

# Vademecum EBLM in the clinical pathology laboratory

Eugenio Cerelli<sup>1</sup>, Simona Maria Pagano<sup>2</sup>, Antonio D’Addiego<sup>3</sup>, Virginia Nocito<sup>4</sup>, Eugenio Gautiero<sup>5</sup>, Pietro Del Fine<sup>6</sup>, Raffaele Ascione<sup>7</sup> and Vincenzo Palumbieri<sup>8\*</sup>

1. ASL Reggio Emilia;
2. Mater Dei Hospital - Malta;
3. ASL Teramo;
4. ASP Cosenza;
5. IRCCS San Gerardo dei Tintori – Monza;
6. AST Pesaro;
7. AOU Federico II – Napoli;
8. ASReM – Campobasso;

\* Correspondence: [vincenzopalumbieri@gmail.com](mailto:vincenzopalumbieri@gmail.com)

## KEYWORDS:

Evidence-Based Laboratory Medicine; Clinical pathology; Diagnostic performance; Analytical validation; Predictive values.

## ABSTRACT

Laboratory medicine plays a central role in clinical decision-making, yet the interpretation of analytical and diagnostic performance parameters is often inconsistent in both scientific literature and routine practice. Evidence-Based Laboratory Medicine (EBLM) provides a structured framework for evaluating the quality, clinical utility, and appropriate use of in vitro diagnostic tests. This vademecum offers a practical and integrated overview of the main concepts required to assess analytical and diagnostic performance in the clinical pathology laboratory. Particular attention is given to analytical sensitivity, selectivity, limits of detection and quantification, measurement range, robustness, accuracy, trueness, precision, measurement uncertainty, reference intervals, and Reference Change Value. The document also examines the diagnostic implications of cut-off selection, sensitivity, specificity, diagnostic accuracy, ROC curves, area under the curve, Youden Index, likelihood ratios, predictive values, pre-test and post-test probability, and the influence of disease prevalence. By clarifying the distinction between prevalence-independent and prevalence-dependent parameters, the article highlights the importance of integrating laboratory data with clinical context and Bayesian reasoning. Sequential and parallel testing strategies are also discussed as tools to improve diagnostic predictive value. Finally, a case study illustrates the selection of optimal cut-off values in quantitative tests through ROC and diagnostic performance curves. This vademecum aims to support laboratory professionals in critically appraising diagnostic evidence, improving test interpretation, and contributing more effectively to patient-centred clinical decision-making.



RESEARCH ARTICLE

| Acronym        | Meaning / Explanation                                |
|----------------|------------------------------------------------------|
| ACC            | Diagnostic accuracy of a test                        |
| AUC            | Area Under the Curve (area below the ROC curve)      |
| Avg            | Arithmetic mean of a dataset                         |
| CI             | Confidence Interval                                  |
| c/o (Cut-off)  | Threshold value used to classify results             |
| CPD-Cut        | Diagnostic Performance Curve by Cut-off              |
| CVa            | Analytical coefficient of variation                  |
| CVg            | Inter-individual biological coefficient of variation |
| CVi            | Intra-individual biological coefficient of variation |
| CVt            | Total coefficient of variation                       |
| RCV            | Reference Change Value                               |
| EBLM           | Evidence-Based Laboratory Medicine                   |
| FN             | False Negative                                       |
| FP             | False Positive                                       |
| IVD            | In vitro diagnostics                                 |
| IVD instrument | Instrumentation used for in vitro diagnostics        |



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).



RESEARCH ARTICLE



| Acronym            | Meaning / Explanation                                   |
|--------------------|---------------------------------------------------------|
| IVD medical device | Medical devices intended for in vitro diagnostic use    |
| IVDD               | In Vitro Diagnostic Directive (Directive 98/79/EC)      |
| IVDR               | In Vitro Diagnostic Regulation (EU Regulation 2017/746) |
| K index            | Kurtosis index                                          |
| LOQ                | Limit of Quantification                                 |
| LOD                | Limit of Detection                                      |
| LR+                | Positive Likelihood Ratio                               |
| LR-                | Negative Likelihood Ratio                               |
| Max                | Maximum value in a dataset                              |
| Min                | Minimum value in a dataset                              |
| MR                 | Measurement Range                                       |
| p                  | Statistical probability                                 |
| Perc               | Percentile                                              |
| p(pre)             | Pre-test probability                                    |
| p(post-)           | Post-test probability after a negative result           |
| p(post+)           | Post-test probability after a positive result           |
| Pr                 | Prevalence                                              |
| Ppre               | Pre-test probability                                    |
| ROC Curve          | Receiver Operating Characteristic curve                 |
| SD                 | Standard Deviation                                      |
| SEM                | Standard Error of the Mean                              |
| Se (TPR)           | Sensitivity / True Positive Rate                        |
| Sk index           | Skewness index                                          |
| Sp (TNR)           | Specificity / True Negative Rate                        |
| TE                 | Total analytical error (bias + imprecision)             |
| UML                | Upper Measurement Limit                                 |
| VR (RV)            | Reference Values                                        |
| VN                 | True Negative                                           |
| VP                 | True Positive                                           |
| NPV                | Negative Predictive Value                               |
| PPV                | Positive Predictive Value                               |
| YI                 | Youden Index                                            |
| Ω                  | Statistical odds                                        |

## INTRODUCTION

Despite the central role of laboratory testing in clinical decision-making, the interpretation of analytical and diagnostic performance parameters is still often inconsistent in both literature and everyday practice. Sensitivity, specificity, predictive values and likelihood ratios are frequently used outside their proper conceptual framework, which can lead to misinterpretation of results and, in some cases, inappropriate clinical decisions. These are not only theoretical issues, but problems with real consequences in pathology-driven diagnostic workflows.

The aim of this article is not to provide a technical manual or a set of procedural recommendations. Instead, it offers a critical and integrated discussion of key Evidence-Based Laboratory Medicine (EBLM) concepts, focusing on the most common interpretative and methodological pitfalls encountered when evaluating analytical and diagnostic performance. The goal is to help bridge the gap between analytical validation, diagnostic accuracy studies and real-world clinical reasoning in laboratory medicine and pathology.

Within the European Union (EU), the responsibility for ensuring the high-quality standards required for the marketing of in vitro diagnostic (IVD) devices is shared among manufacturers, authorized representatives, importers, and distributors. Each of these actors is required to comply with the provisions set by Regulation (EU) 2017/746 on In Vitro Diagnostic Medical Devices (IVDR), which replaced by Directive 98/79/EC (IVDD – In Vitro Diagnostic Medical Devices Directive). The Regulation has been progressively implemented from 26 May 2017 to 26 May 2022, with an extended transition period introduced to address emerging critical issues.

Clinical laboratories, as end-users, are responsible for using each device in accordance with its intended purpose—namely the clinical indication, target population, authorized operators, operating conditions, and stated limitations. This responsibility includes understanding the warnings and instructions described in the device manual and verifying, within the laboratory's own setting, that the analytical and diagnostic performance claimed by the manufacturer is met. Whenever a device is demonstrably “unfit for its intended use,” or when incidents or near-incidents attributable to the device (including embedded software) occur, the device must be withdrawn from use, and the laboratory must activate the established vigilance procedure for in vitro medical devices.

The responsibility for the correct technical and clinical use of IVD tests therefore lies with the laboratory, which must comply with the mandatory regional accreditation requirements and may additionally adhere to the international standard ISO 15189:2022. This obligation entails that the ability to appraise scientific documentation be an integral competency for every biomedical professional, particularly the skills needed to critically evaluate

the literature, systematic reviews, and meta-analytical data arising from studies that assess analytical and diagnostic performance.

Evidence-Based Laboratory Medicine (EBLM), a specialized branch of Evidence-Based Medicine (EBM), focuses on evaluating the quality and clinical utility of in vitro diagnostics to ensure their effective and efficient use in practice. Its scope is broad and includes population-level screening, differential diagnosis, precision and gender-specific medicine, as well as the monitoring of disease progression, therapeutic response, and risk in clinical trials.

Approximately 70% of clinical decisions are supported by in vitro diagnostics, even though laboratory testing accounts for less than 1% of total healthcare expenditure in Italy. This disproportion underscores the importance of robust evidence to guide the introduction and appropriate use of diagnostic tests.

The evaluation of a new diagnostic test follows a hierarchical framework, where each level—addressing both effectiveness and efficiency—must be validated, preferably through standardized procedures, before progressing to the next. Importantly, a test should be introduced into clinical practice only if it provides a tangible improvement to the overall diagnostic–therapeutic pathway.

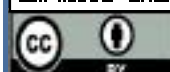
Whether the aim is to assess a new analyte proposed as an alternative to a traditional marker—for example, serum cystatin C in place of, or alongside, creatinine for the evaluation of renal function—to compare different analytical systems (such as high-throughput multichannel analysers versus smaller platforms), or to evaluate alternative analytical methods (for instance, a chemical versus an enzymatic assay), the focus remains on analytical performance, diagnostic yield, and the impact on patient outcomes. In the long term, improvements in organizational efficiency and healthcare costs are also expected.

Each of these interconnected perspectives requires robust supporting evidence; identifying and appraising such evidence lies at the core of EBLM.

In summary, the overall evaluation process can be structured into five phases:

1. Verification of technical-analytical performance
2. Verification of diagnostic performance
3. Assessment of clinical impact (changes in diagnostic or therapeutic strategy, clinical outcomes, etc.)
4. Assessment of organizational impact (changes in patient management, hospital stay, workflow, etc.)
5. Assessment of economic impact

Laboratory professionals have long mastered Total Quality Management (TQM) principles and are generally proficient in improving analytical quality and diagnostic performance. However, the evidence required for phases 3, 4, and 5 is often evaluated less extensively in laboratory settings. These pha-



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

ses involve clinical practice guidelines (analogous to process maps or standard operating procedures), outcome measures (quantitative indicators of treatment performance, like the role of QC in analytical systems), critical appraised topics (CATs), and cost-effectiveness and appropriateness studies.

This document aims to serve as a practical reference for laboratory professionals, providing appropriate terminology and the key theoretical and practical concepts needed to evaluate the essential parameters that define the analytical and diagnostic performance of an IVD test. Tests that directly assess genetic material (chromosomes, DNA, RNA) and microbiological or virological culture-based assays are excluded, as both the material analysed, and the techniques used require additional and distinct evaluation approaches

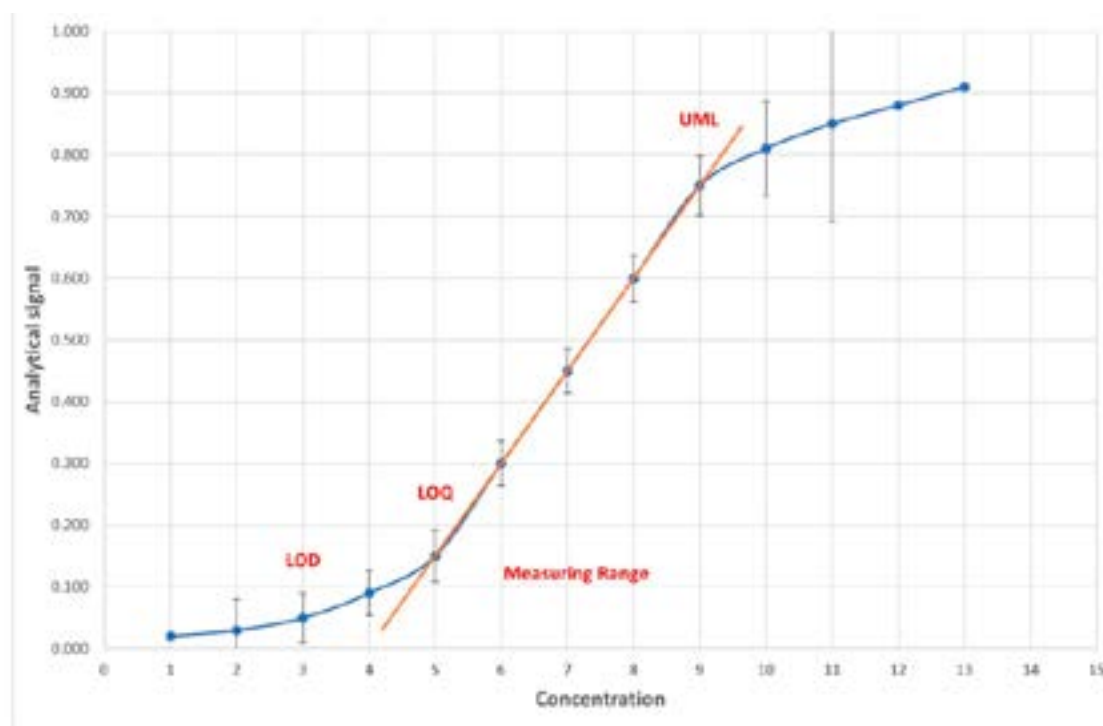
#### ANALYTICAL PERFORMANCE PARAMETERS

Analytical performance parameters describe the chemical and physical characteristics of a diagnostic test. They define the ability of an analytical system to detect and quantify a target substance in complex biological matrices through chemical, enzymatic, immunological, or other types of reactions, and

to convert the resulting detection signal (e.g., color, fluorescence, radioactivity) into an electrical value proportional to the concentration of the analyte, regardless of the measurement unit employed (amount of substance in  $\mu\text{mol}$ , mass in grams, biological activity in International Units, enzymatic activity in Arbitrary Units, reaction time in seconds, and so forth). These parameters apply primarily to quantitative tests, which generate continuous numerical results without a clear discontinuity between healthy and diseased populations. Quantitative assays are crucial both in diagnostic decision-making and in patient monitoring. For this reason, analytical performance—which constitutes an essential prerequisite for any diagnostic evaluation – must be assessed and validated with high methodological rigor.

