Specific performance characteristics of in-vitro diagnostic medical devices

Gianalfredo Scassellati

On behalf of European Diagnostic Manufacturers Association (EDMA)[#]

INTRODUCTION

The purpose of this paper is to focus the attention of in vitro diagnostic (IVD) manufacturers on correct interpretation of the terms employed to define the so-called specific performance characteristics of an in-vitro diagnostic procedure.

This will help the manufacturers to prepare both the technical documentation and the instructions for use for the different devices, as well as to fulfil the requirements of the different CEN Standards and, finally, to find a common understanding with the Notified Bodies.

The correct use of such terms will also be useful to laboratory professionals, by helping the interpretation of the results for product performance evaluation and for internal and external quality assessment.

TERMINOLOGY BACKGROUND

The "specific performance characteristics" of a diagnostic assay procedure are reported in the European Directive 98/79/EC, covering the in vitro diagnostic medical devices, in Annex I (Essential Requirements).

Annex I, Part A, Section 3 reads as follows:

"The devices must be so designed and manufactured ... They must achieve the performances, in particular, where appropriate, in terms of analytical sensitivity, diagnostic sensitivity, analytical specificity, diagnostic specificity, accuracy, repeatability, reproducibility, including control of known relevant interference, and limits of detection, stated by the manufacturer".

Annex I, Part B, Section 8.7 (concerning the contents of the Instructions for use) reads as follows:

"d) the performance referred to in section 3 of part A

h) the specific analytical performance characteristics (e.g. sensitivity, specificity, accuracy, repeatability, reproducibility, limits of detection and measurement range, ... "

The correct meaning of these so-called specific performance characteristics is not always clear in the laboratory medicine area, both on the side of manufacturers and on the side of laboratory professionals, as it is demonstrated by different opinions repeatedly expressed in many scientific meetings, including Wordlab in Florence on June 1999.

In part, these misunderstandings are due to the fact that many new ISO Standards (e.g. VIM, GUM) dealing with metrology issues and including internationally agreed definitions have been published just in the last decade (7-10).

In these basic ISO standards some terms have been more extensively defined, while other terms did not retain the previous meaning.

As a typical example, the term Accuracy was hitherto used to cover the concept of Trueness (as it occurs also in the text of the IVD Directive 98/79/EC), whilst, according to ISO 5725, accuracy includes now both the concepts of trueness and precision.

More recently, several standards in the laboratory medicine area have been mandated to CEN by the European Commission in support of the IVD Directive 98/79/EC. These standards are being developed in the Technical Committees CEN/TC140 and (by parallel work) ISO/TC212 by experts from many different countries. While some of these standards contain extensive parts devoted to "Terms and Definitions", other standards do not address this issue to a similar extent.

A Glossary (reporting the terms and definitions used in the different standards) is in preparation at both CEN and ISO level, with the purpose of harmonizing at least the most relevant terms.

The definitions reported in the present paper are mainly derived from a publication of Dr. Rene Dybkaer (1), who is chairman of CEN/TC140/WG4 and ISO/TC212/WG2. These two WGs have prepared or are preparing several relevant draft standards, e.g. Traceability of values assigned to calibrators and controls, Requirements for reference measurement laboratories (11-14).

Dybkaer's publication (1) includes most relevant definitions drawn from the basic ISO Standards and particularly from VIM (7-10).

Since these terms and definitions are currently used in the draft standards being presently developed in CEN and ISO on metrological issues (11-14), also the definitions reported in the present EDMA paper refer to Dybkaer publication (1) (see quotations like e.g Ref. 1:4.14)

The present paper, which is mainly intended to focus on the so-called specific analytical performance characteristics of a diagnostic test, addresses in detail the following terms: analytical specificity, analytical sensitivity, precision (i.e. repeatibility and reproducibility), trueness and accuracy.

Because of the growing importance of the so-called "yes or no" (i.e positive/ negative) diagnostic tests which are mainly used for detection of infectious diseases and are mostly classified in Annex II of the IVD Directive, this paper also takes into account other diagnostic performance characteristics which are relevant to such assay procedures, namely diagnostic sensitivity and diagnostic specificity.

