (Equation 1.5) $x_0 = 0.380/0.0760$ = 5.00 mg/L, within the range.
2. A second sample with A = 0.950 would give 12.5 mg/L, above the highest standard: the value is not reliable.
3. Diluted 1:5, the second sample gives A = 0.231, i.e. 3.04 mg/L; multiplying by 5 gives 15.2 mg/L.

Result: 5.00 mg/L and 15.2 mg/L. Near the plateau, the direct reading underestimated the concentration by about 18%.

Key points of Chapter 1

- Every measurement has an uncertainty; the maximum tolerable error is established before starting the analysis.
- The result is written as $\bar{x} \pm U$ with its unit; $U = k \cdot u_c$ ($k = 2$, about 95%) is neither an error nor the standard deviation of the replicates.
- The uncertainty has 1–2 significant figures and the central value ends at the same decimal place; leading zeros are not significant.
- Analysis turns data into information, and a model turns information into knowledge.
- Data are quantitative or qualitative, discrete or continuous, on a nominal, ordinal, interval or ratio scale.
- The slope of the calibration curve is the sensitivity; proportionality holds only within the linear range, up to the LOL.

Test yourself

- **Uncertainty and reporting of results**
- **Data and analytical signal**

2. Measurement errors

To improve a method we need to know where errors come from and how they behave. Some shift all the results in the same direction and can be corrected; others scatter them unpredictably and can only be reduced and quantified; still others are blunders to be eliminated. Recognizing the type of error makes it possible to choose the right remedy.

2.1 Error and uncertainty: two distinct concepts

The International Vocabulary of Metrology (VIM) and the GUM distinguish two concepts that everyday language confuses.

- **Measurement error** is a **signed difference**: measured value minus true (or reference) value. Since the true value is generally unknown, so is the error; it can be calculated only by analyzing a material of known value.
- **Measurement uncertainty** is a **dispersion parameter**, always positive, that indicates the width of the interval of values that can be attributed to the measurand ($\bar{x} \pm U$, Chapter 1). It is estimated without knowing the true value.

Expressions such as "error bars" or "an error of ± 0.1 mL" use "error" in the sense of uncertainty: this colloquial usage is best avoided. The two concepts are independent (Figure 2.1). A result with a small uncertainty but far from the true value points to an undetected systematic error.

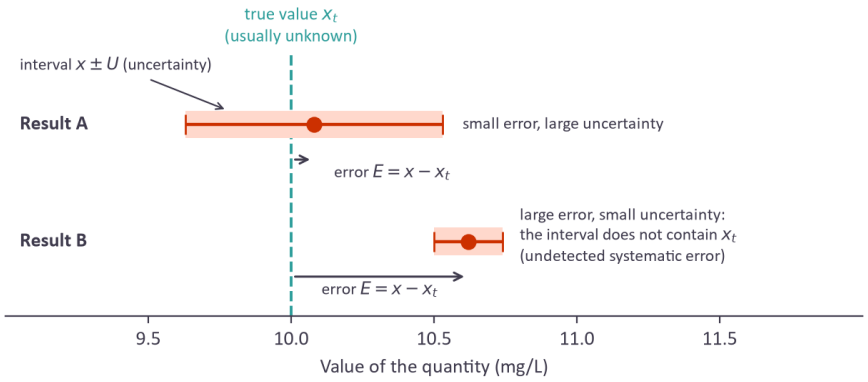


Figure 2.1 — Error and uncertainty are distinct concepts: result A has a small error and a large uncertainty, result B a large error and a small uncertainty.

2.2 Absolute error, relative error and bias

The **absolute error** of a result x_i with respect to the true value x_t, and the percent **relative error**, are:

$$E = x_i - x_t \tag{2.1}$$

$$E_r = \frac{x_i - x_t}{x_t} \times 100 \tag{2.2}$$

A positive error indicates an **overestimate**, a negative one an **underestimate**. The true value is always subtracted from the measured value and, in the relative error, the difference is divided by the true value. For example, a 10.40 mg/L standard analyzed as 10.25 mg/L has $E_r = -0.15/10.40 \times 100 = -1.44\%$.

The error of a single result contains a systematic and a random component. The **bias** is the systematic error of the population:

$$\text{bias} = \mu - x_t \tag{2.3}$$

where μ is the population mean, i.e. the mean of the infinite possible results. The error of a single result therefore splits into the bias plus a random error:

$$x_i - x_t = (\mu - x_t) + (x_i - \mu)$$

In practice the bias is estimated from the mean $\bar{x}$ of replicates performed on a material of known value. It is also expressed in relative form, $(\bar{x} - x_t)/x_t \times 100$, or as recovery, $\bar{x}/x_t \times 100$.

Example 2.1 — Error of a weighing and bias of a method

Data: (a) a calibrated 20.0000 g mass weighed on a balance reads 20.0012 g; (b) a 10.00 mg/L standard analyzed five times gives 10.13, 10.08, 10.15, 10.05 and 10.09 mg/L.

1. (a) Equations 2.1 and 2.2: $E = 20.0012 - 20.0000 = +0.0012$ g = +1.2 mg, i.e. E_r = +0.006%: the balance **overestimates**.
2. (b) $\bar{x} = 50.50/5 = 10.10$ mg/L, hence (Equation 2.3, with $\bar{x}$ in place of μ) bias = $10.10 - 10.00$ = +0.10 mg/L (+1.0%).

Result: E = +1.2 mg; bias = +0.10 mg/L. The errors of the individual results (from +0.05 to +0.15 mg/L) also contain the random component: the bias is estimated from the mean.

2.3 The three types of error

- **Gross errors** are macroscopic, easily detected blunders: wrong sample, swapped tubes, obvious contamination. **They are eliminated** by repeating the analysis; they are not treated statistically.
- **Systematic** (determinate) **errors** act according to **definite laws**: they always have the same sign and cause overestimation or underestimation. **They can be identified and corrected**, and they determine **trueness** (Chapter 3).
- **Random** (indeterminate) **errors** arise from **unpredictable fluctuations** in the experimental parameters and scatter the results around the mean. **They cannot be eliminated**, but they can be reduced and quantified statistically; they determine **precision**.

Increasing the number of replicates and averaging them **reduces the effect of random errors, but not the bias**. Random errors vary in sign and cancel out in the mean: the dispersion of the mean decreases as $s/\sqrt{n}$ (Chapter 5). The bias, instead, shifts all the results in the same direction: a hundred measurements with a miscalibrated pipette give a precise but wrong mean.

2.4 Systematic errors: the sources

Systematic errors are classified according to their **cause**:

- **instrumental**: incorrect calibration, improper use of glassware and instruments, unsuitable instruments. Example: a micropipette that always delivers 2% less than the set volume causes an instrumental error and, since it is a fixed fraction of the volume, also a proportional one (Section 2.5). It is eliminated by calibrating the pipette gravimetrically;
- **method**: non-ideal behavior of reagents and reactions, unsuitable experimental conditions. Examples: incomplete precipitation, unexpected solubility, inadequate ignition times, an indicator that changes color before the equivalence point, incomplete extraction, ionization suppression in LC-MS. They are the hardest to detect;
- **personal**: carelessness, poor knowledge of the procedure, physical limitations and unconscious prejudice. A color-blind analyst who sees the color change late, or the tendency to stop a titration near the expected volume, produce errors that always recur with the same sign: they are systematic, not gross.

2.5 Constant and proportional systematic errors

Systematic errors are also distinguished by how they depend on the amount of analyte (Figure 2.2).

A **constant error** has a magnitude **independent of the amount of analyte**. On the calibration line it produces a **parallel shift**: the intercept changes, not the slope. Examples: solubility losses during washing in gravimetry (negative bias) and extraneous absorption in spectrophotometry (positive bias). The absolute error is fixed, so **the relative error decreases as the amount** of analyte or the sample mass **increases**.

A **proportional error** has a magnitude **proportional to the amount of analyte**. It **changes the slope** of the line and leaves the intercept unchanged. Examples: losses due to extractions with incomplete yield (negative bias)

and a matrix effect that enhances the signal (positive bias). **The relative error is the same at every level**, while the absolute error grows with the amount of analyte. To tell the two cases apart, samples of different mass are compared. A material containing 10.0 mg/g of analyte, analyzed in 0.500 g and 1.000 g portions (5.00 and 10.00 mg of analyte), gives for example 4.50 and 9.00 mg. The absolute errors (−0.50 and −1.00 mg) double with the mass, while the relative error stays at −10%: the error is **negative proportional**. A constant error of −0.50 mg would instead give 9.50 mg with the larger portion (−5%).

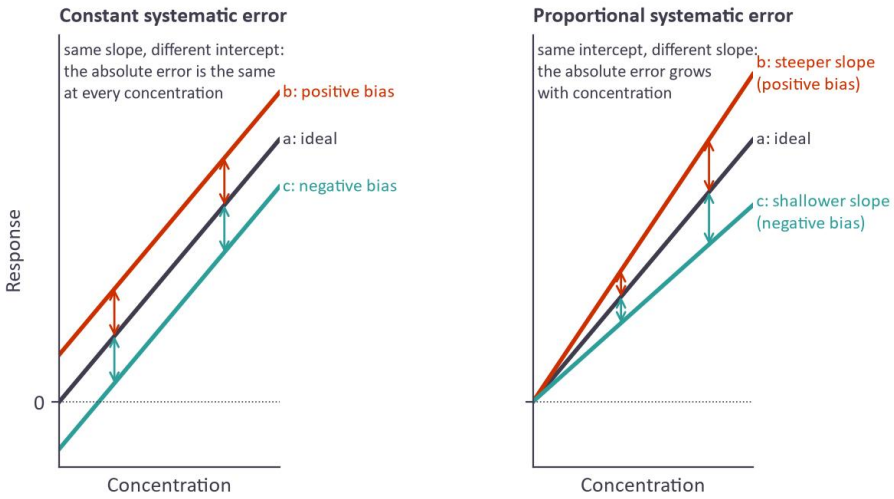


Figure 2.2 — Constant error (left): shifted line, same slope, same absolute error at every concentration. Proportional error (right): same intercept, different slope, increasing absolute error.

Example 2.2 — Constant and proportional errors in practice

Data: (a) washing a precipitate always dissolves 0.50 mg of it; (b) an indicator always requires 0.04 mL of excess titrant; (c) a CRM containing 50.0 µg/kg of lead gives a mean of 47.6 µg/kg, with a proportional error.

1. (a) Constant error: $E_r = -0.50/50.0 \times 100 = -1.0\%$ with 50.0 mg of precipitate and $-0.50/500 \times 100 = -0.10\%$ with 500 mg.
2. (b) For the relative error not to exceed 0.1% (i.e. 0.001), $V \geq 0.04/0.001 = 40$ mL of titrant is needed.
3. (c) Relative bias $= (47.6 - 50.0)/50.0 \times 100 = -4.8\%$: the method returns 95.2% of the true value. For a sample containing 120.0 µg/kg, the expected result is $120.0 \times 0.952 = 114.2$ µg/kg.

Result: −1.0% and −0.10%; at least 40 mL; 114.2 µg/kg. If the error in (c) were constant, 2.4 µg/kg would be subtracted at every level (117.6 µg/kg). To correct a result affected by the proportional error, it is divided by 0.952.

2.6 How to detect and eliminate systematic errors

A systematic error cannot be seen from the scatter of the replicates: an external term of comparison is needed.

1. **Certified reference materials (CRMs)**: samples whose value is certified by competent producers (Chapter 9). The mean of the results is compared with the certified value.
2. **Independent method**: comparison with a reference method or with instrumentation of proven reliability.
3. **Blank analysis**: a solution containing all the components except the analyte. Its signal, due only to contributions unrelated to the analyte, is subtracted from that of the sample and corrects **additive** contributions (constant errors). It does not correct **multiplicative** matrix effects (proportional errors), which require matrix-matched calibration, standard addition or an internal standard (Chapter 10). A true blank matrix also requires an analyte-free matrix, which is not always available.

4. **Recovery tests** (*spike*): addition of a known amount of analyte to the real sample. They are the right choice when no CRM with a similar matrix is available, because they work in the actual matrix.
5. **Varying the sample mass**: reveals constant errors, not proportional ones.

The percent recovery is calculated with Equation 2.4:

$$R\% = \frac{C_{\text{spiked}} - C_{\text{unspiked}}}{C_{\text{added}}} \times 100 \tag{2.4}$$

Example 2.3 — Blank and recovery

Data: (a) blank-corrected line through the origin, slope 0.0760 L/mg; sample A = 0.415, blank A = 0.035. (b) Water containing 2.10 mg/L of nitrate; after a 1.00 mg/L spike, 3.02 mg/L is found.

1. (a) $(0.415 - 0.035)/0.0760 = 5.00$ mg/L. Without subtracting the blank, 5.46 mg/L would be obtained: the positive constant error causes an overestimate.
2. (b) $R\% = (3.02 - 2.10)/1.00 \times 100 = 92\%$.

Result: 5.00 mg/L; recovery 92%. The ratio of total found to total expected (3.02/3.10 = 97%) is wrong, because it dampens the effect of the loss.

2.7 Random errors

Random errors arise from indeterminate fluctuations in temperature, pH, pressure, humidity, ionic strength, detector noise and scale reading. They cause a **symmetric scatter** of the data around the mean value: positive and negative errors are equally likely, and small errors are more frequent than large ones (normal distribution, Chapter 5). **They cannot be eliminated**, but they can be reduced by working carefully and **quantified** statistically from repeated measurements: standard deviation (Chapter 4), standard error and confidence interval (Chapter 5).

Example: reading a burette. A burette is graduated every 0.1 mL; this is the scale division, i.e. the reading resolution, not to be confused with the *sensitivity* of the instrument. In the reading 9.68 mL the tenths are read from the graduation marks and the hundredths are **estimated by eye** between two marks, to about 1/10 of a division. The estimated digit varies from one reading to the next and is subject to random error, but it must be reported because it is significant.

Glassware tolerance is not a random error

For **a given piece** of glassware, the deviation from the nominal value is fixed: always using the same pipette produces a **systematic error**, which can be eliminated by calibration. The stated tolerance (for example ±0.02 mL for a 10 mL class A pipette) instead enters the uncertainty as a Type B contribution (Chapter 6). It becomes a source of random variability only if different pieces are used at random. Small uncontrolled temperature fluctuations, on the other hand, are typically random.

Key points of Chapter 2

- Error (a signed difference, generally unknown) and uncertainty (a dispersion parameter) are distinct concepts.
- $E = x_i - x_t$; E_r is divided by the true value; bias $= \mu - x_t$, estimated from the mean of the replicates.
- Gross errors are eliminated, systematic errors corrected, random errors reduced and quantified. More replicates reduce the effect of random errors, not the bias.
- Systematic errors are instrumental, method or personal errors.
- A constant error changes the intercept and weighs less on large samples; a proportional error changes the slope and has a constant relative error.
- Systematic errors are detected with CRMs, an independent method, the blank (additive contributions only), recovery tests and by varying the sample mass.

Test yourself

- **Types of error and random errors**
- **Absolute error, relative error and bias**
- **Systematic errors**

3. Precision, trueness and accuracy

"Is this method accurate?" is a question that requires two properties to be separated: how close the results are to one another, and how close their mean is to the true value. They depend on different errors (Chapter 2) and are improved by different remedies. The numerical criteria used in validation are given in Chapter 12.

3.1 Definitions according to the VIM and ISO 5725

The definitions follow the VIM and the ISO 5725 standard (*Accuracy (trueness and precision) of measurement methods and results*). **Accuracy** is the broadest concept and includes **trueness** and **precision**.

Precision, trueness and accuracy

- **Precision**: closeness of agreement between results of replicate measurements on the same or similar objects **under specified conditions**. It depends on random errors and is quantified by the standard deviation.
- **Trueness**: closeness of agreement between the mean of a very large number of results and the true value x_t (or a reference value). It corresponds to the absence of systematic errors and is expressed by the **bias**.
- **Accuracy**: closeness of agreement between a result and the true value. **It includes trueness and precision**: an accurate result is both true and precise.

Precision describes the scatter of the results around their mean and **does not require the true value**. It is expressed by the standard deviation s (Chapter 4) or by the relative standard deviation:

$$\mathrm{RSD} = \mathrm{CV}\% = \frac{s}{\bar{x}} \times 100 \tag{3.1}$$

The larger s, the poorer the precision. The definition refers to *specified* conditions: measurements under the same conditions define repeatability, measurements under different conditions define intermediate precision or reproducibility (Section 3.3).

Trueness, by contrast, requires a **reference value**, such as the certified value of a CRM, and is quantified by the bias, $\bar{x} - x_t$ (Equation 2.3), or by the recovery. The smaller the bias, the better the trueness. The difference between the mean and the true value therefore measures trueness, not accuracy.

Accuracy concerns the single result, whose error is the sum of the bias and a random error: it is good only if both are small. A method that is true but imprecise gives a correct mean, but individual results that may lie far from the true value. A method that is precise but not true gives results close to one another, but all shifted.

Accuracy in the guidelines

ICH Q2(R2) (2023) and ICH M10 (2022) use *accuracy* for the closeness of the mean to the nominal value, i.e. for ISO trueness, and assess precision separately (Chapter 12). In these notes: **trueness** = absence of bias, **accuracy** = trueness + precision.

Example 3.1 — Two methods compared on a CRM

Data: CRM with a certified value of 100.0 mg/kg, six analyses per method. Method 1: $\bar{x}$ = 100.4 mg/kg, s = 3.1 mg/kg. Method 2: $\bar{x}$ = 94.8 mg/kg, s = 0.5 mg/kg.

1. Bias: method 1 = +0.4 mg/kg (+0.4%); method 2 = −5.2 mg/kg (−5.2%).
2. RSD (Equation 3.1): method 1 = $3.1/100.4 \times 100 = 3.1\%$; method 2 = $0.5/94.8 \times 100 = 0.53\%$.

Result: method 1 has better trueness, method 2 better precision. Neither method is fully accurate. A small s is not enough: method 2 has tightly grouped shots that are far from the center.

3.2 The four cases: the targets

In the targets of Figure 3.1, the center is the true value and each shot is a result. The distance of the mean of the shots from the center represents the bias, the size of the group the scatter.

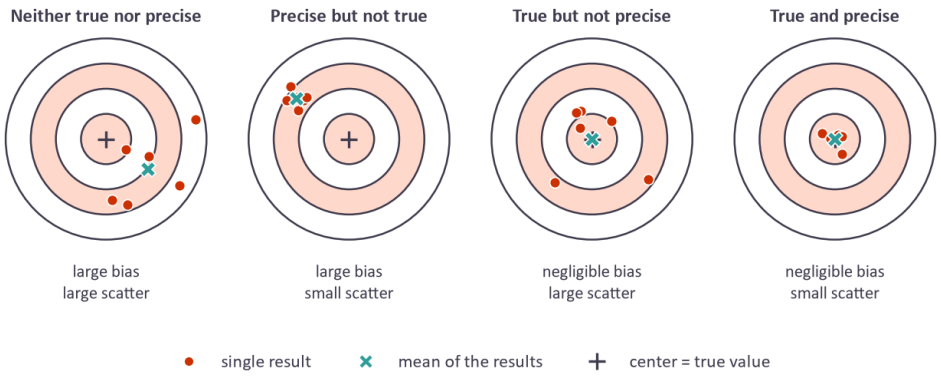


Figure 3.1 — The four cases of trueness and precision. Red dots: individual results; teal cross: mean; center: true value.

1. **Neither true nor precise**: large bias and large scatter; systematic and random errors are both significant.
2. **Precise but not true**: shots tightly grouped but far from the center; the systematic error dominates, the random error is small.
3. **True but not precise**: shots scattered but centered; negligible bias, large random error.
4. **True and precise**: shots grouped and centered. This is the ideal goal and the only accurate case.

Random errors are always present

Saying that case 2 has "only systematic errors" and case 3 "only random errors" is a simplification: random errors are always present, and one speaks of the **dominant** error. Case 3 is "true but not precise", not "accurate but not precise". Moreover, in the laboratory the center cannot be seen: the group of results allows the precision to be judged, but trueness requires an external reference.

There is also a practical difference between the two shortcomings. A known and stable bias can be corrected, for example with a recovery factor (Chapter 2). Imprecision, instead, can only be reduced, by averaging more replicates or improving the method. Correcting the bias is no substitute for finding its cause, and the correction itself adds an uncertainty contribution (Chapter 6).

3.3 Repeatability, intermediate precision and reproducibility

ISO 5725 and the VIM distinguish three precision conditions, with increasing variability.

Precision conditions (ISO 5725)

- **Repeatability** (s_r): same procedure, **same operator, same instrument, same laboratory, short period of time**. It expresses the minimum variability of the method.
- **Intermediate precision** (s_I): **same laboratory**, over an extended period, with different days, operators, instruments or reagent batches. It describes the variability of routine work.
- **Reproducibility** (s_R): **different laboratories** applying the same method to the same material. It is estimated by interlaboratory studies and is essential for the standardization of methods.

The key point is the **laboratory**: as long as the work stays in the same laboratory, changing the operator, instrument or day defines intermediate precision. Reproducibility applies only to different laboratories (Table 3.1). In bioanalysis, repeatability corresponds to the *within-run* precision of ICH M10 and intermediate precision to *between-run* precision.

Table 3.1 — What varies in the three precision conditions

Factor	Repeatability	Intermediate precision	Reproducibility
Laboratory	same	same	**different**
Operator, instrument	same	may vary	different
Time	short	extended (different days)	extended
Expected variability	minimum	intermediate	maximum

In an **interlaboratory study**, several laboratories analyze the same homogeneous material with the same method. The standard deviation calculated over the results of all the laboratories estimates the reproducibility, because it includes the differences in operators, instruments, reagents and environment, as well as the small bias specific to each laboratory. A standard deviation always describes precision, never trueness.

All three are quantified by the standard deviation under the respective conditions, or by the RSD. We expect $s_r \leq s_I \leq s_R$. In the ISO 5725-2 model, the reproducibility variance is the sum of the repeatability variance and the between-laboratory variance, s_L^2:

$$s_R^2 = s_r^2 + s_L^2 \tag{3.2}$$

Variances are added, not standard deviations. For the model parameters, reproducibility is therefore always greater than or equal to repeatability. With real data, however, the estimate of s_R may turn out smaller than s_r by chance; ISO 5725-2 then sets $s_R = s_r$.

Example 3.2 — Reproducibility standard deviation

Data: interlaboratory study with s_r = 0.30 mg/L and s_L = 0.40 mg/L.

1. Equation 3.2: $s_R^2 = 0.30^2 + 0.40^2 = 0.09 + 0.16 = 0.25$ (mg/L)².
2. $s_R = \sqrt{0.25} = 0.50$ mg/L.

Result: s_R **= 0.50 mg/L.** Adding the standard deviations (0.70 mg/L) is wrong; 0.25 is the variance, not the standard deviation.

Key points of Chapter 3

- Precision: agreement among replicate results under specified conditions; it depends on random errors, is expressed by s or the RSD and does not require the true value.
- Trueness: agreement between the mean and the true value; it depends on systematic errors and is expressed by the bias.
- Accuracy = trueness + precision.
- In the targets, the offset of the group is the bias and its size the scatter; random errors are always present.
- Repeatability: same conditions; intermediate precision: same laboratory, varied conditions; reproducibility: different laboratories.
- $s_R^2 = s_r^2 + s_L^2$: variances are added.

Test yourself

- **Precision, trueness and accuracy**
- **Repeatability and reproducibility**

4. Descriptive statistics

An analytical result almost always comes from several replicates of the same sample, which are similar but not identical because of random errors (Chapter 2). Descriptive statistics summarizes the series with a **central value**, which represents the result, and a **measure of dispersion**, which quantifies its precision: these are the quantities on which the confidence interval (Chapter 5), the uncertainty (Chapter 6) and the validation criteria (Chapter 12) are based.

4.1 Sample and population

Population and sample

Population: the set of all possible measurements of the quantity under the conditions considered, theoretically infinite. Its **parameters** are denoted by Greek letters: μ (mean) and σ (standard deviation).

Sample: the finite set of the n measurements actually performed. Its **statistics** are denoted by Latin letters: $\bar{x}$ (mean) and s (standard deviation). The sample statistics are **estimates** of the population parameters.

The parameters μ and σ are fixed but **unknown**; $\bar{x}$ and s are known but vary from one sample to another. Increasing the number of replicates does not turn $\bar{x}$ and s into μ and σ: it makes them more reliable estimates. The distribution of the sample is represented in a **discrete** form, with a **histogram** (bars as high as the

number of measurements in each class of values); that of the population with a **continuous curve**, which for random errors is the Gaussian (Chapter 5). Finally, recall that μ coincides with the true value only in the absence of systematic errors: the difference between the two is the bias (Chapter 2).

4.2 Measures of central tendency

The **arithmetic mean** is the sum of the measurements divided by their number:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{4.1}$$

The mean uses all the data and is the best estimate of μ for normally distributed data; the sum of the deviations from the mean, $\sum(x_i - \bar{x})$, is always zero. Precisely because it uses all the data, however, it is **sensitive to outliers**.

The **median** is the central value of the data **sorted in ascending order**: with odd n it is the value in position $(n+1)/2$; with even n it is the **mean of the two central values** (positions $n/2$ and $n/2+1$). It depends only on the order of the data, not on how far away the extremes are: it is a **robust** measure, preferable to the mean when outliers are suspected or the distribution is skewed.

The **mode** is the most frequent value; a series may have none or more than one. It is rarely used as a result, but comparing the mean, median and mode reveals the shape of the distribution (Chapter 5).

Example 4.1 — Mean and median in the presence of an outlier

Consider the seven values 12, 15, 20, 23, 24, 27, 200.

1. Mean: $(12+15+20+23+24+27+200)/7 = 321/7 = 45.9$.
2. Median: data already sorted, $n = 7$ odd, 4th value = 23.
3. Without the value 200 ($n = 6$, even): mean $121/6 = 20.2$; median = mean of the 3rd and 4th values, $(20+23)/2 = 21.5$.

Result: the value 200 shifts the mean to 45.9, far from all the other data, while the median (23) remains representative; removing it changes the mean by more than 25 units and the median by 1.5.

Common mistakes with the median

Taking the central value of the series in the order of acquisition without sorting it; with even n, choosing only one of the two central values; confusing the median with the mean of the minimum and maximum (which depends precisely on the extremes) or with the mode.

4.3 Measures of dispersion

The **range** $w = x_{\max} - x_{\min}$ is immediate but uses only the two extreme values; it appears in Dixon's Q test (Chapter 5).

To use all the data, we start from the deviations $x_i - \bar{x}$; since their sum is zero, they are squared and averaged, which gives the **variance** (σ^2 for the population, s^2 for the sample). The unit of the variance is the square of the unit of the data, for example (mg/L)², but the variance has a valuable property: **the variances of independent contributions add up** (Chapter 6). Its square root is the **standard deviation**, which has the **same unit as the data**.

Population and sample standard deviation

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}} \tag{4.2}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \tag{4.3}$$

- Equation 4.2 applies when the whole population is known; with a series of replicates, equation 4.3 is **always** used.
- $n - 1 = \nu$ are the **degrees of freedom**.

Why n − 1. The deviations are calculated with respect to $\bar{x}$, which is obtained from the same data; since their sum must be zero, only $n - 1$ deviations are independent: calculating $\bar{x}$ "uses up" one degree of freedom. Moreover, $\bar{x}$ is the value that minimizes the sum of the squared deviations, so the deviations from $\bar{x}$ are on average smaller than those from the unknown μ: dividing by n would **systematically underestimate** σ, and the denominator $n - 1$ compensates for this effect. The difference matters with few data: with three replicates, dividing by n lowers s by 18%. With a single measurement, s cannot be calculated. Beware of software: in Excel the sample function is STDEV.S (not STDEV.P); on calculators it is the s or σ_{n-1} key.

To compare precision at different concentration levels, between methods or between analytes, the standard deviation is divided by the mean.

Relative standard deviation (RSD) and coefficient of variation (CV%)

$$\mathrm{RSD} = \frac{s}{\bar{x}} \tag{4.4}$$

$$\mathrm{CV\%} = \frac{s}{\bar{x}} \times 100 \tag{4.5}$$

- The RSD is **dimensionless** (a ratio between quantities with the same unit); expressed as a percentage it is called CV%.
- The numerator is s, not the standard error $s/\sqrt{n}$. A lower CV indicates **better relative precision**.

For example, an LC-MS/MS method with $s = 0.12$ ng/mL at 2.0 ng/mL and $s = 4.0$ ng/mL at 200 ng/mL has an s more than thirty times larger at the high level, but a CV of 6.0% at the low level and of 2.0% at the high level: in relative terms it is more precise at the high level. This is why the precision criteria of the guidelines are expressed as CV% (for example CV ≤ 15%, ≤ 20% at the LLOQ in ICH M10, Chapter 12).

4.4 A complete exercise: Fe(III) in a water sample

Example 4.2 — Descriptive statistics of seven replicates

The concentration of Fe(III) in a water sample is determined with seven replicates (mg/L): 2.41, 2.38, 2.45, 2.43, 2.39, 2.44, 2.42.

1. Mean (equation 4.1): $\bar{x} = 16.92/7 = 2.41714$ mg/L; 2.4171 is used in the calculations, and 2.417 mg/L is reported as the result.
2. Median: sorted data 2.38, 2.39, 2.41, 2.42, 2.43, 2.44, 2.45; 4th value = 2.42 mg/L.
3. Deviations and their squares (Table 4.1): sum of squares = 0.00394287 (mg/L)².
4. Variance ($\nu = 6$): $s^2 = 0.00394287/6 = 0.00065715$ (mg/L)².
5. Standard deviation (equation 4.3): $s = \sqrt{0.00065715} = 0.02563$ mg/L ≈ 0.026 mg/L.
6. CV% (equation 4.5): $0.02563/2.4171 \times 100 = 1.06\% \approx 1.1\%$ (the RSD of equation 4.4 is 0.0106).

Result: mean 2.417 mg/L, median 2.42 mg/L, s = 0.026 mg/L (n = 7), CV = 1.1%.

Table 4.1 — Deviations from the mean for the seven Fe(III) replicates (mean = 2.4171 mg/L)

Replicate	x_i (mg/L)	$x_i - \bar{x}$ (mg/L)	$(x_i - \bar{x})^2$ (mg/L)²
1	2.41	−0.0071	0.00005041
2	2.38	−0.0371	0.00137641
3	2.45	+0.0329	0.00108241
4	2.43	+0.0129	0.00016641
5	2.39	−0.0271	0.00073441
6	2.44	+0.0229	0.00052441
7	2.42	+0.0029	0.00000841
Sum	16.92	≈ 0	**0.00394287**

Dividing by n instead of n − 1

Dividing the sum of squares by $n = 7$ would give $\sqrt{0.00394287/7} = 0.024$ mg/L and a CV of about 1.0%: this is the population formula (4.2), which does not apply to seven replicates. Other typical mistakes: using $s/\sqrt{n}$ instead of s in the CV; forgetting the factor 100 (RSD = 0.011 corresponds to CV = 1.1%); rounding intermediate results (mean, deviations), which should be done only on the final result.