A central aspect of quantitative testing is the full evaluation of the calibration function (calibration curve)

(Figure 1). This function is typically represented as a straight line or a curve on a Cartesian plot, where the x-axis shows the analyte concentration and the y-axis displays the analytical signal produced by one or more calibrators (e.g., absorbance, fluorescence, turbidity, radioactivity, reaction time).



**(Figure 1).** Example of a calibration curve. The vertical error bars represent  $\pm 3$  standard deviations (SD) of the signal at each concentration level. At the Limit of Detection (LOD), the lower end of the error bar overlaps with the “zero” signal (reagent blank or sample blank). Within the linear range (from the LOQ to the UML), the analytical sensitivity is constant and corresponds to the slope of the calibration line (angular coefficient  $b$  in the equation  $y = a + bx$ ). In the curvilinear portion of the curve, the analytical sensitivity varies at each concentration level.



In contrast, when evaluating dichotomous qualitative tests—which distinguish only two nominal categories (e.g., negative/positive, healthy/diseased)—and semiquantitative tests, which rely on ordered categories with an implicit direction (from lower to higher, or from better to worse), and are typically used in screening settings, manufacturers tend to focus primarily on diagnostic performance parameters. It is important to note, however, that several qualitative assays (for example, ELISA or CMIA screening tests used in transfusion centres) generate an underlying continuous signal that is subsequently converted into a binary or ordinal result through predefined cut-off values. For such systems, certain characteristics typical of quantitative methods—such as the limit of detection—can be validated in a similar manner. Furthermore, the cut-off can be optimised to maximise either sensitivity or specificity, depending on the intended clinical application. The probability of obtaining results that truly reflect the patient’s clinical status (positive results in diseased individuals and negative results in healthy ones) therefore depends closely on the analytical design of the assay. Therefore, the performance of different tests targeting the same analyte may vary considerably across manufacturers and may improve substantially with technological advancements, leading to succes-

sive “generations” of assays. A relevant example is the evolution of several routine tests into assays with “high analytical sensitivity,” such as hs-CRP or high-sensitivity troponins. These platforms can detect very low analyte concentrations with high accuracy and precision, allowing clinicians to identify small variations in circulating levels and thus stratify disease risk even within the low-concentration range. This technological progress enables, for instance, a substantial reduction in the time required to assess troponin kinetics in suspected acute myocardial infarction, leading to earlier diagnosis. Similarly, high-sensitivity CRP can be used not only as a non-specific inflammatory marker but also to distinguish, among hospitalised elderly patients with iron deficiency, those who will respond to oral iron therapy from those who will not. Ultimately, analytical performance parameters have a direct impact on the quality of the diagnostic process. Depending on the clinical context and the intended purpose, it may be appropriate to use tests with differing analytical performances to measure the same analyte, provided that their operational limitations and any restrictions on use are carefully considered.

Analytical Performance Parameters table

| Parameter (abbreviation)                | Definition                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Method of assessment and/or calculation                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-----------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sensitivity [analytical]<br>Sensitivity | Indicates the rate at which the detector signal changes for each increase in concentration. It does not apply to qualitative tests.<br>It represents the slope (trigonometric tangent) of the calibration curve.<br>If the calibration function is linear (first-degree polynomial), analytical sensitivity is constant over the entire concentration range and corresponds to the slope of the line (b). If the function is curvilinear, sensitivity varies with concentration. | The analytical sensitivity is the first derivative of the mathematical function linking signal and concentration:<br>$y = a + bx \rightarrow y' = b$<br>$y = a + bx + cx^2 \rightarrow y' = b + 2cx$<br>$y = a + bx + cx^2 + dx^3 \rightarrow y' = b + x \cdot (2c + 3dx)$                                                                                                                                                                                                                              |
| Selectivity<br>Selectivity              | Ability of the test to measure only the molecule/substance of interest, without being affected by interfering substances or other components different from the analyte under investigation (sample matrix composition).<br>NB: the use of the term “specificity” is not recommended by IUPAC, and this term is not defined in VIM 3.                                                                                                                                            | Comparison studies with a method based on a different and reliable physicochemical principle.<br>Cross-reactivity studies, performed by adding to real samples molecules that are similar or potentially interfering with the method principle, reaction kinetics, or signal generation (mainly drugs, foods, and substances of toxicological interest).<br>In spectrophotometric tests, evaluation of the matrix effect for each sample using the serum index for haemoglobin, bilirubin, and lipemia. |





|                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|---------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Limit of detection (LOD)              | <p>The lowest amount of substance that can be distinguished from instrumental background noise or from the blank sample with a defined probability (usually <math>p = 0.01</math>).</p> <p>A blank sample is a matrix similar to the sample of interest but without the substance to be measured.</p> <p>LOD is important for characterizing the method at its lower concentration limit, even if its clinical usefulness may sometimes be negligible when far below biological lower limits.</p>                                        | <p>The LOD is not easy to determine and several criteria may be used:</p> <p>Signal-to-noise criterion: concentration corresponding to the mean background signal + 3 standard deviations.</p> <p>Limit variance criterion: LOD corresponds to the lowest concentration with a CV% of 33.3%.</p> <p>Calibration curve criterion: extrapolation toward zero of the calibration line, taking into account analytical signal error.</p> |
| Limit of quantification (LOQ)         | <p>The lowest amount of substance that can be measured with an analytical error compatible with predefined goals for bias, precision, and total error, as established by consensus at the First Strategic Conference of the EFLM, Milan, 2014.</p> <p>Depending on the defined objective, LOQ may coincide with LOD but can never be lower.</p> <p>A signal between LOD and LOQ indicates that the analyte is detectable but not reliably quantifiable.</p> <p>(LOQ is sometimes also referred to as the “Lower Measurement Limit”).</p> | <p>Empirical criterion (recommended): identification of the lowest measurable concentration with acceptable precision using the precision profile.</p> <p>Theoretical criterion (not recommended): concentration corresponding to 10 times the standard deviation of the mean background signal (approximately 3.3 times the LOD calculated by the signal-to-noise method).</p>                                                      |
| Upper measurement limit (UML)         | <p>Highest concentration at which the analytical signal begins to deviate from the calibration line/curve.</p> <p>After empirical verification of the dilution effect on the matrix, samples with concentrations higher than the UML may be analyzed after appropriate dilution.</p>                                                                                                                                                                                                                                                     | <p>It may correspond to the highest concentration used to construct the calibration curve or, after empirical verification, to a higher concentration extrapolated from it.</p>                                                                                                                                                                                                                                                      |
| Measurement range (MR)                | <p>The minimum (LOQ) and maximum (UML) concentrations between which analytical signal values can be interpolated using the calibration line/curve.</p>                                                                                                                                                                                                                                                                                                                                                                                   | <p>MR = from LOQ to UML</p>                                                                                                                                                                                                                                                                                                                                                                                                          |
| Robustness (of the analytical method) | <p>Insensitivity of the method to small deviations from standard analytical conditions.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                              | <p>Robustness testing by intentionally introducing errors into the procedure, such as carryover from high samples; instrumental drift over time; small variations in temperature, incubation times, and environmental conditions; evaluation of possible hook effect in ELISA tests at high antigen concentrations.</p>                                                                                                              |

Accuracy, Trueness, and Precision

Accuracy, trueness, and precision are the analytical parameters that most effectively define the performance of a method. Together, they quantify the intrinsic uncertainty of any measurement and ultimately determine the reliability of a result for its intended clinical use.

The accuracy of a laboratory result—its closeness to the “true” value—is inherently uncertain. It depends on the ability of the measurement system (the

combined contribution of human competence, technological instrumentation, and chemical–physical methodology) to be both true and precise.

Trueness describes how closely the measurement approaches the true or reference value of the analyte, while precision reflects the closeness of repeated measurements to each other (Figure 2).

Only when both trueness and precision are satisfactory can a method be considered analytically accurate.



**Figure 2.** Graphical representation of the two components of accuracy: trueness and precision. Adapted from Dr. Jan-Frederik Gürh, University of Munich).

Every laboratory result should ideally be reported together with two key elements: the measured value and its associated measurement uncertainty. The latter—which reflects both inaccuracy and imprecision—represents the inherent error of the laboratory result (without implying any negative connotation). It is expressed using the same unit of measurement as the result itself or as a percentage relative to the measured value.

The expanded uncertainty defines the interval within which the true value of the measurand lies with a predefined level of confidence, typically 95%. The smaller this uncertainty, the more reliable the result. Knowing the expanded uncertainty would allow clinicians to better assess the trustworthiness of a measurement, particularly when comparing it

with reference limits, legal thresholds, or previous values, thereby supporting more informed clinical decision-making.

**Influence of Precision on Reference Values**

Reference intervals or reference values (RV) represent the expected distribution of results in a carefully selected “healthy” population, characterised by specific demographic, genetic, and environmental attributes (such as age, sex, diet, lifestyle, absence of defined pathological conditions, chronobiological factors, etc.). Clinicians use these intervals to interpret patient results obtained from individuals whose characteristics are broadly comparable to those of the reference population. For this reason, they are often—though imprecisely—described as “normal values.”

Measurement Accuracy, Trueness, and Precision table

| Parameter (abbreviation)                           | Definition                                                                                                                                                                                                                                                                                                                                              | Method of assessment and/or calculation                                                                                                                                                                                                                                                                                                                                  |
|----------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Measurement accuracy<br>(of the result)            | Degree of agreement between a measured value and the true value of a measurand (VIM 2.13).<br>Measurement accuracy describes how close a single measurement result is to the true value (VIM 2.11). Accuracy therefore includes the effect of both precision and trueness.                                                                              | Accuracy cannot be assigned a numerical value; measurement results are described as being “more accurate” when measurement errors, and therefore measurement uncertainty, are smaller.<br>Measurement accuracy cannot provide a quantitative indication of the reliability of measurement results. For this purpose, an estimate of measurement uncertainty is required. |
| Measurement trueness<br>(of the analytical method) | Degree of agreement between the average of an infinite number of measured values and a reference value (VIM 2.14).<br>Measurement trueness expresses the hypothetical ability of a measurement procedure to produce results close to expected reference values for that quantity, such as the value of a certified reference material (CRM) (VIM 5.14). | Trueness is not a quantity and therefore cannot be expressed numerically. However, trueness varies inversely with systematic error (VIM 2.17), which can be evaluated as the mean measurement deviation (bias) relative to a reference value (VIM 2.18).                                                                                                                 |



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Measurement precision  
(of the analytical method)

Dispersion of repeated measurements on the same sample around the mean value (Type A uncertainty, obtained by statistical methods from experimental data; this differs from non-statistical Type B methods based on manufacturers' data for instruments and calibrators or on prior experience).

VIM 3 defines three measurement conditions for precision: repeatability conditions (VIM 2.20), intermediate precision conditions (VIM 2.22), and reproducibility conditions (VIM 2.24).

Typically, dispersion across the measurement range varies as a function of concentration. It is therefore important to assess the "precision profile" (Figure 3) of each quantitative method, evaluating how the standard deviation and CV% vary within the established measurement range (MR), with particular attention to extreme values near the LOQ and the UML.