The use of these terms depends on the analytical techniques applied and the analytes to be detected. On the other hand, other characteristics (e.g. linearity, recovery, etc.) may sometimes be taken into consideration.

ANALYTICAL PERFORMANCE CHARACTERISTICS

Analytical Specificity

Ability of a measurement procedure[a] to determine solely the measurable quantity[b] it purports to measure, i.e. the measurand[c] (Ref.1: 4.14)

In practice, the analytical specificity may be defined as the ability of an assay procedure to determine specifically the concentration of the target analyte[d] in the presence of:

- potentially interfering factors in the sample matrix (e.g. haemolysis, anticoagulants, effects of sample treatment).
- potentially cross-reacting analytes like similar substances (e.g. antigens, antibodies), impurities or degradation products.

A typical example of specificity as required in the steroid immunoassay is the cross-reaction of an antibody e.g. anti-testosterone versus other steroids of similar structure e.g. dihydrotestosterone.

---

*[a]Measurement procedure: set of operations, described specifically, used in carrying out a measurement according to a given method of measurement (Ref.1: 4.71, VIM: 2.5)

[b]Measurable quantity: attribute of a phenomenon, body or substance that may be distinguished qualitatively and determined quantitatively (Ref.1: 4.68, VIM: 1.2)

[c]Measurand : a particular measurable quantity subject to measurement (VIM: 2.6). Example : "mass of protein in 24 hour urine" or " concentration of glucose in plasma" (Ref.1: 4.69)

[d]Analyte: component indicated in the name of a measurable quantity. Example (in the example for measurand) : protein or glucose (Ref.1: 4.4)

Analytical sensitivity

Slope of the analytical calibration[e] function (Ref.1: 4.12)

Analytical sensitivity is defined by IUPAC as the ability of an assay procedure to recognize small variations of the measurand (e.g., concentration of target analyte) in the sample to be assayed as a function of significant variations of the measurement signal.

However, this definition does not take into account that the analytical sensitivity is determined by both the slope and the random variation of the calibration function. Even if the slope is steep the analytical sensitivity can be low if the calibration function has a high random variation; on the other hand, an assay with a moderate slope can have a high analytical sensitivity if the calibration function has a very low random variation.

Therefore, in a number of analytical situations, the above concept of analytical sensitivity may not be meaningful (2). More relevant is the

Limit of detection (detection limit), which is the lowest concentration (amount) of target analyte still precisely detectable by the measurement procedure.

The detection limit is usually estimated as the signal obtained from a "blank" plus 3 times the standard deviation observed at that concentration.

Limit of detection should not be used synonymously with analytical sensitivity. However, it is often considered to be the most useful indicator of assay sensitivity, since it reflects the sensitivity in a very low concentration range.

Precision (of measurement)

Closeness of agreement between independent results of measurements obtained under stipulated conditions (Ref.1: 4.85)

The precision of measurement cannot be given a numerical value in terms of the measurand. The degree of precision is usually expressed numerically by statistical measures of imprecision of measurements such as standard deviation (SD) or coefficient of variation (CV), that are inversely related to precision.

Imprecision of measurements depends solely on the dispersion of random error of measurement and does not relate to a true value of the measurable quantity. (Ref. 1:4.47)

Precision of a given measurement procedure is usually subdivided according to specified conditions into:

- **Repeatability**: closeness of agreement between results of successive measurements of the same measurand carried out under the same conditions of measurements (8:3.6).

It is often termed in laboratory medicine "within-assay, within-run, intra-assay, intra-run precision"

- **Reproducibility**: closeness of agreement between results of measurements of the same measurand carried out under changed conditions of measurements (8:3.7), involving e.g.time, operators and measuring systems (including different calibrators and/or reagent batches).

Two types of reproducibility are often used in laboratory medicine:

-"between-assay, between-run,inter-assay, inter-run precision" and

- "inter-laboratory precision".

The operative conditions should always be specified.