The standard deviation is normally reported with two significant figures and always together with n: without the number of data, neither the standard error of the mean ($0.0256/\sqrt{7} = 0.0097$ mg/L), nor the confidence interval (Chapter 5), nor the repeatability contribution to the uncertainty (Chapter 6) can be calculated.

Key points of Chapter 4

- **Population**: all possible measurements, unknown parameters μ and σ, continuous curve; **sample**: the n measurements performed, statistics $\bar{x}$ and s that estimate those parameters, histogram.
- The **mean** uses all the data and is sensitive to outliers; the **median** (central value of the sorted data; for even n, the mean of the two central values) is robust; the **mode** is the most frequent value.
- The **variance** has squared units and is additive for independent contributions; the **standard deviation** has the same unit as the data.
- For a sample, divide by **n – 1** (degrees of freedom), because $\bar{x}$ is estimated from the same data: dividing by n underestimates σ.
- **RSD** $= s/\bar{x}$ (dimensionless) and **CV%** $= s/\bar{x} \times 100$ compare precision across different levels and methods.
- Fe(III): $\bar{x} = 2.417$ mg/L, $s = 0.026$ mg/L, CV = 1.1%.

Test yourself

To practice on this chapter, use the quiz questions on the topics:

- **Sample, population and central tendency**
- **Dispersion: standard deviation, variance and CV%**

5. Distributions, inference and outliers

In the laboratory we need to know what fraction of the results will fall within certain limits, how far the sample mean may lie from the population mean, and whether a replicate that is very different from the others should be kept. Answering these questions requires the normal distribution and the tools of **statistical inference**, which lead from the sample to the population with a stated confidence level.

5.1 The normal distribution

The random errors of a measurement are the sum of many small independent contributions (temperature fluctuations, detector noise, small differences in volume); by the **central limit theorem**, their sum tends to follow the **normal** or **Gaussian distribution**:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] \tag{5.1}$$

where y is the measured value ($-\infty < y < +\infty$), μ the mean and σ the standard deviation of the population. In equation 5.1, μ sets the **position** of the maximum, around which the curve is symmetric (mean = median = mode); σ sets the **width**: the larger σ, the wider and lower the curve. The factor $1/(\sigma\sqrt{2\pi})$ makes the **total area under the curve equal to 1**, that is, to the total probability. The probability that a result falls between two values is the area under the curve between those values.

The areas within a given number of standard deviations from the mean are always the same (Figure 5.1): **68.3%** of the results fall within $\mu \pm 1\sigma$, **95.4%** within $\mu \pm 2\sigma$ and **99.7%** within $\mu \pm 3\sigma$. The curve never reaches zero: even beyond $\pm 3\sigma$ lie 0.27% of the results. The part outside a symmetric interval is split **equally between the two tails**: 4.6% lies outside $\mu \pm 2\sigma$, that is, 2.3% above $\mu + 2\sigma$ and 2.3% below $\mu - 2\sigma$. Exactly 95% corresponds to $\mu \pm 1.96\sigma$, and 99% to $\mu \pm 2.576\sigma$. The distance from the mean is often expressed in units of σ with the standardized variable $z = (y - \mu)/\sigma$.

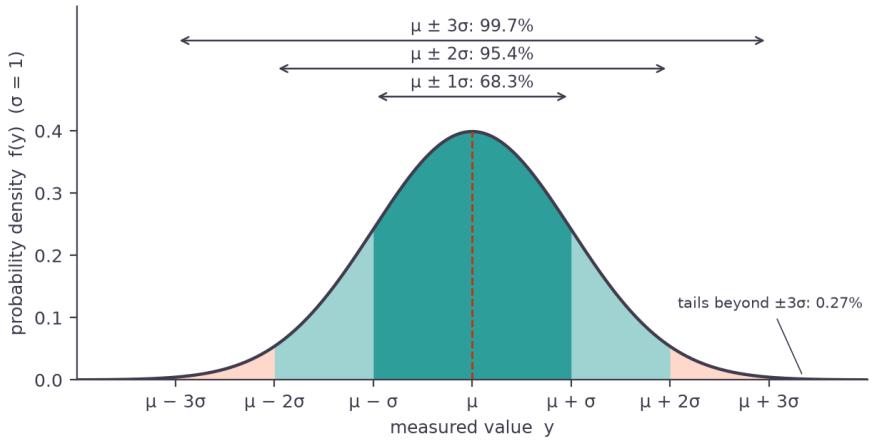


Figure 5.1 — Normal distribution: fraction of results within μ ± 1σ, μ ± 2σ and μ ± 3σ.

Example 5.1 — Percentages of the normal distribution

A control serum gives normally distributed results with $\mu = 5.50$ mmol/L and $\sigma = 0.10$ mmol/L. What fraction of the results is expected (a) between 5.40 and 5.60 mmol/L, (b) above 5.70 mmol/L, (c) below 5.40 mmol/L?

1. (a) 5.40 and 5.60 mmol/L are $\mu \pm 1\sigma$: **68.3%**.
2. (b) $z = (5.70 - 5.50)/0.10 = 2$. Outside $\mu \pm 2\sigma$ lies 4.6%, split between the two tails: **2.3%** lies above $\mu + 2\sigma$.
3. (c) Outside $\mu \pm 1\sigma$ lies 31.7%; half of it falls below $\mu - 1\sigma$, that is, **15.9%**.

Result: 68.3%; 2.3%; 15.9%. The typical mistake is forgetting to split the outside fraction between the two tails (4.6% instead of 2.3%).

5.2 Symmetric and skewed distributions

When the data are not Gaussian, the relative position of the mean, median and mode reveals the shape of the distribution (Figure 5.2):

- **symmetric distribution**: mean = median = mode; this is typical of measurements affected only by random errors;
- **positive skew** (tail to the right): a few very high values pull the mean more than the median, while the mode stays on the peak: **mean > median > mode**;
- **negative skew** (tail to the left): **mode > median > mean**.

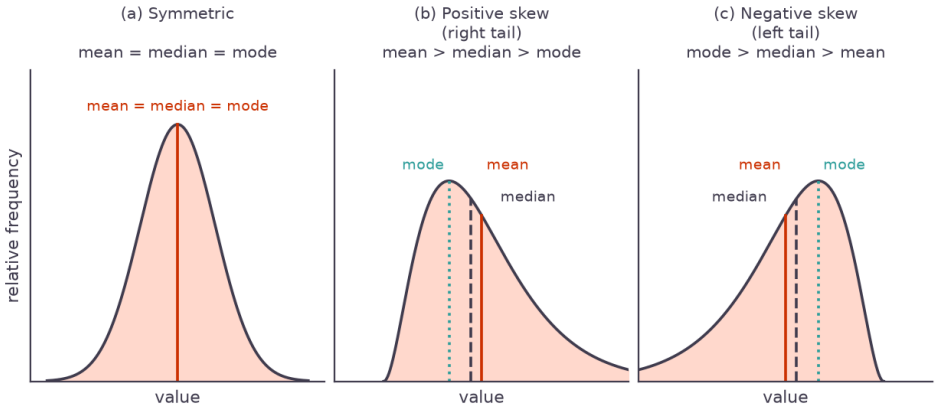


Figure 5.2 — Relative position of the mean, median and mode in a symmetric distribution (a), in a distribution with positive skew (b) and in one with negative skew (c).

Trace contaminants in a series of lots typically show positive skew (non-negative values, close to zero): nine lots of coffee with ochratoxin A at 0.8, 0.9, 0.9, 0.9, 1.0, 1.1, 1.3, 2.2 and 3.5 µg/kg have a mean of 1.40, a median of 1.0 and a mode of 0.9 µg/kg. Extraction recoveries, on the other hand, show negative skew (almost all between 95% and 98%, a few much lower because of losses). The mean–median–mode order is an **empirical rule** for distributions with a single peak, not a general property. With skewed data the median describes the typical value better, and tools based on normality should be used with caution.

5.3 The standard error of the mean

If a series of n replicates is repeated several times, different means are obtained: the **mean is itself a random variable**, distributed around μ more narrowly than the individual measurements, because the random errors partly cancel out in the mean. Its standard deviation is the **standard error of the mean**:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \tag{5.2}$$

($\sigma/\sqrt{n}$ if σ is known). The standard deviation s describes the dispersion of the **individual measurements** and does not decrease as n increases; the standard error describes the uncertainty of the **mean** and decreases with $\sqrt{n}$: more measurements, more reliable estimate. To **halve** it, the number of replicates must be **quadrupled** (with $s = 0.40$ mg/L: 0.20 mg/L with 4 replicates, 0.10 mg/L with 16). Increasing n does not, however, reduce systematic errors: $\bar{x}$ converges to μ, not to the true value.

5.4 Student's t distribution

If σ were known, the mean would fall within $\mu \pm 1.96\,\sigma/\sqrt{n}$ in 95% of cases. In practice σ is unknown and its estimate s, with few data, is itself uncertain: **Student's t distribution** with $\nu = n - 1$ **degrees of freedom** is then used. It has **wider tails** than the normal distribution, so at the same confidence level $t > z$: for $\nu = 3$ and 95%, $t = 3.182$ versus $z = 1.96$. As ν increases, t tends to z (for $\nu = 30$, $t = 2.042$). The often-quoted "$n < 30$" threshold is only a rule of thumb: t is used whenever σ is estimated by s, always with approximately normal data.

The value of t depends on the degrees of freedom and on the confidence level, and tables distinguish **two-tailed** and **one-tailed** values. For a symmetric interval, the probability of error (5% at the 95% level) is split between the two tails: the **two-tailed critical value** ($t_{\alpha/2,\nu}$) is needed. The one-tailed value at 95% coincides with the two-tailed value at 90% and is smaller (for $\nu = 6$: 1.943 instead of 2.447).

Table 5.1 — Two-tailed critical values of Student's t (complete tables in Appendix B)

ν	90%	95%	99%
2	2.920	4.303	9.925
3	2.353	3.182	5.841
4	2.132	2.776	4.604
5	2.015	2.571	4.032
6	1.943	2.447	3.707
7	1.895	2.365	3.499
8	1.860	2.306	3.355
9	1.833	2.262	3.250
10	1.812	2.228	3.169
20	1.725	2.086	2.845
∞ (= z)	1.645	1.960	2.576

5.5 The confidence interval of the mean

Confidence interval of the mean

$$\mu = \bar{x} \pm t \cdot \frac{s}{\sqrt{n}} \tag{5.3}$$

- $\bar{x}$, s: mean and standard deviation of the n replicates; $s/\sqrt{n}$: standard error (equation 5.2)
- t: **two-tailed** value of Student's t at the chosen confidence level (usually 95%) with $\nu = n - 1$
- If σ is known, $z \cdot \sigma/\sqrt{n}$ is used ($z = 1.96$ at 95%).

The 95% interval is constructed so that, if the experiment were repeated many times, **95% of the intervals would contain the population mean** μ. It does not describe where the individual measurements will fall (their dispersion is s, not $s/\sqrt{n}$), and it contains the true value only in the absence of systematic errors, because it accounts for random errors only.

The interval **widens** when the confidence level is raised (for $\nu = 5$, t goes from 2.571 at 95% to 4.032 at 99%: more confidence requires a wider interval) and **narrows** as n increases, for two reasons: $\sqrt{n}$ grows and t decreases. With $s = 0.30$ mg/L, the 95% half-width is 0.75 mg/L with 3 replicates ($t = 4.303$), 0.31 mg/L with 6 ($t = 2.571$) and 0.19 mg/L with 12 ($t = 2.201$).

Example 5.2 — Confidence interval of the mean

25-Hydroxyvitamin D in a serum sample is determined by LC-MS/MS with six replicates (ng/mL): 31.2, 30.4, 32.0, 31.5, 30.8, 31.7.

1. Mean: $\bar{x} = 187.6/6 = 31.267$ ng/mL.
2. Sum of squared deviations: 1.7533 (ng/mL)²; $s = \sqrt{1.7533/5} = 0.592$ ng/mL.
3. Standard error: $0.592/\sqrt{6} = 0.2418$ ng/mL; degrees of freedom $\nu = 5$.
4. At 95% (equation 5.3): $t = 2.571$; half-width $2.571 \times 0.2418 = 0.62$ ng/mL.
5. At 99%: $t = 4.032$; half-width $4.032 \times 0.2418 = 0.97$ ng/mL.

Result: μ = (31.27 ± 0.62) ng/mL at 95% and (31.27 ± 0.97) ng/mL at 99% (n = 6).

Typical mistakes in the confidence interval

With the data of Example 5.2 (correct value at 95%: ± 0.62 ng/mL): using $\nu = n = 6$ gives ± 0.59; using the one-tailed t (2.015) gives ± 0.49; using $z = 1.96$ gives ± 0.47; multiplying t by s instead of $s/\sqrt{n}$ gives ± 1.52 ng/mL.

5.6 Outliers

An **outlier** is a result that appears inconsistent with the others in the series. It may come from a gross error (Chapter 2) or be a legitimate value in the tail of the distribution: in 20 normally distributed measurements, the probability that at least one falls beyond $\pm 2\sigma$ is about 61%. Objective criteria are therefore needed. The following tests assume normally distributed replicates of the same quantity and examine **a single suspect value**, the one farthest from the others; if the test statistic exceeds the tabulated **critical value** (a function of n and of the confidence level), the value is a **statistical outlier** at that level.

Dixon's Q test and Grubbs test

$$Q = \frac{|x_{\text{susp}} - x_{\text{near}}|}{x_{\max} - x_{\min}} \tag{5.4}$$

$$G = \frac{|x_{\text{susp}} - \bar{x}|}{s} \tag{5.5}$$

- **Dixon's Q**, for small samples ($3 \leq n \leq 10$): sorted data; x_{susp} is the suspect value and x_{near} the value adjacent to it (its nearest neighbor); the denominator is the range of **all** the data.
- **Grubbs**: distance from the mean in units of standard deviation; $\bar{x}$ and s are calculated on all the data, **including** the suspect value. It uses all the data and is more powerful than the Q test for $n > 10$.
- The value is an outlier if $Q > Q_{\text{crit}}$ or $G > G_{\text{crit}}$ (Table 5.2), read for the **total** number of data n.

Table 5.2 — Critical values at 95% and 99% for Dixon's Q test (Rorabacher, 1991) and for the two-sided Grubbs test

n	Q 95%	Q 99%	G 95%	G 99%
3	0.970	0.994	1.154	1.155
4	0.829	0.926	1.481	1.496
5	0.710	0.821	1.715	1.764
6	0.625	0.740	1.887	1.973
7	0.568	0.680	2.020	2.139
8	0.526	0.634	2.127	2.274
9	0.493	0.598	2.215	2.387
10	0.466	0.568	2.290	2.482
12			2.412	2.636
15			2.548	2.806
20			2.708	3.001

For Grubbs, the **two-sided** test is used when the suspect value, identified by looking at the data, could be either the minimum or the maximum (the usual case).

Example 5.3 — Dixon's Q test and Grubbs test

(a) Five determinations of zinc in a control serum (µmol/L): 15.2, 15.6, 15.4, 16.9, 15.3. Is the value 16.9 an outlier?

1. Sorted data: 15.2, 15.3, 15.4, 15.6, 16.9; neighbor of the suspect value: 15.6; range: $16.9 - 15.2 = 1.7$.
2. Equation 5.4: $Q = (16.9 - 15.6)/1.7 = 0.765$.
3. For $n = 5$: $0.765 > 0.710$ (outlier at 95%) but $0.765 < 0.821$ (not an outlier at 99%).

(b) Twelve determinations of a paracetamol control solution (mg/L): 98.6, 99.1, 98.9, 99.4, 98.8, 99.0, 101.3, 99.2, 98.7, 99.3, 99.0, 98.9. Is the value 101.3 an outlier?

1. Over all 12 values: $\bar{x} = 1190.2/12 = 99.183$ mg/L; $s = 0.707$ mg/L.
2. Equation 5.5: $G = (101.3 - 99.183)/0.707 = 2.99$.
3. For $n = 12$: $2.99 > 2.412$ (95%) and > 2.636 (99%).

Result: (a) Q = 0.765, outlier at 95% but not at 99%; (b) G = 2.99, outlier at both 95% and 99%. In both cases, an experimental cause is sought before the value is excluded.

Common mistakes in outlier tests

- Q test: dividing by the range of the remaining values only, or reading the table with $n - 1$ instead of n.
- Grubbs: calculating $\bar{x}$ and s **without** the suspect value (in Example 5.3b this would give $G = 9.4$) or dividing by $s/\sqrt{n}$ ($G = 10.4$).
- Applying the test to the points of a calibration curve: they are at different concentrations and are not replicates; an aberrant calibrator is identified from the regression residuals (Chapter 8) and the curve acceptance criteria (Chapter 16).

The test provides **statistical evidence**: it says that the value is improbable given the distribution of the other data, not why it is different. The rule is: **a value is removed only with statistical evidence and, preferably, an identified experimental cause** (transcription error, contamination, wrong dilution, faulty injection); **never just because it "looks far off"**. In practice:

- positive test: before deciding, **look for the cause**; if one is found, the value is excluded, documenting the data, the test and the cause; otherwise, a procedure established in advance is followed (for example, reporting the result with and without the value, or using the median; in interlaboratory studies ISO 5725-2 normally excludes outliers at the 1% level);
- negative test: the value is kept;
- the test is applied **once**, to the most extreme value; replacing the value with the mean of the others, or removing the opposite extreme "for symmetry", distorts the mean and the dispersion;
- a documented gross error (wrong sample) invalidates the result regardless of any test.

Key points of Chapter 5

- Gaussian: μ sets the position, σ the width, total area 1; 68.3%, 95.4% and 99.7% of the results fall within $\mu \pm 1\sigma$, 2σ and 3σ; the fraction outside is split between the two tails.
- Symmetric: mean = median = mode; positive skew (right tail): mean > median > mode; negative skew (left tail): mode > median > mean (empirical rule).
- Standard error of the mean $s/\sqrt{n}$: halving it requires four times as many replicates; it does not reduce systematic errors.
- When σ is estimated by s, Student's t with $\nu = n - 1$ is used; $t > z$ and $t \to z$ as ν increases.
- CI of the mean: $\bar{x} \pm t \cdot s/\sqrt{n}$ with the two-tailed t; it contains μ with the stated confidence; it widens from 95% to 99% and narrows as n increases.
- Dixon's Q test ($3 \leq n \leq 10$): gap from the neighbor / range; Grubbs: $|x_{\text{susp}} - \bar{x}|/s$ with the suspect value included, more powerful for $n > 10$.
- A value is removed only with statistical evidence and, preferably, an identified cause.

Test yourself

To practice on this chapter, use the quiz questions on the topics:

- **Normal distribution and skewness**
- **Standard error and confidence interval**
- **Outlier tests**

6. Measurement uncertainty

A result without an uncertainty cannot be compared with a limit or with the result of another laboratory. If lead in a water sample is found at 9.6 µg/L and the limit is 10 µg/L, with an expanded uncertainty of 0.3 µg/L the water is compliant; with 0.8 µg/L the interval (8.8–10.4 µg/L) includes the limit and compliance cannot be stated. This is why ISO/IEC 17025 (2017) requires laboratories to evaluate the uncertainty of their results.

6.1 Error and uncertainty

Error, uncertainty and related quantities (VIM, GUM)

Measurement error: the difference between the measured value and the true value; since the true value is not known, the error of a result is in general **unknown** (if we knew it, we would correct it).

Measurement uncertainty: a non-negative parameter characterizing the **dispersion of the values** that can reasonably be attributed to the measurand; unlike the error, it is **estimated**.

Standard uncertainty u: uncertainty expressed as a standard deviation. **Combined standard uncertainty** u_c: the standard uncertainty of the result, obtained by combining all the contributions. **Expanded uncertainty** $U = k \cdot u_c$, with **coverage factor** k. **Relative uncertainty**: $u(x)/x$, often expressed as a percentage.

Known systematic effects (for example the blank contribution or a documented recovery) are corrected before the uncertainty is evaluated; the uncertainty of the correction remains in the budget.

6.2 The GUM: Type A and Type B contributions

The **GUM** (*Guide to the Expression of Uncertainty in Measurement*, JCGM 100:2008) is the international guide for evaluating and expressing measurement uncertainty (not to be confused with the VIM, the vocabulary of metrology, with ISO 5725 on trueness and precision, or with ICH M10 on bioanalytical method validation). It distinguishes two categories of contributions according to the **method of evaluation**:

- **Type A**: evaluated by **statistical analysis of series of** repeated **observations** (for example, the repeatability uncertainty);
- **Type B**: evaluated by **means other than statistical analysis**: calibration certificates, manufacturer's specifications, certified values, literature data, expert judgment.

The distinction does **not** coincide with that between random and systematic errors, nor does it indicate which contribution is larger or more reliable: both types are expressed as standard deviations and combined in the same way.

The evaluation proceeds as follows: write the equation from which the result is calculated, identify the sources of uncertainty (Table 6.1), express each one as a standard uncertainty (Type A or B), combine them into u_c and calculate $U = k \cdot u_c$.

Table 6.1 — Typical sources of uncertainty in an analytical determination

Source	Type	Origin of the information
Repeatability	A	series of replicates: $u_A = s/\sqrt{n}$
Instrument calibration	B	calibration certificate of the instrument
Weighing	B	calibration and resolution of the balance
Final volume	B	tolerance of the volumetric glassware, temperature
Internal standard and calibration solutions	B	purity and preparation of the standards
Calibration curve	A	scatter of the points around the line (Chapter 7)

6.3 Type A evaluation

From n repeated observations, the standard deviation s is calculated (Chapter 4). If the reported result is the **mean** of the n replicates, the repeatability uncertainty is the standard error of the mean (Chapter 5):

$$u_A = \frac{s}{\sqrt{n}} \tag{6.1}$$

with $\nu = n - 1$ degrees of freedom. If instead a **single determination** is reported, the repeatability contribution is s itself, preferably estimated from many data (validation, quality control). For example, with $s = 0.24$ mg/L and the mean of 4 replicates, $u_A = 0.24/2 = 0.12$ mg/L; for a single determination the contribution would be 0.24 mg/L.

6.4 Type B evaluation

In a Type B evaluation, the available information must be converted into a standard uncertainty by choosing a distribution consistent with what is known (Table 6.2, Figure 6.1).

Table 6.2 — Converting Type B information into a standard uncertainty

Available information	Distribution	Standard uncertainty
certificate with U and factor k	normal	$u = U/k$
certificate with U at 95%, without k	normal	$u = U/1.96$
tolerance $\pm a$, no information on the shape	rectangular	$u = a/\sqrt{3}$
tolerance $\pm a$, central values more likely	triangular	$u = a/\sqrt{6}$
digital resolution d	rectangular (half-width $d/2$)	$u = d/(2\sqrt{3})$

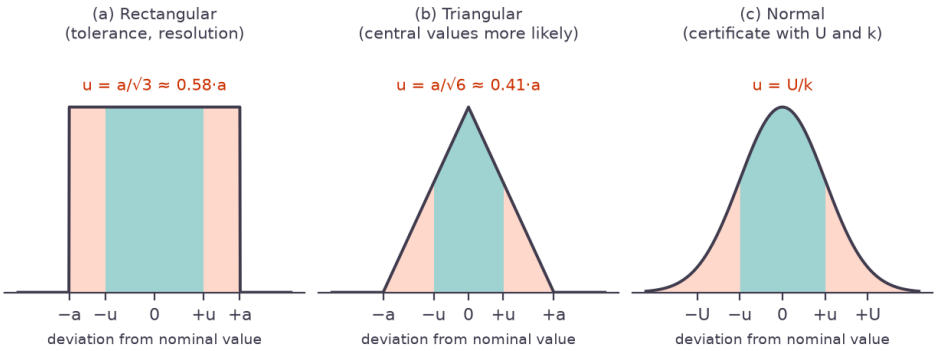


Figure 6.1 — Distributions used in Type B evaluations: rectangular (a), triangular (b) and normal (c); the darker area shows the interval ±u.

For example: a balance certificate with $U = 0.10$ mg ($k = 2$) gives $u = 0.050$ mg; a 100 mL class A volumetric flask with a tolerance of ± 0.10 mL gives $u = 0.10/\sqrt{3} = 0.058$ mL with the rectangular distribution, which is the cautious choice when nothing else is known (0.041 mL with the triangular one); a standard with a purity of $(99.5 \pm 0.3)\%$ gives $u = 0.3/\sqrt{3} = 0.17\%$.

6.5 Combined standard uncertainty

Independent contributions are combined **in quadrature**, because for independent quantities **the variances add up** (u^2), not the standard deviations: the contributions do not all act at their maximum and in the same direction at the same time, and they partly cancel out. A linear sum of the uncertainties would assume exactly this worst case and would **overestimate** u_c.

Combined standard uncertainty

Sums and differences ($y = x_1 + x_2 - x_3$), contributions expressed in the unit of the result:

$$u_c = \sqrt{u_A^2 + u_{B1}^2 + u_{B2}^2 + \ldots} \tag{6.2}$$

Products and quotients ($y = x_1 \cdot x_2 / x_3$), relative uncertainties:

$$\frac{u_c(y)}{y} = \sqrt{\left(\frac{u(x_1)}{x_1}\right)^2 + \left(\frac{u(x_2)}{x_2}\right)^2 + \left(\frac{u(x_3)}{x_3}\right)^2} \tag{6.3}$$

- Both apply only to **independent** (uncorrelated) contributions.
- Equation 6.2 requires contributions expressed in the unit of the result and transferred to it "one to one"; equation 6.3 applies to most analytical results (mass × purity / volume; serial dilutions, Chapter 16).

Because the squares are added, large contributions weigh much more than small ones: a contribution equal to one third of the largest one increases u_c by only about 5%. To reduce the uncertainty, one must therefore act on the **dominant contribution**.

6.6 Expanded uncertainty

Expanded uncertainty

$$U = k \cdot u_c \tag{6.4}$$

- The result is expressed as $y \pm U$.
- $k = 1$: confidence level of about 68%; $k = 2$: about 95%; $k = 3$: about 99.7%.
- These correspondences hold if the distribution of the result is **approximately normal** and the **effective degrees of freedom** are sufficient; with a dominant Type A contribution estimated from few replicates, the two-tailed 95% t value is used as k.

Example 6.1 — Combined and expanded uncertainty with contributions in the same unit

Copper in a water sample is determined with five replicates (mg/L): 0.842, 0.851, 0.838, 0.847, 0.852. The result is the mean; the Type B contributions (purity and preparation of the standards, glassware), already combined, amount to $u_B = 0.0050$ mg/L.

1. Mean and standard deviation: $\bar{x} = 0.846$ mg/L; $s = \sqrt{142 \times 10^{-6}/4} = 0.0060$ mg/L.
2. Type A (equation 6.1): $u_A = 0.00596/\sqrt{5} = 0.00266$ mg/L.
3. Combined (equation 6.2): $u_c = \sqrt{0.00266^2 + 0.0050^2} = \sqrt{(7.1 + 25.0) \times 10^{-6}} = 0.00567$ mg/L.
4. Expanded (equation 6.4): $U = 2 \times 0.00567 = 0.011$ mg/L.

Result: Cu = (0.846 ± 0.011) mg/L (k = 2, confidence level of about 95%).

Typical mistakes in uncertainty calculations

With the data of Example 6.1: using s instead of $s/\sqrt{n}$ gives $U = 0.016$ mg/L; adding the contributions linearly gives 0.015 mg/L; forgetting k gives 0.0057 mg/L. Other mistakes: omitting the square root; combining absolute uncertainties in a model made of products and quotients.

6.7 The uncertainty budget

The **uncertainty budget** is the table that lists, for each source, its type, its uncertainty and its share of u_c^2: it documents the evaluation and shows **where to intervene**.

Example 6.2 — Uncertainty budget for the assay of an active ingredient

Paracetamol in tablets is determined by HPLC-UV: a portion of powder (125.0 mg) is extracted and made up to volume in a 100 mL volumetric flask; calibration uses a standard with a purity of $(99.5 \pm 0.3)\%$. The result, the mean of six independent preparations, is 498.6 mg per tablet. The model is a product/quotient, so relative uncertainties are combined (Table 6.3).

1. Repeatability (A): CV of the six preparations 0.90%; for the mean, $0.90/\sqrt{6} = 0.367\%$.
2. Calibration curve (A): 0.30% from the regression (Chapter 7).
3. Standard solution (B): 0.20%, a contribution dominated by the purity (0.17%).
4. Final volume (B): tolerance (0.058 mL) and temperature (0.048 mL), $\sqrt{0.058^2 + 0.048^2} = 0.075$ mL on 100 mL, that is, 0.075%.
5. Weighing (B): balance certificate, $0.050/125.0 = 0.040\%$.
6. Combined (equation 6.3): $\sqrt{0.367^2 + 0.30^2 + 0.20^2 + 0.075^2 + 0.040^2} = \sqrt{0.272} = 0.52\%$.
7. Expanded: $U = 2 \times 0.52\% = 1.04\%$, that is, $0.0104 \times 498.6 = 5.2$ mg.

Result: (498.6 ± 5.2) mg per tablet (k = 2, confidence level of about 95%).

Table 6.3 — Uncertainty budget for Example 6.2

Source	Type	Relative u (%)	Share of u_c^2
Repeatability (mean of 6)	A	0.367	49.6%
Calibration curve	A	0.30	33.1%
Standard solution	B	0.20	14.7%
Final volume	B	0.075	2.1%
Weighing	B	0.040	0.6%
Combined / expanded ($k = 2$)		**0.52 / 1.04**	

Repeatability and calibration account for more than 80% of the variance; removing volume and weighing altogether would lower U only from 1.04% to 1.03%. It is therefore better to increase the number of preparations (with 12, U falls to 0.90%) than to buy more accurate glassware; the linear sum would give 1.96%, almost twice as much. For volume and weighing, only calibration, tolerance and temperature have been included: their random part is already contained in the repeatability of independent preparations and **must not be counted twice**.

6.8 How to report the result

The result is written as $y \pm U$ with its unit, the value of k and the approximate confidence level, for example "(498.6 ± 5.2) mg, $k = 2$, about 95%", or "498.6 mg ± 1.0% ($k = 2$)". As seen in Chapter 1, the uncertainty is

rounded to one or two significant figures and the result to the same decimal place. The value of k must always be stated, otherwise u_c cannot be recovered. Uncertainty is not a validation characteristic in ICH Q2(R2) (2023) or ICH M10 (2022), but it is a requirement of ISO/IEC 17025; validation data (intermediate precision, trueness) are the natural basis for estimating it (Chapter 12).