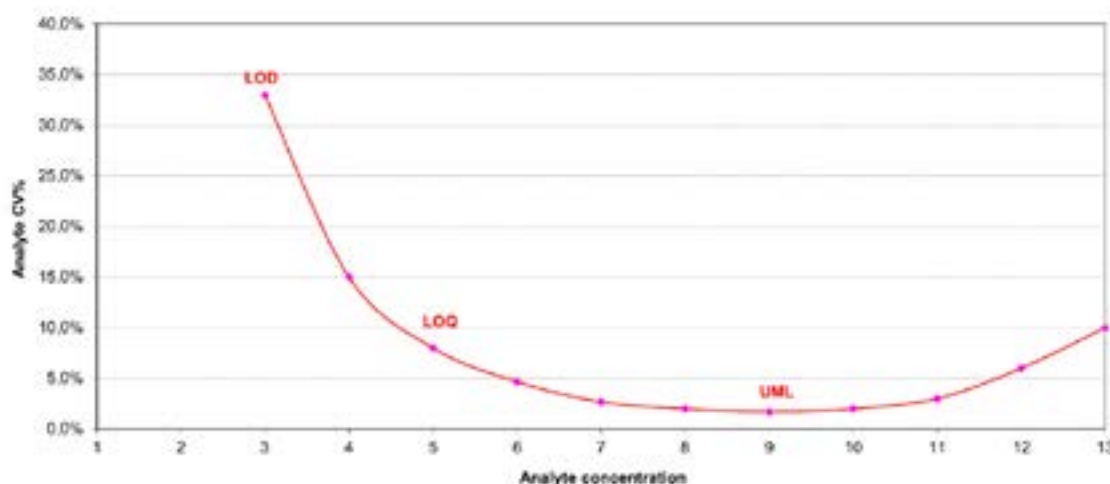


**Figure 3.** Example of a precision profile corresponding to the calibration curve shown in Figure 1. The LOQ is positioned at the minimum acceptable analytical target (in this example around 8%), and across the entire linearity range (from LOQ to UML) the precision remains better than this minimum target. For simplicity, the LOD is placed at a CV of approximately one-third of 100%. Any CV value greater than 33.3% prevents distinguishing, with 99% statistical confidence, background noise (absence of analyte) from the signal of any sample, because the lower confidence limit at  $p = 0.01$  for any result  $x$  would equal zero:  $x - (3 \times 0.333 \times x) = 0$

Three fundamental considerations are essential when interpreting reference intervals:

1. **Inter-laboratory variability:** A result considered "normal" in one laboratory may be judged abnormal in another. This means that all clinical interpretations must rely on the reference intervals provided by the same laboratory that performed the analysis. Despite high levels of trueness and precision in modern analytical methods, variability persists across laboratories due to differences in reference materials, measurement principles, reagent formulations, analytical platforms, and calibration procedures. Although international harmonisation efforts are ongoing, universally applicable reference intervals are not yet available for many laboratory tests.
2. **Interpretation of "normal" results:** A value within the reference interval does not necessarily indicate the absence of disease. For most laboratory tests, the distributions of results in healthy and diseased individuals partially overlap. When the Harris index of individuality (the

ratio of intra-individual to inter-individual biological variation) is low ( $K < 0.5$ ), an individual's results may remain within the reference limits even in the presence of early or mild pathology.

3. **Interpretation of "abnormal" results:** A value outside the reference interval is not automatically indicative of disease. Reference intervals are established through statistical analyses performed on healthy populations under standardised conditions; therefore, some healthy individuals may naturally fall outside these limits, particularly when values lie close to the boundaries (2.5th and 97.5th percentiles).

An additional, less frequently examined aspect concerns the effect of analytical precision on the determination of reference intervals. These intervals depend on both the biological variability between individuals (VB), which cannot be removed, and the analytical variability (Va), which is intrinsic to the measurement system in use. Although analytical variability cannot be eliminated, it can be kept within acceptable limits to ensure appropriate clinical application—this is the basis of the "analytical perfor-



mance goal” or “analytical target.”

According to the law of error propagation, the square root of the sum of the variances of these two independent components determines the total variability ( $V_t$ ), which in turn defines the width of the reference interval (Figure 4). Unfortunately, many studies on biological variability do not report the analytical imprecision, making comparisons across laboratories that use different analytical systems difficult.

When the sources of variability are expressed as coefficients of variation (CV), the total CV can be calculated as follows:

$$CV_t = (CV_a^2 + CV_g^2)^{1/2}$$

**Note.** The total variance should ideally include not only the analytical variance and the biological variance, but also the pre-analytical variance—namely, all factors preceding the actual analytical process that may increase overall variability. These include patient preparation, timing and method of sample collection, sample type and container, post-collection handling, environmental conditions, and other similar influences.

When these variables are properly controlled, their effect can be considered negligible or, in practical terms, included within the analytical variance.

### Impact of the Pre-analytical Phase on Analytical Performance

The pre-analytical phase represents the main source of error in the clinical laboratory and may directly affect the reliability of the final analytical result. Before the measurement itself is performed, several variables can modify the concentration, stability, detectability or interpretability of the measurand. These variables include patient preparation, fasting

status, posture, physical activity, timing of collection, circadian variation, drug intake, venous stasis, sample type, tube selection, anticoagulant or additive, order of draw, sample volume, mixing, clotting time, centrifugation, transport temperature, storage time and possible haemolysis, icterus or lipaemia.

From an EBLM perspective, sample quality is therefore an essential component of analytical performance. Even when a method shows adequate trueness, precision and measurement uncertainty under controlled conditions, an unsuitable or unstable sample may generate a result that is analytically measurable but clinically misleading. Pre-analytical variability may widen the overall uncertainty associated with the result, increase the probability of false positive or false negative classifications around decision limits, and reduce the comparability of serial measurements used for monitoring or for the calculation of the Reference Change Value.

For this reason, verification of analytical performance should be accompanied by procedures aimed at controlling pre-analytical quality: clear patient preparation instructions, standardised collection and transport protocols, defined sample acceptance and rejection criteria, monitoring of haemolysis/icterus/lipaemia indices where applicable, documentation of non-conformities, and communication with clinicians when sample quality may affect interpretation. In this sense, the pre-analytical phase is not merely an operational prerequisite but a determinant of the clinical validity of the laboratory information provided to the patient-care pathway.

In summary, reference intervals are essential tools in clinical practice, yet their interpretation must always be contextualised by considering the characteristics of the reference population, the individual patient, the analytical performance of the measurement system, and the inherent limitations of the methods employed.

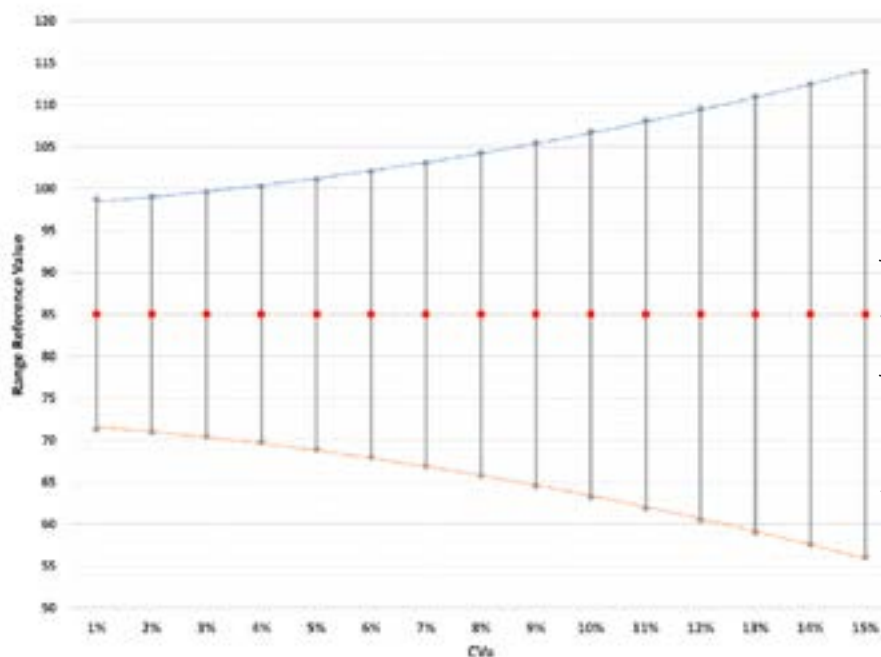


**Figure 4.** Illustrative simulation of how increasing analytical imprecision affects the width of the reference interval for plasma glucose (mean value fixed at 85 mg/dL; biological CV for glucose = 8.0%). As the analytical CV increases, the reference interval widens substantially—from approximately 71–99 mg/dL at low analytical variability to about 56–114 mg/dL at higher variability—thereby markedly complicating clinical interpretation.



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## Influence of Precision on the Reference Change Value

Analytical precision also affects the calculation of the Reference Change Value (RCV), which is essential in follow-up assessment to determine whether the difference between two results obtained at different times in the same individual is statistically significant (Figure 5).

The RCV between two measurements corresponds to twice the standard error of the mean ( $\pm$ SEM), multiplied by the z-value associated with the selected confidence level.

$$RCV = z \times \sqrt{2} \times \sqrt{(CVA^2 + CVi^2)} \quad \text{for } p = 0.05$$

$$\text{and } z = 1.96 \quad RCV = 2.77 \times \sqrt{(CVA^2 + CVi^2)}$$

The example below refers to the measurement of plasma glucose, for which the median intra-individual biological variability (CVi) is 4.7%.

In other words, when the analytical CV is no greater than half of the biological CV (2.35% in this example, indicated by the orange arrow), the additional contribution of analytical variability to the total variability remains modest—only an 11.8% increase (from 4.7%, due solely to CVi, to 5.3%). Beyond this point, the total variability increases more rapidly. This is why an analytical CV not exceeding half of the intra-individual biological CV is generally considered acceptable.

## Premise to Diagnostic Performance Parameters: The Cut-off

The probability that a test will generate clinically meaningful information—distinguishing diseased from healthy individuals, assessing treatment effectiveness, and so on—depends primarily on how the

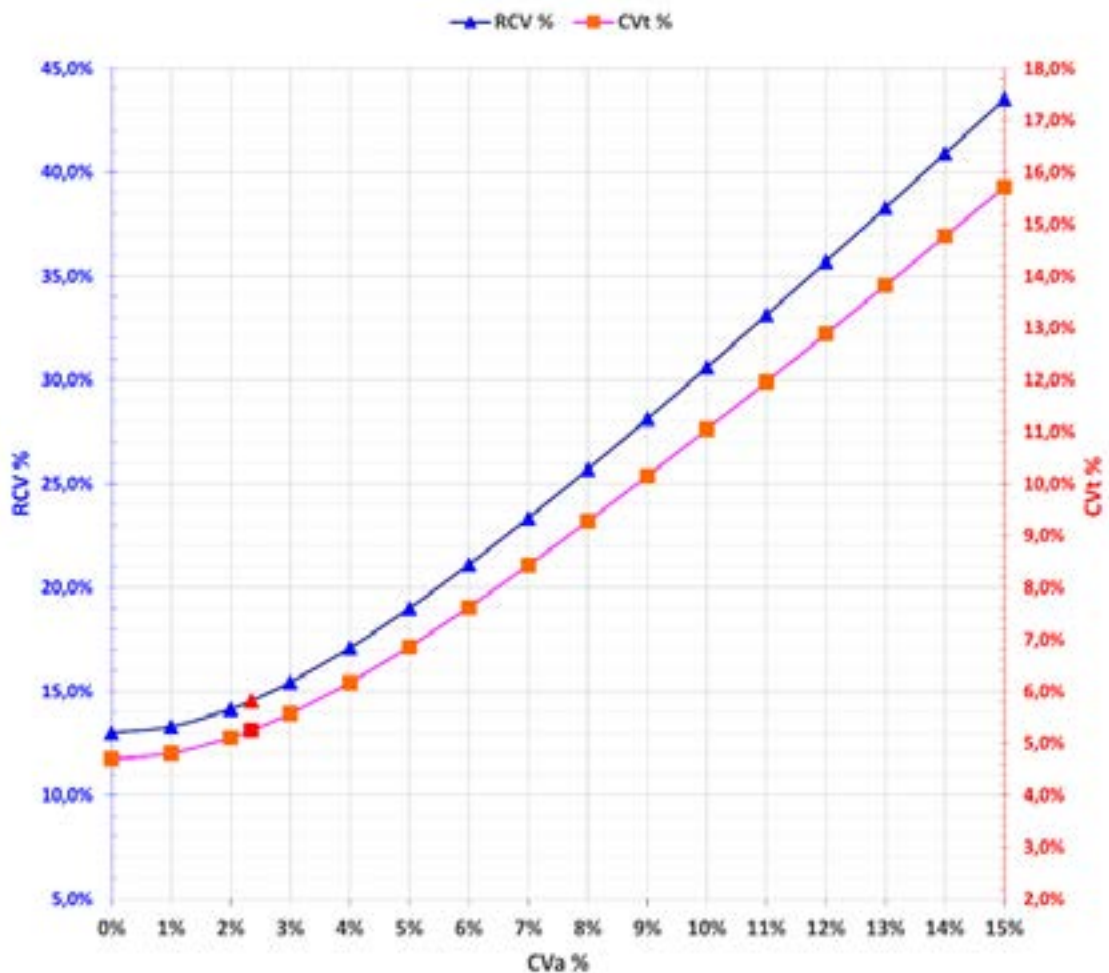


**Figure 5.** Variation of the Reference Change Value (RCV) and total coefficient of variation (CVt) as a function of analytical imprecision. With a constant intra-individual biological variability (CVi) of 4.7%, each increase in analytical variability (CVA) ranging from 0 to 15% affects both parameters illustrated:

- the blue line (triangles, left axis) shows the RCV required for a change between two results to be statistically significant;
- the red line (squares, right axis) shows the corresponding rise in total variability (CVt), which increases from 4.7% to 15.7%.

test is designed, the fundamental mechanism through which it operates, and the nature of what it measures. In other words, diagnostic usefulness is rooted in the test’s intrinsic analytical performance and in how reliably it is executed in the laboratory. That said, understanding the parameters that quantify diagnostic performance is essential for using any test appropriately in real clinical settings. It also enables an objective comparison of tests that may appear similar but differ in manufacturing, technology, or analytical characteristics. Regardless of whether a test is qualitative, semi-quantitative, or quantitative, a cut-off value must be established. This threshold defines the boundary between healthy and diseased individuals, or between improvement and deterioration, depending on the clinical context.

