Eurachem
A focus for analytical chemistry in Europe

CITAC
Cooperation on International Traceability in Analytical Chemistry

Lisbon, 11 and 12 May 2026

Validation and Quality Control of Qualitative Analysis

EURACHEM/CITAC WORKSHOP ON QUALITY IN ANALYTICAL MEASUREMENTS

Paulo Pereira, Ph.D.,
Portuguese Institute of Blood and Transplantation

1

The Majority of Clinical Decisions Are Driven by Lab Results

Quality Control Protects Patients - One wrong result can change care

Lab-Driven
70–80% of clinical decisions rely on labs

Quality Control
Protects patients from wrong results

Diagnosis
Informs diagnosis, treatment, monitoring

Patient Safety
QC equals patient safety

2

Why Validation and QC Matter

False Negative

The patient may be falsely reassured when the condition is truly present — the most clinically dangerous error.

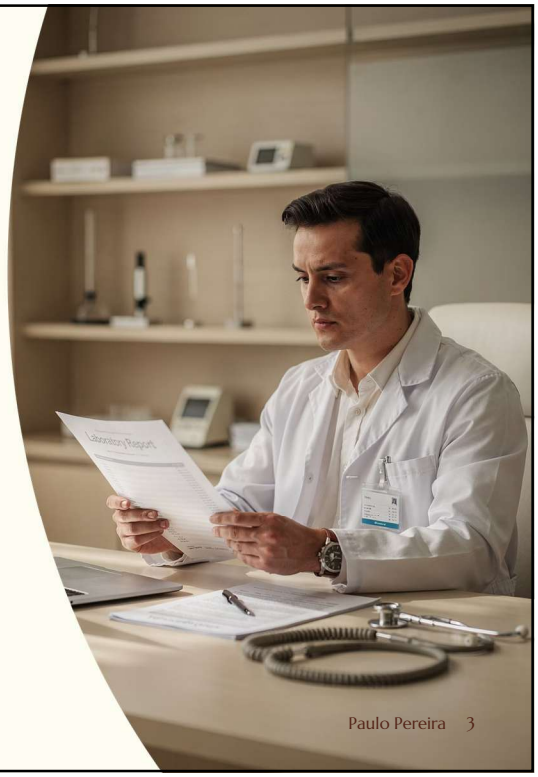
False Positive

Unnecessary follow-up testing, anxiety, cost, or inappropriate clinical action can follow.

Valid Result

Low misclassification risk, transparent performance data, and better-informed clinical decisions.

Key idea: we do not validate a word such as "positive" — we validate the **probability that the word is correct** for the intended use.



What Counts as Qualitative Analysis?

Definition from AQA 2021, translated into medical laboratory language: **"Classification according to specified criteria."**

Qualitative Criteria

Classification arises from **direct observation** of a feature — a color change, agglutination, or the presence or absence of a DNA target. No numerical threshold is required.

Quantitative Criteria

A **numerical signal is converted into a class** by predefined rules or thresholds. Medical example: an immunoassay signal index is compared to a cutoff to yield reactive or non-reactive.

Same metrological question on both sides: **How often do we classify correctly, and how certain are we about that estimate?**

Validation and Routine QC Are Different Jobs

1. **Define the Property**
   What exactly is being classified? Establish the measurand and clinical scope.
2. **Develop the Method**
   Define the signal, classification criteria, cutoff, and workflow.
3. **Validate**
   Estimate false result rates and uncertainty before routine use.
4. **Test Under QC**
   Check each analytical run against the validated state.
5. **Report**
   State the class and, when useful, communicate confidence to the user.

Validation = evidence before routine use. QC = confidence during routine use. QC must challenge the same decision logic established during validation.

The Core Validation Model: The 2×2 Table

	Disease +	Disease −	Totals
Reported +	tp	fp	p
Reported −	fn	tn	n
Totals	pc	nc	N

Validation is not just "how many positives." It separates correct positives, missed positives, false alarms, and correct negatives.

Metrics Used Most Often

Clinical Sensitivity (SS)
tp / (tp + fn) — proportion of true positives correctly identified

Clinical Specificity (SP)
tn / (tn + fp) — proportion of true negatives correctly identified

False Positive Rate (FP)
fp / (fp + tn) — proportion of negatives incorrectly classified as positive

False Negative Rate (FN)
fn / (fn + tp) — proportion of positives missed by the method

Point Estimate vs. Confidence Interval

A rate without its uncertainty can be seriously misleading — especially for high-stakes screening.