Key points of Chapter 6

- **Error** (difference from the true value, unknown) and **uncertainty** (dispersion of the values attributable to the measurand, which can be estimated) are distinct concepts.
- **GUM** (JCGM 100:2008): **Type A** = statistical analysis of repeated observations; **Type B** = other sources (certificates, specifications, literature, expert judgment). The distinction concerns the method of evaluation, not the nature of the error.
- Repeatability: $u_A = s/\sqrt{n}$ if the mean of n replicates is reported; s for a single determination.
- Type B: U/k from certificates; $a/\sqrt{3}$ (rectangular) or $a/\sqrt{6}$ (triangular) from tolerances.
- Independent contributions are combined **in quadrature** (the variances add up): absolute uncertainties for sums and differences, relative uncertainties for products and quotients; the linear sum overestimates.
- $U = k \cdot u_c$; $k = 2$ corresponds to about 95% for approximately normal distributions with sufficient degrees of freedom.
- The uncertainty budget identifies the dominant contributions, which are the ones worth acting on.

Test yourself

To practice on this chapter, use the quiz questions on the topics:

- **Measurement uncertainty (GUM)**
- **Uncertainty propagation in dilutions** (covered again in Chapter 16)

7. Linear regression and the calibration curve

Instrumental techniques (spectrophotometry, HPLC, LC-MS/MS, ICP-MS) measure a signal, not a concentration. The calibration curve converts the signal into a concentration, and linear regression is the statistical method used to build it; Chapter 8 checks whether it is adequate.

7.1 Calibration and regression

Measuring the signal of a series of **standards** of known concentration (Chapter 9) gives n pairs (x_i, y_i), from which **regression** estimates an equation that predicts one variable from the other. The roles are fixed: the **independent variable** X is the concentration of the standards, chosen by the analyst; the **dependent variable** Y is the instrumental signal, which is affected by random error. Least squares in fact attributes all the error to Y: the concentrations of the standards, prepared by weighing and dilution with calibrated glassware, are much less uncertain than the signal. Only afterwards, by **inverse prediction** (Section 7.4), is the concentration obtained from the signal.

We speak of **simple regression** when there is a single explanatory variable X (the calibration curve) and of **multiple regression** when several variables $X_1, X_2, \ldots, X_p$ determine Y. **Correlation** is something else: it only measures the degree of association between two quantitative variables, with an index (r, Section 7.6), without building an equation for prediction. Neither of them proves a cause-and-effect relationship: in calibration, that is guaranteed by the chemistry of the measurement (for example, the Beer–Lambert law). The relationship is linear only up to the limit of linearity (LOL), beyond which the detector saturates (Chapter 11).

In the slides and in the quiz the variables are written in capitals (X_i, Y_i, $\hat{Y}_i$); here we use lowercase letters (x_i, y_i, $\hat{y}_i$), with the same meaning.

7.2 The simple linear regression model

Each observation is the sum of a straight line and a random error:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{7.1}$$

Model parameters

- β_0 — **intercept**: expected value of Y for X = 0; in calibration it is the blank signal.
- β_1 — **slope**: change in Y for each unit increase in X; in calibration it is the **analytical sensitivity**, in units of signal per unit of concentration (for example L/mg for an absorbance as a function of mg/L). Its sign indicates whether the relationship is direct ($\beta_1 > 0$) or inverse ($\beta_1 < 0$).
- ε_i — **random error** of the i-th observation, which cannot be observed.

Greek letters denote population parameters, which are unknown and of which the data provide the **estimates** b_0 and b_1. In the same way, the **residual** $e_i = y_i - \hat{y}_i$, calculated from the data, is the estimate of the error ε_i, which cannot be observed; since b_0 and b_1 do not coincide with β_0 and β_1, in general $e_i \neq \varepsilon_i$. Sensitivity must not be confused with the limit of detection, which also depends on the noise of the blank (Section 8.4). Since the blank almost always gives a small signal, the intercept is estimated from the data and is normally not forced to zero (Chapter 10).

The estimates are reliable if the relationship is linear over the range, the errors are independent, have zero mean, are normally distributed and have constant variance (**homoscedasticity**), and the error in X is negligible; linearity and constant variance are checked with the residuals (Chapter 8).

7.3 The method of least squares

The **ordinary least-squares** (OLS) method chooses, among all possible lines, the one that minimizes the sum of the squared residuals, that is, of the **vertical** distances between the points and the line (Figure 7.1a): the error sum of squares $\mathrm{SSE} = \sum(y_i - \hat{y}_i)^2 \rightarrow$ minimum. The plain sum of the residuals is of no use: positive and negative residuals cancel out, and the sum is zero for any line passing through the centroid of the data.

Least-squares estimates

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}} \tag{7.2}$$

$$b_0 = \bar{y} - b_1\bar{x} \tag{7.3}$$

$$\hat{y}_i = b_0 + b_1 x_i \tag{7.4}$$

- $\bar{x}$, $\bar{y}$: means of the concentrations and of the signals of the standards; n: number of standards.
- S_{xy}: sum of cross-products (sum of the products of the deviations from the means); S_{xx}: sum of squares of X (sum of the squared deviations).
- $\hat{y}_i$: value predicted by the line for x_i.

Three properties are useful.

- **The sign of b_1 is that of the numerator**: S_{xx} is a sum of squares, always positive, so $b_1 > 0$ if Y increases with X and $b_1 < 0$ if it decreases.
- **The line always passes through the centroid ($\bar{x}$, $\bar{y}$)**, as can be seen by rewriting Equation 7.3 as $\bar{y} = b_0 + b_1\bar{x}$: the value predicted for $x = \bar{x}$ is $\bar{y}$, without calculating the intercept.
- **The sum of the residuals is zero**: this is a check on the calculations, not a proof that the model is adequate.

Common mistakes are calculating the slope as $\sum x_i\,y_i / \sum x_i^2$ (line forced through the origin) or as $\bar{y}/\bar{x}$, swapping numerator and denominator, or writing $b_0 = \bar{y} + b_1\bar{x}$.

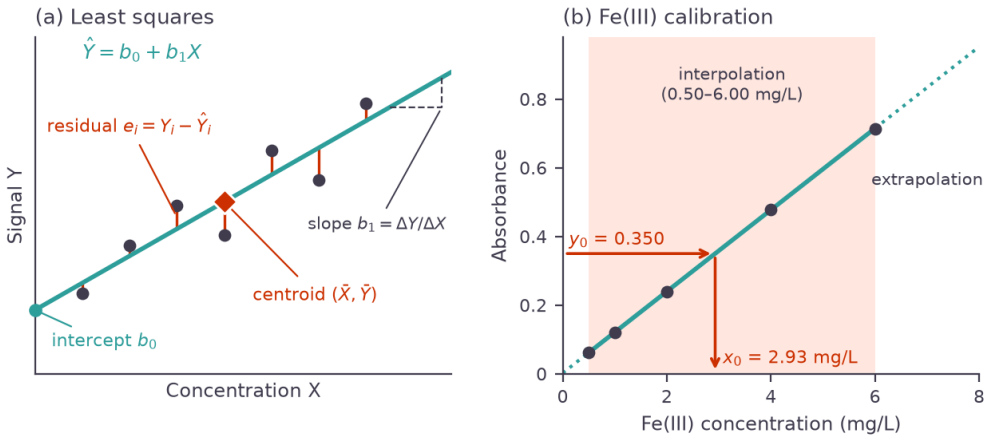


Figure 7.1 — (a) The least-squares line minimizes the sum of the squared residuals (red vertical segments), passes through the centroid of the data and crosses the Y axis at the intercept. (b) Fe(III) calibration of Example 7.3: between the lowest and the highest standard (shaded area) one works by interpolation, above 6.00 mg/L one would be extrapolating; the arrows show the inverse prediction for an absorbance of 0.350.

7.4 Interpolation, extrapolation and inverse prediction

Interpolation occurs when the prediction concerns a value of X **within** the interval used to estimate the line (between the lowest and the highest standard): it is reliable and recommended, because linearity has been

verified there. **Extrapolation** occurs when X lies **outside** the interval: it is risky, because beyond the range linearity may not hold (saturation at the top, noise at the bottom), and not even an r² very close to 1 guarantees it. In signal units the interpolation interval runs from $\hat{y}(x_{min}) = b_0 + b_1 x_{min}$ to $\hat{y}(x_{max}) = b_0 + b_1 x_{max}$, intercept included. A sample whose signal is too high is diluted until it falls within the range, and the result is then multiplied by the dilution factor (Chapter 14); a signal below the lowest standard is reported as "< LLOQ" (Chapter 11).

In routine work the signal y_0 of the sample is known and the concentration is obtained by solving $y_0 = b_0 + b_1 x_0$: this is **inverse prediction**, the fundamental operation of quantitative analysis (Figure 7.1b).

Inverse prediction

$$x_0 = \frac{y_0 - b_0}{b_1} \tag{7.5}$$

- y_0: signal of the sample (or mean of replicate readings); the intercept (blank signal) is **subtracted** and the result is **divided** by the slope, which converts the net signal into concentration.
- If the sample was diluted, x_0 must be multiplied by the dilution factor.

Checking the units avoids the most common mistakes (adding b_0, multiplying by b_1, inverting the fraction, using the direct equation $b_0 + b_1 y_0$): $(y_0 - b_0)$ is a signal and b_1 is signal per concentration, so their ratio is a concentration.

Example 7.1 — Interpolation interval and a sample to be diluted

An HPLC calibration with standards from 1.0 to 50.0 ng/mL gives the line Area = 850 + 1520 · X (X in ng/mL, area in arbitrary units, a.u.). A sample gives 92 000 a.u.

Step 1 — Limits in signal units. $\hat{y}(1.0)$ = 850 + 1520 × 1.0 = 2370 a.u.; $\hat{y}(50.0)$ = 850 + 1520 × 50.0 = 76 850 a.u.

Step 2 — Check. 92 000 a.u. is outside the interval: (92 000 − 850)/1520 = 60.0 ng/mL would be an extrapolation and must not be reported.

Step 3 — 1:5 dilution and new measurement. Area = 19 500 a.u., within the interval: x_0 = (19 500 − 850)/1520 = 12.27 ng/mL.

Result: 12.27 × 5 = 61.3 ng/mL. The extrapolated value was 2% too low: above the highest standard the response begins to saturate.

7.5 The scatter plot and the choice of model

Before calculating a line, one looks at the **scatter plot** (Y versus X): the shape of the cloud of points suggests the model. The typical situations are:

- **direct linear**: Y increases in proportion to X (almost all calibrations within the linear range);
- **inverse linear**: Y decreases linearly, negative slope (for example decolorization assays such as DPPH);
- **no relationship**: random points around a horizontal line, slope and r close to zero;
- **increasing curvilinear** (called direct polynomial): Y increases rapidly and then more and more slowly toward a plateau (detector saturation); a straight line forced through the data gives systematic residuals;
- **U-shaped curvilinear**: Y decreases and then increases (for example the van Deemter curve in chromatography); even though the relationship is strong, r is close to zero;
- **decreasing curvilinear** (called inverse polynomial): Y decreases rapidly and then more and more slowly, like an analyte that degrades over time.

Curved relationships are described with polynomials, exponentials or saturation functions; in calibration the straight line is preferred within the linear range, otherwise the simplest model that adequately describes the data (Chapter 10).

7.6 Decomposition of variability, r^2 and r

7.6.1 SST = SSR + SSE

The deviation of each point's signal from the mean is split into a part **explained** by the line and a **residual** (Figure 7.2): $y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$. Squaring and summing over all the points, for the least-squares line the cross term vanishes:

Decomposition of variability

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2$$

$$\text{SST} = \text{SSR} + \text{SSE} \qquad (7.6)$$

- **SST** (total sum of squares): total variability of the signals around their mean; it depends only on the data.
- **SSR** (regression sum of squares): variability **explained** by the line, due to the different concentrations of the standards.
- **SSE** (error sum of squares, or residual sum of squares): **unexplained** variability, $\sum e_i^2$, due to random error and to any shortcomings of the model.

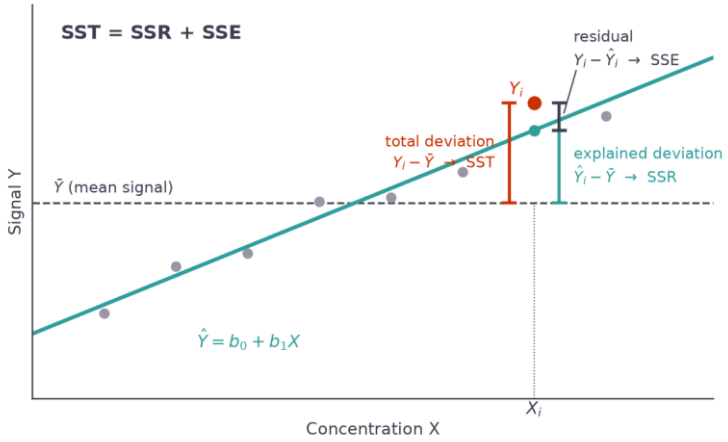


Figure 7.2 — Decomposition of the deviation of a point from the mean signal into a deviation explained by the line and a residual; summing the squares over all the points gives SST = SSR + SSE.

If all the points lie on the line (with a non-zero slope), the residuals are zero: SSE = 0 and SSR = SST. If the line is horizontal (b_1 = 0), every $\hat{y}_i$ equals $\bar{y}$: SSR = 0 and SSE = SST. These are **sums** of squares, not variances, and SSR can be obtained by difference (SST − SSE). The standard deviation of the residuals, $s_{y/x} = \sqrt{\text{SSE}/(n-2)}$, measures the scatter of the points around the line. Beware of the notation: here, as in the slides and in the quiz, SSR is the sum of squares due to *regression*, but some English-language books use SSR for the *sum of squared residuals* (our SSE).

7.6.2 Coefficient of determination and correlation coefficient

Coefficients r² and r

$$r^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}} \tag{7.7}$$

$$r = \frac{\sum(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \cdot \sum(y_i - \bar{y})^2}} \tag{7.8}$$

- r²: fraction of the variability of Y explained by the regression on X, between 0 and 1 (r² = 0.999: 99.9% explained, 0.1% unexplained).
- r (Pearson): strength and direction of the linear association, between −1 and +1; it has the same numerator as b_1, and therefore the same sign as the slope.
- In simple regression r² is the square of r: given r², $|r| = \sqrt{r^2}$ and the sign is taken from the slope.

What r² and r do not tell you

- **A high r² does not prove linearity**: a slight systematic curvature can give r² > 0.99 (Example 8.1). The residual plot must always be examined.
- **r² does not measure the slope**: lines with very different sensitivities can have the same r².
- **r measures only linear association**: in a symmetric U-shaped relationship the sum of cross-products is about zero and r ≈ 0, even though the relationship is strong.
- **Over wide ranges r² is dominated by the highest standards** and can exceed 0.9999 even with 50% errors on the lowest standards (Example 8.2). This is why ICH M10 (2022) judges the curve on the back-calculated concentrations, not on r² (Chapter 10).

Example 7.2 — Decomposition of variability, r² and r

Five standards (1–5 mg/L) give absorbances 0.12; 0.22; 0.35; 0.44; 0.57; the line is $\hat{y}$ = 0.004 + 0.112 x.

Step 1 — SST. $\bar{y}$ = 0.340; deviations −0.22; −0.12; +0.01; +0.10; +0.23; SST = 0.0484 + 0.0144 + 0.0001 + 0.0100 + 0.0529 = 0.1258.

Step 2 — Residuals and SSE. $\hat{y}$ = 0.116; 0.228; 0.340; 0.452; 0.564; residuals +0.004; −0.008; +0.010; −0.012; +0.006 (the 4 mg/L standard lies below the line). SSE = 0.000016 + 0.000064 + 0.000100 + 0.000144 + 0.000036 = 0.000360.

Step 3 — SSR, r² and r. SSR = 0.1258 − 0.00036 = 0.12544; r² = 0.12544/0.1258 = 0.9971; r = +√0.9971 = +0.9986 (positive slope).

Result: r² = 0.997: 99.7% of the variability is explained by the line. With a negative slope and the same r², r would be −0.9986.

7.7 A complete example: spectrophotometric calibration of Fe(III)

Example 7.3 — Fe(III) calibration line and an unknown sample

Five Fe(III) standards give the absorbances in Table 7.1; a sample gives y_0 = 0.350.

Table 7.1 — Fe(III) calibration: data, intermediate calculations, predicted values and residuals

x_i (mg/L)	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$\hat{y}_i$	e_i
0.50	0.062	−2.20	−0.2608	0.57376	4.84	0.06153	+0.00047
1.00	0.121	−1.70	−0.2018	0.34306	2.89	0.12091	+0.00009
2.00	0.238	−0.70	−0.0848	0.05936	0.49	0.23967	−0.00167
4.00	0.479	1.30	0.1562	0.20306	1.69	0.47719	+0.00181
6.00	0.714	3.30	0.3912	1.29096	10.89	0.71471	−0.00071
Sum	1.614			**2.47020**	**20.80**		**0.00000**

Step 1 — Means. $\bar{x}$ = 13.50/5 = 2.70 mg/L; $\bar{y}$ = 1.614/5 = 0.3228.

Step 2 — Slope and intercept. b_1 = 2.4702/20.80 = 0.11876 L/mg; b_0 = 0.3228 − 0.11876 × 2.70 = 0.00215.

Step 3 — Residuals. Calculated with the unrounded coefficients, they sum to zero, are small (at most 0.0018) and change sign without forming arches or funnels; the back-calculated concentrations (for example (0.062 − 0.00215)/0.11876 = 0.504 mg/L) differ from the nominal values by less than 1%.

Step 4 — r². SST = $\sum(y_i - \bar{y})^2$ = 0.293367; SSE = $\sum e_i^2$ = 6.80 × 10^{-6}; SSR = 0.293360; r² = 0.99998 (r = +0.99999).

Step 5 — Inverse prediction. y_0 = 0.350 lies between $\hat{y}$(0.50) = 0.0615 and $\hat{y}$(6.00) = 0.7147: we interpolate. x_0 = (0.350 − 0.00215)/0.11876 = 0.34785/0.11876 = 2.929 mg/L.

Result: x_0 = 2.93 mg/L. Linearity is adequate because r² = 0.99998 **and** the residuals are random: r² alone would not be enough.

Rounding: 2.92 or 2.93 mg/L?

With the rounded coefficients b_1 = 0.119 L/mg and b_0 = 0.002, one obtains x_0 = 0.348/0.119 = 2.92 mg/L. The difference comes mainly from the slope: 0.119 is 0.2% higher than 0.11876, so x_0 comes out 0.2% lower. With the rounded line the residuals no longer sum to zero either. Rule: keep at least 4 significant figures in the slope (0.1188 L/mg) and round only the final result (Chapter 1). If, however, an exercise gives the line already rounded ($\hat{y}$ = 0.002 + 0.119 x), those coefficients are used and the correct result of the calculation is 2.92 mg/L.

Key points of Chapter 7

- In calibration X is the concentration of the standards (independent variable) and Y the signal (dependent variable): $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$. Regression is used for prediction, correlation measures association; neither proves causation.
- β_0 is the blank signal, β_1 the sensitivity; the residual $e_i = y_i - \hat{y}_i$ estimates the error ε_i.
- Least squares minimizes $\sum e_i^2$: $b_1 = S_{xy}/S_{xx}$, $b_0 = \bar{y} - b_1\bar{x}$; the line passes through ($\bar{x}$, $\bar{y}$) and the sign of b_1 is that of the sum of cross-products.
- Inverse prediction $x_0 = (y_0 - b_0)/b_1$ only for signals within the interval (interpolation); above the range the sample is diluted, below it the result is reported as "< LLOQ".
- SST = SSR + SSE, with SSR the regression sum of squares; r^2 = SSR/SST (0–1); r between −1 and +1 with the sign of the slope; a high r^2 does not prove linearity and does not measure sensitivity.
- Round only at the end: Fe(III), 2.93 mg/L with the unrounded coefficients.

Test yourself

In the self-assessment quiz, practice with the topics:

- Regression model and scatter plot
- Least squares method
- Interpolation, extrapolation and inverse prediction
- Variability decomposition, r^2 and r

8. Regression diagnostics and the link to validation

A least-squares line can be calculated for any data set, even when the model is not appropriate. Its assumptions must therefore be checked with the residuals, a non-constant variance must be handled correctly, and one must know how systematic errors distort the line: these are the same tools used in method validation (Part II).

8.1 Residual analysis

The **residual** is the estimate of the error made in predicting y_i with $\hat{y}_i$:

$$e_i = y_i - \hat{y}_i \tag{8.1}$$

A positive residual indicates a point above the line, a negative one a point below it. For example, with the line $\hat{y}$ = 0.010 + 0.0950 x, a 2.00 mg/L standard with absorbance 0.196 has $\hat{y}$ = 0.200 and e = −0.004, and contributes (−0.004)² = 0.000016 to SSE (the error sum of squares). The deviations $\hat{y}_i - \bar{y}$ and $y_i - \bar{y}$ enter SSR (the regression sum of squares) and SST, and are not residuals. The sum of the residuals is zero by construction and does **not** prove that the model is adequate.

The **residual plot** (e_i versus X) is the fundamental diagnostic tool, because it magnifies deviations that are invisible in the plot of the line; relative residuals or the percentage deviations of the back-calculated concentrations (Chapter 10) are also used. If the model is adequate, the residuals are distributed **randomly** around zero in a uniform band (Figure 8.1a). An **arch-shaped** pattern (Figure 8.1b) indicates **curvature**: the linear model is not adequate, and the range must be restricted to the linear region or a polynomial (usually quadratic) or non-linear model must be used. A **funnel-shaped** pattern (Figure 8.1c) indicates heteroscedasticity (Section 8.2). A single very large residual signals a possible outlier to be investigated (Chapter 5).

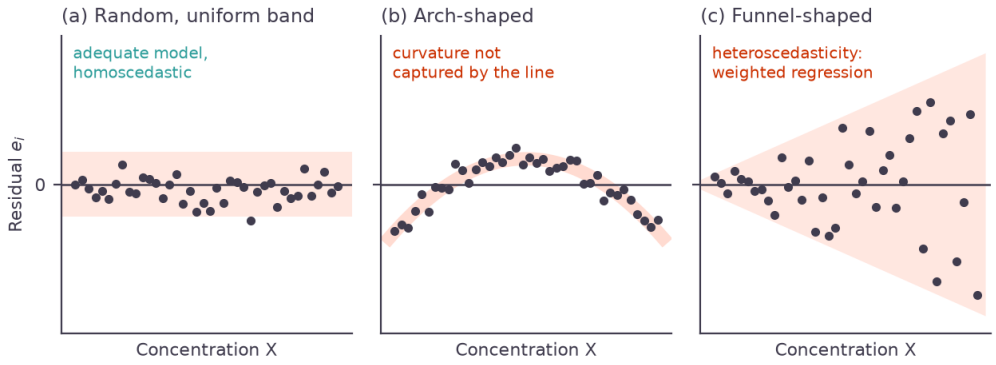


Figure 8.1 — Typical patterns of the residual plot. (a) Random, in a uniform band: adequate model, constant variance. (b) Arch-shaped: curvature not described by the line. (c) Funnel-shaped: variance increasing with concentration (heteroscedasticity).

Example 8.1 — A curvature hidden by a high r²

Eight standards (1, 2, …, 8 mg/L) give absorbances 0.098; 0.189; 0.278; 0.360; 0.437; 0.511; 0.577; 0.640. The least-squares line is $\hat{y}$ = 0.0375 + 0.0775 x, with r² = 0.9958.

Step 1 — Residuals. −0.0170; −0.0035; +0.0080; +0.0125; +0.0120; +0.0085; −0.0030; −0.0175: the signs (− − + + + + − −) trace an arch.

Step 2 — Effect on the result. The 1 mg/L standard is back-calculated as (0.098 − 0.0375)/0.0775 = 0.78 mg/L, that is, −21.9%.

Step 3 — Alternative model. The curve $\hat{y}$ = 0.100 x − 0.0025 x^2 leaves residuals within ±0.001, with no pattern.

Result: despite r² = 0.9958 the linear model is not adequate; only the residuals (or the back-calculation) reveal the curvature.

8.2 Homoscedasticity, heteroscedasticity and weighted regression

Homoscedasticity means that the variance of the errors is **constant** across values of X: it is an assumption of ordinary least-squares (OLS) regression, and in the residual plot it appears as a uniform band. **Heteroscedasticity** means that the variance **changes** with X, usually increasing: the residuals form a **funnel**. It is common over wide concentration ranges, for example in LC-MS/MS, where the CV% of the signal is roughly constant, so that the standard deviation increases with concentration.

With heteroscedastic data, OLS is dominated by the most concentrated standards, which have the largest absolute errors: the intercept is poorly estimated and low concentrations are affected by large relative errors, while r² remains extremely high. **Weighted regression** (*weighted least squares*) is then used: it minimizes $\sum w_i\,(y_i - \hat{y}_i)^2$ with weights w_i inversely proportional to the variance of the signal, so that the most variable points, that is, the most concentrated ones, count less. The most common weights are $w_i = 1/x_i^2$, suitable when the standard deviation is proportional to concentration (constant CV%, the typical case in LC-MS/MS), and $w_i = 1/x_i$, suitable when it is the variance that increases in proportion to concentration. The ratio between the weights of the extreme standards (1 with OLS) is:

$$\frac{w(x_{min})}{w(x_{max})} = \left(\frac{x_{max}}{x_{min}}\right)^k \tag{8.2}$$

with k = 1 for $1/x$ weights and k = 2 for $1/x^2$ weights. The simplest weighting is chosen that makes the relative residuals homogeneous and minimizes the sum of the percentage deviations of the back-calculated concentrations (Chapter 10).

Example 8.2 — Ordinary and weighted regression in LC-MS/MS

An LC-MS/MS calibration in plasma (1–500 ng/mL, eight standards, signal = analyte/SIL-IS peak area ratio) has a roughly constant signal CV%.

Step 1 — Ratio between the weights (Equation 8.2). 500/1 = 500: with $1/x$ weights the 1 ng/mL standard weighs 500 times as much as the 500 ng/mL one; with $1/x^2$ weights, 500² = 250 000 times.

Step 2 — Lines (from a spreadsheet): OLS $\hat{y}$ = 0.00638 + 0.009778 x (r² = 0.9999); $1/x^2$ weights $\hat{y}$ = 0.00119 + 0.009970 x.

Table 8.1 — Percentage deviations of the back-calculated concentrations

x_i (ng/mL)	1	2	5	10	50	100	250	500
OLS (%)	−50.7	−27.3	−5.7	−0.2	−0.9	+3.8	−1.0	+0.1
$1/x$ weights (%)	−4.8	−4.5	+3.2	+4.0	−0.4	+3.8	−1.2	−0.2
$1/x^2$ weights (%)	+0.4	−2.6	+2.9	+3.1	−1.7	+2.3	−2.7	−1.7

Result: $1/x^2$ weighted regression is chosen. With OLS, despite r² = 0.9999, the lowest standard is back-calculated at −50.7%, because the intercept (0.00638) is more than half of its signal (0.0112). With weighting, all deviations fall within ±5%, and $1/x^2$ gives the smallest sum of absolute deviations (17.4% versus 22.1% with $1/x$ and 89.7% with OLS).

8.3 Effect of systematic errors on the calibration curve

A **constant error** (Chapter 2) adds or subtracts the same amount of signal at every level: the line shifts parallel to itself and the **intercept** changes. A **proportional error** alters the signal by a fixed fraction: the **slope** (apparent sensitivity) changes, not the intercept. Neither changes r² or the pattern of the residuals.

The direction of the bias depends on **where** the error acts, because the sample signal is compared with the line. With β_0, β_1 the coefficients of the correct line and x_{true} the true concentration:

- **constant error $+c$ on the standards only** (for example a contaminated diluent used only for the standards): higher intercept and

$$x_0 = x_{\text{true}} - \frac{c}{\beta_1} \tag{8.3}$$

that is, an underestimate (an overestimate if $c < 0$); the signal error becomes a concentration error when divided by the slope, and it weighs more at low concentrations;

- **constant error on the sample only** (color or turbidity of the matrix, if the sample blank is not subtracted): $x_0 = x_{\text{true}} + c/\beta_1$, an overestimate;
- **proportional error on the standards only** (stock solution more or less concentrated than stated): the slope becomes $\beta_1(1 + p)$ and

$$x_0 = \frac{x_{\text{true}}}{1 + p} \tag{8.4}$$

so a higher slope ($p > 0$) leads to an underestimate and a lower one to an overestimate; with a +10% slope, x_0 decreases by 9.1% (factor 1/1.10), not by 10%;

- **proportional error on the sample only** (incomplete extraction, that is, recovery R < 100%, or ionization suppression with standards in solvent): $x_0 = R \cdot x_{\text{true}}$, an underestimate; signal enhancement gives an overestimate (Chapter 13);
- **same error on standards and sample** (reagent signal, matrix-matched standards extracted like the samples, compensation with a SIL-IS): no bias. An extraction that is inefficient but equal for all lowers the sensitivity, not the trueness.

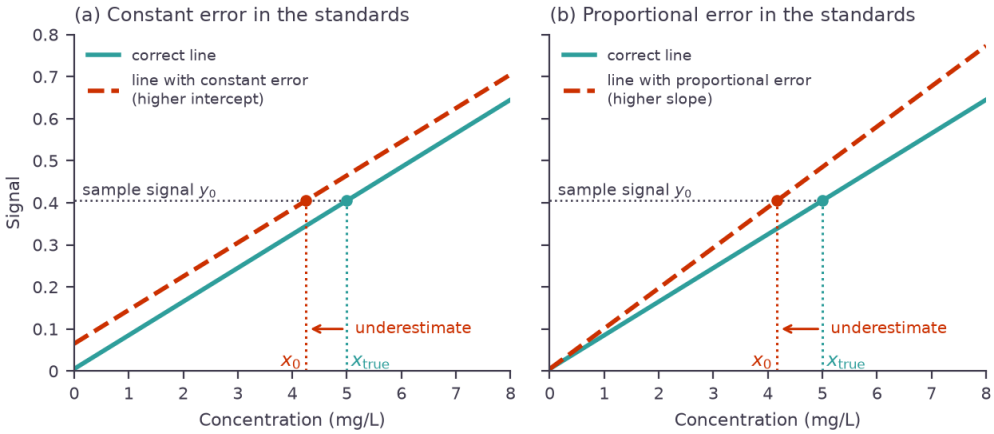


Figure 8.2 — Systematic error affecting only the standards, with the sample signal unchanged (errors exaggerated for clarity): (a) a positive constant error raises the intercept; (b) a positive proportional error increases the slope. In both cases the sample signal meets the biased line at a concentration lower than the true one.

Example 8.3 — Standards or sample?