Diagnostic Cut-off – Table

| Parameter     | Definition (EN)                                                                                                                                                                                                                                                                                                 | Study / Calculation Method (EN)                                                                                                                                                                                                                      |
|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Cut-off (c/o) | In quantitative tests, the cut-off is the concentration that determines a medical decision in a specific clinical context. In qualitative tests, it is the analytical or sensory signal below which the result is considered negative and above which it is considered positive for a given clinical condition. | Statistical evaluation of data from healthy and diseased populations; ROC and CPD Cut analyses; efficiency analysis; clinical cost effectiveness assessment; social cost benefit evaluation; mandatory legal thresholds (e.g., forensic toxicology). |

In qualitative tests, the cut-off is determined by the test design itself and specified by the manufacturer in accordance with the EU Regulation on in vitro diagnostic medical devices (IVDR 2017/746), with particular attention to higher-risk tests (Class D). These assays may have a substantial impact on patient health and must therefore comply with very strict Common Specifications, whose conformity must be verified by a European Union Reference Laboratory. Such highly “sensitive” tests are used for applications with major clinical implications, including blood-group typing and the detection of infectious agents such as hepatitis viruses, HIV, SARS-CoV-2, and other serious conditions. In contrast, for quantitative tests—once reference values have been locally validated in healthy individuals—the optimal cut-off may be selected by the laboratory according to the specific clinical and operational context, based on the medical objectives to be achieved. This choice is guided by an understanding of the natural history of the disease and the clinical and economic consequences of incorrect classification: missed diagnoses leading to lack of necessary treatment, or, conversely, over-diagnosis with unnecessary and potentially harmful interventions.

In summary (if higher values indicate deterioration or disease):

- Lowering the cut-off increases sensitivity at the expense of specificity (fewer false negatives and more true positives, but also more false positives);
- Raising the cut-off increases specificity at the expense of sensitivity (fewer false positives and more true negatives, but also more false negatives)

As a rule, a highly sensitive test is preferable when failing to detect a disease would have serious consequences—such as in epidemic screening or situations where missed or delayed diagnoses pose significant risks. Conversely, a highly specific test is preferable when a false-positive result could expose the patient to harm, for example by triggering unnecessary risky treatments or invasive procedures.

Diagnostic Performance Parameters Independent of Prevalence

Prevalence-independent parameters are extremely useful for comparing analytical methods with one another, but they are less informative from a diagnostic standpoint. Except in the rare case of a test that provides perfectly correct classifications (Se and/or Sp = 1), these parameters do not indicate the probability that, within a given population, a positive result truly corresponds to a diseased individual or a negative result to a healthy one. In other words, they have no predictive value. Their calculation relies on a 2 × 2 contingency table (Table 1), in which individuals classified as “healthy” or “diseased” (based on clinical and instrumental gold-standard criteria independent of the test being evaluated) are further subdivided into “negative” and “positive” according to the test result and the selected cut-off. In dichotomous qualitative tests, the cut-off is established by the manufacturer and reported in the technical specifications of the assay. In quantitative tests, which yield a continuous numerical result, the cut-off may be selected by the user according to the specific clinical objective, even independently of the reference values provided by the manufacturer.



RESEARCH ARTICLE



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).



Table 1. Distribution of Test Results

|               | Diseased Subjects           | Healthy Subjects           | Total                                             |
|---------------|-----------------------------|----------------------------|---------------------------------------------------|
| Positive test | TP                          | FP                         | TP + FP<br>(total positive tests)                 |
| Negative test | FN                          | TN                         | FN + TN<br>(total negative tests)                 |
| Total         | TP + FN<br>(total diseased) | FP + TN<br>(total healthy) | TP + FP + FN + TN<br>(total individuals examined) |

Table 1. TP = True Positives; FP = False Positives; FN = False Negatives; TN = True Negatives.

Diagnostic Parameters table

| Parameter (abbreviation)                                | Definition                                                                                                                                          | Method of assessment and/or calculation                                                                        |
|---------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|
| Diagnostic Sensitivity (Se)<br>True Positive Rate (TPR) | Ability of the test to provide positive results in diseased sub-jects; proportion of true posi-tives (TP) over all diseased indi-viduals (TP + FN). | In subjects confirmed to have the disease: Se = TP / (TP + FN) Desired value = 1 (when FN = 0)                 |
| Diagnostic Specificity (Sp)<br>True Negative Rate (TNR) | Ability of the test to provide negative results in healthy sub-jects; proportion of true nega-tives (TN) over all healthy indi-viduals (TN + FP).   | In healthy subjects or those with con-founding conditions: Sp = TN / (TN + FP) Desired value = 1 (when FP = 0) |
| Diagnostic Accuracy (ACC)                               | Number of correct results (TP + TN) over the total number of tests performed.                                                                       | ACC = (TP + TN) / (TP + FP + TN + FN) Desired value = 1 (when FP = 0 and FN = 0)                               |

ROC Curve – Receiver Operating Characteristic Curve (ROC)

| Parameter (abbreviation)                                                                                                                       | Definition                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Method of assessment and/or calculation                                                                                                                                                                                                                                                                                                                                                                            |
|------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ROC Curve Receiver Operating Characteristic Curve (ROC) (The name originates from radar signal detection terminology used during World War II) | The ROC curve is a two-dimensio-nal plot that relates Sensitivity and Specificity for a quantitative test as the cut-off value changes. Each point on the curve corresponds to a specific combination of Se and Sp associ-ated with a given threshold. The ROC curve provides a compact visual summary of the global diagnostic performance of a method, independently of dis-ease prevalence. By comparing ROC curves, it is possible to evaluate different diagnostic methods. The ROC curve must be inter-preted carefully, especially when class sizes are unbalanced. It does not directly indicate clin-ically opti-mal cut-offs. This limita-tion may be overcome by pairing the ROC curve with a CPD-Cut (Diagnostic Performance Curve by Cut-off). | X-axis: 1 – Sp [1 – (TN / (FP + TN))] (false-positive rate among healthy indi-viduals)<br>Y-axis: Se [TP / (TP + FN)] (true-positive rate among dis-eased individuals) (Figure 6)<br>An ideal model rises vertically to the point (0,1) and then runs horizontally, indicating perfect Sensitivity and Specifi-city. Real ROC curves deviate toward the diagonal depending on model quality and data distribution. |



|                                                    |                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                         |
|----------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Area Under the Curve<br>Area Under the Curve (AUC) | The area under the ROC curve provides a numerical measure of a test's overall ability to distinguish between healthy and diseased individuals across all possible cut-off values.<br>It represents the probability (from 0.5 to 1) that a diseased individual will have a higher (more positive) test result compared to a healthy individual. | The AUC is calculated by integrating—graphically or mathematically—the area under the ROC curve.<br>Maximum value = 1.0 (perfect test)<br>Minimum value = 0.5 (no discrimination, equivalent to random classification). |
| Youden Index<br>Youden Index (YI)                  | The Youden Index identifies the point on the ROC curve that lies furthest from the diagonal and minimises the combined number of FP and FN. It conceptually reflects the same idea as diagnostic accuracy.                                                                                                                                     | The Youden Index can be assessed graphically or computed using:<br>$YI = \max [Se - (1 - Sp)]$<br>The highest value determines the optimal cut-off.                                                                     |

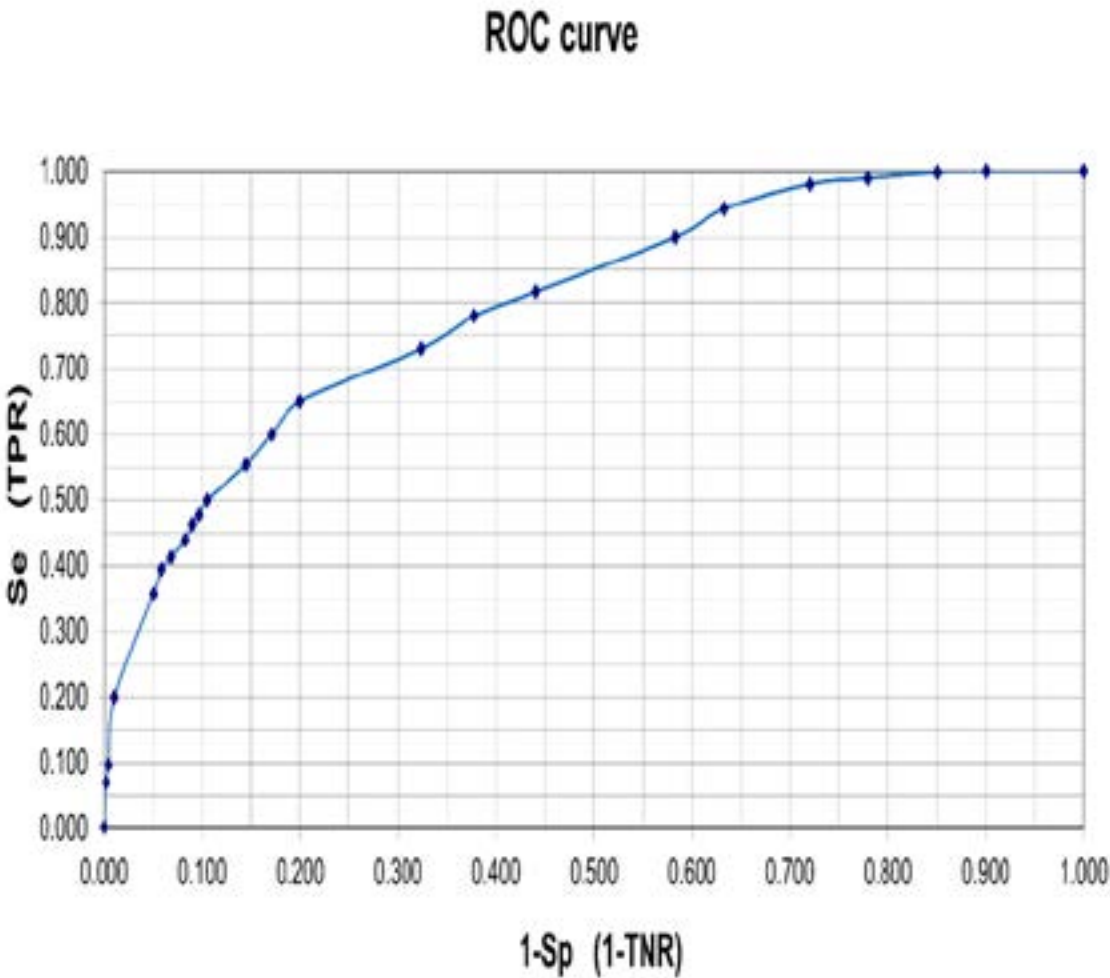


**Figure 6.** Example of a real ROC curve. The area between the curve and the diagonal (AUC, or diagnostic efficiency area) can range from 0.5—when the curve overlaps the orange diagonal in the graph, indicating a useless test that does not discriminate between positives and negatives—to 1, when the curve forms an inverted right isosceles triangle whose hypotenuse is the graph's diagonal and whose two legs meet at the point (0,1). In this ideal case, the test shows perfect discrimination between healthy and diseased subjects (FN and FP = 0).