	Sensitivity	95% CI (lower)	Cases tested
Case A	100.0%	57.0%	5 / 5
Case B	100.0%	99.0%	400 / 400

Both cases report 100% sensitivity. Only Case B provides enough evidence to claim the lower 95% confidence limit is also close to 100%.

What This Means in Practice

For high-impact screening examinations, AQA 2021 recommends setting targets not only for SS or SP, but also for the **lower 95% confidence limit (LLSS.95 and LLSP.95)**.

Small sample sizes produce wide confidence intervals that expose the weakness of a point estimate — a laboratory can achieve apparent perfection with only five samples.

The number of cases studied is as important as the result itself.

From Validation Data to Confidence in One Reported Result

Plain-language reading: **the same test can mean different things in different populations.** The pathway from raw rates to posterior probability is the metrological bridge between laboratory performance and clinical interpretation.

Simple Bayes

Same test performance. Different prevalence. Dramatically different posterior probability.

Maria — Island (Prevalence 1%)

Sensitivity (TP) = 81% | False Positive Rate (FP) = 0.5%

62.0%

Posterior probability of infection after a positive test result

Gabriel — Ship (Prevalence 7%)

Same rapid test, same analytical performance parameters.

92.4%

Posterior probability of infection after a positive test result

Main message: a positive result is never interpreted in a vacuum. **Prevalence fundamentally changes what "positive" means** for the individual patient.

9

Practical Case: Anti-HCV Antibody Examination

1. **Sample**
   Serum or plasma specimen
2. **Signal**
   Immunoassay response or index value
3. **Rule**
   Predefined cutoff and acceptance criteria
4. **Class**
   Reactive / Non-reactive / Equivocal
5. **Action**
   Report, repeat, or refer to follow-up testing

Binary output: **reactive / non-reactive**, with an equivocal zone managed by a predefined repeat or follow-up pathway.

10

Illustrative Validation Data for the Anti-HCV Case

	Disease +	Disease −	Totals
Reactive	196	2	198
Non-reactive	2	398	400
Totals	198	400	598

Clinical Sensitivity

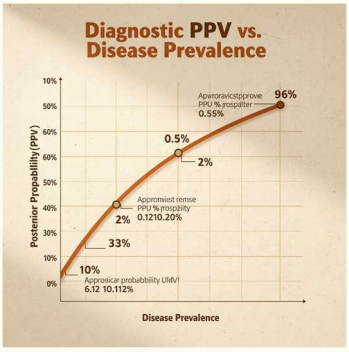
95% CI: 96.4–99.7%

Clinical Specificity

95% CI: 98.2–99.9%

If laboratory targets were **LLSS.95 ≥ 95%** and **LLSP.95 ≥ 95%**, this illustrative method passes both criteria.

Same Method, Different Prevalence, Different Meaning of "Reactive"



Reading the PPV Curve

Posterior probability connects method performance to the population being tested.
Using hypothetical validation figures: SS = 99.0%, SP = 99.5%.

The PPV curve illustrates why follow-up confirmatory testing is essential in low-prevalence populations.

0.5% Prevalence

PPV ≈ **49.9%** — a reactive result is barely better than a coin toss in this low-prevalence screening setting.

2% Prevalence

PPV ≈ **80.2%** — meaningful uplift, but one in five reactive results is still a false positive.

10% Prevalence

PPV ≈ **95.7%** — high confidence; method performance dominates over prevalence effect.

What Routine Quality Control Should Protect

Expected Classification

Do the control materials still yield the expected class? Systematic misclassification indicates method drift.

Threshold Stability

Is the decision threshold stable, especially near the cutoff or equivocal zone where classification risk is highest?

Change Management

After lot, reagent, or instrument changes, does performance still match the validated claim?

For anti-HCV screening, **near-cutoff control samples are essential.** If near-cutoff behavior shifts, false positive and false negative risk shifts with it. QC must challenge the decision point — not only strong positives and strong negatives.

Reporting: Say the Class, and When Useful, Say the Confidence

Minimal Report

> **Anti-HCV antibodies: Reactive**

States the classification clearly. Appropriate when the clinical context is well established. Does not communicate how much confidence to place in the result.

More Informative Report

> **Anti-HCV antibodies: Reactive** Illustrative method performance: SS 99.0% (95% CI 96.4–99.7) SP 99.5% (95% CI 98.2–99.9) Follow-up testing indicated.

Provides the metrological context needed for informed clinical action. Especially valuable when the result has significant consequences.

Choice of metric depends on the user: SS/SP, likelihood ratio, or posterior probability may all be defensible — **if clearly explained**.

Take-Home Messages

1. **Validation of qualitative analysis is validation of classification risk**
   Not only of analytical signal. The goal is to quantify how often the reported class is correct.