Trueness (of measurement)

Closeness of agreement between the average value obtained from a large series of results of measurement and a true value[f] (Ref.1:4.137, ISO 5725: 3.7)

Trueness of measurement cannot be given a numerical value. What is usually expressed numerically is the statistical measure **bias** that is inversely related to trueness

[e]Calibration: set of operations that establish, under specified conditions, the relationship between values of measurable quantities indicated by a measuring system and the corresponding values of quantities realized by measurement standards (REF.1: 4.22, VIM: 6.11)

[f]The indefinite article "a" rather than the definite article "the" is used in conduction with "true value" because there may be many values consistent with the definition of a given particular quantity (VIM 1993: 1.19: Note 3).

and is the difference between the expectation of the results of measurements and a true value of the measurand (Ref.1: 4.19)

Unfortunately, there is still controversy in the community of laboratory professionals both in Europe and USA regarding the replacement of the old term "accuracy" with the new term "trueness".

EDMA recommends using the term trueness instead of accuracy (although the latter is used also in the EU Directive 98/79/EC) since the ISO concept (8) is conclusive and by following the updated international standards the current confusion will be avoided.

Accuracy must only be used with the meaning illustrated below.

Accuracy (of measurement)

Closeness of the agreement between the result of a measurement and a true value of the measurand (Ref.1: 4.1, ISO 5725:3.6)

The general term Accuracy is used in VIM (1993) and ISO 5725 (1994) to cover both Trueness and Precision, whereas this term was used in the past to cover only the one component now named Trueness; this different interpretation also occurs in the IVD Directive 98/79/EC.

Indeed, the concept of accuracy of measurement is related to both trueness of measurement and precision of measurement, as illustrated in figure I, redrafted from Büttner (3) and Franzini (4).

While **trueness**, affected by systematic error, is normally expressed in terms of bias, and **precision**, affected by random error, is naturally expressed in terms of standard deviation, **accuracy** is affected by a combination of systematic and random effects that contribute as individual components of the total error of measurement. Accuracy is then inversely related to the "**uncertainty**" of measurement, which may be calculated according to GUM (Guide to the expression of uncertainty of measurement, ISO 1993).

As with trueness and precision, accuracy cannot be given a numerical value in terms of the measurand.

ANALYTICAL PERFORMANCE CHARACTERISTICS	KIND OF ERROR	TERM OF EXPRESSION
TRUENESS Justesse Esattezza Richtigkeit Veracidad	SYSTEMATIC ERROR Erreur systematique Errore sistematico Systematischer Fehler Error sistemático	BIAS Ecart systematique Scostamento sistematico Systematische Abweichung Apartamiento sistemático
ACCURACY Exactitude Accuratezza Genauigkeit Exactitud	TOTAL ERROR Erreur totale Errore totale Gesamtfehler Error total	UNCERTAINTY Incertitude Incertezza Unsicherheit Incertitud
PRECISION Fidélité Precisione Präzision Precisión	RANDOM ERROR Erreur casuelle Errore casuale Zufälliger Fehler Error casual	STANDARD DEVIATION Ecart type Deviazione standard Standardabweichung Desviación estándar

(from Büttner, 1994 and Franzini, 1996)

Figure I
Analytical performance characteristics. *Terms are expressed in English, French, Italian, German and Spanish languages.*

DIAGNOSTIC PERFORMANCE CHARACTERISTICS

The previously defined "analytical performance characteristics" refer to analytical systems where different amounts of the target analyte can be measured quantitatively; it is then of paramount importance to determine the different concentrations of analyte in the different samples to be tested, with adequate analytical sensitivity and specificity.

In other diagnostic systems, whose application is today continuously growing, the assay procedure is intended mainly to detect the presence or absence of a target marker of a specific disease.

Examples of this kind of tests include the tests for infectious diseases (e.g. for detection of AIDS, Viral Hepatitis, Rubella, CMV, Toxoplasmosis) or for tumour markers like e.g. PSA, CEA, CA-125.