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).



| Parameter (abbreviation)                                     | Definition                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Method of assessment and/or calculation                                                                                                                                                                                                                          |
|--------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Positive Likelihood Ratio<br>Positive Likelihood Ratio (LR+) | <p>Ratio between the probability of obtaining a positive result in a diseased individual and the probability of obtaining a positive result in a healthy individual.</p> <p>For example, with Sensitivity and Specificity both at 90%, the test will yield a positive result nine times more frequently (0.9 / 0.1) in diseased subjects than in healthy subjects.</p> <p>(Note: ratios based on the number of cases refer to odds, not probabilities.)</p> | <p><math>LR+ = Se / (1 - Sp)</math></p> <p><math>LR+ = \Omega_{post+} / \Omega_{pre}</math></p> <p>The desirable value tends toward infinity.</p> <p>(Note: LR+ may also be inferred graphically from the ratio between the y- and x-axis on the ROC curve.)</p> |
| Negative Likelihood Ratio<br>Negative Likelihood Ratio (LR-) | <p>Ratio between the probability of obtaining a negative result in a diseased individual and the probability of obtaining a negative result in a healthy individual.</p> <p>For example, with Sensitivity and Specificity both at 90%, the test will yield a negative result nine times less frequently (0.1 / 0.9) in diseased subjects than in healthy subjects.</p>                                                                                      | <p><math>LR- = (1 - Se) / Sp</math></p> <p><math>LR- = \Omega_{post-} / \Omega_{pre}</math></p> <p>The desirable value tends toward 0.</p>                                                                                                                       |

The likelihood ratio represents the connecting element between diagnostic performance parameters that are independent of disease frequency in the tested population (non-predictive parameters) and predictive parameters, which are strongly influenced by the prevalence of the condition in the studied group.

In summary, the positive (LR+) and negative (LR-) likelihood ratios are extremely useful because, although they are not direct predictive instruments (they do not in themselves express a “statistical probability”), they synthetically quantify the **diagnostic strength** of a test in relation to a positive or negative result.

The likelihood ratio provides a practical alternative to the ROC curve in the preliminary evaluation of a diagnostic test. However, its main value lies in making Bayes’ statement operational:

“The probability that a test result corresponds—or does not correspond—to the presence of a disease depends on the pre-test probability and on the diagnostic strength of the test itself.”

By knowing these two elements, it becomes possible to estimate the probability of having—or not having—the disease after the test has been performed.

### Interpretation Criteria for Positive and Negative Likelihood Ratios

The table 2 provides a general indication of how much the diagnostic strength of a test modifies the pre-test probability toward the post-test probability.

| LR+    | LR-       | Increase of pre-test probability |
|--------|-----------|----------------------------------|
| 1 – 2  | 1 – 0.5   | Not clinically significant       |
| 2 – 5  | 0.5 – 0.2 | Slight increase                  |
| 5 – 10 | 0.2 – 0.1 | Moderate increase                |
| > 10   | < 0.1     | Strong increase                  |

Table 2. Diagnostic “strength” of a positive or negative result in modifying the pre-test probability.

In summary:

- A test with a likelihood ratio of 1.0 provides no information, and its results do not modify the post-test probability of disease, which therefore remains identical to the pre-test probability.
- A likelihood ratio greater than 1.0 increases the post-test probability of disease; the larger the likelihood ratio, the more informative a positive test result becomes (the higher the probability of being diseased after the test).
- A likelihood ratio less than 1.0 decreases the post-test probability of disease; the smaller the likelihood ratio, the more informative a negative test result becomes (the lower the probability of being diseased after the test).

Finally, it is worth emphasizing that when dealing with quantitative tests—where each possible result has its own sensitivity, specificity, and likelihood ratio—it is possible to build useful comparison tables showing the diagnostic “strength” (LR+) associated with individual results across the measurement range of the method, independently of disease prevalence (Table 3).

| CPK Values (U/L) | Cut-off (UI/l) | Sensitivity | Specificity | LR+  | LR-  |
|------------------|----------------|-------------|-------------|------|------|
| <80              | 80             | 98%         | 12%         | 1.11 | 0.17 |
| 80-150           | 150            | 94%         | 60%         | 2.35 | 0.10 |
| 151-220          | 220            | 86%         | 84%         | 5.38 | 0.17 |
| 221-280          | 280            | 66%         | 96%         | 16.5 | 0.35 |

Table 3. Sensitivity, Specificity, LR+, and LR- for four cut-off values of serum creatine phosphokinase (CPK) levels in the diagnosis—prior to the troponin era—of acute myocardial infarction. (Example adapted from: Katz DL, Clinical Epidemiology & Evidence-Based Medicine: Fundamental Principles of Clinical Reasoning, SAGE Publications, Inc., 2001  
– DOI: <https://doi.org/10.4135/9781452232638>)

### Diagnostic Performance Parameters Dependent on Prevalence Impact of Prevalence on Diagnostic Predictive Value

As briefly outlined earlier, all analytical performance parameters—and some diagnostic performance

parameters—are independent of the proportion of healthy and diseased individuals in the population. Their main use lies in comparing tests produced by different manufacturers and in selecting the most suitable assays for screening campaigns or for biochemical or epidemiological studies. Other diagnostic performance parameters, however, are heavily influenced by the prevalence of the disease or condition within the population being studied, and their use becomes particularly relevant in everyday clinical practice. Prevalence is the proportion of individuals with the disease in a population, during a given period, relative to the total population.

WARNING: The apparent prevalence (calculated based on a test result and therefore affected by the number of false positives) must not be confused with the true prevalence (point prevalence), which is determined through other types of studies using gold-standard methods (Figure 7).



RESEARCH ARTICLE



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).



Figure 7. Difference between apparent prevalence and true prevalence (Prof. Ezio Bottarelli: *Quaderno di Epidemiologia*)

The prevalence of a disease within a homogeneous group of individuals, as used in epidemiological studies, is logically related to the pre-test probability applied in clinical diagnostics.

In fact, the probability that a disease (or event) will occur in a patient whose clinical information is unknown corresponds to the frequency with which that disease or event typically occurs in the population to which the patient belongs.

Every clinician should aim to produce the most reliable possible estimate of the pre-test probability; however, owing to its inherently subjective nature, no universally valid calculation method can be recommended. In practice, the more extensive the clinician's understanding of the disease—its prevalence in the general population, its frequency within specific subgroups defined by ethnicity, age, sex, or associated risk factors—the more accurate this estimate will be.

By way of example, in Italy an adult presenting to the Emergency Department for an injury has a very low pre-test probability, below 0.1%, of testing positive for hepatitis B or C if they are a regular blood donor (reflecting the prevalence in the donor population). If the same individual is not a habitual donor, the probability increases to roughly 1–2% (the estimated prevalence in the general population), representing at least a twenty-fold difference.

To these sources of information—which are not always readily obtainable from the literature—one must add the indispensable contribution of clinical experience. Elements such as medical history, physical examination, and symptom assessment can substantially increase or decrease the clinician's pre-test probability estimate.

Unfortunately, even today prevalence studies are insufficient, often conducted using non-standardized and outdated methods, and generally limited to a few diseases of major interest or social impact (epidemiological studies).

## Prevalence and Diagnostic Accuracy Studies

As discussed so far, sensitivity, specificity, and likelihood ratios are intrinsic characteristics of a diagnostic test and are not expected to change with variations in the prevalence of the target condition in the study sample. The logic behind this assumption lies in the mathematical calculation of these measures derived from the classic  $2 \times 2$  diagnostic table, in which prevalence does not appear.

However, recent meta-analyses of diagnostic accuracy have shown that sensitivity and specificity do change—beyond their confidence intervals—as disease prevalence varies higher disease prevalence is associated with higher estimated sensitivity and lower estimated specificity.

The explanation for this phenomenon, which may appear to contradict the principles of EBLM, can be found in the heterogeneity of the disease spectrum (the variety of clinical manifestations and range of possible outcomes of a given pathological condition) within the patient groups used in diagnostic accuracy studies. Different prevalence levels may therefore reflect different disease spectra (essentially, a different “target condition” for the test), resulting in uncertainty in sensitivity and specificity estimates. This uncertainty also affects likelihood ratios—derived from sensitivity and specificity—and can ultimately influence predictive values.

In conclusion, these recent findings do not contradict the theoretical foundations of EBLM. Instead, they should encourage researchers conducting original diagnostic accuracy studies to pay close attention to the disease spectrum in their sample, ensuring that it matches the disease spectrum of the population that will undergo testing in clinical practice. Additionally, when the spectrum or prevalence of the condition is highly variable, researchers should plan stratified analyses by disease severity and spectrum, following the rigorous methodologies outlined in the most recent (2015) version of the STARD recommendations (Standards for Reporting of Diagnostic Accuracy).

Predictive parameters of disease (PPV) or absence of disease (NPV)

| Parameter (abbreviation)                             | Definition                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Method of study and/or calculation                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Prevalence (Pr)                                      | Given a population composed of healthy and diseased individuals, prevalence indicates the rate of disease cases within the total population during a defined time period.                                                                                                                                                                                                                                                                                                                                                       | $Pr = \text{Diseased} / (\text{Healthy} + \text{Diseased})$<br>Pr is a probability and can range from 0 to 1 (or from 0% to 100%).                                                                                                                                                                                                                                                                                                                                                               |
| Pre-test Odds ( $\Omega_{pre}$ )                     | Given a population composed of healthy and diseased individuals, $\Omega_{pre}$ represents, as a ratio, the share of diseased individuals relative to healthy individuals, as known or estimated before performing the test.<br>It is a different way of expressing disease prevalence: not as a probability but as a ratio between proportions.<br>Like prevalence, it can be known or estimated before performing the laboratory test.                                                                                        | $\Omega_{pre} = \text{Diseased} / \text{Healthy}$<br>(The denominator does not include the total population [healthy + diseased], as in prevalence, but only the number of healthy individuals.)<br>$\Omega_{pre} = Pr / (1 - Pr)$ We have 1 diseased and 4 healthy subjects: an odds of 0.25 indicates that the diseased proportion is one quarter that of healthy individuals ( $1/4 = 0.25$ ).<br>In probability terms (prevalence), the probability is 1 in 5 $\rightarrow 1/(4+1) = 0.20$ . |
| Post-test Odds ( $\Omega_{post+} / \Omega_{post-}$ ) | $\Omega_{post+}$ indicates the share of diseased individuals who test positive (TP) relative to healthy individuals who test positive (FP). (odds of being diseased if the test is positive)<br>$\Omega_{post-}$ indicates the share of diseased individuals who test negative (FN) relative to healthy individuals who test negative (TN). (odds of being diseased if the test is negative)<br><br>( <b>Note:</b> for both $\Omega_{post+}$ and $\Omega_{post-}$ , the numerator always includes diseased subjects, TP or FN.) | In case of<br>$LR+: \Omega_{post+} = TP / FP \quad \Omega_{post+} = \Omega_{pre} \times LR+$<br><br>In case of<br>$LR-: \Omega_{post-} = FN / TN \quad \Omega_{post-} = \Omega_{pre} \times LR-$<br><br>( <b>Note:</b> $\Omega_{post}$ represents a line whose slope is LR.)                                                                                                                                                                                                                     |
| p <sub>post-</sub>                                   | Number of FN over the total number of negative results.<br>Probability that a subject is diseased given a negative result (FN).                                                                                                                                                                                                                                                                                                                                                                                                 | $p_{post-} = FN / (TN + FN)$<br>$p_{post-} = \Omega_{post-} / (\Omega_{post-} + 1)$<br>$p_{post-} = 1 - NPV$ Range: $1 \rightarrow 0$<br>Desirable value = 0.0                                                                                                                                                                                                                                                                                                                                   |
| p <sub>post+</sub>                                   | Number of TP over the total number of positive results.<br>Probability that a subject is diseased given a positive result (TP).                                                                                                                                                                                                                                                                                                                                                                                                 | $p_{post+} = TP / (TP + FP)$<br>$p_{post+} = \Omega_{post+} / (\Omega_{post+} + 1)$<br>$p_{post+} = PPV$ Range: $0 \rightarrow 1$<br>Desirable value = 1.0                                                                                                                                                                                                                                                                                                                                       |



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).



|                                 |                                                                                                                                                                                                                                                                            |                                                                                                                                                                                                     |
|---------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Negative Predictive Value (NPV) | Number of TN over the total number of negative results.<br>Probability of obtaining a negative result in a healthy subject (TN).<br>(Probability of being healthy if the test is negative.)<br><br><b>Note:</b> <i>NPV is the complement of ppost- (NPV = 1 - ppost-).</i> | $NPV = TN / (TN + FN)$<br>$NPV = (Pr \times LR-) / [(Pr \times LR-) + (1 - Pr)]$<br>$NPV = Sp \times (1 - Pr) / [Sp \times (1 - Pr) + (1 - Se) \times Pr]$<br>Range: 0 → 1<br>Desirable value = 1.0 |
| Positive Predictive Value (PPV) | Number of TP over the total number of positive results.<br>Probability of obtaining a positive result in a diseased subject (TP).<br>(Probability of being diseased if the test is positive.)<br><br><b>Note:</b> <i>PPV coincides with ppost+ (PPV = ppost+).</i>         | $PPV = TP / (TP + FP)$<br>$PPV = (Pr \times LR+) / [(Pr \times LR+) + (1 - Pr)]$<br>$PPV = Se \times Pr / [Se \times Pr + (1 - Sp) \times (1 - Pr)]$<br>Range: 0 → 1<br>Desirable value = 1.0       |

“Odds” and “statistical probability” are not the same thing

The term odds originate from Anglo-Saxon betting language—where bookmakers express probability as the ratio between potential winnings and the stake—and can be translated as “quota” or “ratio.” It is essential to note that odds is used only in the plural. Both odds and prevalence refer to the same underlying concept: an estimate of the probability that a specific condition will occur (for example, a positive or negative test result) within a defined population. Although they describe the same phenomenon, the two terms must not be confused.