The correct line is $\hat{y}$ = 0.005 + 0.0800 x (x in mg/L); a 5.00 mg/L sample gives y_0 = 0.405.

Case 1 — Constant error +0.020 on the standards only. x_0 = (0.405 − 0.025)/0.0800 = 4.75 mg/L: an underestimate of 0.020/0.0800 = 0.25 mg/L (−5%); for a 1.00 mg/L sample it would be −25%.

Case 2 — Same error on the sample only. x_0 = (0.425 − 0.005)/0.0800 = 5.25 mg/L (+5%); if present in both, (0.425 − 0.025)/0.0800 = 5.00 mg/L.

Case 3 — Slope of the standards +5%. x_0 = 0.400/0.0840 = 4.76 mg/L: −4.8% (= 1/1.05 − 1).

Case 4 — 80% recovery on the sample only. y_0 = 0.005 + 0.0800 × 0.80 × 5.00 = 0.325; x_0 = 0.320/0.0800 = 4.00 mg/L (−20%). With standards extracted in the same way, slope 0.0640 L/mg and x_0 = 5.00 mg/L.

Result: the same error gives an underestimate, an overestimate or no bias depending on whether it acts on the standards, on the sample or on both.

8.4 Link to method validation

- **Linearity and range**: r², residual plot, back-calculation of the standards and the **lack-of-fit** test, which, with replicates at each level, compares the distance of the means from the line with the scatter of the replicates (Chapters 10 and 11).
- **LOD and LOQ**: from the standard deviation of the blank s_B and the slope, for example $\text{LOD} = 3.3 \cdot s_B/b_1$ and $\text{LOQ} = 10 \cdot s_B/b_1$ (ICH Q2(R2), 2023). The slope converts the noise, in signal units, into concentration: for the same noise, a larger slope gives a lower LOD (with s_B = 0.0020 and b_1 = 0.119 L/mg, LOD = 0.055 mg/L; with twice the slope, 0.028 mg/L; Chapter 11).
- **Trueness and precision**: bias and recovery with respect to a reference value; s and CV% under repeatability, intermediate precision and reproducibility conditions (Chapters 3 and 12).
- **Recovery, selectivity, matrix effect**: spiking experiments, interferents, comparison of responses or slopes in solvent and in matrix (Chapters 11 and 13).

- **Measurement uncertainty**: it is not an ICH Q2(R2) validation characteristic, but a requirement of ISO/IEC 17025 (2017); it is estimated according to the GUM (Type A and Type B contributions) using the precision and trueness data from validation, as indicated by the Eurachem guide (2025) (Chapter 6).

8.5 Summary of the key concepts of Part I

Table 8.2 — Key concepts of statistics applied to analytical methods

Concept	In brief
Uncertainty	Every measurement has an uncertainty: result = value ± U, with correct units and significant figures (Chapters 1 and 6).
Errors	Systematic: bias = $\mu - x_t$, detectable and correctable. Random: can be reduced and quantified, not eliminated (Chapter 2).
Precision, trueness, accuracy	Precision = scatter (s, CV%); trueness = absence of bias; **accuracy = trueness + precision** (Chapter 3).
Gaussian distribution	68.3%, 95.4% and 99.7% of values within ±1σ, ±2σ, ±3σ; total area = 1; confidence interval $\bar{x} \pm t \cdot s/\sqrt{n}$ (Chapter 5).
GUM	Type A (statistical) and Type B (other information) contributions; $U = k \cdot u_c$, with k = 2 for about 95% (Chapter 6).
Regression	$\hat{y} = b_0 + b_1 x$ by least squares; r² **and** residuals; $x_0 = (y_0 - b_0)/b_1$ by interpolation (Chapters 7 and 8).

Key points of Chapter 8

- The residual $e_i = y_i - \hat{y}_i$ estimates the error; if the model is adequate the residuals are random around zero. A zero sum of residuals and a high r² do not prove adequacy; arch-shaped residuals indicate curvature.
- Homoscedasticity = constant variance (an assumption of OLS); funnel-shaped residuals = heteroscedasticity, to be handled with weighted regression ($1/x$ or $1/x^2$), which reduces the influence of the most concentrated standards.
- Constant error → intercept; proportional error → slope. On the standards only: a higher intercept or slope underestimates the sample; on the sample only the direction is reversed; on both the error cancels out.
- Validation: r², residuals, lack of fit and back-calculation for linearity; s_B/b_1 for LOD and LOQ; uncertainty is estimated from validation data but is not an ICH Q2(R2) characteristic.
- Accuracy = trueness + precision.

Test yourself

In the self-assessment quiz, practice with the topics:

- Residual analysis and homoscedasticity
- Systematic errors and the calibration line
- Link to method validation

Part II — Calibration and Validation of Analytical Methods

An analytical method gives useful results only if it is calibrated correctly and if documented evidence shows that it is fit for purpose. This part follows the path from reference standards to the calibration curve, up to the validation parameters and acceptance criteria of the international guidelines, with special attention to LC-MS/MS bioanalysis.

9. Standards and reference materials

In instrumental methods the concentration is obtained by comparing the signal of the sample with that of standards of known concentration: a result can therefore be no more reliable than the standards used to obtain it. An error in purity, in weighing or in a dilution is a **systematic error** (Chapter 2) that is passed on unchanged to every sample and that no number of replicates can reveal.

9.1 The hierarchy of analytical standards

Standards are organized in a **hierarchy** in which each level receives its value by comparison with the level above: at the top, a few materials with the smallest uncertainty; at the bottom, the solutions used every day (Figure 9.1).

Analytical standards

- **Primary standard** (PS): a substance of very high, certified purity, stable, non-hygroscopic and of known stoichiometry, whose amount is established directly by weighing. It is the highest-level standard in the laboratory and is not calibrated against other standards of the same quantity.
- **Reference material** (RM): a material sufficiently homogeneous and stable with respect to specified properties, fit for its intended use (VIM 5.13).
- **Certified reference material** (CRM): an RM accompanied by a certificate issued by an authoritative body, which states one or more property values with their **uncertainty** and **traceability**, obtained using valid procedures (VIM 5.14).
- **Secondary standard**: a substance or solution that does not meet all the requirements of a PS and whose content is established by comparison with a PS or a CRM.
- **Working standard**: a diluted solution, prepared from the stock solution, used in routine work to build the calibration curve.

9.1.1 Primary standards and their requirements

A primary standard is weighed, dissolved in a known volume, and directly yields a solution of known concentration. Classic examples are potassium hydrogen phthalate (KHP, used to standardize bases), sodium carbonate (for acids), sodium oxalate (for permanganate), potassium dichromate and potassium iodate. The expression "absolute reference" should be taken with caution: the purity of a primary standard has also been determined analytically and **has an uncertainty** (for example 99.98% ± 0.02%). To be used as such, it must meet **all four** requirements **at the same time** (Table 9.1).

Table 9.1 — The requirements of a primary standard

Requirement	Why it is needed
Purity ≥ 99.9%, with documented certification	the weighed mass must correspond to the mass of analyte
Stability over time and under the storage conditions	the composition must not change during storage (degradation, oxidation, reaction with CO_2)
Non-hygroscopicity	it must not absorb moisture during weighing; otherwise water is weighed too
Known stoichiometry (defined formula, exact molar mass)	the amount of substance is calculated as $n = m/M$

Stability and non-hygroscopicity are distinct requirements: the first concerns storage over time, the second the moment of weighing (a hygroscopic substance makes the balance reading creep upward). NaOH, for example, has a defined formula but is hygroscopic and absorbs CO_2: it is not a primary standard, and a defined formula alone is not enough.

9.1.2 Reference materials and certified reference materials

An **RM** must be homogeneous and stable, but it does not necessarily have a certified value (for example, a plasma *pool* used as an in-house control). Only a **CRM** has **certified values with uncertainty** (usually expanded, with k = 2: Chapter 6) and a statement of traceability. There are **pure-substance** CRMs and **certified calibration solutions**, which are used for calibration, and **matrix CRMs** (sediment, serum, milk powder), the only ones that reproduce the matrix of real samples. The latter are used to **verify the trueness of the entire procedure** (Chapter 12). To check a method for lead in sediments, digestion included, the appropriate material is a sediment CRM with certified lead content. A pure lead solution only checks the calibration; a non-certified sediment allows precision to be estimated, but not trueness.

Major producers include NIST (USA), JRC (EU), LGC (United Kingdom) and BAM (Germany). The competence requirements for **reference material producers** are set by **ISO 17034** (2016); ISO/IEC 17025 (2017), by contrast, applies to testing and calibration laboratories.

9.1.3 Secondary standards

A secondary standard receives its value by **standardization** against a PS or a CRM. An NaOH solution prepared by approximate weighing becomes a secondary standard after it has been titrated with KHP (Example 9.1). In bioanalysis one usually starts from commercial reference standards with a **certificate of analysis** (CoA), which ICH M10 (2022) requires for the analyte: the purity stated in the CoA must always be used in the calculations.

Example 9.1 — Standardizing NaOH with KHP

Data. 0.5124 g of KHP (M = 204.22 g/mol, purity 99.95%) requires 25.43 mL of "approximately 0.1 mol/L" NaOH; the reaction is 1:1.

1. $n(\mathrm{KHP}) = \frac{0.5124 \times 0.9995}{204.22} = 2.508 \times 10^{-3}$ mol $= n(\mathrm{NaOH})$.
2. $c(\mathrm{NaOH}) = \frac{2.508 \times 10^{-3}\ \mathrm{mol}}{0.02543\ \mathrm{L}} = 0.09862$ mol/L.

Result: 0.09862 mol/L. Using the nominal value of 0.1000 mol/L, all subsequent titrations would be overestimated by 1.4%.

9.2 Metrological traceability

Metrological traceability (VIM 2.41)

Property of a **measurement result** whereby the result can be related to a reference (usually an SI unit) through a **documented unbroken chain of calibrations**, each contributing to the measurement uncertainty.

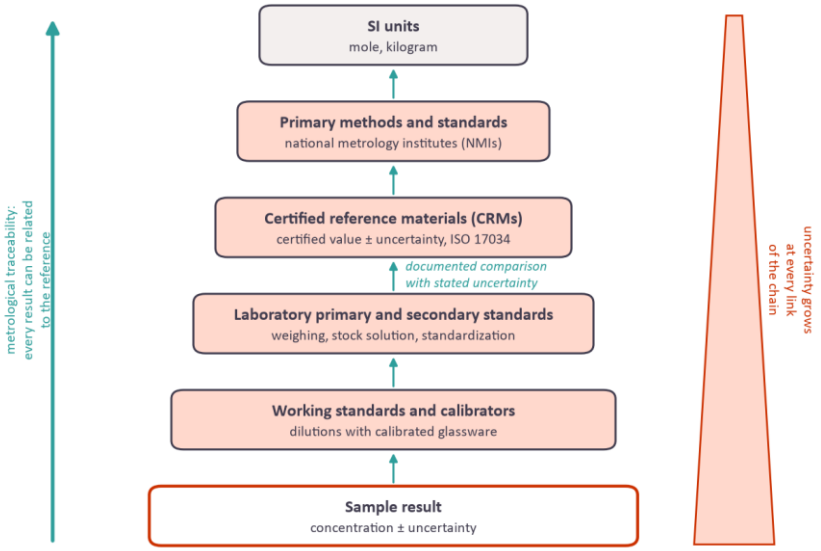


Figure 9.1 — Hierarchy of standards and traceability chain: each link is a documented comparison with stated uncertainty; moving up the chain, the result is related to SI units; moving down, the uncertainty grows.

Every link in the chain (CRM certificate, balance calibration, glassware class, preparation record) must have a **stated uncertainty**. Because each step adds its own contribution, **the uncertainty grows with distance from the reference**, and an undocumented link **breaks** traceability. Traceability is a property of the **result**, not of the material: the CRM is the tool with which the laboratory establishes it. It makes results obtained in different laboratories and at different times comparable, but it does not guarantee the trueness of the method: an incomplete extraction produces a result that is traceable but underestimated (Chapter 13).

9.3 From the PS to working standards

Preparation follows the sequence **weighing of the PS → stock solution → diluted solution → working standards** (Figure 9.2). Weighing is done on a calibrated analytical balance (preferably by difference, avoiding very small masses: with a readability of 0.01 mg, 10 mg gives a resolution of 0.1% and 1 mg of 1%); the material is made up to volume in a **volumetric flask** and diluted with **calibrated pipettes and volumetric flasks**:

$$c_{\text{stock}} = \frac{m \cdot P \cdot f}{V} \tag{9.1}$$

$$c_{\text{working}} = c_{\text{stock}} \cdot \frac{V_{\text{aliquot}}}{V_{\text{final}}} = \frac{c_{\text{stock}}}{\text{DF}} \tag{9.2}$$

where m is the weighed mass, P the purity as a fraction, V the volume of the flask, f the ratio between the molar mass of the analyte and that of the weighed form (for example salt → free base; f = 1 if they are the same) and DF the dilution factor; with several dilutions, the overall factor is the product of the individual factors.

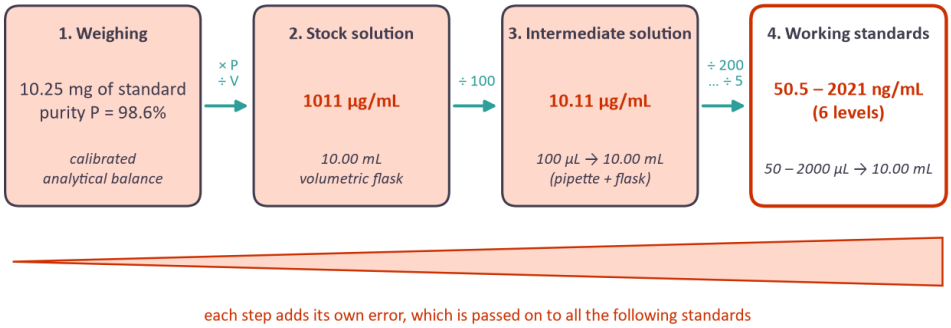


Figure 9.2 — From weighing to working standards (data from Example 9.2): the error of each step adds to the previous ones and is passed on to all subsequent standards.

Error propagates through dilutions. The relative error of the weighing is passed on unchanged to all derived solutions (dilutions do not reduce it), and each pipetting step adds its own: pipetting precision is as critical as weighing precision. It is advisable to limit the number of steps and to avoid very small volumes; the propagation calculation is given in Chapter 16. So that an error in the stock solution can be detected, ICH M10 recommends preparing **calibrators and QCs from separate stock solutions** (Chapter 12).

Example 9.2 — From weighing to working standard

Data. 10.25 mg of standard (purity 98.6%) in a 10.00 mL volumetric flask; 100 µL of stock solution made up to 10.00 mL; 500 µL of this intermediate solution made up to 10.00 mL.

1. Stock solution (Eq. 9.1): $c = 10.25 \times 0.986/10.00 = 1.0107$ mg/mL = 1011 µg/mL.
2. Intermediate solution (DF = 100): 10.11 µg/mL.
3. Working standard (DF = 20): $10.107/20 = 0.5053$ µg/mL.

Result: 505 ng/mL. Ignoring the purity would give 512.5 ng/mL (+1.4%); dividing by the purity instead of multiplying, 519.8 ng/mL. Errors of this kind are systematic: they shift all the sample results in the same direction.

Key points of Chapter 9

- Hierarchy: **primary standards**, **RMs/CRMs**, **secondary standards**, **working standards**; each level receives its value by comparison with the level above.
- A PS must meet **all** the requirements: certified purity ≥ 99.9%, stability, non-hygroscopicity, known stoichiometry; even its purity has an uncertainty.
- An **RM** is homogeneous and stable; only a **CRM** has certified values with uncertainty and traceability. Producers follow **ISO 17034**. Matrix CRMs verify the trueness of the entire procedure.
- A **secondary standard** (e.g. NaOH) is standardized against a PS (e.g. KHP) or a CRM.
- **Traceability**: a documented unbroken chain of comparisons, each with a stated uncertainty; the uncertainty grows with distance from the reference.
- Preparation: weighing → stock → diluted → working, with calibrated glassware; correct for purity (and salt form); each step adds an error that is passed on downstream.

Test yourself

- **Standards and metrological traceability**
- **Standard preparation; absolute and relative methods** (questions on standard preparation)

10. Calibration methods

To calibrate means to establish, with standards of known concentration, the relationship between signal and amount of analyte, and then to apply it to the samples. If standards and samples do not respond in the same way (a matrix that suppresses the signal, a variable injected volume), a statistically perfect curve can give systematically wrong results: this is why the choice of calibration method matters as much as the regression, which is covered in Chapters 7 and 8.

10.1 Absolute and relative methods

In **absolute methods** the relationship between the measured quantity and the amount of analyte is given by **physical laws or stoichiometry**: gravimetry (mass of a precipitate of known composition), titrimetry (volume of titrant and stoichiometric ratio), coulometry (Faraday's law). No calibration curve is needed, but the "known" quantities must really be known: a calibrated balance and glassware and, in titrations, a **titrant standardized against a primary standard** (in an acid–base titration, NaOH against KHP: Example 9.1). An error in the titrant concentration is a proportional systematic error on all results. The exception is a titrant that is itself a primary standard prepared by direct weighing (potassium dichromate or potassium iodate).

In **relative methods** the constant linking signal and concentration is not known a priori, because it depends on the instrument, the conditions and the analyte: it must be determined with standards, by building a **calibration curve**. Most instrumental methods belong to this group: spectroscopy (UV-Vis, AAS, ICP-OES, ICP-MS), chromatography (HPLC, GC) and mass spectrometry (LC-MS/MS).

10.2 The analytical blank

The **blank** is a sample that contains everything except the analyte and is measured like the samples: its signal estimates the contribution not due to the analyte (Table 10.1).

Table 10.1 — Types of analytical blank

Type of blank	Composition	What it corrects
Solvent	the sample solvent only	the instrumental background of the solvent
Reagent	solvent and all the reagents of the preparation	the additive interferences of the reagents
Procedural	taken through the whole procedure, usually without matrix (e.g. reagent water)	contamination introduced during treatment
Matrix	analyte-free matrix, treated like the sample	all contributions not due to the analyte, matrix included

The **matrix blank** is the best approximation of the "true blank", because it reproduces both the composition of the sample and the entire treatment; the procedural blank, on the other hand, goes through the whole procedure but does not necessarily contain the matrix. An analyte-free matrix does not always exist (endogenous analytes). In bioanalysis, ICH M10 (2022) distinguishes the **blank sample** (matrix with neither analyte nor internal standard) from the **zero sample** (matrix with the internal standard only).

What blank subtraction corrects. An **additive** contribution (background, reagents, matrix components that give a signal) adds to the analyte signal without changing the slope: subtracting the blank removes it. A **multiplicative** effect (ionization suppression or enhancement in LC-MS/MS, different nebulization in ICP, incomplete extraction), by contrast, changes the **slope**, that is, the sensitivity: in the blank, where the analyte is absent, it produces no signal, so subtraction **does not correct it**. Matrix-matched calibration, an internal standard or standard additions are needed (Sections 10.6–10.7; constant and proportional errors: Chapters 2 and 8). If the calibrators are prepared and treated like the samples, the blank contribution is already contained in the intercept of the curve and must not be subtracted a second time.

10.3 Choice of signal: peak area or height

The **peak area** collects the entire response of the analyte and is **less sensitive to changes in shape**: if the peak broadens or shows *tailing*, for the same injected amount the height decreases while the area remains practically constant. For this reason, in chromatography and LC-MS/MS, area is the **de facto standard** for quantification unless the method states otherwise. **Height** is easier to measure manually, but it is acceptable only if the peak shape is **constant and reproducible**.

10.4 External standard calibration

In **external standard calibration** (ES), standards of known concentration prepared **separately** from the sample are measured, the curve is determined by regression, and the sample concentration is obtained by **interpolation** (Figure 10.1):

$$C_{\text{sample}} = \frac{S - b_0}{b_1} \cdot \text{DF} \tag{10.1}$$

where S is the sample signal, b_0 and b_1 the intercept and slope of the curve, and DF the dilution factor applied to the sample (1 if undiluted).

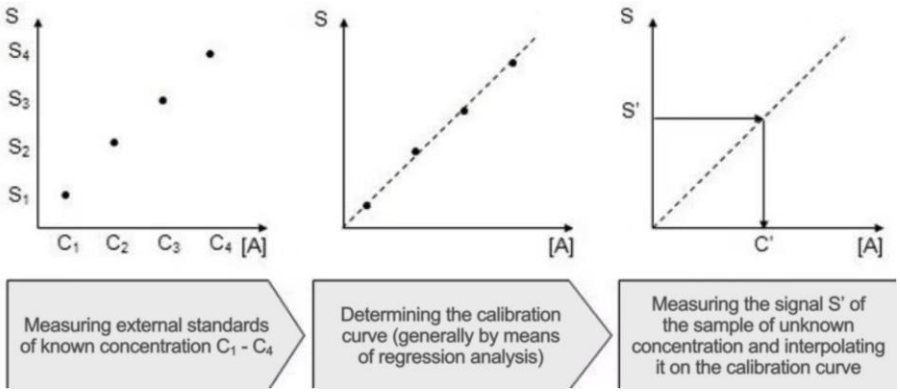


Figure 10.1 — External standard calibration: measurement of the standards, determination of the curve by regression, interpolation of the sample signal S′. The diagram is simplified: a real curve has at least 6 levels, and the intercept must be estimated, not forced to zero.

It is a **simple and fast** method: a single curve serves all samples, and the number of standards is small. Its **critical assumption** is that the response to the analyte is **identical in the standard and in the sample**: same chemical form and no matrix interferences. ES **compensates neither for the matrix effect nor for injection variability**: if the matrix suppresses the signal by 20%, the result is underestimated by 20%. Preparing the calibrators in blank matrix (*matrix-matched*) compensates only for the average matrix effect, not for differences between samples. ES is therefore the choice for routine analysis of many samples with a **simple and reproducible matrix**. The curve is valid only for interpolation: a sample above the ULOQ must be diluted and reanalyzed (Chapter 14).

Example 10.1 — External standard: dilution and matrix effect

Data. Curve from 2.00 to 50.0 ng/mL: $A = 1850 \cdot C + 950$ (A in a.u., C in ng/mL). A sample diluted 1:10 (0.100 mL made up to 1.00 mL) gives A = 46 200 a.u.

1. Measured solution: $C = (46\,200 - 950)/1850 = 24.46$ ng/mL, within the calibration range.
2. Original sample (Eq. 10.1, DF = 10): $24.46 \times 10 = 244.6$ ng/mL.
3. If the sample matrix reduced the response by 20% (multiplicative effect), the net signal would be 0.80 times the expected one and the result would be $0.80 \times 244.6 = 195.7$ ng/mL: blank subtraction would change nothing, because the blank contains no analyte.

Result: 245 ng/mL. Typical errors: forgetting DF (24.5 ng/mL), dividing by DF (2.45 ng/mL), ignoring the intercept (249.7 ng/mL).

10.5 The calibration curve: operational requirements

The following requirements are those of ICH M10 for chromatographic methods (within its scope, the guideline has superseded EMA 2011 and FDA 2018).

1. **At least 6 concentration levels** from the LLOQ (the lowest calibrator) to the ULOQ (the highest; Chapter 11), **plus a blank sample and a zero sample**, which do not count among the levels and are not included in the regression.
2. **Least-squares regression** (Chapter 7), weighted if necessary ($1/x$, $1/x^2$) when the data are heteroscedastic (Chapter 8). **The intercept must not be forced through zero**: it contains the blank signal, and setting it to zero introduces a systematic error, especially near the LLOQ (Chapter 16).
3. **Back-calculation**: the concentration of each calibrator is recalculated from the curve and compared with the nominal value (Eq. 10.2). It must be within **±15%** (±20% at the LLOQ), and **at least 75% of the calibrators**, with **at least 6 levels**, must meet this criterion. A non-conforming calibrator is excluded and the regression is recalculated; the percentage is calculated on the total number of calibrators. For *ligand-binding* methods the limits are ±20% (±25% at the LLOQ and ULOQ).
4. **A new curve for every analytical run**, because the sensitivity, background and performance of the instrument change over time (source, column, reagents); the slope, the blank signal and the QCs are monitored (Chapters 12 and 16).

$$\text{deviation}\,\% = \frac{C_{\text{back}} - C_{\text{nom}}}{C_{\text{nom}}} \times 100 \quad \text{with} \quad C_{\text{back}} = \frac{S_i - b_0}{b_1} \tag{10.2}$$

The two conditions in point 3 must be read together: with 8 levels, 6 conforming calibrators are enough (6/8 = 75%); with 6 levels all must conform, because 5/6 = 83% exceeds 75% but leaves only 5 levels.

Why r^2 is not enough. An r^2 close to 1 can hide systematic deviations from linearity, especially at the ends of the range, and two models with the same r^2 can have very different residuals. Over a wide range (1–1000 ng/mL) the total variability is dominated by the high calibrators: an error of a few tenths of a ng/mL at the LLOQ does not change r^2, but at 1 ng/mL it amounts to 20–50%. What is needed, therefore, is the **residual plot**, with residuals randomly distributed around zero (Chapter 8), and the **back-calculated concentrations**, which express the deviation of each calibrator as a percentage.

Example 10.2 — Back-calculation and the 75% rule

Data. 8-level curve (one replicate per level). Deviations of the back-calculated values: LLOQ (1.00 ng/mL → 1.17 ng/mL) +17.0%; 10.0 ng/mL → 8.38 ng/mL; all others within ±5%.

1. LLOQ: +17.0%, within ±20%: conforming.
2. 10.0 ng/mL level (Eq. 10.2): $(8.38 - 10.0)/10.0 \times 100 = -16.2\%$, beyond ±15%: non-conforming.
3. Conforming: 7/8 = 87.5% ≥ 75%, with 7 levels ≥ 6.

Result: curve acceptable; the non-conforming level is excluded and the regression is recalculated. Watch the denominator: a calibrator with nominal 40.0 ng/mL back-calculated at 46.6 ng/mL gives $(46.6 - 40.0)/40.0 \times 100 = +16.5\%$ (non-conforming); dividing by 46.6 would give +14.2%, and it would be wrongly accepted, whatever the r^2.

10.6 Internal standard calibration

In **internal standard calibration** (IS), a known amount of a reference species is added **as early as possible** in the process, at the **same concentration** C_s, to **standards, QCs, samples and the zero sample**, but **not to the blank**. The response is the **ratio** of the analyte signal to the IS signal:

$$R = \frac{S_{\text{analyte}}}{S_{\text{IS}}} = b_0 + b_1 C \quad \Rightarrow \quad C = \frac{R - b_0}{b_1} \tag{10.3}$$

Variations that act **proportionally** on both signals cancel out in the ratio (not in the difference). If, in GC with a microsyringe, 10% less volume is injected, with ES the result is underestimated by 10%, whereas with IS both signals fall by 10% and the ratio does not change. An IS added before extraction also compensates for losses during preparation; added only before injection, it compensates only for instrumental variability.

An IS is used when **injection reproducibility is critical** (microsyringe in GC), when the **signal source is unstable** (flame, ICP plasma), when sample preparation involves **variable losses**, and in LC-MS/MS, where ionization depends on the matrix. The matrix effect, however, is compensated **only if the IS undergoes exactly the same effect** as the analyte; an IS that elutes at a different time may undergo a different effect, and the ratio does not correct additive interferences. The correction is therefore **partial**, and it becomes **optimal** with a **SIL-IS** (stable isotope-labeled internal standard: the analyte labeled with ^{2}H, ^{13}C or ^{15}N), which co-elutes with the analyte, undergoes the same losses and the same suppression or enhancement, but can be distinguished by *m/z* (details and limitations in Chapter 16).

Requirements for the IS: **absent from the sample** (otherwise its total amount would not be known), **non-reactive** with analyte and matrix, a **signal well separated** from all others, and **physicochemical properties similar** to those of the analyte.

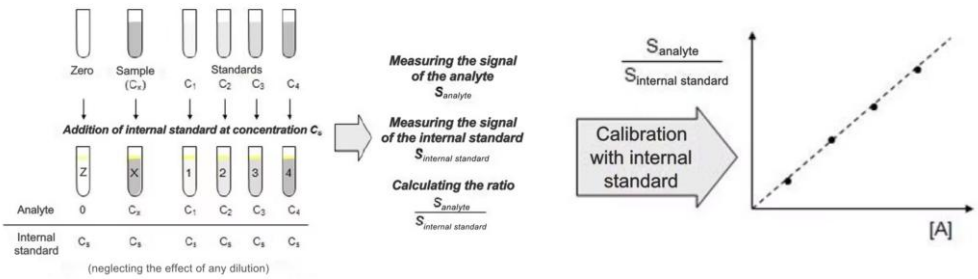


Figure 10.2 — Operating scheme of the internal standard: the IS is added at the same concentration Cs to the zero sample, the sample and the standards (not to the blank, not shown); for each solution the ratio S(analyte)/S(internal standard) is calculated, and the line of this ratio versus concentration is constructed. The diagram shows 4 levels for simplicity.

As Figure 10.2 shows, the blank sample contains neither analyte nor IS and the zero sample contains only the IS: neither counts among the 6 levels of the curve. ICH M10 also requires the IS response in the samples to be monitored, because an anomalous value signals a problem in that sample. A worked calculation with the line $R = 0.0250 \cdot C + 0.0080$ (C in ng/mL): a sample with A_{analyte} = 18 450 and A_{IS} = 92 300 gives $R = 18\,450/92\,300 = 0.1999$ and $C = (0.1999 - 0.0080)/0.0250 = 7.68$ ng/mL (ignoring the intercept would give 8.00 ng/mL; with the inverted ratio $A_{\text{IS}}/A_{\text{analyte}}$, an absurd value).

10.7 The method of standard additions

In the **method of standard additions** (standard addition), calibration is performed **in the sample matrix**: equal aliquots of the sample with unknown concentration C_x are spiked with **increasing, known** amounts of pure analyte, brought to the same volume and measured (Figure 10.3). If the response is linear and blank-corrected, $S = k(C_x + C_{\text{add}})$: regressing the signal on the added concentration gives $b_1 = k$ and $b_0 = kC_x$, and the extrapolated line crosses the x-axis at $-C_x$:

$$C_x = \frac{b_0}{b_1} \tag{10.4}$$

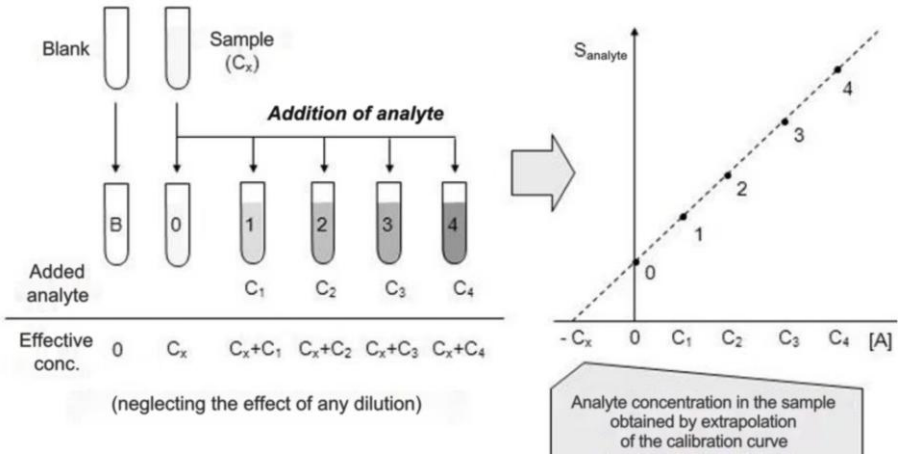


Figure 10.3 — Method of standard additions: aliquots of the sample spiked with increasing amounts of analyte; the extrapolated signal line crosses the x-axis at −Cx (diagram neglecting dilution).