2. **The 2×2 table is the central model**
   SS, SP, FP, and FN are the basic language. All downstream metrics derive from this foundation.

3. **Confidence intervals matter**
   A point estimate without its uncertainty is weak evidence — especially for high-stakes screening decisions.

4. **Prevalence changes the meaning of a positive result**
   Bayesian reasoning is especially valuable in medical laboratories operating across different patient populations.

5. **Routine QC should protect the validated decision rule**
   Especially around the cutoff and equivocal zone — where misclassification risk is greatest.

Two Companion Leaflets Make the Ideas Easier

Should Gabriel visit Maria?

A short introduction to Bayes' Theorem

Maria and Gabriel are old friends from Porto Seguro, a small island in the middle of the Atlantic Ocean. Maria misses Gabriel, who joined the Navy for a long trip on the Sergas school ship. Maria is happy to have received a postcard from Gabriel, posted from Rio de Janeiro, informing her that he would visit her when the Sergas moored at Porto Seguro for three days on its way to Lisbon.

Oh no! COVID-19 outbreaks in Porto Seguro and on the Sergas, forcing both Maria and Gabriel to be tested for the disease. Both friends used the same rapid test, COVRAPID, and tested positive! The outbreak is stronger on the ship than on the island, with an estimated percentage of infected crew members and island population being 7 % and 1 %, respectively.

Gabriel sent an email to Maria saying that they could meet because they had both received positive results from the COVRAPID test, so the chance of them both actually having the disease should be the same.

Maria had just studied Bayes' theorem, which states that the probablity is related to the infection prevalence as well as the COVRAPID test performance (true positive rate, *TP*, or sensitivity, and false positive rate, *FP*). Unfortunately, COVID-19 is much more common on the ship than on the island.

Picture generated in Vello AI

Eurachem

A FOCUS FOR ANALYTICAL CHEMISTRY IN EUROPE

Performance Assessment of Binary Output Examinations in Medical Laboratories

1 - What is fitness for purpose?

The test meets the necessary performance criteria for a clinical binary examination **to ensure it is suitable for making accurate and reliable clinical decisions in the intended patient population.** Among others, performance criteria include clinical sensitivity and clinical specificity.

2 - How to assess clinical performance?

Assessing the clinical performance of a binary examination, stating results as 'positive' or 'negative,' involves calculating the examination's **Clinical Sensitivity and Clinical Specificity**. From the physician's point of view, clinical performance is evaluated by using predictive values, i.e., how likely is it that a positive or negative result corresponds to the presence or absence of the condition, which is also influenced by the prevalence of the condition in the population being tested.

3 - What is Clinical Sensitivity?

	Clinical Diagnosis		
	Disease, D_+	Non-Disease, D_-	Result Totals
Positive, *p*	*tp*	*fp*	*tp* + *fp*
Negative, *n*	*fn*	*tn*	*tn* + *fn*
Case Totals	*tp* + *fn*	*tn* + *fp*	*N*

Table 1. 2x2 Contingency table for clinical diagnosis.

Clinical sensitivity (*SS*), also known as the True Positive Rate, measures the proportion of true positives correctly identified by the test.

Data from a group of subjects who have been confirmed to have the disease or condition through diagnosis or a gold standard test (D_+ in Table 1) are needed to calculate clinical sensitivity.

The formula for clinical sensitivity is:

$SS = tp / (tp + fn)$, with:

tp, True Positives are the cases where the test correctly identifies the presence of the disease.

fn, False Negatives are the cases where the test incorrectly identifies the absence of the disease when it is actually present.

Table 1 represents the components of clinical sensitivity and specificity. Figure 1 illustrates an example of the overlapping of the number of components. The symbol "∩" represents the intersection of groups. For example, $D_+ \cap n$ denotes the set of negative results obtained for known positive cases, i.e., false negatives.

4 - What is Clinical Specificity?

Clinical Specificity (SP), also known as the True Negative Rate, measures the proportion of true negatives correctly identified by the test.

Data from a group of subjects who have been confirmed not to have the disease or condition through a gold standard test (D_- in Table 1) are needed to calculate clinical specificity.

Eurachem

A FOCUS FOR ANALYTICAL CHEMISTRY IN EUROPE

Eurachem
A focus for analytical chemistry in Europe

CITAC
Cooperation on international Traceability in Analytical Chemistry

Lisbon, 11 and 12 May 2026

Validation and Quality Control of Qualitative Analysis

EURACHEM/CITAC WORKSHOP ON QUALITY IN ANALYTICAL MEASUREMENTS

Paulo Pereira, Ph.D.,
Portuguese Institute of Blood and Transplantation

17