More precisely, odds represent, separately within the groups of healthy and diseased individuals, the ratio between true and false classifications: in the diseased group, the ratio of true positives to false negatives (TP/FN), and in the healthy group, the ratio of false positives to true negatives (FP/TN). (Note: the numerator always refers to the diseased cases.)

Prevalence, by contrast, represents the proportion of diseased or healthy individuals relative to the total population under study—for example, the proportion of diseased subjects in a mixed population.

Thus, odds compare different categories (test positives vs. test negatives) within a homogeneous group (healthy or diseased), whereas prevalence represents the probability of a condition in the entire population of interest, composed of both healthy and diseased individuals.

Another distinguishing feature is the measurement scale. Although both quantities increase monotonically (that is, as one increase, the other always increases), their relationship is not linear: statistical probability ranges from 0 to 1, whereas odds range from 0 to infinity.

This property—especially when probability approaches its upper limit (1 or 100%)—makes differences

in odds easier to detect. Indeed:

- When probability is low (near 0), odds tend to be very close to the probability itself.
- When probability is high (near 1), odds tend to diverge sharply from the probability value.

Furthermore, odds allow for more sophisticated mathematical modeling than proportions, such as in logistic regression.

Having clarified the differences between the two concepts—and since medicine typically reasons in terms of probability (a more intuitive and familiar concept) rather than odds—odds can be easily converted to probability (and vice versa) using the following formulas:

**Probability = odds / (odds + 1)**  
**Odds = probability / (1 - probability)**

A final caution: odds (and certainly not LR values) must not be confused with the better-known odds ratio, which is frequently used in epidemiology and in case-control studies to evaluate the association between exposure to a risk factor and an outcome in two groups of individuals—exposed and non-exposed.

Within the framework of EBLM, it is essential to fully understand the positive predictive value (PPV) and negative predictive value (NPV) of a diagnostic test, as these allow the correct interpretation of results in relation to the actual probability of the presence or absence of a condition or disease.

By applying Bayes’ theorem, the post-test probabilities of a positive or negative result can be computed directly on a simple spreadsheet. This approach relies, on one hand, on estimating the pre-test probability of disease in the population of interest, and on the other hand, on applying the diagnostic sensitivity and specificity (or the likelihood ratios derived from them), as calculated in validation studies of

commercial methods and reported in the corresponding technical inserts.

information on their diagnostic performance (including the estimation of confidence intervals) is strongly

$$PPV = \frac{Se \cdot Pr}{Se \cdot Pr + (1 - Sp) \cdot (1 - Pr)} = \frac{Pr \cdot LR+}{1 - Pr + Pr \cdot LR+}$$

$$NPV = \frac{Sp \cdot (1 - Pr)}{Sp \cdot (1 - Pr) + (1 - Se) \cdot Pr} = \frac{1 - Pr}{1 - Pr + Pr \cdot LR-}$$

Ultimately, a laboratory test can be considered clinically valid and effective when it is able to provide a substantial amount of useful information that contributes meaningfully to the clinician’s diagnostic reasoning, and when the therapeutic decisions based on its results lead to significant improvements for the patient.

From this standpoint, the usefulness of a test is specific to each individual situation the clinician must address in daily practice. For this reason, the clinician can never overlook—without jeopardizing the entire diagnostic and therapeutic process—the understanding of the diagnostic performance and efficiency parameters that every laboratory test can offer.

The use of in vitro tests that lack adequate infor-

discouraged, as this would lead to an interpretation of results that overlooks the essential probabilistic component that characterizes the entire diagnostic process.

**Fagan’s Nomogram**

Another tool that can be used for the approximate calculation of PPV and NPV—although unfortunately rarely used in daily practice—is Fagan’s nomogram (Figure 8): a simple and useful “instrument” that every clinician should ideally keep in the pocket of their lab coat.

This graphical device allows the estimation of ppost+ (PPV) or ppost– (1 – NPV) starting from ppre (prevalence) and from the likelihood ratios (LR+ or LR–), which can be easily derived from the values of sensitivity and specificity.

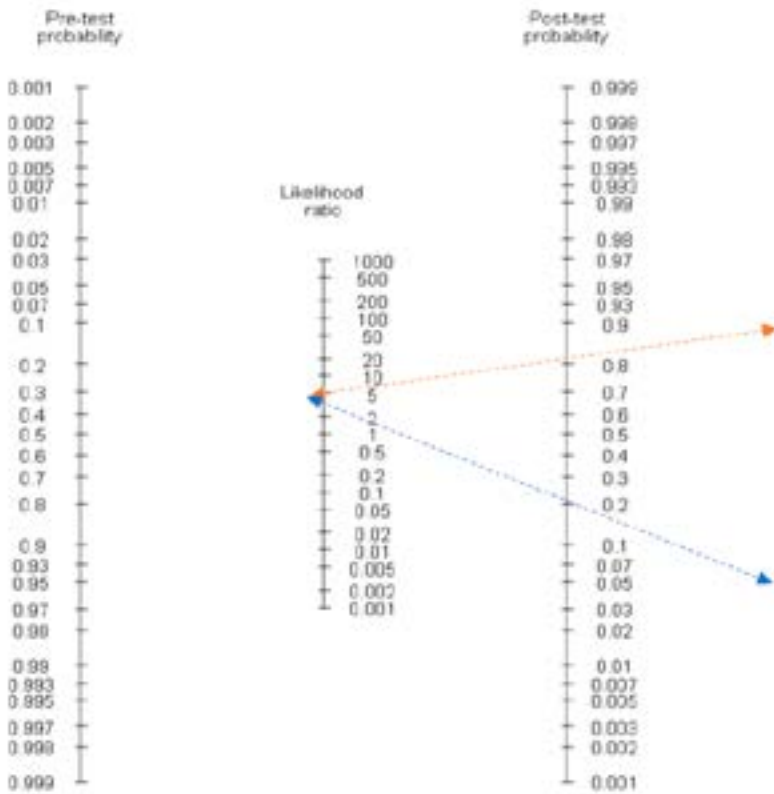


**Figure 8.** Fagan’s nomogram. Draw a straight line connecting the value of ppre on the left-hand scale with the value of LR+ or LR– in the centre, in order to obtain, at the point of intersection with the right-hand scale, the posterior probability, that is, the prior probability updated on the basis of the likelihood ratio that characterises the test for Pr = 30%; if LR+ = 20 → Ppost+ (PPV) = 90%; if LR– = 0.1 → Ppost– = 5% (NPV = 95%)



RESEARCH ARTICLE



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).



It should be noted that, precisely to overcome the issue of expressing values in odds—which are uncommon and of limited practical use in the clinical setting—Fagan’s nomogram expresses all quantities, both input and output, directly in terms of probabilities of disease: ppre as the input, and ppost+ and ppost– as the outputs.

This may create some confusion for those who are accustomed to working with predictive values, because there are certain “terminological” differences between ppost+, ppost–, PPV, and NPV that must be clarified.

Indeed, the calculation of ppost values and of predictive values depends on the conventional definitions assigned to each term, and these definitions do not always overlap.

While ppost+ and PPV share the same definition (the probability that a subject is diseased given a positive result), ppost– and NPV are defined differently:

- ppost– is the probability that a subject is diseased given a negative result (FN);
- NPV is the probability that a subject does not have the disease (is healthy) given a negative result (TN).

Therefore, once these definitions are established, NPV and ppost– are complementary probabilities. That is, each represents 1 minus the other (if ppost– = 5%, then NPV = 95%, and vice versa), and they must be interpreted accordingly.

## How to Improve Diagnostic Predictive Value

To increase the probability of correctly identifying healthy and diseased patients, multiple independent tests may be performed either sequentially (progressively increasing the pre-test probability, ppre) or in parallel, combining their respective sensitivity and specificity using Boolean logic (AND or OR).

The choice between these two approaches depends on the diagnostic objective:

- If the goal is to avoid missing any cases (emphasizing sensitivity), the preferred approach is parallel testing with OR logic.
- If the aim is to reduce false positives (emphasizing specificity), the preferred approach is sequential testing or parallel testing with AND logic.

For tests to be considered truly independent, they must—while referring to the same population of healthy and diseased subjects—be based on different analytical principles and be supported by separate diagnostic accuracy studies.

Therefore, repeating the same test to confirm or exclude a diagnosis does not constitute an appropriate method for improving diagnostic predictive validity. Such repetition may be useful to rule out gross analytical errors or, in the case of quantitative tests, to reduce the standard error of the mean (SEM) and thus narrow the confidence interval of the averaged result.

www.jahc.it

## Two independent tests performed in sequence

The calculation of predictive value in sequential testing is the simplest: the prevalence estimated through the first examination (the so-called screening test, with high sensitivity) becomes the ppre for the second test, and so on (the confirmatory or diagnostic test, with high specificity). In this way, if all tests are positive, the PPV increases progressively.

### Practical Example No. 1

A first-level screening test ( $Se_1 = 0.99$ ;  $Sp_1 = 0.82$ ;  $LR^+ = 5.5$ ) is performed on a patient with suspected disease, where the prevalence among patients with identical symptoms is 10%.

The result of the test is positive.

A negative result would have almost certainly ruled out the disease ( $NPV = 99.9\%$ ), whereas a positive result informs us that the PPV increases from 10% to 37.9% (ppost+): still not sufficient to establish a diagnosis.

By performing a second-level test—less sensitive but preferably more specific ( $Se_2 = 0.90$ ;  $Sp_2 = 0.90$ ;  $LR^+ = 9.0$ )—and using the previous ppost+ value (0.379) as the new ppre, the sequential NPV ( $NPV_{SEQ}$ ) decreases to 93.6%, but the sequential PPV ( $PPV_{SEQ}$ ) rises to 84.6%, allowing a more reliable confirmation of the disease.

## Why was the screening test not bypassed by performing the more specific test immediately?

Aside from issues related to higher costs and greater technical complexity—both common characteristics of second-level or confirmatory tests—if Test 2 had been performed first, without knowing the result of Test 1 (the screening test), starting from the initial prevalence value  $Pr = 10\%$ , the  $PPV_2$  would have increased only to 50.0%, a level of probability still insufficient for a reliable diagnosis.