Because the slope k is measured in the sample matrix, the **multiplicative matrix effect** cancels out in the ratio b_0/b_1. C_x is the concentration in the measured solution: for the original sample it is multiplied by the dilution factor.

Requirements: **linearity** over the whole range of the additions (and in the extrapolated segment) and **blank correction before extrapolation**, because the intercept must contain only the analyte signal. **Limitations**: the method **does not correct constant additive interferences**, which increase b_0 and lead to overestimation of C_x; it requires **one curve per sample**, so it is slow and costly; extrapolation increases the uncertainty. It is suited to **complex, non-reproducible matrices** when no suitable IS is available.

Example 10.3 — Standard additions with blank correction

Data. Regressing the **raw** signals on the added concentration gives b_0 = 0.210 and b_1 = 0.0700 L/mg; the reagent blank gives 0.035. The measured solution was obtained by diluting the sample 1:5.

1. The blank is a constant contribution and changes only the intercept: corrected b_0 = 0.210 − 0.035 = 0.175.
2. Eq. 10.4: $C_x = 0.175/0.0700 = 2.50$ mg/L in the measured solution.
3. In the original sample: $2.50 \times 5 = 12.5$ mg/L.

Result: 2.50 mg/L in the measured solution (12.5 mg/L in the sample). Without the correction one would obtain 3.00 mg/L (+20%); inverting the ratio (b_1/b_0), a meaningless value.

10.8 Comparison of calibration methods

Table 10.2 — Comparison of calibration methods

Feature	External standard	Internal standard	Standard additions
Simplicity	High	Medium	Low
Correction of the matrix effect	No (with matrix-matched calibrators, only the average effect)	Partial, if the IS undergoes the same effect (SIL-IS: optimal)	Yes (multiplicative)
Correction of injection variability	No	Yes	No
Number of standards per sample	Low	Medium (same calibrators, plus the IS in every solution)	High (1 curve per sample)
Typical application	Simple matrices	LC-MS/MS, GC-MS	Complex, non-reproducible matrices

No method corrects additive interferences, which require the blank. For example, for a contaminant in wastewater from different industrial plants, with a highly variable matrix and no suitable IS, the appropriate strategy is standard additions on each sample: ES does not correct the matrix effect, and a curve built on a single sample is not valid for the others.

Key points of Chapter 10

- **Absolute methods** (gravimetry, titrations): no curve, but a titrant standardized against a PS. **Relative methods** (almost all instrumental methods): calibration curve.
- The **blank** corrects only **additive** contributions, not the **multiplicative** matrix effect; the matrix blank is the best approximation of the true blank.
- Quantify with peak **area**; height only if the shape is constant.
- **External standard**: simple, one curve for all samples; assumes an identical response in standards and samples; corrects neither the matrix effect nor injection.
- **Curve** (ICH M10): ≥ 6 levels from LLOQ to ULOQ plus blank and zero; intercept not forced; back-calculation within ±15% (±20% at the LLOQ) of nominal; ≥ 75% conforming with ≥ 6 levels; a new curve for every run. r^2 **is not enough**.
- **Internal standard**: same amount in standards, QCs, samples and zero (not in the blank), added as early as possible; ratio $S_{\text{analyte}}/S_{\text{IS}}$; partial correction of the matrix effect, optimal with a SIL-IS.
- **Standard additions**: $C_x = b_0/b_1$; correct the multiplicative effect, not the additive one (blank subtracted first); one curve per sample.

Test yourself

- **Standard preparation; absolute and relative methods** (absolute and relative methods)
- **Analytical blank and signal selection**
- **External standard calibration**
- **Calibration curve construction and evaluation**
- **Internal standard calibration**
- **Standard addition method**
- **Calibration methods and instrumental methods** (comparison of the methods)

11. Validation: purpose, selectivity, sensitivity, LOD/LOQ and linear range

The results of an analytical method are used to make decisions: the dose of a drug, the compliance of a food, the release of a batch. Before using them, one must demonstrate that the method delivers them reliably: this is **validation**. This chapter presents its purpose and the first parameters (selectivity, sensitivity, LOD/LOQ, working range); trueness, precision and QC are covered in Chapter 12, recovery and matrix effect in Chapter 13, stability and other parameters in Chapter 14, and the regulatory framework in Chapter 15.

11.1 Why validate an analytical method

Validation

Validation is the process that demonstrates, through documented experimental evidence, that a method is suitable for its intended purpose (*fit for purpose*).

The required performance depends on the purpose: a presence/absence screening mainly requires a low limit of detection, while a method for a pharmacokinetic study requires known trueness and precision over a wide concentration range. Validation follows a **protocol** written in advance, with predefined experiments and **acceptance criteria**, and ends with a report stating whether the method is fit for purpose.

It must not be confused with **metrological traceability** (an unbroken chain of calibrations up to a reference, Chapter 9), with the **system suitability test** (a check of instrument performance before each run) or with the **uncertainty evaluation** of each result (Chapter 6). A new method requires full validation; a modification, including transfer to another laboratory, requires partial validation (Chapter 16).

The reference guideline depends on the sector (Table 11.1; details in Chapter 15). For example, for an ICP-MS method for metals in surface waters, which is not a pharmaceutical, bioanalytical or food context, the appropriate reference is the Eurachem guide. In this chapter and in Chapter 12, the numerical criteria quoted are those of ICH M10 for chromatographic methods.

Table 11.1 — Main reference documents for validation

Document	Scope
ICH M10 (2022)	Bioanalysis (drugs and metabolites in biological matrices); adopted by EMA and FDA, it supersedes EMA 2011 and FDA 2018 within this scope
ICH Q2(R2) and Q14 (2023)	Analytical procedures for pharmaceutical products
Eurachem, *The Fitness for Purpose of Analytical Methods* (3rd ed., 2025)	Chemical laboratories in general
SANTE/11312/2021 v2026	Pesticide residues in food and feed
Regulation (EU) 2021/808	Residues of veterinary drugs in food of animal origin

11.2 Validation parameters

The main parameters are: **selectivity/specificity** (11.3), **calibration curve and range** (Chapter 10 and 11.6), **LOD/LOQ** (11.5), **trueness** and **precision** (Chapter 12), **recovery** and **matrix effect** (Chapter 13), and **stability** (Chapter 14). To these are added **carry-over**, **dilution integrity**, **robustness** and **ruggedness** (Chapter 14). Not all guidelines require them in the same way: ICH M10, for example, does not include recovery and robustness among the elements of full validation (recovery is dealt with during method development).

No single parameter, on its own, demonstrates that a method is valid. A method with an excellent calibration line may be unusable because a metabolite interferes or because the results at the lowest level are too scattered.

A high r^2 is not a validation

"The method is validated because the line has $r^2 = 0.999$" is a wrong conclusion. The coefficient of determination says nothing about bias, precision, interferences or stability, and ICH M10 sets no minimum value for r^2: all the parameters must be documented.

11.3 Selectivity and specificity

Selectivity and specificity

Selectivity is the ability of a method to measure the analyte in the presence of other components of the matrix (endogenous substances, metabolites, co-administered drugs) without significant interference. **Specificity** is absolute selectivity (a response to the analyte only); in practice the two terms are often used as synonyms.

ICH M10 uses them for two distinct tests, with the same criteria: selectivity concerns the components of the blank matrix, specificity concerns substances related to the analyte (metabolites, isomers, degradation products, concomitant drugs). In LC-MS/MS, selectivity comes from the chromatographic separation and the MRM transition; in HPLC-UV, a substance that co-elutes and absorbs at the same wavelength adds to the signal. An interference usually produces an additive error, which weighs most at low levels (Chapter 12).

Selectivity is checked on **blank matrix samples** (matrix processed without analyte or internal standard, IS) from **at least 6 individual sources or lots**, for example plasma from 6 different donors. Individual lots are needed because an interference may be present in a single donor: six replicates of the same lot would not reveal it, and a pool would dilute it. According to ICH M10, interfering responses must be **no more than** (≤) **20%** of the analyte response at the LLOQ, at the retention time of the analyte, and **no more than 5%** of the IS response in the LLOQ sample, at the retention time of the IS.

Selectivity criteria (ICH M10)

$$\frac{A_{\text{blank}}(t_{R,\text{analyte}})}{A_{\text{analyte}}(\text{LLOQ})} \times 100 \leq 20\%$$

$$\frac{A_{\text{blank}}(t_{R,\text{IS}})}{A_{\text{IS}}(\text{LLOQ})} \times 100 \leq 5\%$$

- A_{blank}: area measured in the blank at the retention time t_R of the analyte or of the IS;
- $A_{\text{analyte}}(\text{LLOQ})$, $A_{\text{IS}}(\text{LLOQ})$: areas of the analyte and of the IS in the LLOQ sample.

The limit for the IS is stricter because the result is calculated from the analyte/IS ratio: an interference on the IS alters that ratio across the whole range.

Example 11.1 — Selectivity in plasma lots

At the LLOQ the analyte area is 4 200 and the IS area 150 000. Three lots of blank plasma give the following interfering areas (analyte; IS): lot 1: 650; 2 900. Lot 2: 905; 1 800. Lot 3: 120; 8 100.

1. Limits: 20% of 4 200 = 840; 5% of 150 000 = 7 500.
2. Lot 1: 650/4 200 = 15.5% and 2 900/150 000 = 1.9%: conforming.
3. Lot 2: 905/4 200 = 21.5% > 20%: non-conforming for the analyte.
4. Lot 3: 120/4 200 = 2.9%, but 8 100/150 000 = 5.4% > 5%: non-conforming for the IS.

Result: lots 2 and 3 exceed the limits; selectivity is not demonstrated until the interferences are eliminated. Mistakenly comparing 905 with the IS area would give 0.6%, and lot 2 would appear to conform.

11.4 Calibration sensitivity and analytical sensitivity

Sensitivity

Calibration sensitivity is the slope of the calibration line, that is, the change in signal per unit of concentration; for a linear curve it is constant:

$$b_1 = \frac{\Delta S}{\Delta C} \tag{11.1}$$

Analytical sensitivity is the ratio of the slope to the standard deviation s_S of the signal at the concentration considered:

$$\gamma = \frac{b_1}{s_S} \tag{11.2}$$

The slope does not take noise into account: if the signal is amplified, both b_1 and the noise increase. Analytical sensitivity corrects this shortcoming, but it **depends on concentration even for linear curves**, because s_S varies with the signal level (heteroscedasticity, Chapter 8) while b_1 remains constant.

The term "sensitivity" is often used loosely to mean the LOD, but they are distinct quantities: the LOD depends on both the slope and the noise of the blank (11.5). In the ICH M10 glossary, *sensitivity* denotes the LLOQ: when reading a document, it is worth checking in which sense the term is used.

Example 11.2 — A steeper slope does not mean a lower LOD

Method A: b_1 = 2 500 area units per ng/mL, s_B = 450. Method B: b_1 = 1 200, s_B = 120 (formula 11.3).

1. LOD_A = 3 · 450/2 500 = 0.54 ng/mL.
2. LOD_B = 3 · 120/1 200 = 0.30 ng/mL.

Result: B has the smaller slope but the better LOD (0.30 ng/mL), thanks to its lower noise. Forgetting the factor 3 would give 0.10 ng/mL for B; using 10 would give the LOQ (1.0 ng/mL).

11.5 Limit of detection and limit of quantification

The **LOD** (limit of detection) is the lowest concentration that gives a signal distinguishable from the blank with a defined level of confidence: its use is **qualitative** (presence/absence). The **LOQ** (limit of quantification) is the lowest concentration that can be determined with acceptable precision and trueness: **quantitative** use starts from there.

A blank is measured several times (usually at least ten replicates), and its mean $\bar{y}_B$ and standard deviation s_B are calculated. The **limit signals** are $y_{\mathrm{LOD}} = \bar{y}_B + 3\,s_B$ and $y_{\mathrm{LOQ}} = \bar{y}_B + 10\,s_B$, and they divide the signal axis into three zones: not detectable, detectable but not quantifiable, quantifiable (Figure 11.1A). The factor 3 is the multiple of s_B by which the signal must exceed the blank in order to state, with a defined level of confidence, that the analyte is present (for a normal distribution, a blank exceeds $\bar{y}_B + 3\,s_B$ in only about 0.13% of cases). At the y_{LOQ} level the net signal equals $10\,s_B$, so its relative standard deviation is about 10%, which is compatible with quantification.

The limit signals are converted into concentrations with the calibration line, whose intercept approximately coincides with $\bar{y}_B$: the net signal is divided by the slope (Figure 11.1B).

LOD and LOQ (concentrations)

$$\mathrm{LOD} = \frac{3\,s_B}{b_1} \tag{11.3}$$

$$\mathrm{LOQ} = \frac{10\,s_B}{b_1} \tag{11.4}$$

- s_B: standard deviation of the blank signal; b_1: slope of the line.
- $\bar{y}_B + 3\,s_B$ and $\bar{y}_B + 10\,s_B$ are **signals** (y_{LOD}, y_{LOQ}), not concentrations.

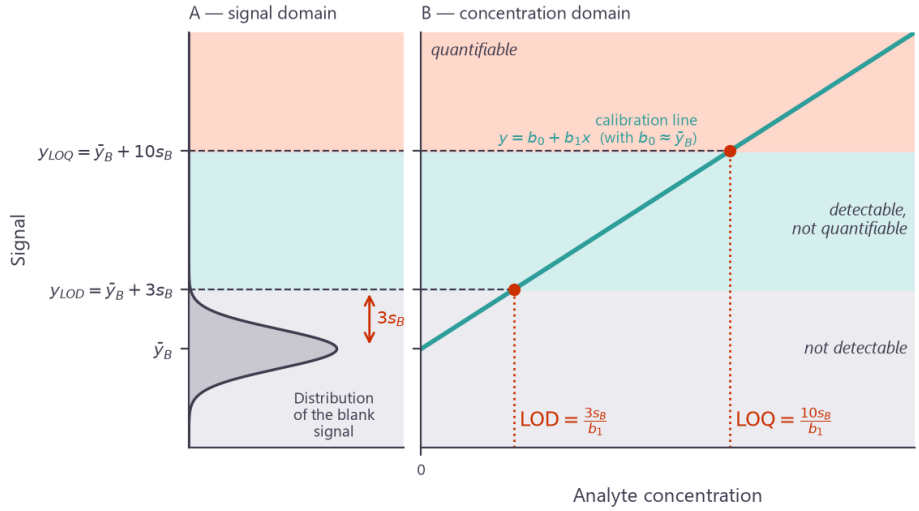


Figure 11.1 — LOD and LOQ. A: signal domain; the distribution of the blank signals defines the limit signals (blank mean plus 3 and plus 10 standard deviations) and three zones: not detectable, detectable but not quantifiable, quantifiable. B: the calibration line converts the limit signals into the LOD and LOQ concentrations (formulas 11.3 and 11.4).

In chromatography the **signal-to-noise ratio** is often used: the LOD corresponds to S/N ≈ 3, the LOQ to **S/N ≥ 10**. The formulas give estimates: the LOQ must then be verified by analyzing samples at that concentration.

Example 11.3 — LOD and LOQ from blank replicates

Ten replicates of a blank (HPLC-UV) give areas (mAU·s): 152, 138, 165, 147, 171, 140, 158, 149, 133, 160. Slope b_1 = 1 250 mAU·s per µg/mL.

1. $\bar{y}_B$ = 1 513/10 = 151.3 mAU·s; s_B = 12.3 mAU·s ($n - 1$ = 9).
2. Limit signals: y_{LOD} = 151.3 + 3 · 12.3 = 188.2 mAU·s; y_{LOQ} = 151.3 + 10 · 12.3 = 274.3 mAU·s.
3. LOD = 3 · 12.3/1 250 = 0.0295 µg/mL ≈ 30 ng/mL.
4. LOQ = 10 · 12.3/1 250 = 0.098 µg/mL ≈ 98 ng/mL.
5. A sample with a signal of 230 mAU·s falls between y_{LOD} and y_{LOQ}: detected, but not quantifiable.

Result: LOD ≈ 30 ng/mL and LOQ ≈ 98 ng/mL. The values 188.2 and 274.3 are signals: confusing them with the LOD and LOQ is the most common mistake.

A result below the LOD is reported as "not detected"; between LOD and LOQ as "detected, < LOQ"; above the LOQ as a numerical value. The LOQ must not be confused with the highest standard of the curve (ULOQ).

In bioanalysis the operational reference is the **LLOQ**, the lowest calibrator, at which ICH M10 requires trueness within **±20%** and **CV ≤ 20%**. **The LLOQ can never lie below the LOQ**. If a study requires concentrations lower than the LOQ, one does not lower the LLOQ or widen the criteria: the method must be improved (more signal, less noise, cleaner extraction, more concentrated sample).

11.6 Linear dynamic range and calibration range

The **linear dynamic range** is the interval in which the signal is linearly related to the concentration: it extends from the **LOQ** to the **LOL** (limit of linearity). With a non-zero intercept the relationship is linear but not proportional. Beyond the LOL the detector saturates and the response levels off. Its width depends on the analyte and the detector; linearity allows least-squares regression with few standards (Chapters 7 and 8).

The **calibration range** (ICH M10) extends from the **LLOQ** (lowest standard, never below the LOQ) to the **ULOQ** (highest standard). It is built with a **blank sample** (matrix with neither analyte nor IS), a **zero sample** (matrix with the IS only) and **at least 6 levels** of calibrators (Chapter 10). For a linear model it must lie within the linear range ($\mathrm{LOQ} \leq \mathrm{LLOQ}$ and $\mathrm{ULOQ} \leq \mathrm{LOL}$): if the ULOQ exceeded the LOL, the highest calibrators would fall

on the curved segment. ICH M10 also allows justified non-linear models. The **QCs** cover the whole range: LLOQ, low (≤ 3 × LLOQ), mid (30–50%), high (≥ 75% of the ULOQ) (Figure 11.2; criteria in Chapter 12).

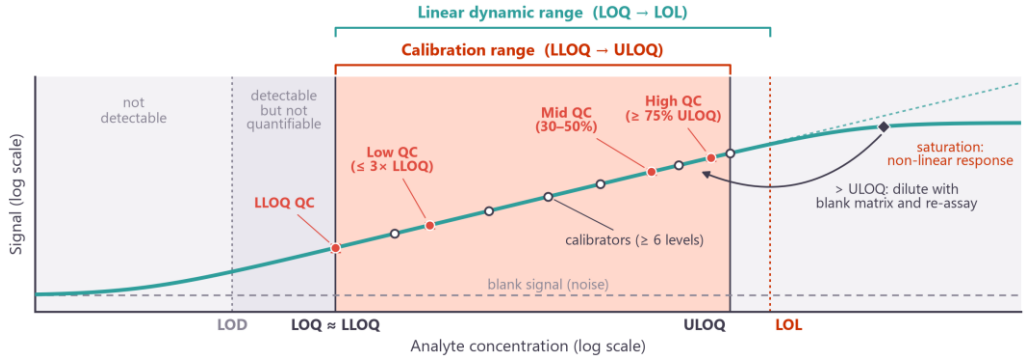


Figure 11.2 — Linear dynamic range (from LOQ to LOL) and calibration range (from LLOQ to ULOQ, with at least 6 calibrators); position of the QCs; beyond the LOL the response levels off because of saturation. Logarithmic axes.

Trueness and precision are demonstrated only between the LLOQ and the ULOQ, so **one never extrapolates** beyond the curve, even if the detector would still be linear:

- **above the ULOQ** the sample is diluted with **blank matrix** (not with solvent, because it must behave like the calibrators) and reanalyzed; the dilution must be validated (dilution integrity, Chapter 14). For example, with ULOQ = 200 ng/mL, a sample diluted 1 + 4 that gives 51.8 ng/mL contains 51.8 · 5 = 259 ng/mL;
- **below the LLOQ** the result is reported as **"< LLOQ"** (for example "< 0.50 ng/mL"), not as an extrapolated value or as zero, because the analyte may be present even if it cannot be quantified; nor can a lower calibrator be added after the fact.

Key points of Chapter 11

- Validation demonstrates with documented evidence that a method is fit for purpose; the guideline depends on the sector (ICH M10, ICH Q2(R2)/Q14, Eurachem, SANTE, Regulation (EU) 2021/808). A high r^2 is not a validation.
- Selectivity: blanks from at least 6 individual lots; interference ≤ 20% of the analyte response at the LLOQ and ≤ 5% of the IS response.
- Calibration sensitivity = slope b_1; analytical sensitivity = b_1/s_S, which varies with concentration; neither of them is the LOD.
- $\mathrm{LOD} = 3s_B/b_1$ (presence/absence, S/N ≈ 3) and $\mathrm{LOQ} = 10s_B/b_1$ (quantification, S/N ≥ 10); $\bar{y}_B + 3s_B$ and $\bar{y}_B + 10s_B$ are signals.
- LLOQ: ±20%, CV ≤ 20%, never below the LOQ.
- Linear range: LOQ → LOL; calibration range: LLOQ → ULOQ, at least 6 levels plus blank and zero, within the linear range for a linear model.
- Above the ULOQ, dilute with blank matrix (validated dilution); below the LLOQ, report "< LLOQ"; never extrapolate.

Test yourself

Quiz topics: Validation: purpose and parameters; Selectivity and specificity; Sensitivity, LOD and LOQ; Linear dynamic range and calibration range.

12. Trueness, precision and quality control

Trueness and precision tell us how far a result can be trusted: how close the mean is to the true value, and how scattered repeated results are. This chapter shows how they are measured, which criteria ICH M10 (2022) sets, and how quality control (QC) samples verify them across the whole calibration range. The general definitions are in Chapter 3, standard deviation and CV% in Chapter 4.

12.1 Trueness and bias

Trueness and bias

Trueness is the closeness of agreement between the **mean** of a large number of results and a reference value μ (ISO 5725, VIM). It is expressed by the **bias**, the difference between the mean and the reference value.

The reference value is the nominal concentration of a QC (blank matrix spiked with a known amount of analyte), the certified value of a CRM (Chapter 9) or the result of a reference method.

Absolute and relative bias

$$\text{bias} = \bar{x} - \mu \tag{12.1}$$

$$\text{bias}\ (\%) = \frac{\bar{x} - \mu}{\mu} \times 100 \tag{12.2}$$

- $\bar{x}$: mean of the replicates; μ: nominal or certified value.
- The sign matters: negative = underestimation, positive = overestimation.

Bias is calculated on the **mean**: the difference $x_i - \mu$ of a single result is its measurement error, which also includes the random error. The ICH M10 criterion likewise applies to the mean of the replicates: bias within **±15%** of nominal at all QC levels, **±20% at the LLOQ**. ICH M10 and ICH Q2(R2) call precisely this trueness *accuracy*, also expressed as a percentage of nominal ($\bar{x}/\mu \times 100$: 92.4% of nominal is equivalent to a bias of −7.6%).

Example 12.1 — Bias of a high QC

A high QC (nominal 400 ng/mL) gives in 5 replicates: 372, 365, 380, 358, 370 ng/mL.

1. $\bar{x}$ = 1 845/5 = 369.0 ng/mL.
2. Bias = 369.0 − 400 = −31.0 ng/mL.
3. Bias% = −31.0/400 × 100 = −7.75% ≈ −7.8%, within ±15%.

Result: bias = −7.8% (underestimation), conforming. Dividing by the measured value (369.0) would give −8.4%: the denominator is always the reference value. Other typical errors are losing the sign, confusing the absolute bias (in ng/mL) with the relative bias (in %), and calculating the bias on a single result instead of on the mean.

12.2 Accuracy = trueness + precision

According to ISO 5725 and the VIM, **accuracy** is the general term describing how close a single result is to the true value: since the error of a result is the sum of bias and random error, accuracy includes both trueness (absence of bias) and precision (absence of significant random error). It is not a synonym for either of them. For example, five measurements of a CRM certified at 25.0 mg/kg give 22.9; 23.1; 23.0; 22.8; 23.2 mg/kg: mean 23.0 mg/kg, s = 0.16 mg/kg, bias −8.0%. The results are **precise but not true**: such a small scatter cannot explain a deviation of 2 mg/kg, which is systematic. The method is therefore not accurate.

12.3 Precision: repeatability, intermediate precision, reproducibility

Precision is the closeness of agreement between independent results under specified conditions, and it is expressed as the CV% of replicate measurements (not as bias%, r² or S/N):

$$\mathrm{RSD\%} = \mathrm{CV\%} = \frac{s}{\bar{x}} \times 100 \tag{12.3}$$

with s calculated with $n - 1$ in the denominator. The three precision conditions (Chapter 3) are:

- **repeatability** (intra-day): same procedure, same operator, same instrument, same day; in ICH M10 this is *within-run* precision (within the same run);
- **intermediate precision** (inter-day): same laboratory, but different days, operators and instruments; it describes the variability in routine work and is the *between-run* precision of ICH M10. A QC analyzed on 5 days by two analysts on two instruments gives the intermediate precision;
- **reproducibility** (inter-laboratory): different laboratories using the same procedure; it is the basis for standardization and regulatory recognition of the method.

Each condition adds sources of variability, so one expects $\mathrm{CV}_r \leq \mathrm{CV}_I \leq \mathrm{CV}_R$ (repeatability, intermediate, reproducibility).

ICH M10 reasons in terms of runs, not strictly in terms of days: within-run precision is assessed with at least **5 replicates** per QC level in each run, between-run precision with each level in at least **3 runs** on at least **2 days**. The criterion applies to **both**: **CV ≤ 15%**, **≤ 20% at the LLOQ**. A mid QC with a within-run CV of 5% and a between-run CV of 17% does not conform: the two CVs are not averaged, and the 20% limit applies only at the LLOQ.

Example 12.2 — RSD% of six replicates

Six replicates of a mid QC (nominal 250 ng/mL): 241, 258, 236, 262, 249, 244 ng/mL.

1. $\bar{x}$ = 1 490/6 = 248.3 ng/mL.
2. $\sum(x_i - \bar{x})^2$ = 505.3; $s = \sqrt{505.3/5}$ = 10.05 ng/mL.
3. RSD% = 10.05/248.3 × 100 = 4.0%.

Result: RSD% = 4.0%, within 15%. Typical errors: dividing by n (3.7%), using $s/\sqrt{n}$ (1.7%), forgetting the factor 100 (0.040%).

12.4 The Horwitz curve

From the analysis of numerous interlaboratory studies, Horwitz observed that **interlaboratory reproducibility** deteriorates predictably as concentration decreases, almost independently of analyte, matrix and technique (Figure 12.1).

Horwitz equation

$$\mathrm{CV}_R\ (\%) = 2^{(1-0.5\log C)} \tag{12.4}$$

- CV_R: expected interlaboratory reproducibility CV;
- C: **dimensionless mass fraction** (1 = 100%; 10^{-6} = 1 mg/kg; 10^{-9} = 1 µg/kg);
- $\log$: **base-10** logarithm.

Each time the concentration decreases by a factor of 100, the exponent increases by 1 and the expected CV doubles: 2% for a pure analyte (C = 1), 4% at 1%, 8% at 100 mg/kg, 16% at 1 mg/kg, about 45% at 1 µg/kg. For example, at 100 mg/kg: $C = 10^{-4}$, $\log C$ = −4, $\mathrm{CV}_R = 2^{(1+2)}$ = 8%. At 0.5 mg/kg: $C = 5 \times 10^{-7}$, $\log C$ = −6.301, $\mathrm{CV}_R = 2^{4.151}$ = 17.8%. The curve is mainly used as a benchmark: in interlaboratory studies the **HorRat** ratio between the observed and the predicted reproducibility CV is considered acceptable between about 0.5 and 2 (for example, an observed 12% at 0.5 mg/kg gives HorRat = 12/17.8 ≈ 0.7).

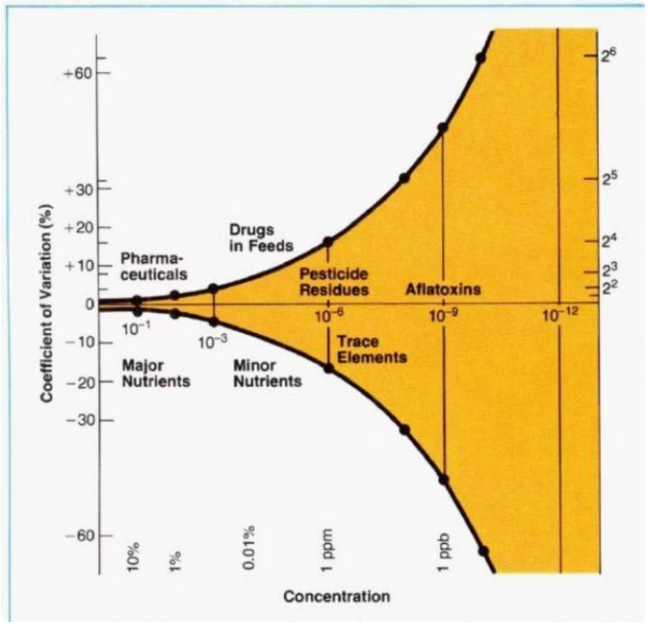


Figure 12.1 — The Horwitz curve: expected interlaboratory CV (with ± sign) as a function of concentration, expressed as a mass fraction on a logarithmic scale; the CV doubles every two orders of magnitude (right-hand scale: 2^2, 2^3, …, 2^6 %).

How not to use the Horwitz curve

- C must be expressed as a mass fraction: entering 100 (for 100 mg/kg) would give 1% instead of 8%; forgetting the factor 0.5 would give 32%.
- The curve describes **interlaboratory reproducibility**, not the repeatability or intermediate precision of a single laboratory, which are normally smaller. The 2% value applies only for $C \approx 1$.
- Its trend is **consistent** with the wider ICH M10 criteria at the LLOQ (±20%, CV ≤ 20%), because precision deteriorates near the lower limit; but the ICH M10 criteria are intra-laboratory and are not derived from the equation.
- Below about 1.2×10^{-7} (120 µg/kg) the equation overestimates the observed CV.