In such cases the assay is designed to distinguish between negative and positive responses, i.e. between results below or above a pre-determined cut-off value. Therefore, terms like sensitivity and specificity refer to the reliability of the diagnostic test from a clinical standpoint and may be termed diagnostic sensitivity and diagnostic specificity.

These parameters depend on the value used for differentiating between non-healthy and non-diseased subjects but are independent of the prevalence of the disease.

Such performance parameters are evaluated according to a well-known statistical approach proposed by Galen and Gambino for medical diagnoses (5) and later adopted by an NCCLS guideline for immunological testing of infectious diseases (6).

Diagnostic sensitivity[g]

Probability of the assay procedure of scoring positive in infected (or pathological) samples (coming from subjects known to have the disease).

Diagnostic sensitivity is expressed as the ratio of true positive (TP) samples over the total number of samples which should give positive results, i.e. true positive (TP) plus false negative (FN) samples.

$$Diagnostic\ sensitivity = \frac{TP}{TP+FN}$$

This sensitivity may also be expressed as percentage, after multiplication by 100.

This kind of sensitivity is also called clinical sensitivity or nosographical sensitivity (Ref.1: 4.12. Note 3).

Diagnostic specificity[g]

Probability of the assay procedure of scoring negative in non infected (or non-pathological) samples (coming from subjects known to be free of the disease).

Diagnostic specificity is expressed as the ratio of the True negative (TN) samples over the total number of samples which should give negative results, i.e. True negative (TN) plus False positive (FP) samples.

$$Diagnostic\ specificity = \frac{TN}{TN+FP}$$

This specificity may also be expressed as percentage, after multiplication by 100.

This kind of specificity is also called clinical specificity or nosographical specificity (Ref.1 : 4.14 Note 2).

REFERENCES

1. Dybkaer. R. Vocabulary for use in Measurement Procedures and Description of Reference Materials in Laboratory Medicine - Eur. J. Clin. Chem. Clin. Biochem, (1997): 35(2) 141-173
2. Ekins R, Edwards P. On the meaning of "sensitivity". Clin.Chem. (1997)43:1824-1831
3. Büttner. J. Reference Materials and Reference Methods in Laboratory Medicine: A Challenge to International Cooperation. - Eur.J.Clin.Chem.Clin.Biochem. (1994)32(8):571-577

[g]Diagnostic sensitivity and diagnostic specificity are the performance parameters taken into account in the Common Technical Specifications (CTS), which are being developed for Annex II List A devices, according to the requirements of IVD Directive 98/79/EC (Art. 5.3).

4. Franzini C. Esattezza (?) dei risultati analitici - Biochim. Clin. (1996) 20: 434-435
5. Galen R.S. Gambino SR. Beyond normality: the predictive value and efficiency of medical diagnoses (1975) John Wiley & Sons, New York
6. NCCLS Document I/LA 18-A - Specifications for Immunological Testing for Infectious. Disease; Approved Guideline - NCCLS, Vol. 14, No. 18, Dec. 1994

ISO basic metrological standards

7. ISO 3534-1: 1993 - Statistics - Vocabulary and symbols. Part 1. Probability and general statistical terms.
8. ISO 5725-1: 1994 - Accuracy (trueness and precision) of measurement methods and results. Part 1: General principles and definitions
9. ISO (VIM) 1993 - International vocabulary of basic and general terms in metrology; issued by BIPM, IEC, IFCC, ISO, IUPAC, IUPAP and OIML
10. ISO (GUM) 1993 - Guide to the expression of uncertainty in measurement; issued by BIPM, IEC, IFCC, ISO, IUPAC, IUPAP and OIML

CEN and ISO draft standards on metrological issues

11. EN 12286: 1998 - Measurement of quantities in samples of biological origin - Presentation of reference measurement procedures.
12. EN 12287: 1999 - Measurement of quantities in samples of biological origin - Description of reference materials
13. EN/ISO 17511 - Measurement of quantities in samples of biological origin. Metrological Traceability of values assigned to calibrators and control materials.
14. ISO/EN 15195 - Requirements for Reference Measurement Laboratories