## Two independent tests performed in parallel (OR mode)

A subject is considered negative only if both tests are negative, while they are considered positive if at least one (OR) of the two tests is positive.

$$SeOR = Se_1 + Se_2 - (Se_1 \cdot Se_2)$$

$$SpOR = Sp_1 \cdot Sp_2$$

## Two independent tests performed in parallel (AND mode)

A subject is considered negative if both tests are negative, while they are considered positive only if both (AND) tests are positive.

$$SeAND = Se_1 \cdot Se_2;$$

$$SpAND = Sp_1 + Sp_2 - (Sp_1 \cdot Sp_2)$$

### Practical Example No. 2

Using the same  $Se$  and  $Sp$  values from the two previous sequential tests, with  $Pr = 10\%$ , we obtain:

OR method:

$$Se^{OR} = 0.999; Sp^{OR} = 0.738 \rightarrow NPV^{OR} = 100\%; PPV^{OR} = 29.8\%$$

AND method:

$Se^{AND} = 0.891; Sp^{AND} = 0.982 \rightarrow NPV^{AND} = 98.8\%; PPV^{AND} = 84.6\%$

It is possible to verify the calculations of the previous examples using the Excel sheet attached to this vademecum; however, in practice, dedicated statistical software, apps, or probability tables are often used to simplify computations—especially when the tests are not independent from one another and covariance must be taken into account, or when more than two tests are involved, or when they are combined in a more complex manner.

CONCLUSIONS

Naturally, if the patient’s condition were already known—as occurs in studies assessing diagnostic accuracy—there would be no need to perform any test. In routine clinical practice, however, the purpose is the reverse of research: the patient’s status must be inferred from the test result itself. Within this framework, the predictive values, both negative and positive, become fundamental.

Predictive values are not intrinsic properties of a diagnostic test. Although they are influenced by the test’s sensitivity and specificity—and improve as these parameters increase—they also depend substantially on a factor that is entirely external to the test itself: the prevalence of the disease in the population under investigation.

The observation that the positive predictive value increases as prevalence rises discourages the use of screening programmes for rare diseases, where prevalence is extremely low. In such situations, screening offers limited benefit—only a very small number of individuals will profit from early detection, simply because the disease is uncommon—while a large proportion (the false positives) may be harmed by being incorrectly classified as diseased.

Nonetheless, there are circumstances in which an exception to this general rule is justified: namely, when the disease or condition carries significant implications for public health, as in the case of severe but rare disorders that are readily transmissible.

Summary formulary  
Diagnostic Parameters Table

| Parameter | Calculation Method                                           | Description                                    | Range         |
|-----------|--------------------------------------------------------------|------------------------------------------------|---------------|
| Se        | $TP / (TP + FN)$<br>$TP / \text{Diseased}$                   | % of diseased individuals correctly identified | 0–1<br>0–100% |
| Sp        | $TN / (TN + FP)$<br>$TN / \text{Healthy}$                    | % of healthy individuals correctly identified  | 0–1<br>0–100% |
| ACC       | $(TP+TN) / (TP+FP+TN+FN)$<br>$Se \cdot Pr + Sp \cdot (1-Pr)$ | Correct tests / (Healthy + Diseased)           | 0–1<br>0–100% |

For certain infectious diseases with high social impact (as recently illustrated by the COVID-19 pandemic), even a single false negative may pose a substantial risk by enabling further transmission. In these scenarios, the diagnostic test must prioritise maximal sensitivity, even at the cost of producing a higher number of false positives.

A similar rationale applies to screening performed in blood donors. Protecting recipients is imperative, and no margin for error can be tolerated in identifying potentially infectious units. Consequently, it becomes acceptable to discard a proportion of blood bags that are uninfected but have tested falsely positive in safety assays.

Conversely, when disease prevalence is high, a highly specific test is generally preferable: false positives must be minimised to avoid overwhelming diagnostic and therapeutic resources with an excessive number of individuals incorrectly identified as affected.

In summary, within clinical practice:

- A highly sensitive test should be selected when the consequences of a missed diagnosis are particularly serious—for example, diseases that are typically fatal unless treated promptly, or conditions capable of causing major outbreaks. Sensitive tests are also valuable in the early diagnostic phase (first-level tests), when the aim is to narrow a broad set of possible conditions in a differential diagnosis. In this context, the sensitive test is used primarily to exclude one or more diseases, as it is most informative for the clinician when it yields a negative result.
- A highly specific test, by contrast, is especially useful for confirming a diagnosis that has already been suggested by other means (second-level tests), since it is rarely positive in the absence of disease. Tests with very high specificity (third-level or confirmatory tests, which are often more complex and costly) are particularly important when a false-positive result would be harmful—clinically, emotionally, or economically. In essence, a highly specific test provides the greatest clinical value when it yields a positive result.



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

|                          |                                                                                                                     |                                                                                                                |               |
|--------------------------|---------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------|---------------|
| LR+                      | $(TP/Diseased) / (FP/Healthy) \\ Se / (1-Sp)$                                                                       | Ratio between the probability of obtaining a positive result in diseased (TP) and in healthy individuals (FP). | 1-∞           |
| LR-                      | $(FN/Diseased) / (TN/Healthy) \\ (1-Se) / Sp$                                                                       | Ratio between the probability of obtaining a negative result in diseased (FN) and in healthy individuals (TN). | 1-0           |
| Pr                       | $(TP + FN) / (TP + FP + TN + FN) \\ Diseased / (Diseased + Healthy)$                                                | Probability/percentage of diseased subjects in the total population (pre-test).                                | 0-1<br>0-100% |
| $\Omega_{pre}$           | $(TP+FN)/(TN+FP) \\ Diseased / Healthy \\ Pr/(1-Pr)$                                                                | Ratio between Diseased and Healthy (pre-test).                                                                 | 0-1           |
| $\Omega_{post-}$         | $FN/TN \\ LR- \cdot \Omega_{pre}$                                                                                   | Ratio between diseased testing negative and healthy testing negative.                                          | 1-0           |
| $\Omega_{post+}$         | $TP/FP \\ LR+ \cdot \Omega_{pre}$                                                                                   | Ratio between diseased testing positive and healthy testing positive.                                          | 1-∞           |
| p <sub>post-</sub>       | $FN / (TN+FN) \\ LR- \cdot \Omega_{pre}$                                                                            | Probability of a negative result in a diseased subject.                                                        | 1-0<br>100%-0 |
| p <sub>post+</sub> ; PPV | $TP / (TP+FP) \\ (Pr \cdot LR+) / [(Pr \cdot LR+) + (1-Pr)] \\ (Se \cdot Pr) / [Se \cdot Pr + (1-Sp) \cdot (1-Pr)]$ | Probability of a positive result in a diseased subject.                                                        | 0-1<br>0-100% |
| NPV                      | $TN / (TN+FN) \\ (1-Pr) / [(1-Pr) + (Pr \cdot LR-)] \\ Sp \cdot (1-Pr) / [Sp \cdot (1-Pr) + (1-Se) \cdot Pr]$       | Probability of a negative result in a healthy subject.                                                         | 0-1<br>0-100% |
| TP                       | $Se \cdot Diseased \\ (Se \cdot Pr) \cdot Total$                                                                    | Number of positive results among diseased.                                                                     |               |
| FN                       | $(1-Se) \cdot Diseased \\ (1-Se) \cdot Pr \cdot Total$                                                              | Number of negative results among diseased.                                                                     |               |
| FP                       | $(1-Sp) \cdot Healthy \\ (1-Sp) \cdot (1-Pr) \cdot Total$                                                           | Number of positive results among healthy.                                                                      |               |
| TN                       | $Sp \cdot Healthy \\ Sp \cdot (1-Pr) \cdot Total$                                                                   | Number of negative results among healthy.                                                                      |               |

**Healthy:** number of individuals without the disease or condition of interest

**Diseased:** number of individuals with the disease or condition of interest

**Total:** sum of healthy and diseased individuals



CASE STUDY:  
SELECTION OF THE OPTIMAL CUT-OFF IN QUANTITATIVE TESTS

A detailed numerical and graphical analysis of the characteristics of a quantitative test used for disease diagnosis can be performed by appropriately adjusting the cut-off values in a cohort composed of healthy and diseased individuals, and by observing how these changes affect the counts of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN).  
For each participant included in the study, the true clinical status—healthy or diseased for the condi-

tion under investigation—must be established using gold-standard clinical and instrumental criteria that are independent of the test being evaluated.  
This case study employs fabricated data with no medical relevance, provided solely for illustrative purposes. These data are available in the Excel file accompanying this document, which should be consulted for the specific calculations.  
The figures below (generated with the open-source statistical software Jamovi) show the distribution of test results: a histogram with an overlaid Gaussian curve (Figure 9) and a box plot (Figure 10). Table 4 reports the basic descriptive statistics.

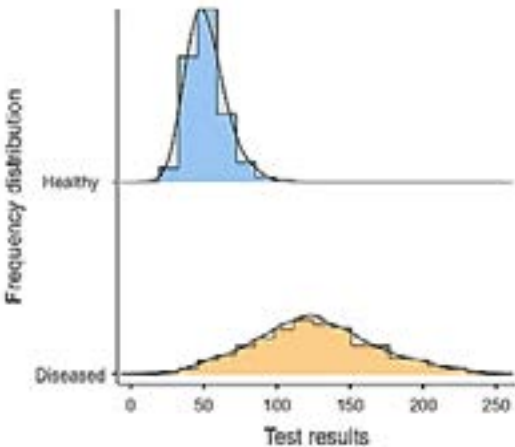


Figure 9. Distribution of test results in healthy and diseased individuals.

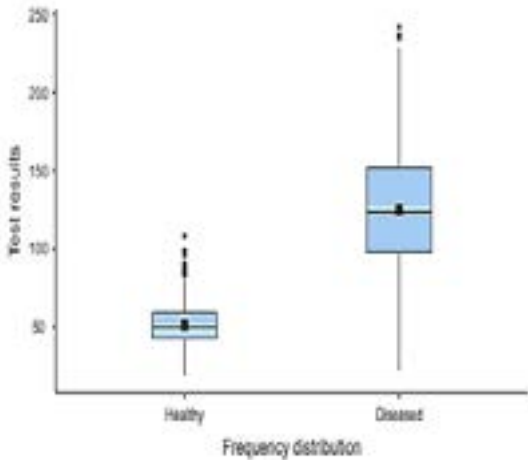


Figure 10. Box-plot of the distribution of test values.

Table 4. Basic statistics

|           | HEALTHY | DISEASED | HEALTHY+ DISEASED |
|-----------|---------|----------|-------------------|
| n         | 853     | 644      | 1497              |
| Pr        |         | 43.0%    |                   |
| 50 Perc   | 50.0    | 124.0    | 63.0              |
| 97,5 Perc | 82.7    | 212.9    | 195.6             |
| 2,5 Perc  | 32.0    | 49.0     | 33.0              |
| Min       | 19      | 22       | 19                |
| Max       | 108     | 242      | 242               |
| Avg       | 51.7    | 125.6    | 83.5              |
| SD        | 12.61   | 41.26    | 46.53             |
| SEM       | 0.43    | 1.63     | 1.20              |
| X+2SD     | 76.9    | 208.2    | 176.5             |
| X-2SD     | 26.4    | 43.1     | -9.6              |
| K index   | 0.78    | 0.19     | 1.04              |
| Sk index  | 0.92    | -0.19    | 0.14              |



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

The differing shape and distribution of values—both in terms of central tendency and dispersion—across the two populations indicate, even on initial visual inspection, that the data are approximately Gaussian and that the test can discriminate between healthy and diseased individuals in a substantial proportion of cases.

However, a wide zone of overlap exists between the two groups, within which this distinction cannot be made (the so-called grey zone).

It is therefore useful to examine the ROC curve (Figure 11) to assess both the overall diagnostic accuracy of the method and the behaviour of sensitivity

and specificity across different concentration thresholds.

A limitation of ROC curves is that they do not graphically display the cut-off values used in their construction. This omission can complicate the selection of decision thresholds, which may prioritise sensitivity, specificity, or overall diagnostic accuracy depending on the clinical objective and the prevalence of the disease in the population being tested. The absence of this information—which cannot be directly inferred from the graph—may make it difficult to determine how diagnostic a given quantitative result truly is.