12.5 Quality control (QC) samples

QCs are blank matrix samples spiked with a known amount of analyte, prepared from a **different stock solution** from that of the calibrators (so that an error in preparing the standards does not remain hidden) and treated like real samples. **They are not used to build the curve**: they put it to the test. They are used in validation and, in routine work, to accept or reject each run (Chapter 16).

Table 12.1 — QC levels and ICH M10 criteria (chromatographic methods)

QC	Position	Trueness	Precision
LLOQ QC	at the lower limit of quantification	±20%	CV ≤ 20%
Low QC	$\leq 3 \times$ LLOQ	±15%	CV ≤ 15%
Mid QC	30–50% of the calibration range	±15%	CV ≤ 15%
High QC	$\geq 75\%$ of the ULOQ	±15%	CV ≤ 15%

QCs refer to the calibration range (LLOQ–ULOQ), not to the LOD or the LOL; for a linear model, the high QC must also fall on the linear segment (Figure 11.2). The low QC checks the region where noise and matrix effect begin to matter, the high QC the region where saturation and non-linearity appear, and the mid QC the center; since the LLOQ is usually much smaller than the ULOQ, 30–50% of the range corresponds in practice to about 30–50% of the ULOQ.

Only the LLOQ QC has wider criteria: near the lower limit the signal barely exceeds the noise and the relative error increases. The LLOQ, however, is not the LOD: reliable quantification is required there too. The high QC, although close to saturation, does not have wider criteria. The criteria apply to trueness and to both within-run and between-run precision. In routine work the low, mid and high QCs are analyzed at least in duplicate, and the run is accepted if at least two-thirds of all QCs, and at least half of those at each level, are within ±15% of nominal (Chapter 16).

Example 12.3 — Do the QCs meet ICH M10?

LLOQ = 1.00 ng/mL, ULOQ = 500 ng/mL. Means and CVs of 5 replicates: LLOQ QC 1.00 → 1.17 ng/mL (CV 18.0%); low QC 2.50 → 2.11 ng/mL (CV 8.5%); mid QC 200 → 212 ng/mL (CV 15.0%); high QC 400 → 372 ng/mL (CV 6.2%).

1. Positions: 2.50 ≤ 3 × 1.00; 200/500 = 40%; 400/500 = 80% ≥ 75%: correct.
2. LLOQ: (1.17 − 1.00)/1.00 × 100 = +17.0% (≤ 20%), CV 18.0% (≤ 20%): conforming.
3. Low: (2.11 − 2.50)/2.50 × 100 = −15.6%: beyond ±15%, non-conforming.
4. Mid: +6.0%; CV 15.0%, which does not exceed 15%: conforming.
5. High: (372 − 400)/400 × 100 = −7.0%, CV 6.2%: conforming.

Result: the low QC does not conform (−15.6%); the cause of the underestimation must be investigated. Applying ±20% to it "because it is close to the LLOQ" is a mistake (Figure 12.2).

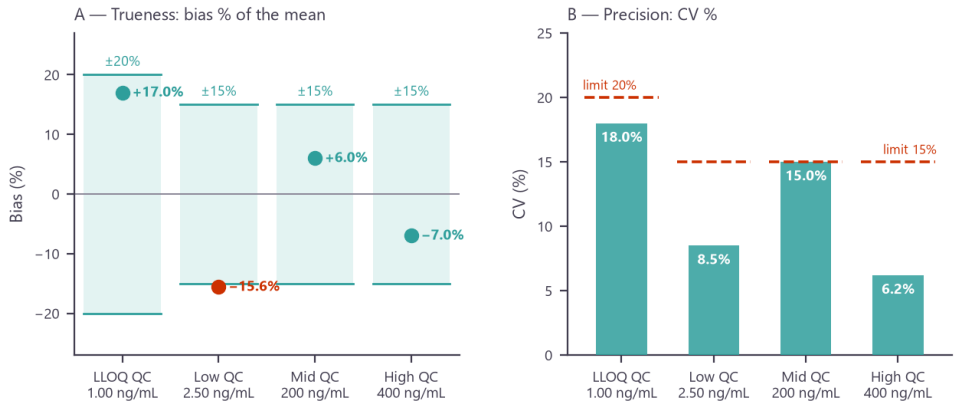


Figure 12.2 — Evaluation of the QCs of Example 12.3. A: bias % of the mean against the limits (±20% at the LLOQ, ±15% at the other levels); the low QC is outside. B: CV % against the limits (20% at the LLOQ, 15% elsewhere).

12.6 Performance at low and high levels

Near the LLOQ the signal is small, and the following factors weigh more: high instrumental noise relative to the signal, baseline drift, a more pronounced matrix effect, co-eluting or co-extracted interferences, instability or adsorption of the analyte. Bias therefore tends to increase approaching the LLOQ, and precision deteriorates. A constant error weighs much more at the low end: an interference equivalent to +0.08 ng/mL gives +16% at an LLOQ of 0.50 ng/mL (0.08/0.50 × 100) but +0.05% at a high QC of 160 ng/mL, whereas a proportional error would give the same bias% at all levels.

Near the ULOQ, detector saturation and deviations from linearity (the response increases less than proportionally: negative bias) and imprecise dilutions come into play; the bias can therefore be either positive or negative. A high QC with a bias of −18%, while the others are within ±5% and the highest calibrators increase less than expected, indicates saturation near the ULOQ.

Carry-over originates from concentrated samples but mainly distorts the low samples injected afterwards: a sample near the LLOQ overestimated by 30% after a very concentrated sample, and correct when reinjected after a blank, is its typical sign. It is evaluated by injecting a blank after the ULOQ (Chapter 14). Some sources of imprecision, such as injection variability or mobile-phase instability, instead act at all levels.

12.7 Consistency across the whole range

A valid method must be:

- **stable**: trueness, precision and recovery do not change significantly from one concentration level to another (this is not about the chemical stability of the analyte, Chapter 14);
- **reliable at the extremes**: acceptable performance even at the LLOQ and ULOQ;
- **constant over time**: results in agreement across days and operators (intermediate precision);
- **correct across the whole curve**: a "perfect" curve with QCs that fail at low levels indicates a **non-valid** method.

For example, with $r^2 = 0.9995$ and calibrators within ±10%, a low QC with a constant bias of about −22% makes the method non-valid: the limit at the low QC is ±15%, and r^2, dominated by the high points, does not reveal relative errors at low levels. This is why the QCs must cover the whole range and not just the center of the curve.

Key points of Chapter 12

- Trueness = bias of the **mean**: bias% = $(\bar{x} - \mu)/\mu \times 100$, with its sign; ICH M10: ±15% (±20% at the LLOQ). In ICH M10, *accuracy* denotes this trueness.
- Accuracy = trueness + precision (ISO 5725, VIM).
- Precision = CV% (with $n - 1$); repeatability (same conditions), intermediate precision (same laboratory, different days/operators/instruments), reproducibility (different laboratories); $\mathrm{CV}_r \leq \mathrm{CV}_I \leq \mathrm{CV}_R$.
- ICH M10: CV ≤ 15% (≤ 20% at the LLOQ) for both within-run and between-run precision.
- Horwitz: $\mathrm{CV}_R = 2^{(1-0.5\log C)}$, with C a mass fraction and log base 10; it describes interlaboratory reproducibility.
- QCs: LLOQ (±20%, CV ≤ 20%), low ≤ 3 × LLOQ, mid 30–50%, high ≥ 75% of the ULOQ (±15%, CV ≤ 15%); separate stock solution; not included in the curve.
- At the low end, noise, matrix effect, interferences and carry-over weigh most; at the high end, saturation, non-linearity and dilutions. The method must be consistent across the whole range.

Test yourself

Quiz topics: Trueness, precision and accuracy; QC levels and ICH M10 criteria; Recovery, trueness and precision across the range (questions on trueness, precision and consistency at the different levels).

13. Recovery, matrix effect and process efficiency

The analyte signal in a real sample almost never matches the signal of the same amount in pure solvent: part of the analyte is lost during sample preparation, and matrix components alter its response in the instrument. Separating the two contributions shows where to intervene and demonstrates that the result remains correct in every sample.

13.1 Definitions and the Matuszewski scheme

Recovery, matrix effect and process efficiency

- **Recovery (RE)**: fraction of the analyte that survives the extraction and sample treatment steps; it describes extraction losses.
- **Matrix effect**: alteration of the analyte response caused by interfering, often unidentified, components of the matrix (ICH M10). It is quantified with the **matrix factor (MF)**.
- **Process efficiency (PE)**: overall effect of extraction losses and matrix effect.

To separate them, the Matuszewski scheme (2003) is used: three sets of samples at the **same final nominal concentration** in the vial (Figure 13.1).

- **Set 1 — standard solution**: analyte in neat solvent, without matrix and without extraction. It provides the reference signal.
- **Set 2 — post-extraction spike**: processed blank matrix, with the analyte added to the final extract; it contains the matrix, but the analyte has not undergone extraction.
- **Set 3 — pre-extraction spike**: blank matrix spiked and then processed like a real sample; it undergoes both extraction and matrix effect.

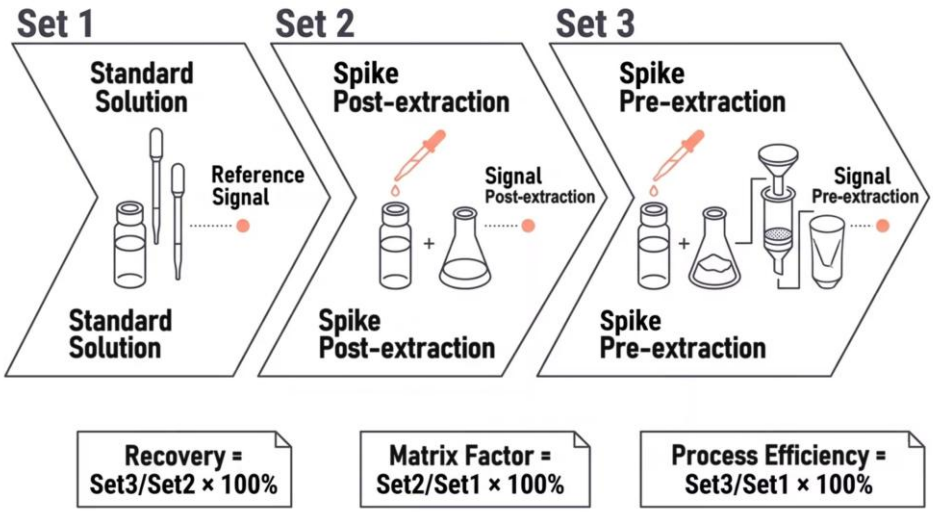


Figure 13.1 — The Matuszewski scheme. Set 1: standard solution in neat solvent (reference signal). Set 2: blank matrix extract spiked after extraction. Set 3: blank matrix spiked before extraction.

13.2 The formulas

Recovery, matrix factor and process efficiency

$$\mathrm{RE}\% = \frac{A_3}{A_2} \times 100 \tag{13.1}$$

$$\mathrm{MF}\% = \frac{A_2}{A_1} \times 100 \tag{13.2}$$

$$\mathrm{PE}\% = \frac{A_3}{A_1} \times 100 \tag{13.3}$$

- A_1, A_2, A_3: mean analyte peak areas in Set 1, Set 2 and Set 3.
- MF < 100%: ion suppression; MF > 100%: ion enhancement.

- **RE = Set 3/Set 2.** Both contain the same matrix extract and differ only in when the analyte is added: the matrix effect cancels out and the extraction loss remains. Recovery is therefore not "the analyte found after the whole process" (that is PE).
- **MF = Set 2/Set 1.** Neither has undergone extraction; they differ only in the presence of the matrix.
- **PE = Set 3/Set 1.** It includes both effects.

The three quantities are related: multiplying and dividing by A_2,

$$\frac{A_3}{A_1} = \frac{A_3}{A_2} \times \frac{A_2}{A_1} \Rightarrow \mathrm{PE} = \mathrm{RE} \times \mathrm{MF} \tag{13.4}$$

with the quantities expressed as fractions: the two effects act in sequence, so they multiply.

Example 13.1 — RE, MF and PE from peak areas

Mean areas at the low QC: Set 1 = 410 000; Set 2 = 307 500; Set 3 = 264 450.

1. Recovery: $\mathrm{RE\%} = 264\,450/307\,500 \times 100 = 86.0\%$.
2. Matrix factor: $\mathrm{MF\%} = 307\,500/410\,000 \times 100 = 75.0\%$ (25% suppression).
3. Process efficiency: $\mathrm{PE\%} = 264\,450/410\,000 \times 100 = 64.5\%$.
4. Check: $0.860 \times 0.750 = 0.645$.

Result: RE = 86.0%, MF = 75.0%, PE = 64.5%. The signal loss depends more on the matrix than on the extraction: it pays to improve the clean-up of the extract or the chromatography. Given PE and MF, RE = PE/MF.

Common mistakes

- Inverting the ratios: Set 1/Set 2 turns a suppression (MF = 80%) into an apparent enhancement (125%).
- Confusing recovery (Set 3/Set 2) with process efficiency (Set 3/Set 1).
- Adding or dividing RE and MF instead of multiplying them.

13.3 The matrix effect in LC-MS/MS

In the electrospray ionization (ESI) source, analyte ions are formed from charged droplets whose charge and surface are limited: matrix components that **co-elute** with the analyte **compete with it for ionization**, and poorly volatile ones (salts, phospholipids) hinder droplet evaporation. The most frequent case is suppression (MF < 100%); more rarely the matrix promotes ionization (MF > 100%). Atmospheric pressure chemical ionization (APCI), which takes place in the gas phase, is generally less susceptible, but not immune.

The matrix effect is therefore a **change in signal not due to the analyte concentration**: the molecules in the vial are the same, but the fraction that becomes ions changes. It is usually **multiplicative** (it changes the slope) and must be distinguished from an **additive** interference, which gives a signal at the m/z and retention time of the analyte even in the blank (selectivity, Chapter 11). Typical culprits are plasma phospholipids, salts, excipients, and hemolyzed or lipemic samples; since the effect changes from one individual to another, it is evaluated on several matrix lots (Section 13.5).

The consequences depend on the calibration (Chapter 10). With **calibrators in solvent and no internal standard**, the samples undergo the effect and the calibrators do not: with MF < 100% the concentration is **underestimated**, with MF > 100% **overestimated**, by roughly the same factor if recovery is complete (MF = 130% gives about +30%). With **matrix-matched calibrators** the average effect is compensated; with an **internal standard (IS)** what matters is the effect left on the signal ratio (Section 13.4). Even when compensated, strong suppression lowers the signal-to-noise ratio and worsens the LLOQ.

13.4 The IS-normalized matrix factor

IS-normalized matrix factor

$$\mathrm{MF_{norm}} = \frac{\mathrm{MF_{analyte}}}{\mathrm{MF_{IS}}} \tag{13.5}$$

- $\mathrm{MF_{analyte}}$, $\mathrm{MF_{IS}}$: area in the extract spiked after extraction / area in solvent, for the analyte and for the IS.
- $\mathrm{MF_{norm}} \approx 1$: the IS undergoes the same effect as the analyte and compensates for it in the ratio.

With an IS, the concentration is obtained from the signal ratio: what matters is the matrix effect left on that ratio. If the IS reaches the source together with the same co-eluting components, it undergoes the same suppression and the normalized MF is close to 1. The ideal IS is the **stable isotope-labeled internal standard (SIL-IS)**, which co-elutes with the analyte and differs only in m/z (ICH M10 recommends it whenever possible); a deuterated analog may elute slightly earlier and compensate imperfectly, and a structural analog elutes elsewhere and may undergo a different effect. Normalization does not eliminate the matrix and does not concern extraction losses.

Example 13.2 — Normalized MF with a SIL-IS and with a structural analog

Areas in solvent and in a blank plasma extract spiked after extraction: analyte 320 000 and 208 000; SIL-IS 150 000 and 98 250; structural analog 210 000 and 197 400.

1. $\mathrm{MF}_{\mathrm{analyte}} = 208\,000/320\,000 = 0.650$ (35% suppression).
2. $\mathrm{MF}_{\mathrm{IS}}$ = 98 250/150 000 = 0.655 for the SIL-IS; 197 400/210 000 = 0.940 for the analog.
3. $\mathrm{MF}_{\mathrm{norm}}$ = 0.650/0.655 = 0.99 with the SIL-IS; 0.650/0.940 = 0.69 with the analog.

Result: with the SIL-IS the suppression, although marked, is compensated (0.99); with the analog an effect of −31% remains on the ratio and would be transferred to the result.

13.5 Evaluation in the guidelines

ICH M10 (2022) evaluates the matrix effect on the final result: low and high QCs prepared in matrix from **at least 6 different sources or lots**, in at least 3 replicates, quantified against the regular calibration curve. For **each lot**, trueness must be within **±15%** of nominal and **CV ≤ 15%**. The mean across lots does not count: if one lot (for example a lipemic one, +24%) is outside, the test fails, because its very purpose is to reveal differences between matrices. An anomalous lot is not excluded after the fact: on the contrary, M10 recommends, case by case, tests on hemolyzed or lipemic matrices and on special populations; for rare matrices fewer lots are sufficient. Not to be confused with selectivity: blanks from 6 lots, signal ≤ 20% of the LLOQ response (Chapter 11).

The **EMA guideline (2011)**, replaced by M10 since 2023, instead calculated the MF of the analyte and of the IS in 6 lots, at a low and a high concentration, and required the **CV of the IS-normalized MF** to be ≤ 15%. The MF remains the most widely used tool during method development.

13.6 Detecting and reducing the matrix effect

To see **where** the suppression zones lie along the chromatogram, **post-column infusion** is used (Figure 13.2): a syringe pump continuously infuses the analyte, through a T-piece, between the column outlet and the source, while a blank matrix extract is injected. The signal of the infused analyte should remain constant; a **dip** at a given time indicates that components suppressing ionization co-elute there, a rise indicates enhancement. The infused analyte does not pass through the column, and carry-over or contamination would increase the signal, not decrease it.

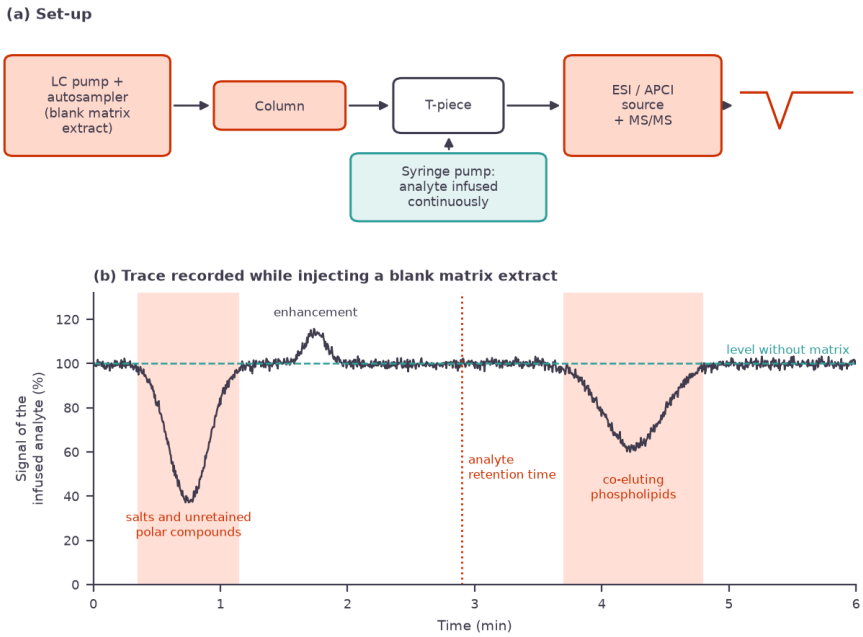


Figure 13.2 — Post-column infusion. (a) The analyte is infused continuously between the column and the source while a blank matrix extract is injected. (b) The dips indicate suppression zones (salts at the start of the run, phospholipids), the rise an enhancement zone; the analyte should elute where the signal stays at the matrix-free level.

To **reduce the effect at the source** (fewer co-eluting components reaching the source), sample clean-up is made more selective (protein precipitation, PPT, leaves phospholipids and salts; liquid–liquid extraction, LLE, and above all solid-phase extraction, SPE, give cleaner extracts), the chromatographic separation is improved, the extract is diluted, or ESI is replaced by APCI. To **compensate for it**: SIL-IS, matrix-matched calibrators, the method of standard additions. The SIL-IS does not reduce the matrix effect: it undergoes it like the analyte and cancels it out in the ratio.

13.7 Recovery at the different concentration levels

The test verifies that extraction works **equally well at all QC levels** (low, mid, high). ICH M10 treats recovery among the additional considerations (§7.3), as an aspect of method development: it is evaluated for the analyte and the IS, typically at three concentrations, and **it need not be 100%, but it must be constant and reproducible**. A constant loss affects calibrators and samples in the same way and is compensated by the calibration; a recovery that changes with concentration produces bias at the extremes. A profile of 62%, 60%, 63% is acceptable; 55%, 78%, 95% is not, even though the mean is higher; a recovery measured only at the mid QC does not demonstrate that it is constant.

The causes of variation differ at the two extremes:

- **low levels** (nearly fixed losses, which weigh more in percentage terms): adsorption on container walls, inefficiency of SPE at low concentrations, volatilization during evaporation, analyte instability;
- **high levels**: saturation of the SPE stationary phase and non-linear effects; co-extraction of interferences affects selectivity and the matrix effect more than recovery.

Key points of Chapter 13

- Matuszewski scheme: Set 1 = standard in solvent; Set 2 = blank matrix extract spiked after extraction; Set 3 = matrix spiked before extraction.
- RE = Set 3/Set 2 (extraction losses; the matrix effect cancels out); MF = Set 2/Set 1 (matrix effect); PE = Set 3/Set 1 = RE × MF.
- MF < 100%: suppression; MF > 100%: enhancement. The mechanism is competition for ionization, in the ESI or APCI source, between the analyte and co-eluting components; without compensation it leads to underestimation or overestimation.
- Normalized MF = analyte MF/IS MF; with a co-eluting SIL-IS it is close to 1.
- ICH M10: low and high QCs in ≥ 6 lots, trueness ±15% and CV ≤ 15% in each lot. EMA 2011 (superseded): CV of the IS-normalized MF ≤ 15%.
- Post-column infusion maps the suppression zones; sample clean-up and chromatography reduce the effect, SIL-IS and matrix-matched calibrators compensate for it.
- Recovery must be checked at the low, mid and high QCs: it need not be 100%, but it must be constant and reproducible.

Test yourself

Use the quiz questions on these topics:

- **Recovery, matrix effect and process efficiency**
- **Matrix effect**
- **Recovery, trueness and precision across the range** (the questions on recovery at the different levels)

14. Stability, robustness, ruggedness, carry-over and dilution integrity

A method that is true and precise in validation can fail in routine work: the analyte degrades while the sample waits, a small pH drift makes an interference co-elute, the method transferred to another laboratory does not work, a concentrated sample contaminates the next one, a sample above the ULOQ has to be diluted. The tests in this chapter verify that the method remains reliable in these real situations.

14.1 Analyte stability

The result must represent the concentration present in the sample **at the time of collection**. Until injection, however, the analyte may degrade (enzymatic hydrolysis, oxidation, light, pH-dependent interconversions) or adsorb onto container walls; in stock solutions, solvent evaporation may instead increase the concentration. Stability tests reproduce, one by one, the conditions to which the study samples will be exposed (Figure 14.1).

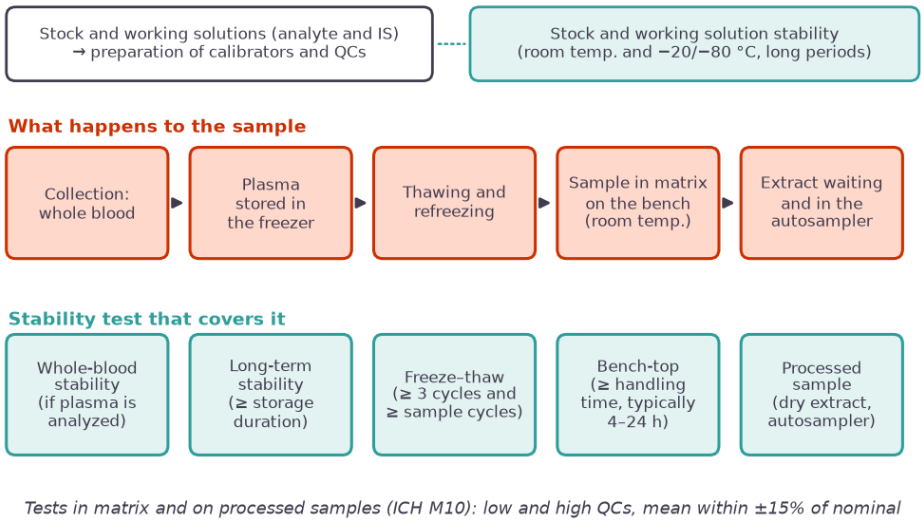


Figure 14.1 — The path of the sample and the stability tests that cover it. Each step in which the analyte could degrade (top) corresponds to a test (bottom); the stability of stock and working solutions is evaluated separately.

General rules of ICH M10 (2022):

- the test conditions (temperatures, times, matrix, anticoagulant, containers) must **reflect those of the study samples**; literature data are not sufficient;
- in matrix, **low and high QCs** are used, each divided into at least 3 aliquots, analyzed against freshly prepared calibrators;
- **criterion**: the mean concentration at each level must be within **±15% of nominal**.

Table 14.1 — Stability tests

Test	What it simulates	Conditions and duration
Stock and working solutions	storage and use of the analyte and IS solutions	room temperature and −20 °C/−80 °C, over extended periods; comparison with fresh solutions
Bench-top (short-term)	sample **in matrix** on the bench, before and during processing	bench temperature, at least the handling time (typically 4–24 h), continuous
Freeze–thaw cycles	repeated thawing	**at least 3 cycles** and no fewer than the samples will undergo; ≥ 12 h frozen between cycles
Long-term	storage in the freezer	same temperature as the samples, duration **≥ storage** of the samples (weeks, months)
Processed sample	extract waiting and in the autosampler	conditions of use, until the end of the analysis
Whole blood	interval between collection and centrifugation	if plasma is analyzed

Some clarifications. **Stock solutions** are used to prepare calibrators and QCs for the whole study: if they degrade, the error is passed on to all results. The **bench-top** test concerns the sample in matrix, not the extract: the stability of the extract in the autosampler is a separate test. Times must be **continuous**: for example, the hours on the bench from the different freeze–thaw tests are not added up. For **freeze–thaw**, a reanalyzed sample accumulates cycles: if it may undergo 5, then 5 cycles are validated. For **long-term** stability, if the samples will remain at −80 °C for 8 months, stability at −80 °C for at least 8 months is needed (for chemical drugs, stability at −20 °C can be extrapolated to −70/−80 °C); no other test replaces it. If the analyte is unstable on the bench, samples are handled in an ice bath, enzyme inhibitors are added, or the handling time is limited to the time demonstrated.

14.2 Robustness

Robustness

A measure of the capacity of an analytical procedure to meet the expected performance requirements during normal use; it is tested by **small, deliberate variations of method parameters** (ICH Q2(R2) and Q14, 2023). Key question: *does the method remain reliable even if something changes slightly?*

Deliberate means that the analyst introduces the variations on purpose, in a planned way (one parameter at a time or with an experimental design), within the range that can occur in routine work. Typical examples:

- column temperature ±2–5 °C; mobile-phase pH ±0.1;
- mobile-phase composition (from 98/2 to 97/3); flow rate ±0.1 mL/min;
- extraction volume or time, shaking or centrifugation time;
- columns from different lots and/or suppliers (ICH Q2 lists them among robustness variations);
- settable detector parameters: UV wavelength, temperatures, gas flows and MS source voltages.

A robust method keeps trueness, precision, retention times and resolution within acceptable limits. The variations must be **small**: pH ±2 units, a HILIC column instead of a C18 or a different detector mean a new method, not a robustness test. The aim is to identify the **critical parameters** and set their tolerances in the procedure and in the system suitability checks (for example, if at pH −0.1 the resolution from an interference drops below 1.5, the pH is critical and must be controlled more tightly). This is why robustness is studied **during development** (ICH Q2(R2) and Q14); ICH M10 does not include it among the elements of full validation.

14.3 Ruggedness

Ruggedness assesses whether the method gives similar results when **context factors** change, which are not parameters written in the procedure. Key question: *does the method give similar results if who uses it, where or how changes?* Examples: different but equivalent instruments, different laboratories (interlaboratory studies), operators with different experience, reagents and consumables of different brands, days far apart, different environmental conditions. If changing the operator or the instrument is enough to make the results collapse, the method is not rugged and cannot be transferred to other laboratories; according to ICH M10, transfer between laboratories requires a partial validation (Chapter 16).

The term *ruggedness* came from the United States Pharmacopeia (USP); the ICH guidelines do not use it, and today the concept coincides with **intermediate precision** (same laboratory, different days, operators or instruments) and **reproducibility** (different laboratories), according to ISO 5725 (Chapters 3 and 12). Some examples are borderline: for ICH, columns from different lots are a robustness variation. In this course: **robustness** = small deliberate variations of method parameters; **ruggedness** = changes of context. A method can be robust but not rugged: it withstands variations in flow rate and mobile phase, but performs clearly worse in a partner laboratory, with other instruments and operators.

14.4 Carry-over

Carry-over

Alteration of the measured concentration due to **residual analyte from a previous sample** left in the instrument (autosampler needle, valve and loop, column, source) (ICH M10).

Carry-over mainly harms **low-concentration samples injected after a concentrated sample**: a residue negligible for a high sample can be a large fraction of the signal of a sample near the LLOQ, which is then overestimated. The test reproduces the worst case: a **blank matrix sample is injected right after the ULOQ calibrator**. ICH M10 requires the signal in the blank not to exceed **20% of the analyte response at the LLOQ** and **5% of the internal standard response**.

$$\text{carry-over}\% = \frac{A_{\text{blank}}}{\bar{A}_{\text{LLOQ}}} \times 100 \leq 20\% \tag{14.1}$$

where A_{blank} is the analyte area in the blank after the ULOQ and $\bar{A}_{\text{LLOQ}}$ the mean area at the LLOQ. The reference is the LLOQ, not the ULOQ. Carry-over is minimized already during development (needle washes with strong solvents, column wash gradient); if it is unavoidable, samples are not randomized and blanks are injected after highly concentrated ones. A signal in a blank injected *before* the ULOQ indicates contamination instead.

Example 14.1 — Evaluating carry-over

Mean analyte area at the LLOQ = 6 250; at the ULOQ = 2 480 000; mean IS response = 152 000. In the blank injected after the ULOQ: 1 040 at the analyte position, 3 800 at the IS position.

1. Analyte (Equation 14.1): $1\,040/6\,250 \times 100 = 16.6\%$, within 20%.
2. IS: $3\,800/152\,000 \times 100 = 2.5\%$, within 5%.
3. Mistake to avoid: $1\,040/2\,480\,000 \times 100 = 0.042\%$ uses the ULOQ as the reference and says nothing about low samples.

Result: carry-over compliant. An area of 1 490 in the blank would give 23.8%: non-compliant.

14.5 Dilution integrity

A sample **above the ULOQ** is not quantified by extrapolating the curve: it must be **diluted with blank matrix** (same species and same type as the QCs, not solvent, so that recovery and matrix effect remain those validated), and the result is multiplied by the dilution factor f:

$$C_{\text{sample}} = C_{\text{measured}} \times f \tag{14.2}$$

Dilution integrity verifies that dilution does not compromise trueness and precision. According to ICH M10, **dilution QCs** are prepared in matrix **above the ULOQ** and diluted with blank matrix by the factor or factors that will be used for the samples (2×, 5×, 10× are only examples), and **at least 5 replicates per factor** are analyzed. Criteria: mean trueness **±15%** and **CV ≤ 15%**. The factors and concentrations applied to the samples must fall within those validated.

Example 14.2 — Verifying dilution integrity

ULOQ = 1 000 ng/mL. A dilution QC at 8 000 ng/mL is diluted 20-fold with blank plasma; the 5 replicates give 384; 401; 392; 377; 395 ng/mL.

1. Values × 20: 7 680; 8 020; 7 840; 7 540; 7 900 ng/mL; mean 7 796 ng/mL.
2. Trueness: $(7\,796 - 8\,000)/8\,000 \times 100 = -2.6\%$.
3. Precision: s = 188 ng/mL; $\text{CV}\% = 188/7\,796 \times 100 = 2.4\%$.
4. Mistake to avoid: comparing the mean without the factor (389.8 ng/mL) with 8 000 ng/mL gives −95.1%.

Result: 20-fold dilution integrity demonstrated. A sample at about 45 000 ng/mL, diluted 20-fold, would remain above the ULOQ (2 250 ng/mL): a larger factor is needed (for example 100-fold, about 450 ng/mL), to be validated before use with a suitable dilution QC.

Key points of Chapter 14

- Stability: stock solutions (room temperature and −20/−80 °C, extended periods), bench-top (sample in matrix on the bench, typically 4–24 h), freeze–thaw (≥ 3 cycles and no fewer than those of the samples), long-term (duration ≥ storage), processed sample, whole blood. In matrix: low and high QCs, mean within ±15%.
- Robustness: small deliberate variations of method parameters (pH ±0.1, flow rate ±0.1 mL/min, temperature ±2–5 °C); it identifies the critical parameters and is studied during development.
- Ruggedness: changes of context (operator, instrument, laboratory); today it coincides with intermediate precision and reproducibility. Robust does not mean rugged.
- Carry-over: it affects low samples after high ones; blank after the ULOQ ≤ 20% of the LLOQ response and ≤ 5% of the IS response.
- Above the ULOQ there is no extrapolation: the sample is diluted with blank matrix by a validated factor (QCs above the ULOQ, ≥ 5 replicates, ±15%, CV ≤ 15%).

Test yourself

Use the quiz questions on these topics:

- **Stability**
- **Robustness and ruggedness**
- **Carry-over and dilution integrity**

15. The regulatory framework

Which tests to perform, and with which criteria, depends on the sector and on how the results will be used: a method for a pharmacokinetic study, one for the batch release of a medicinal product, one for pesticides in fruit and one for antibiotics in milk are governed by different documents. Choosing the right document is the first step of the validation plan; the individual criteria are explained in the previous chapters.

15.1 The bodies and the nature of the documents

The **FDA** (*Food and Drug Administration*, United States) and the **EMA** (*European Medicines Agency*, European Union) are the authorities that assess the data needed to authorize medicinal products. The **ICH** (*International Council for Harmonisation*) brings together regulatory authorities and the pharmaceutical industry to write **harmonized** technical guidelines, which the individual authorities then adopt. For food and feed the reference is the European Commission (DG SANTE); **Eurachem** and **CITAC** publish guides that apply to any chemical laboratory.

A **guideline** is not legally binding: alternative approaches are acceptable if scientifically justified, although in practice the authorities expect it to be followed. A **European Union regulation**, by contrast, is binding and directly applicable in all Member States.

15.2 Bioanalysis: ICH M10 (2022)

For years, bioanalytical laboratories followed two similar but not identical documents: the FDA guidance *Bioanalytical Method Validation* (**2018**) and the EMA guideline (**2011**). Harmonization came with **ICH M10** (*Bioanalytical Method Validation and Study Sample Analysis*, **2022**), adopted by both the FDA and the EMA: it **replaced the EMA guideline in 2023** and **supersedes the 2018 FDA guidance for the methods within its scope**. An LC-MS/MS method for a study to be submitted to both agencies is therefore validated only once, according to M10.

Scope. M10 covers the quantification of chemical and biological drugs and their metabolites in **biological matrices** (blood, plasma, serum, urine, tissues) for nonclinical **toxicokinetic** and **pharmacokinetic** studies and for all phases of clinical trials, including bioequivalence studies. Biomarkers and immunogenicity are excluded, as are, obviously, the quality control of medicinal products, pesticides in food and veterinary drugs in food, which have their own documents.

Method families. M10 provides specific criteria for **chromatographic methods** (LC or GC, typically with mass spectrometry) and for **ligand-binding assays** (LBA, such as immunoassays). For chromatographic methods, trueness is ±15% (±20% at the LLOQ) and CV ≤ 15% (≤ 20% at the LLOQ); for LBAs ±20% and CV ≤ 20% (±25% and CV ≤ 25% at the LLOQ and the ULOQ).

Parameters. Full validation of chromatographic methods comprises selectivity and specificity (Chapter 11), matrix effect (Chapter 13), calibration curve and range (Chapter 10), trueness and precision (Chapter 12), carry-over, dilution integrity and stability (Chapter 14), plus reinjection reproducibility (a reinjected extract, for example after an instrument failure, must give the same result). **Recovery** is not among these elements: M10 deals with it in the additional considerations (§7.3), as an aspect of method development, and asks for it to be evaluated (typically at three levels) and to be consistent, not necessarily 100%. Robustness is not part of full validation either.

What is new compared with the previous guidelines:

- a single harmonized document for the FDA, the EMA and the other ICH authorities;
- with MS detection, **SIL-IS recommended whenever possible**;
- **matrix effect with low and high QCs in at least 6 lots**, with trueness ±15% and CV ≤ 15% in each lot, instead of the CV of the normalized matrix factor of the EMA guideline;
- criteria for **LBAs and chromatographic methods** in the same document.

15.3 ICH Q2(R2) and ICH Q14 (2023): methods for pharmaceutical products

ICH Q2(R2) is the revision of the long-established guideline on the validation of analytical procedures (Q2(R1), 2005). It covers methods for the **release and stability testing of drug substances and drug products** (for example the assay of the active ingredient in tablets, or of impurities), not bioanalysis. It describes the characteristics to be validated: specificity, range and response (linearity, but also non-linear models), accuracy, precision, LOD, LOQ and robustness, the latter addressed mainly during development. Two clarifications: Q2(R2) **does not set universal numerical criteria** (the applicant sets them according to the intended use), and in the ICH documents *accuracy* denotes **trueness** (the ISO term), not accuracy in the sense of trueness + precision (Chapter 3).

ICH Q14 (*Analytical Procedure Development*) is the new guideline on method **development**, based on method understanding and risk management (*Analytical Quality by Design*, AQbD): systematic study of the parameters, often with experimental designs, a control strategy, and change management throughout the life cycle.

Analytical Target Profile (ATP)

Description, established **before** development, of what the procedure must measure and with what performance (analyte, matrix, range, required trueness and precision), independent of the technique. It is introduced by ICH Q14.

The two documents are **complementary**: Q2(R2) says **what to validate**, Q14 **how to develop** the method; Q2(R2) links to the ATP, because validation demonstrates that the requirements of the ATP are met. Neither replaces the other.

15.4 Food, veterinary drugs and laboratories in general

Document **SANTE/11312/2021** (version v2026, applicable from 1 January 2026) is the European guidance on quality control and method validation for **pesticide residues in food and feed**, in particular for **multi-residue** methods (for example QuEChERS with LC-MS/MS and GC-MS/MS). The general criteria are **mean recovery 70–120%** at each spiking level and **RSD ≤ 20%**; the LOQ is the lowest level that meets them. Mean recoveries outside 70–120% are accepted only in exceptional cases, if consistent (RSD ≤ 20%) and with a well-documented cause. Here recovery is the ratio between the concentration found with the whole method and the concentration added, that is, a measure of trueness, different from the extraction recovery of M10 (Chapter 13). Version v2026 superseded v2 of 2024, leaving these criteria unchanged.

$$R\% = \frac{C_{\text{found}}}{C_{\text{added}}} \times 100 \tag{15.1}$$

Example 15.1 — Checking the SANTE criteria

An apple sample free of residues is spiked with a pesticide at 0.010 mg/kg; 5 replicates give 0.0064; 0.0069; 0.0071; 0.0066; 0.0068 mg/kg.

1. Recoveries (Equation 15.1): 64; 69; 71; 66; 68%.
2. Mean 67.6%; s = 2.70; $\text{RSD} = 2.70/67.6 \times 100 = 4.0\%$.
3. Precision is good, but the mean recovery is below 70%.

Result: non-compliant with the general criteria. For another pesticide with recoveries of 105, 78, 122, 84 and 131%, the mean (104.0%) would be acceptable, but the RSD (22.2%) exceeds 20%: in this case too the method is non-compliant.

Regulation (EU) 2021/808 governs the analytical methods for the **official control of veterinary drug residues in food of animal origin** (meat, milk, eggs, honey). It **replaced Commission Decision 2002/657/EC** and **updates its criteria for CCα and CCβ**. The **decision limit (CCα)** is the concentration above which it can be concluded, with an error probability α (false non-compliant), that the sample is non-compliant; the **detection capability (CCβ)** is the smallest concentration that the method can detect, identify and/or quantify with an error probability β (false compliant).

The Eurachem guide *The Fitness for Purpose of Analytical Methods* (3rd ed., 2025) is a general guide for chemical laboratories in **any sector** (environment, food, clinical, industry) and a reference for accreditation to ISO/IEC 17025: it explains when and how to validate a method and how to evaluate its characteristics. For **measurement uncertainty** it refers to the **Eurachem/CITAC QUAM** guide (*Quantifying Uncertainty in Analytical Measurement*), which applies the GUM to chemistry and also makes use of validation data (Chapter 6).

15.5 Comparison of the documents

Table 15.1 — The main documents on validation

Document	Body, year	Scope	Key elements
Bioanalytical Method Validation	FDA, 2018	bioanalysis (USA)	superseded by ICH M10 for the methods within its scope
Guideline on Bioanalytical Method Validation	EMA, 2011	bioanalysis (EU)	replaced by ICH M10 since 2023; CV of the normalized MF ≤ 15%
ICH M10	ICH, 2022 (FDA and EMA)	drugs and metabolites in biological matrices, PK/TK studies	chromatographic methods and LBAs; ±15% (±20% at the LLOQ); matrix effect in ≥ 6 lots; SIL-IS
ICH Q2(R2)	ICH, 2023	methods for drug substances and drug products	what to validate; *accuracy* = trueness
ICH Q14	ICH, 2023	development of pharmaceutical methods	how to develop; ATP; AQbD
SANTE/11312/2021 v2026	European Commission, 2026	pesticide residues in food and feed	recovery 70–120%, RSD ≤ 20%
Regulation (EU) 2021/808	European Commission, 2021	official control of veterinary drug residues	replaces Decision 2002/657/EC; CCα and CCβ
Fitness for Purpose (3rd ed.)	Eurachem, 2025	chemical laboratories in every sector	validation and verification; uncertainty with Eurachem/CITAC QUAM

Key points of Chapter 15

- ICH M10 (2022) is the harmonized guideline on bioanalytical method validation, adopted by the FDA and the EMA: it replaced EMA 2011 (since 2023) and supersedes FDA 2018 for the methods within its scope.
- M10: drugs and metabolites in biological matrices for PK/TK studies; criteria for chromatographic methods and LBAs; SIL-IS recommended; matrix effect with QCs in ≥ 6 lots; recovery is dealt with among the additional considerations (§7.3).
- ICH Q2(R2) says what to validate for pharmaceutical methods, ICH Q14 how to develop them, and Q14 introduces the ATP; in the ICH documents *accuracy* = trueness.
- SANTE/11312/2021: pesticides in food and feed, recovery 70–120%, RSD ≤ 20%.
- Regulation (EU) 2021/808: veterinary drug residues in food of animal origin; it replaces Decision 2002/657/EC and updates CCα and CCβ.
- Eurachem *Fitness for Purpose*: validation in any sector; uncertainty with Eurachem/CITAC QUAM.

Test yourself

Use the quiz questions on this topic:

- **Guidelines and regulatory framework**

16. From method to routine: practical aspects

A validated method must then be used every day without losing reliability. This chapter brings together the practical aspects that link calibration and validation to routine work, with particular attention to bioanalysis by LC-MS/MS.

16.1 Instrumental methods and main techniques

Gravimetry and titrimetry are *absolute* methods (Chapter 10). Instrumental methods offer three advantages:

- **much lower limits of detection and quantification**, suited to trace analysis (ng/mL or pg/mL in LC-MS/MS). This advantage is often called "higher sensitivity", but strictly speaking sensitivity is the slope of the curve, which is distinct from LOD and LOQ (Chapter 11);
- **automation**: serial analyses with autosamplers, fewer human errors and higher throughput;
- **data management**: acquisition, processing and storage on a PC with dedicated software.

However, they are **relative methods**: they require calibration, standards and QCs, and their precision worsens at low concentrations. Automation does not eliminate systematic errors or matrix effects, and it does not allow quantification outside the validated range.

Table 16.1 — The four families of instrumental techniques

Family	Principle	Examples
Spectroscopy	Absorption or emission of radiation by molecules or atoms	UV-Vis, IR (molecular absorption); fluorescence (molecular emission); AAS (atomic absorption); ICP-OES (atomic emission)
Electrochemical methods	Measurement of potential, current or conductance	Potentiometry (potential), voltammetry (current), conductometry (conductance), electrolysis
Chromatography	Separation between a mobile phase and a stationary phase, followed by spectroscopic or electrochemical detection	GC, HPLC, ion chromatography
Mass spectrometry	Separation of ions according to their mass-to-charge ratio (m/z); high specificity and sensitivity	MS, LC-MS/MS, GC-MS

In a triple quadrupole, the precursor ion → product ion transition is characteristic of the analyte: this is why LC-MS/MS is the reference technique in bioanalysis. Its weak point is the ion source, which is sensitive to the matrix effect (Chapter 13).

16.2 The analytical process

The analytical process comprises five steps: **sampling → standards and blank → calibration → analysis → data processing** (Figure 16.1). Calibration comes before analysis because it is the curve that converts the sample signals into concentrations.

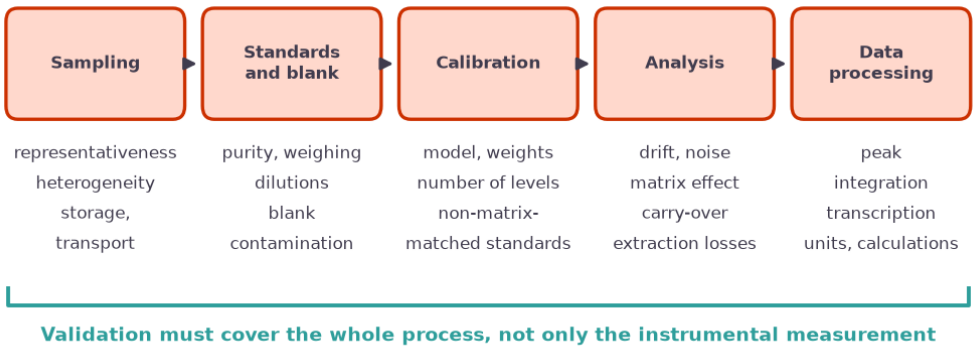


Figure 16.1 — The five steps of the analytical process, with some typical sources of error for each step.

Each step introduces potential errors, systematic or random. For this reason **validation must cover the whole process, not only the instrumental step**: if only standards in solvent were injected, extraction losses, the ma-

trix effect and sample instability would remain invisible. Sampling escapes even the QCs, which are prepared in the laboratory from a homogeneous matrix. If two portions of the same lot (for example a feed contaminated with mycotoxins, which are unevenly distributed) differ by a factor of 3 while the QCs and repeated injections comply, the cause is non-representative sampling.

16.3 Uncertainty propagation in serial dilutions

A working solution is obtained through a chain: weighing → stock solution → dilutions (Chapter 9). The final concentration is a **product of quotients**, for example $c = \frac{m \cdot P}{V_0} \cdot \frac{V_{p,1}}{V_{m,1}} \cdot \frac{V_{p,2}}{V_{m,2}}$. For products and quotients of independent quantities, the **relative uncertainties are combined in quadrature** (GUM, Chapter 6):

$$\frac{u_c(c)}{c} = \sqrt{\sum_{i=1}^{n} \left(\frac{u_i}{x_i}\right)^2} \tag{16.1}$$

Equation 16.1

- x_i: each quantity in the calculation (mass, purity, volume of each pipette and flask); u_i: its standard uncertainty; u_i/x_i: relative uncertainty.
- Relative uncertainties are used because a 1% error on any factor gives 1% on c; absolute uncertainties are combined in this way only for sums and differences.
- The linear sum overestimates; the sum of squares without the square root is in %². With n steps of equal uncertainty u_{rel} the result is $u_{\mathrm{rel}}\sqrt{n}$: it grows with n (there is no division by $\sqrt{n}$, which applies to the standard error of a mean).
- A glassware tolerance $\pm a$ is converted into a standard uncertainty with $u = a/\sqrt{6}$ (triangular distribution, Eurachem/CITAC) or $a/\sqrt{3}$ (rectangular).

Example 16.1 — A working solution in two dilutions

Data. Stock solution: 25.00 mg of standard (purity 99.2%) in 25 mL, 0.992 mg/mL, with a relative uncertainty of 0.32% (weighing, purity, flask). Dilutions: 1.000 mL → 100.0 mL, then 5.00 mL → 50.0 mL. Relative uncertainties: 1 mL pipette 0.327%; 100 mL flask 0.041%; 5 mL pipette 0.122%; 50 mL flask 0.049%.

1. c = 0.992 mg/mL × (1.000/100.0) × (5.00/50.0) = 992 ng/mL.
2. Sum of squares: 0.32² + 0.327² + 0.041² + 0.122² + 0.049² = 0.228 %².
3. $u_c(c)/c = \sqrt{0.228}$ = 0.48%; u_c = 4.7 ng/mL; U (k = 2) = 9.5 ng/mL.

Result: c = (992.0 ± 9.5) ng/mL. The 1 mL pipette (0.327%) and the stock solution (0.32%) dominate: it pays to use larger volumes, not better flasks.

The practical implications are: **keep serial dilutions to a minimum** (for the same uncertainty per step; using tiny volumes to save steps is counterproductive); use **calibrated pipettes** with known uncertainty and **Class A volumetric flasks** (not graduated cylinders); **document every step** with its uncertainty; prepare the **stock solution by direct weighing** of a sufficient mass, corrected for purity. An error in the stock solution is passed on to all calibrators without distorting the curve: this is why ICH M10 requires calibrators and QCs to be prepared from **separate stock solutions**.

16.4 Monitoring the curve over time

In regulated bioanalysis (ICH M10) **a new curve is built for each analytical run**: samples are never quantified with the curve of a previous run. Where, instead, the procedure allows a curve to be reused (for example in

clinical chemistry analyzers), when the **reagent lot** changes the curve must be **verified with 1–2 standards** before it is used any further. If the standards fall outside the tolerance (for example −20%), the response has changed and a new curve is built. Results are not corrected with an empirical factor, and the verification is not repeated until it "passes".

The **slope** is the sensitivity of the method and must remain constant over time: it is good practice to record it in every run on a control chart. A gradual decrease (for example 25% over a few weeks) must be investigated even if each curve is acceptable, with a high r^2 and compliant calibrators: it points to instrumental problems (dirty source, detector, column) or to problems with reagents and solutions. Blank contamination, by contrast, acts on the intercept. **QCs** at several levels, prepared independently of the calibrators, confirm in every run the performance of **procedure, instrument and operators**. They are not included in the regression and do not correct for the matrix effect.

16.5 The intercept and forcing through the origin

The intercept b_0 is the signal predicted at zero concentration, that is, the **blank signal**: instrumental background plus **additive interferences**. An intercept significantly different from zero may indicate contamination of the blank or the reagents, uncorrected additive interferences (including additive matrix interferences) or an instrumental problem such as baseline drift. The matrix effect in the strict sense (ion suppression or enhancement) is instead multiplicative and acts on the **slope**, as do proportional errors (for example an error in the stock solution). Detector saturation bends the line at the top.

The significance of b_0 is checked with a Student's t test (Chapter 5): $t = b_0 / s_{b_0}$, where s_{b_0} is the standard deviation of the intercept provided by the regression software, with $\nu = n - 2$. Equivalently, one checks whether the confidence interval of b_0 includes zero. Forcing the line through the origin means imposing b_0 = 0; the slope then becomes:

$$b_1' = \frac{\sum x_i y_i}{\sum x_i^2} \tag{16.2}$$

Example 16.2 — Effect of forcing with a real intercept

Data. Calibrators at 0.5, 1, 2, 5, 10 and 20 µg/mL; exact response $y = 3.0 + 10.0x$ (intercept due to the background).

1. Equation 16.2: b_1' = 5418/530.25 = 10.218 (only +2.2% compared with 10.0).
2. A 1.00 µg/mL sample gives y = 13.0; with the forced line x = 13.0/10.218 = 1.272 µg/mL, a bias of +27.2%.
3. At 0.50 µg/mL the bias is +56.6%; at 20.0 µg/mL it is −0.7%.

Result: forcing introduces an artificial systematic error concentrated at the low levels (Figure 16.2).

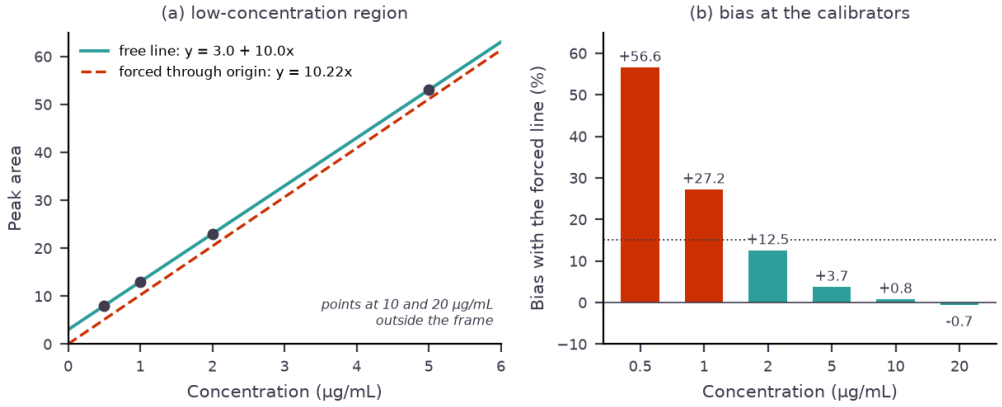


Figure 16.2 — Forcing through the origin with a real intercept (Example 16.2): (a) the forced line passes below the low-concentration points; (b) bias with the forced line, beyond ±15% at the low levels.

The **operating rule** is: the intercept is **estimated freely**; if it is clearly different from zero, one assesses whether it is **physically justifiable** (for example a known instrumental background); otherwise the **analytical procedure is revised**. The line is not forced, the low calibrators are not excluded, and a high r^2 is not taken as sufficient. According to Eurachem and ISO 8466-1, regression through the origin is admissible only if the intercept is **not statistically different from zero** and the choice is justified and documented.

16.6 Outliers and curve acceptance

Outlying points are looked for **after the regression**: residual plot, **standardized residuals** ($e_i/s_{y/x}$; a point is suspect if the absolute value exceeds 3) and back-calculated concentrations. The **Grubbs and Dixon tests** apply only to **replicates of the same level**, not to calibrators at different concentrations.

For each calibrator, the bias of the back-calculated concentration $x_{\text{back}} = (y_i - b_0)/b_1$ is calculated:

$$\text{bias}\% = \frac{x_{\text{back}} - x_{\text{nom}}}{x_{\text{nom}}} \times 100 \qquad (16.3)$$

For example, with y = 0.0042 + 0.0315x, a calibrator at the LLOQ (1.00 ng/mL) with y = 0.0381 gives x_{back} = 1.076 ng/mL and a bias of +7.6% (Equation 16.3), which is compliant. Forgetting the intercept would give +21.0%.

The ICH M10 rules are:

- a calibrator with a bias beyond **±15%** (**±20% at the LLOQ**) is **non-compliant and is rejected**, with no need for a cause; the regression is recalculated without it;
- a **compliant** calibrator is excluded **only with a documented justification** (preparation error, injection problem). Excluding it to raise r^2 is unacceptable, even though a high standardized residual calls for an investigation;
- the curve is accepted if at least **75% of all calibrators** (rejected ones included) are compliant and the compliant ones cover **at least 6 levels**. Round up: with 10 calibrators 0.75 × 10 = 7.5, so 8 are needed; with 8, 6 are needed.

16.7 The stable isotope-labeled internal standard (SIL-IS)

The SIL-IS is the analyte labeled with ^2H, ^13C or ^15N: it behaves like the analyte but is distinguished by its m/z. Quantification uses the ratio $R = A_{\text{analyte}}/A_{\text{IS}}$, in which any factor that multiplies both signals to the same extent cancels out. The SIL-IS therefore compensates, **after it has been added**, for extraction losses, injection variability, the matrix effect (if it co-elutes) and MS source instability.

It does not compensate for: errors in the calibrators (a wrong stock solution distorts all results proportionally), anything that happens before it is added, errors in the volume of IS added, and additive interferences in the analyte channel.

Practical requirements:

- added as early as possible, ideally before extraction;
- constant concentration in samples, calibrators and QCs;
- no cross-contributions: according to ICH M10, the contribution of the IS to the analyte channel must be ≤ 20% of the response at the LLOQ, and that of the analyte to the IS channel ≤ 5% of the IS response;
- high isotopic purity: "≥ 98%" is a rule of thumb, not an ICH M10 criterion.

If the blank matrix is clean but the **zero sample** (same matrix and same reagents, plus the IS) shows a peak in the analyte channel, the cause is the IS itself (isotopic impurity or cross-talk).

Behavior "identical to the analyte" holds fully for ^13C and ^15N. **Deuterated** analogs may elute slightly earlier (deuterium isotope effect) and compensate for the matrix effect imperfectly; moreover, deuterium on –OH or –NH exchanges with the solvent. The main limitation of the SIL-IS is its **high cost**, together with **availability that is not guaranteed** for every analyte.

Example 16.3 — Analyte/IS ratio under ion suppression

Data. Slope 0.0300 mL/ng (area ratio), negligible intercept; IS area in the calibrators about 120 000. A lipemic plasma contains 8.00 ng/mL and undergoes 50% suppression.

1. Measured areas: analyte 14 400, IS 60 000; R = 0.240; c = 0.240/0.0300 = 8.00 ng/mL. The suppression cancels out in the ratio and must not be corrected again.
2. Calibrating on the analyte area alone (3600 per ng/mL without suppression): 14 400/3600 = 4.00 ng/mL (−50%).
3. Deuterated IS eluting where the suppression is only 35%: IS = 78 000; R = 0.185; c = 6.15 ng/mL (−23%).

Result: only an IS that co-elutes exactly returns 8.00 ng/mL.

16.8 Summary of the ICH M10 criteria

Table 16.2 — Main ICH M10 criteria (chromatographic methods)

Parameter	General criterion	At the LLOQ
Trueness (bias%)	±15%	±20%
Precision (CV%)	≤ 15%	≤ 20%
Compliant calibrators	≥ 75% of the total, ≥ 6 levels	—
Selectivity (blanks from ≥ 6 lots)	≤ 20% of the LLOQ response and ≤ 5% of the IS response	—
Matrix effect (low and high QC)	±15% and CV ≤ 15% in each of ≥ 6 lots	—
Carry-over (blank after the ULOQ)	≤ 20% of the LLOQ response and ≤ 5% of the IS response	—
Run acceptance	≥ 2/3 of QCs within ±15% and ≥ 50% at each level	—

Selectivity and carry-over share the same limit but use different samples: **blank matrices from different lots** for selectivity, **a blank injected right after the ULOQ** for carry-over. For example, an area of 480 in the blank against 2850 at the LLOQ gives 16.8%, which is compliant. The matrix effect is evaluated on several lots because matrix composition varies between individuals (lipemic, hemolyzed plasma), and the criterion must be met lot by lot. For ligand-binding methods the limits are ±20% (±25% at the LLOQ and the ULOQ).

16.9 Full validation, partial validation and cross-validation

- **Full validation**: a method that is **new or significantly modified**. For chromatographic methods, ICH M10 requires selectivity, specificity, matrix effect, calibration curve and range, trueness, precision, carry-over, dilution integrity, stability and reinjection reproducibility. **Robustness** is not part of it: it is studied during method development (ICH Q14). Recovery is evaluated, but without a numerical criterion.
- **Partial validation**: **changes to a validated method**, evaluating only the parameters that may be affected. Examples: **transfer of the method to another laboratory**, a new equivalent instrument, a new range (for a new ULOQ: curve, trueness and precision at the new levels, carry-over), a change of matrix. A new reagent lot only requires verification of the curve (Section 16.4).
- **Cross-validation**: comparison of two validated methods, or of different laboratories, on the same samples, when the data are to be **combined or compared** (for example in a multicenter study). **ICH M10 does not set acceptance criteria**: the bias is assessed statistically. The **Bland–Altman plot** shows the differences against the means and gives the mean bias $\bar{d}$ and the limits of agreement $\bar{d} \pm 1.96\, s_d$. **Deming regression**, unlike ordinary least squares, takes into account the measurement error on both axes: a slope different from 1 indicates proportional bias, an intercept different from 0 constant bias. The "±20% for two-thirds of the samples" criterion concerns instead incurred sample reanalysis (ISR).