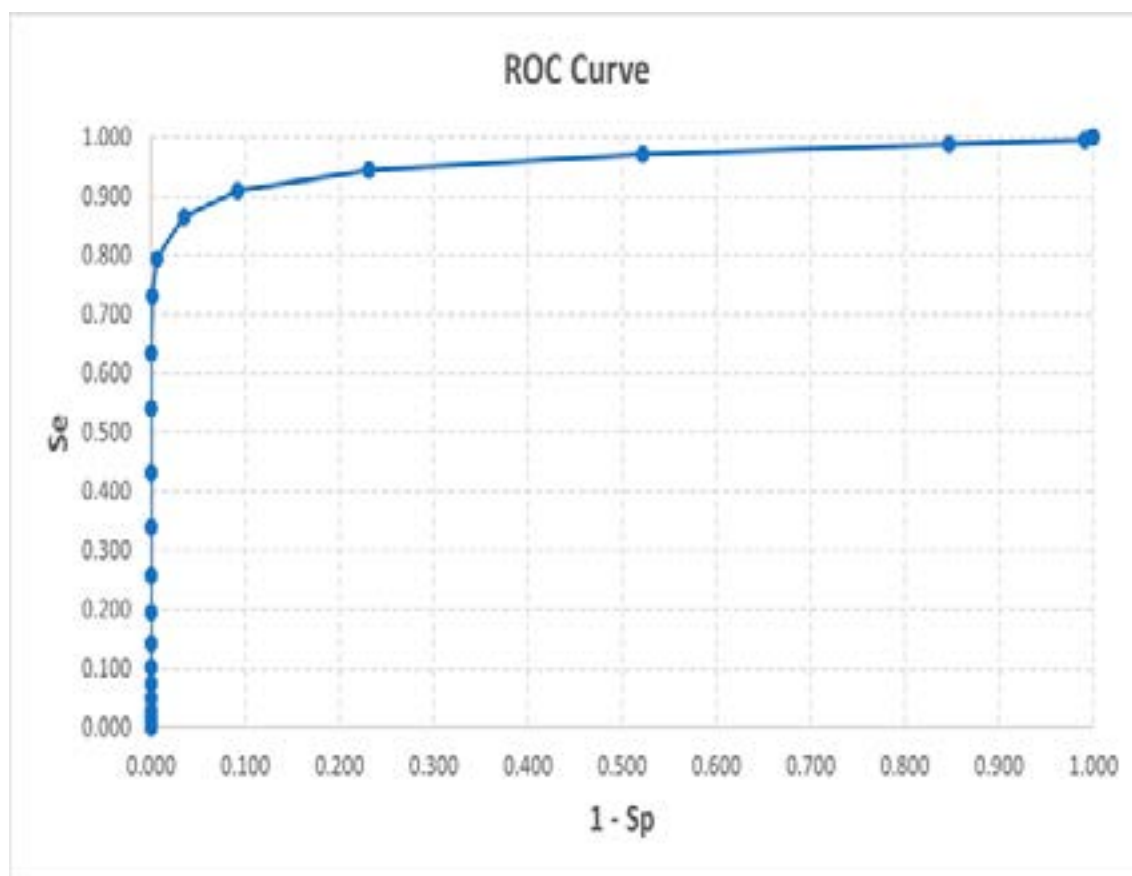


**Figure 11.** ROC curve for the case study. The AUC approaches 1, indicating that the test has good discriminative ability between healthy and diseased individuals. The dashed orange line marks the approximate position of the Youden Index—the point on the ROC curve that lies farthest from the diagonal reference line and minimises misclassifications (FP and FN), thereby maximising the diagnostic accuracy of the test (TP and TN).



## Diagnostic Performance Curve as a Function of the Cut-off (CPD-Cut)

To address this limitation, sensitivity, specificity, and diagnostic accuracy can be visualised as functions of the chosen cut-off value, as shown in Figure 12. Constructing a CPD-Cut (Diagnostic Performance Curve by Cut-off) makes it possible to identify, clearly and immediately, the most appropriate threshold values, selecting them according to predefined objectives—whether clinical or otherwise. This approach facilitates the selection of cut-off levels that are best suited to the specific diagnostic context.

This curve, like the ROC curve, does not change with variations in prevalence and can therefore be applied across different clinical, epidemiological, or operational settings. What does change substantially as prevalence varies, of course, are the negative and positive predictive values.

It should be recalled that only when specificity equals 1 does a positive result always correspond to a true positive ( $LR^+ = \infty$ ), and only when sensitivity equals 1 does a negative result always correspond to a true negative ( $LR^- = 0$ ). Under these ideal conditions, test results become independent of disease prevalence.

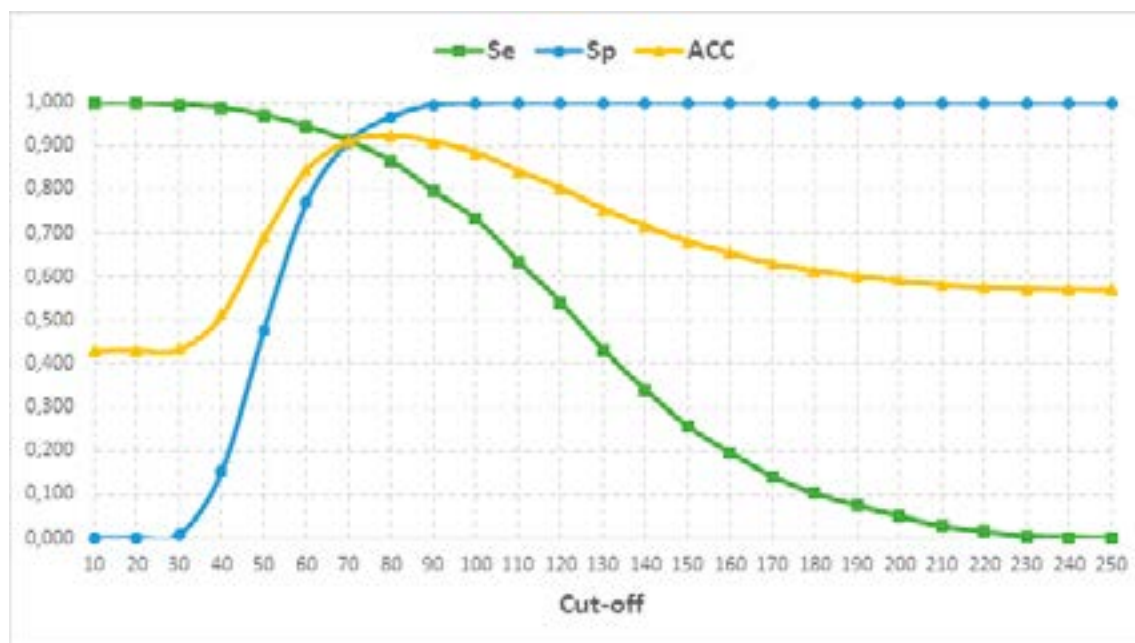


**Figure 12.** Diagnostic Performance Curve as a Function of the Cut-off (CPD-Cut). The yellow accuracy curve (triangles) identifies 80 as the optimal cut-off for minimising the combined number of false negatives and false positives. The blue specificity curve (circles) indicates that values above 100 make the presence of disease highly probable, whereas the green sensitivity curve (squares) shows that values below 30 make the absence of disease almost certain. Values between 30 and 100 constitute a zone of overlap between the two groups—the so-called grey zone—which requires additional non-analytical criteria to distinguish healthy from diseased individuals.

## REFERENCES

1. Istituto Superiore di Sanità. Guida Eurachem. Terminologia per le misurazioni analitiche – Introduzione al VIM 3. (Eurachem Guide. Terminology for Analytical Measurements – Introduction to VIM 3). Traduzione italiana a cura di Maria Belli, Marina Patriarca e Michela Segà 2013, vi, 41 p. Rapporti ISTISAN 13/41
2. V. J. Barwick (Ed.), Eurachem Guide: Terminology in Analytical Measurement – Introduction to VIM 3 (2nd ed.2023). ISBN 978-0-948926-40-2.
3. SIPMEL – Commissione Nazionale Qualità e Accreditamento. Recommendations for the glossary in the new ISO standards for medical laboratories on calibration, labels and quality control (ISO 17511, ISO 18113, and 15198) (National Commission for Quality and Accreditation Recommendations for the glossary in the new ISO standards for medical laboratories on calibration, labelling, and quality control (ISO 17511, ISO 18113, and ISO 15198) 2020. Version 1.
4. Istituto Superiore di Sanità. Guida Eurachem. Idoneità per lo scopo dei metodi analitici. Traduzione italiana a cura di E. Gregori, M. Patriarca e M. Segà 2016, v, 65 p. Rapporti ISTISAN 16/39 (Eurachem Guide. Fitness for Purpose of Analytical Methods Italian translation edited by E. Gregori, M. Patriarca, and M. Segà 2016, v, 65 pp. Rapporti ISTISAN 16/39)



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).



5. Cerelli E, D'Addiego A, Iannello E, Nocito V, Ricotta A, Rossini I, Santarcangeli DF. L'incertezza di misura nel laboratorio clinico: casi particolari di stima della precisione analitica nei test quantitative. Raccomandazioni dell'Associazione Tecnico Scientifica – S.I.T.La.B Gruppo di lavoro Biochimica clinica e Patologia clinica N.055/23 *Journal SITLaB News* – 2023 (Measurement Uncertainty in the Clinical Laboratory: Specific Cases for Estimating Analytical Precision in Quantitative Tests. Recommendations of the Technical-Scientific Association S.I.T.La.B., Clinical Biochemistry and Clinical Pathology Working Group No. 055/23. *SITLaB News*. 2023.)
6. Bossuyt PM, Reitsma JB, Bruns DE, Constantine A et al. – STARD Group STARD 2015. Check list aggiornata per il reporting degli studi di accuratezza diagnostica. Traduzione italiana a cura della Fondazione GIMBE *Evidence* 2016;8(1): e1000129 STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies
7. Mandell BF, Wasserman MR. Capire i test medici e i risultati dei test. Manuale MSD – versione italiana ONLINE per i professionisti. Merck & Co, Inc., Rahway, NJ, USA (Mandell BF, Wasserman MR. *Understanding Medical Tests and Test Results*)
8. Dorizzi RM, Caputo M. Gli strumenti della Evidence Based Medicine nella Medicina di Laboratorio. *Riv Med Lab - JLM*, Vol. 1, N. 1-2, 2000 (SIRSE Srl ed.) 59-67. Atti EVIDENCE BASED LABORATORY MEDICINE Vicenza, 25-26 maggio 2000
9. Horvath AR. Quale evidenza abbiamo degli esami di laboratorio? *Riv Med Lab - JLM*, Vol. 5, N. 4, 2004 (Evidence-based laboratory medicine)
10. Canali C, Trenti T. La Medicina di Laboratorio basata sull'evidenza: gli strumenti utili per una valutazione razionale degli esami di laboratorio, (Evidence-Based Laboratory Medicine: Useful Tools for a Rational Evaluation of Laboratory Tests) *Biochimica Clinica*, 2008, vol. 32, n. 4, 243-250
11. Carobene A. La variabilità biologica: le basi teoriche e l'esperienza dei Gruppi di Lavoro della Federazione Europea di Chimica Clinica e Medicina di Laboratorio (Biological Variation: Theoretical Background and the Experience of the Working Groups of the European Federation of Clinical Chemistry and Laboratory Medicine). *Biochimica Clinica*, 2018, vol. 42, n. 1, 15-25
12. Aarsand AK, Webster C, Fernandez-Calle P, Jonker N, Diaz-Garzon J, Coskun A et al. The EFLM Biological Variation Database. <https://biologicalvariation.eu>
13. Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G et al. Defining analytical performance specifications: Consensus Statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine *Clin Chem Lab Med* 2015; 53(6): 833–835
14. Fraser CG. Le specifiche di qualità (Quality Specifications) *Riv Med Lab - JLM*, Vol. 4, N. 3-4, 2003 (Traduzione a cura di Dorizzi RM, Esposito E, Menozzi I)
15. Murad MH, Lifeng Lin, Haitao Chu, Bashar Hasan et al. The association of sensitivity and specificity with disease prevalence: analysis of 6909 studies of diagnostic test accuracy *CMAJ* 2023 July 17;195: E925-31. doi: 10.1503/cmaj.221802
16. Clerico A, Stenner E, Cappelletti P. Troponine ad alta sensibilità: cosa si sa e cosa si dovrebbe sapere. Confronto tra Laboratorio, Pronto Soccorso e Cardiologia
17. Atti del “Simposio Azienda Sanitaria Universitaria Integrata di Trieste”. Trieste, 9 maggio 2018 (Clerico A, Stenner E, Cappelletti P. High-Sensitivity Troponins: What Is Known and What Should Be Known. Comparison between the Laboratory, Emergency Department, and Cardiology. *Proceedings of the Symposium of the University Health Authority of Trieste*. Trieste, 9 May 2018)
18. Wang B, Wirth R, Bergmann E. et al. Impact of inflammatory status on intestinal iron absorption in older hospitalized patients. *Eur J Clin Nutr* (2025)
19. Besozzi MM. <https://www.bayes.it/bayesarc/> “Sito di Medicina di Laboratorio” (Pur non essendo più aggiornato dal 2023, contiene preziose informazioni sul teorema di Bayes, EBLM, statistica e molto altro).

(Internet links verified on 1 September 2025).