16.10 Validation plan and report

The **validation plan** (VP) is drawn up **before the experiments**: objectives, parameters, **acceptance criteria**, experimental design, responsibilities and timelines. Setting the criteria in advance prevents them from being adapted to the results. The **validation report** (VR) presents all the experimental data, the statistical calculations, the comparison with the criteria and the conclusion on the validity of the method; it must be **signed and archived** according to GLP/GMP rules. In a regulatory setting both are **inspectable documents**, and **traceability of every data point is mandatory**. A failed run must also be reported together with its cause: it does not in itself invalidate the validation, but omitting it compromises the integrity of the study.

16.11 Common mistakes in building the curve

1. **Forcing the line through the origin** without having verified that the intercept is not significant: this introduces a systematic bias.
2. **Relying on r^2 alone**, which does not detect systematic deviations: residuals and back-calculated concentrations must be evaluated. ICH M10 sets no r^2 thresholds. A curve with r^2 = 0.992, random residuals and all calibrators compliant is acceptable.
3. **Excluding compliant calibrators without justification**: this invalidates the curve; every exclusion must be documented.
4. **Reusing the curve with new reagents without verifying it** with at least 1–2 standards.

16.12 The analytical sequence in LC-MS/MS

The injection order is standardized to minimize carry-over and to ensure that the curve remains valid throughout the sequence (Figure 16.3):

1. **blank matrix**, first, so that any signal in it cannot come from carry-over;
2. **zero sample** (matrix + IS), which checks the contribution of the IS;
3. **calibrators** from the LLOQ to the ULOQ, followed by a blank for carry-over;
4. **QCs**, then the **samples** interspersed with further QCs, and finally the **end-of-sequence QCs**, which check the stability of the instrumental response.

The blank and the zero sample are not included in the regression.

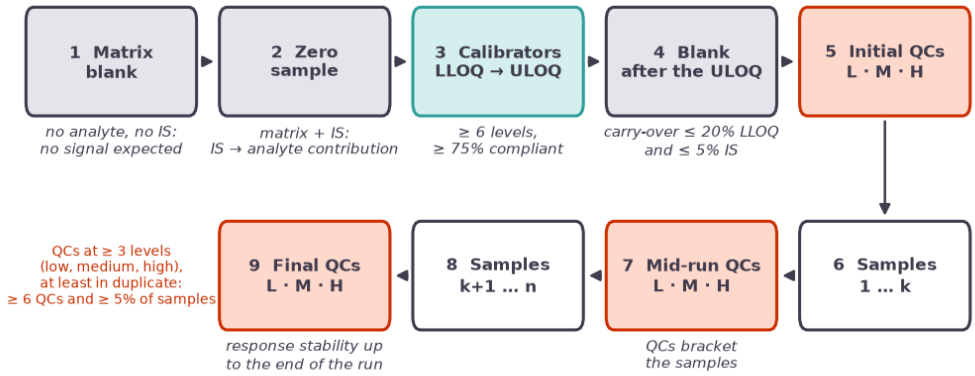


Figure 16.3 — Typical analytical sequence in LC-MS/MS according to ICH M10, with the QCs distributed along the sequence.

"Triplicate QCs" should be understood as **QCs at three levels**. ICH M10 requires QCs at **at least 3 levels** (low, mid, high), **at least in duplicate**, for a total of at least 6 QCs and at least 5% of the number of samples, distributed throughout the sequence. With 150 samples, 5% gives 7.5, so at least 8 QCs are needed: in practice 9, three per level. The run is accepted if at least 6 of them (2/3) comply, with at least 2 per level (50%): two low QCs out of limits cause it to be rejected even with 7 out of 9 compliant.

Key points

- Instrumental methods offer low LOD/LOQ, automation and data management, but they are relative; validation must cover the whole process, from sampling to data processing.
- In dilutions, relative uncertainties add in quadrature: few dilutions, Class A glassware, stock solution by direct weighing.
- One curve per run; with new reagents a reused curve must be verified; slope and QCs are monitored over time.
- The intercept (blank signal) is estimated freely; forcing through the origin is legitimate only if the intercept is not significant and the choice is justified.
- Non-compliant calibrators are rejected, compliant ones only with a documented cause; ≥ 75% of the total compliant and ≥ 6 levels.
- The SIL-IS compensates for extraction, injection, matrix effect and source, but not for errors in the calibrators; deuterated analogs may not co-elute.
- Full validation for new methods, partial validation for changes and transfers, cross-validation without fixed criteria; plan and report ensure traceability.

Test yourself

Quiz topics: *Calibration methods and instrumental methods*; *Instrumental techniques and the analytical process*; *Uncertainty propagation in dilutions*; *Curve monitoring and intercept*; *Curve acceptance and outliers*; *Stable isotope-labeled internal standard (SIL-IS)*; *ICH M10 criteria and analytical sequence*; *Validation types and documentation*.

17. Summary: from calibration to validation

This chapter summarizes Part II: how the tools of calibration and the parameters of validation lead to a method that is ready for routine use, which messages are worth remembering, and where to read further.

17.1 The big picture

Figure 17.1 summarizes Part II in two blocks.

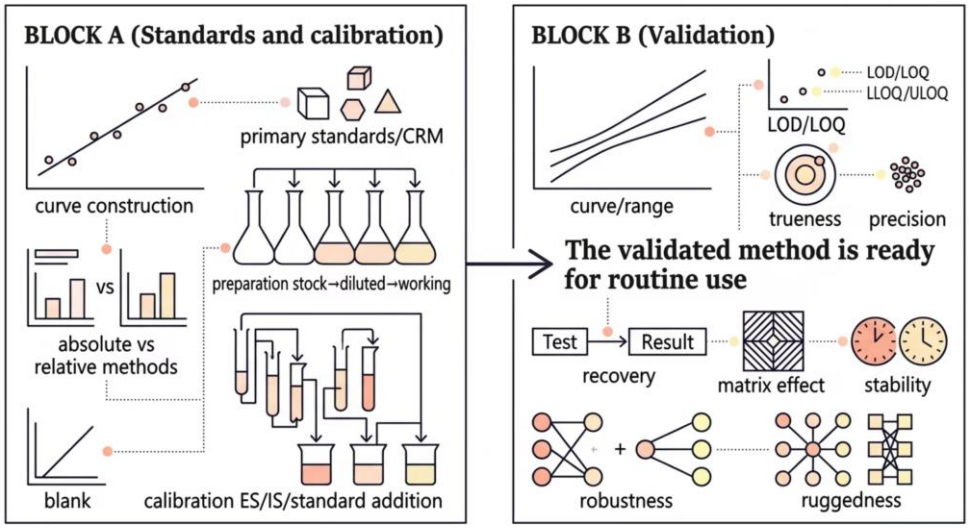


Figure 17.1 — Summary of Part II. Block A: the tools for building a reliable curve. Block B: the parameters that demonstrate that the method is fit for purpose. Only after both is the method ready for routine use.

Block A, standards and calibration, includes:

- primary standards and CRMs, which provide metrological traceability (Chapter 9);
- the preparation of the stock → diluted → working solution chain (Chapters 9 and 16);
- the distinction between absolute and relative methods and the role of the blank (Chapter 10);
- the construction of the curve and the choice between external standard, internal standard and standard addition (Chapter 10, with the statistical basis in Chapters 7–8).

Block B, validation, brings together the parameters that demonstrate that the method is fit for purpose:

- curve and range, LOD/LOQ and LLOQ/ULOQ (Chapter 11);
- trueness and precision (Chapter 12);
- recovery and matrix effect (Chapter 13);
- stability, robustness and ruggedness (Chapter 14).

Robustness and ruggedness appear in the diagram as part of the validation pathway. In ICH M10 and in ICH Q2(R2)/Q14, however, robustness is studied during method development, while the concept of ruggedness is absorbed into intermediate precision and reproducibility (Chapter 14).

Once both blocks have been completed, the method enters routine use, where it remains under control. Each run has its own curve and its own QCs; slope and QCs are monitored over time. Changes require a partial validation; data from different methods or laboratories, if they are to be combined, require a cross-validation (Chapter 16).

17.2 The key messages

1. **Metrological traceability is fundamental.** Every standard must be linked to the SI through a documented chain of comparisons, with the uncertainty stated at each level. A perfect regression cannot reveal an er-

ror in the value assigned to the standard (Example 17.1). Calibrated glassware makes the volumes traceable, not the purity of the standard.

2. r^2 **is not enough.** The curve is evaluated with the residuals and the back-calculated concentrations. A high r^2 can hide systematic deviations; ICH M10 sets no r^2 threshold.
3. **Consistency across the whole range is mandatory.** Trueness, precision and recovery must meet the criteria from the LLOQ to the ULOQ, not only in the middle of the curve. A "perfect" curve with QCs that fail at the low levels indicates a method that is not valid.
4. **The SIL-IS is the optimal choice in LC-MS/MS.** It compensates at the same time for extraction losses, injection variability and the matrix effect, provided that it co-elutes with the analyte. However, it does not correct errors in the calibrators, and deuterated analogs may co-elute imperfectly.

Example 17.1 — A perfect curve with a wrong standard

Data. A curve built with a standard of uncertified purity, assumed to be 100% pure, has r^2 = 0.9999 and random residuals. The actual purity is 95%.

1. Each calibrator actually contains 95% of the nominal concentration; the apparent slope is therefore 0.95 times the true one.
2. A sample at concentration c gives the signal expected for c, but it is quantified as $c/0.95$ = 1.053 c.

Result: all results are overestimated by 5.3%, with no sign in the residuals. Only a traceable standard (a CRM) or QCs prepared from an independent standard reveal the error.

17.3 Further reading

The guidelines to know are **ICH M10 (2022)** for bioanalysis, **ICH Q2(R2) and Q14 (2023)** for pharmaceutical methods, the **Eurachem** guide *The Fitness for Purpose of Analytical Methods* (3rd ed., 2025) for laboratories in every sector, **SANTE/11312/2021 v2026** for pesticide residues and **Regulation (EU) 2021/808** for veterinary drug residues. For further study: Skoog, Holler and Crouch, *Principles of Instrumental Analysis* (instrumental techniques); Harvey, *Modern Analytical Chemistry* (fundamentals of quantitative analytical chemistry); Matuszewski et al., *Anal. Chem.* 2003, 75, 3019–3030, the paper that introduced the three-set scheme (Chapter 13); ISO 5725 on trueness and precision. Full references are in Appendix D.

Key points

- Block A (standards and calibration) and block B (validation) together make a method ready for routine use.
- Traceability to the SI is the basis of trueness: a good regression alone does not guarantee it.
- The curve is judged by residuals and back-calculation, not by r^2.
- The criteria apply across the whole range, from the LLOQ to the ULOQ.
- The SIL-IS is the optimal choice in LC-MS/MS, within the limits of what the ratio can compensate for.
- In routine use the method stays under control with curves, QCs and, when something changes, partial validation or cross-validation.

Test yourself

Quiz topic: *Summary and key messages*. For a final review, take a mixed test on all the calibration and validation topics.

Appendices

Formula sheet, statistical tables, Italian–English glossary and references for study and quick consultation.

Appendix A — Formula sheet

The formulas of these notes, grouped by topic. The chapter shown in parentheses contains the full explanation and the worked examples.

A.1 Descriptive statistics (Chapter 4)

Table A.1 — Measures of location and dispersion

Quantity	Formula	Notes
Mean	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	same unit as the data
Median	middle value of the ordered data	for even n: mean of the two middle values
Sample variance	$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$	squared units
Standard deviation	$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$	$\nu = n - 1$ degrees of freedom
Relative standard deviation	$\mathrm{RSD} = \mathrm{CV}\% = \frac{s}{\bar{x}} \times 100$	dimensionless
Standard error of the mean	$s_{\bar{x}} = \frac{s}{\sqrt{n}}$	decreases with $\sqrt{n}$

A.2 Measurement errors (Chapters 2–3)

Table A.2 — Errors, bias and recovery

Quantity	Formula	Notes
Absolute error	$E = x_i - \mu$	μ = true or reference value
Relative error	$E_{\mathrm{rel}}\% = \frac{x_i - \mu}{\mu} \times 100$	with sign
Bias (trueness)	$b = \bar{x} - \mu$	calculated on the mean
Relative bias	$b\% = \frac{\bar{x} - \mu}{\mu} \times 100$	ICH M10 criterion: within ±15% (±20% at the LLOQ)
Recovery of a spike	$R\% = \frac{C_{\mathrm{spiked}} - C_{\mathrm{unspiked}}}{C_{\mathrm{added}}} \times 100$	check of trueness

A.3 Inference and outliers (Chapter 5)

$$\bar{x} \pm t_\nu \cdot \frac{s}{\sqrt{n}}$$

Confidence interval of the mean, with two-tailed t_ν and $\nu = n - 1$ (Table B.1).

$$Q = \frac{|x_{\text{susp}} - x_{\text{neighbor}}|}{x_{\max} - x_{\min}}$$

Dixon's Q test: the value is an outlier if $Q > Q_{\text{crit}}$ (Table B.2).

$$G = \frac{|x_{\text{susp}} - \bar{x}|}{s}$$

Grubbs test, with $\bar{x}$ and s calculated including the suspect value: the value is an outlier if $G > G_{\text{crit}}$ (Table B.3).

A.4 Measurement uncertainty (Chapter 6)

Table A.3 — Evaluation and combination of uncertainties

Quantity	Formula	Notes
Type A (mean of n replicates)	$u_A = \frac{s}{\sqrt{n}}$	if the result is the mean of the n replicates
Type B, rectangular distribution	$u = \frac{a}{\sqrt{3}}$	a = half-width (tolerance ±a)
Type B, triangular distribution	$u = \frac{a}{\sqrt{6}}$	central values more probable
Combined uncertainty (sums and differences)	$u_c = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$	independent contributions, same unit
Combined uncertainty (products and quotients)	$\frac{u_c(y)}{y} = \sqrt{\sum \left(\frac{u_i}{x_i}\right)^2}$	relative uncertainties are added in quadrature
Expanded uncertainty	$U = k \cdot u_c$	$k = 2$: about 95% (Table B.4)

A.5 Linear regression (Chapters 7–8)

$$b_1 = \frac{\sum (x_i - \bar{x})\,(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \qquad b_0 = \bar{y} - b_1 \bar{x}$$

$$\hat{y}_i = b_0 + b_1 x_i \qquad e_i = y_i - \hat{y}_i$$

$$x_0 = \frac{y_0 - b_0}{b_1}$$

Concentration of a sample from its signal y_0 (inverse prediction), valid only within the calibration range.

Table A.4 — Partition of variability and goodness-of-fit indices

Quantity	Formula	Meaning
Total sum of squares	$\mathrm{SST} = \sum (y_i - \bar{y})^2$	total variability of the signals
Regression sum of squares	$\mathrm{SSR} = \sum (\hat{y}_i - \bar{y})^2$	variability explained by the line (here SSR is the regression sum of squares, not the sum of squared residuals)
Error (residual) sum of squares	$\mathrm{SSE} = \sum (y_i - \hat{y}_i)^2$	residual variability
Relationship	$\mathrm{SST} = \mathrm{SSR} + \mathrm{SSE}$	holds for the least-squares line with intercept
Coefficient of determination	$r^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}}$	fraction of variability explained
Standard deviation of the residuals	$s_{y/x} = \sqrt{\frac{\mathrm{SSE}}{n-2}}$	$n-2$ degrees of freedom
Weights (weighted regression)	$w_i = \frac{1}{x_i^2}$ or $\frac{1}{x_i}$	heteroscedastic data

A.6 Calibration (Chapters 9–10 and 16)

Table A.5 — Calibration, dilutions and standards

Quantity	Formula	Notes
Dilution	$C_1 V_1 = C_2 V_2$	same volume units
Internal standard	$R = \frac{A_{\text{analyte}}}{A_{\text{IS}}}$	the curve is built with R
Standard addition	$C_x = \frac{b_0}{b_1}$	x-intercept at $-C_x$
Back-calculation bias	$\frac{C_{\text{back}} - C_{\text{nom}}}{C_{\text{nom}}} \times 100$	±15% (±20% at the LLOQ)
Propagation in serial dilutions	$\frac{u_c(c)}{c} = \sqrt{\sum \left(\frac{u_i}{x_i}\right)^2}$	each step adds uncertainty

A.7 Validation parameters (Chapters 11–14)

Table A.6 — Sensitivity, limits and expected precision

Quantity	Formula	Notes
Calibration sensitivity	$b_1 = \frac{\Delta S}{\Delta C}$	slope of the line
Analytical sensitivity	$\gamma = \frac{b_1}{s_S}$	depends on concentration
Limit of detection	$\mathrm{LOD} = \frac{3 \cdot s_B}{b_1}$	S/N ≈ 3
Limit of quantification	$\mathrm{LOQ} = \frac{10 \cdot s_B}{b_1}$	S/N ≈ 10
Horwitz curve	$\mathrm{CV\%} \approx 2^{(1-0.5\log C)}$	C = dimensionless mass fraction, base-10 log

Table A.7 — Recovery, matrix effect and process efficiency (Matuszewski scheme)

Quantity	Formula	Meaning
Recovery	$\mathrm{RE\%} = \frac{\text{Set 3}}{\text{Set 2}} \times 100$	losses during extraction
Matrix factor	$\mathrm{MF\%} = \frac{\text{Set 2}}{\text{Set 1}} \times 100$	< 100%: suppression; > 100%: signal enhancement
Process efficiency	$\mathrm{PE\%} = \frac{\text{Set 3}}{\text{Set 1}} \times 100 = \frac{\mathrm{RE} \times \mathrm{MF}}{100}$	overall effect
Carry-over	$\frac{A_{\text{blank after ULOQ}}}{A_{\mathrm{LLOQ}}} \times 100$	≤ 20% (≤ 5% for the IS)

A.8 ICH M10 acceptance criteria (chromatographic methods)

Table A.8 — Summary of the ICH M10 criteria

Parameter	Criterion
Calibrators (back-calculation)	within ±15% of the nominal value (±20% at the LLOQ); at least 75% compliant, with at least 6 levels
Trueness of QCs	mean within ±15% of the nominal value (±20% at the LLOQ QC)
Precision of QCs	CV ≤ 15% (≤ 20% at the LLOQ QC), within-run and between-run
Acceptance of an analytical run	at least 2/3 of QCs within ±15% and at least 50% at each level
Selectivity	response in the blank ≤ 20% of the LLOQ (analyte) and ≤ 5% (internal standard), in at least 6 lots
Matrix effect	low and high QC in at least 6 lots: trueness ±15%, CV ≤ 15%
Carry-over	blank after the ULOQ ≤ 20% of the LLOQ and ≤ 5% of the IS
Dilution integrity	trueness ±15%, CV ≤ 15%
QC levels	LLOQ; low ≤ 3 × LLOQ; mid 30–50% of the range; high ≥ 75% of the ULOQ
Cross-validation	no fixed acceptance criterion: the bias between the methods is assessed statistically

Appendix B — Statistical tables

The tables give the critical values used in the examples and in the quiz. The values of t, G and the coverage factor are computed from the theoretical distributions; the critical values for Dixon's Q test are taken from Rorabacher's table (1991).

B.1 Critical values of Student's t (two-tailed)

Use the column for the desired confidence level and the row for the degrees of freedom $\nu = n - 1$ (for the confidence interval of the mean). For $\nu \to \infty$ the values coincide with those of the normal distribution.

Table B.1 — Values of t_ν for two-sided confidence intervals

ν	90%	95%	99%
1	6.314	12.706	63.657
2	2.920	4.303	9.925
3	2.353	3.182	5.841
4	2.132	2.776	4.604
5	2.015	2.571	4.032
6	1.943	2.447	3.707
7	1.895	2.365	3.499
8	1.860	2.306	3.355
9	1.833	2.262	3.250
10	1.812	2.228	3.169
11	1.796	2.201	3.106
12	1.782	2.179	3.055
13	1.771	2.160	3.012
14	1.761	2.145	2.977
15	1.753	2.131	2.947
16	1.746	2.120	2.921
17	1.740	2.110	2.898
18	1.734	2.101	2.878
19	1.729	2.093	2.861
20	1.725	2.086	2.845
21	1.721	2.080	2.831
22	1.717	2.074	2.819
23	1.714	2.069	2.807
24	1.711	2.064	2.797
25	1.708	2.060	2.787
26	1.706	2.056	2.779
27	1.703	2.052	2.771
28	1.701	2.048	2.763
29	1.699	2.045	2.756
30	1.697	2.042	2.750

ν	90%	95%	99%
40	1.684	2.021	2.704
60	1.671	2.000	2.660
120	1.658	1.980	2.617
∞	1.645	1.960	2.576

B.2 Critical values for Dixon's Q test

The suspect value is an outlier if $Q_{\text{calc}} > Q_{\text{crit}}$ at the chosen confidence level (usually 95%).

Table B.2 — Critical values of Q (two-tailed test)

n	$Q_{90\%}$	$Q_{95\%}$	$Q_{99\%}$
3	0.941	0.970	0.994
4	0.765	0.829	0.926
5	0.642	0.710	0.821
6	0.560	0.625	0.740
7	0.507	0.568	0.680
8	0.468	0.526	0.634
9	0.437	0.493	0.598
10	0.412	0.466	0.568

B.3 Critical values for the Grubbs test

$G = |x_{\text{susp}} - \bar{x}|/s$, with mean and standard deviation calculated including the suspect value. Values for the two-tailed test.

Table B.3 — Critical values of G (two-tailed test)

n	$G_{95\%}$	$G_{99\%}$
3	1.154	1.155
4	1.481	1.496
5	1.715	1.764
6	1.887	1.973
7	2.020	2.139
8	2.127	2.274
9	2.215	2.387
10	2.290	2.482
11	2.355	2.564
12	2.412	2.636
13	2.462	2.699
14	2.507	2.755
15	2.548	2.806
16	2.586	2.852
17	2.620	2.894

n	$G_{95\%}$	$G_{99\%}$
18	2.652	2.932
19	2.681	2.968
20	2.708	3.001

B.4 Coverage factor and confidence level

For a normal distribution (or an approximately normal one with enough degrees of freedom) the coverage factor k corresponds to the confidence level shown.

Table B.4 — Coverage factor k and confidence level

k	**Confidence level**
1	68.27%
1.645	90.00%
1.960	95.00%
2	95.45%
2.576	99.00%
3	99.73%

Appendix C — Italian–English glossary

The terms are in English alphabetical order. The third column gives a brief meaning; the chapter shown contains the full explanation.

Table C.1 — Glossary of technical terms

English	**Italiano**	**Meaning (chapter)**
Absolute / relative method	Metodo assoluto / relativo	method that does not require / requires a calibration curve (10)
Accuracy	Accuratezza	closeness of agreement between a single result and the true value; it comprises trueness and precision (3)
Bias	Bias	difference between the mean of the results and the reference value (2–3)
Blank matrix sample	Bianco di matrice	matrix free of analyte and internal standard, processed like a sample (10)
Calibration curve	Curva di calibrazione	relationship between signal and concentration obtained from the standards (7, 10)
Calibration range	Intervallo di calibrazione	range from the LLOQ to the ULOQ covered by the standards (11)
Carry-over	Carry-over	transfer of analyte from one injection to the next (14)
Certified reference material (CRM)	Materiale di riferimento certificato (CRM)	material with certified values, uncertainty and traceability (9)

English	Italiano	Meaning (chapter)
Coefficient of determination (r^2)	Coefficiente di determinazione	fraction of the variability of the signals explained by the line (7)
Coefficient of variation (CV%)	Coefficiente di variazione	relative standard deviation expressed as a percentage (4)
Confidence interval	Intervallo di confidenza	interval that contains the true value at the chosen confidence level (5)
Coverage factor (k)	Fattore di copertura	multiplier that turns the combined standard uncertainty into the expanded uncertainty (6)
Cross-validation	Cross-validazione	comparison of two validated methods on the same samples (16)
Degrees of freedom (ν)	Gradi di libertà	number of independent data available to estimate the dispersion (4–5)
Dilution integrity	Integrità della diluizione	demonstration that dilution with blank matrix does not alter the result (14)
Dixon's Q test	Test Q di Dixon	outlier test for small data sets (5)
Expanded uncertainty (U)	Incertezza estesa	combined standard uncertainty multiplied by the coverage factor (6)
Gross error	Errore grossolano	major error that invalidates the measurement (2)
Grubbs test	Test di Grubbs	outlier test based on $G = \lvert x - \bar{x} \rvert / s$ (5)
Heteroscedasticity	Eteroschedasticità	variance of the signal that changes with concentration (8)
Intermediate precision	Precisione intermedia	precision within the same laboratory with different days, operators or instruments (3, 12)
Internal standard (IS)	Standard interno (SI)	compound added at constant concentration to correct for variations (10)
Least squares	Minimi quadrati	method that chooses the line by minimizing the sum of the squared residuals (7)
Limit of detection (LOD)	Limite di rivelabilità (LOD)	lowest concentration that can be distinguished from the blank (11)
Limit of linearity (LOL)	Limite di linearità (LOL)	concentration above which the response is no longer linear (11)
Limit of quantification (LOQ)	Limite di quantificazione (LOQ)	lowest concentration that can be determined with acceptable precision and trueness (11)
Linear dynamic range	Range dinamico lineare	range from the LOQ to the LOL over which the response is linear (11)
Lower / upper limit of quantification (LLOQ / ULOQ)	LLOQ / ULOQ	lowest and highest standard of the calibration curve (11)
Matrix effect	Effetto matrice	change in the signal caused by the components of the matrix (13)
Matrix factor (MF)	Matrix factor	ratio of the signal in matrix to the signal in solvent (13)
Mean	Media	sum of the data divided by their number (4)

English	Italiano	Meaning (chapter)
Measurement uncertainty	Incertezza di misura	parameter that characterizes the dispersion of the values attributed to the measurand (1, 6)
Median	Mediana	middle value of the ordered data (4)
Metrological traceability	Tracciabilità metrologica	relation of a result to a reference through an unbroken chain of calibrations (9)
Precision	Precisione	closeness of agreement between repeated measurements (3, 12)
Primary standard	Standard primario	substance of high and known purity used as a reference (9)
Process efficiency	Efficienza di processo	combined effect of recovery and matrix effect (13)
Quality control sample	Controllo qualità (QC)	sample of known concentration used to accept an analytical run (12)
Random error	Errore casuale	error that varies unpredictably between repeated measurements (2)
Reagent blank	Bianco dei reagenti	solvent with all the reagents, without the sample (10)
Recovery	Recupero	fraction of analyte recovered after extraction (2, 13)
Repeatability	Ripetibilità	precision under the same conditions over a short period of time (3)
Reproducibility	Riproducibilità	precision between different laboratories (3)
Residual	Residuo	difference between the measured signal and the signal predicted by the line (7–8)
Robustness	Robustezza	insensitivity of the method to small, deliberate changes in its parameters (14)
Ruggedness	Solidità	consistency of the results when the context changes (laboratory, instrument, operator) (14)
Selectivity	Selettività	ability to measure the analyte without interference from the matrix (11)
Sensitivity	Sensibilità	change in signal per unit change in concentration (slope) (11)
Specificity	Specificità	absolute selectivity: response to the analyte only (11)
Stability	Stabilità	preservation of the analyte concentration under the conditions of use (14)
Stable isotope-labeled internal standard (SIL-IS)	Standard interno marcato con isotopi stabili	isotopic analog of the analyte, the ideal internal standard in LC-MS/MS (10, 16)
Standard addition (method of standard additions)	Aggiunte multiple (aggiunte standard)	calibration performed by adding known amounts of analyte to the sample itself (10)
Standard deviation	Deviazione standard	measure of the dispersion of the data around the mean (4)
Standard error of the mean	Errore standard della media	standard deviation of the mean, $s/\sqrt{n}$ (5)
Systematic error	Errore sistematico	error that remains constant or varies in a predictable way (2)

English	Italiano	Meaning (chapter)
Trueness	Esattezza	closeness of agreement between the mean of many results and the true value (3, 12)
Validation	Validazione	documented demonstration that a method is fit for its intended purpose (11)
Zero sample	Campione zero	blank matrix with the internal standard only (10)

Appendix D — References

D.1 Guidelines and standards

- ICH M10 — *Bioanalytical Method Validation and Study Sample Analysis*, International Council for Harmonisation, 2022.
- ICH Q2(R2) — *Validation of Analytical Procedures*, International Council for Harmonisation, 2023.
- ICH Q14 — *Analytical Procedure Development*, International Council for Harmonisation, 2023.
- EMA — *Guideline on Bioanalytical Method Validation*, European Medicines Agency, 2011 (superseded by ICH M10 since 2023).
- FDA — *Bioanalytical Method Validation, Guidance for Industry*, U.S. Food and Drug Administration, 2018 (superseded by ICH M10 for the methods within its scope).
- Eurachem — *The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and Related Topics*, 3rd edition, 2025.
- Eurachem/CITAC — *Quantifying Uncertainty in Analytical Measurement* (QUAM guide, CG4), 3rd edition, 2012.
- SANTE/11312/2021 v2026 — *Analytical Quality Control and Method Validation Procedures for Pesticide Residues Analysis in Food and Feed*, European Commission, applicable from 1 January 2026 (supersedes v2 of 2024).
- Commission Implementing Regulation (EU) 2021/808 on the performance of analytical methods for residues of pharmacologically active substances used in food-producing animals.
- JCGM 100:2008 — *Evaluation of measurement data — Guide to the expression of uncertainty in measurement* (GUM).
- JCGM 200:2012 — *International vocabulary of metrology — Basic and general concepts and associated terms* (VIM, 3rd edition).
- ISO 5725 (series) — *Accuracy (trueness and precision) of measurement methods and results*.
- ISO 8466-1 — *Water quality — Calibration and evaluation of analytical methods — Part 1: Linear calibration function*.
- ISO 17034:2016 — *General requirements for the competence of reference material producers*.
- ISO/IEC 17025:2017 — *General requirements for the competence of testing and calibration laboratories*.

D.2 Articles

- Matuszewski B. K., Constanzer M. L., Chavez-Eng C. M. — Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC–MS/MS. *Analytical Chemistry* 75 (2003) 3019–3030.
- Horwitz W. — Evaluation of analytical methods used for regulation of foods and drugs. *Analytical Chemistry* 54 (1982) 67A–76A.
- Rorabacher D. B. — Statistical treatment for rejection of deviant values: critical values of Dixon's "Q" parameter and related subrange ratios at the 95% confidence level. *Analytical Chemistry* 63 (1991) 139–146.

- Grubbs F. E. — Procedures for detecting outlying observations in samples. *Technometrics* 11 (1969) 1–21.
- Bland J. M., Altman D. G. — Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* 327 (1986) 307–310.

D.3 Reference textbooks

- Skoog D. A., Holler F. J., Crouch S. R. — *Principles of Instrumental Analysis*, 7th edition, Cengage Learning.
- Harvey D. — *Modern Analytical Chemistry*, McGraw-Hill.
- Miller J. N., Miller J. C., Miller R. D. — *Statistics and Chemometrics for Analytical Chemistry*, 7th edition, Pearson, 2018.