

VERIFICATION OF QUANTITATIVE ANALYTICAL METHODS IN MEDICAL LABORATORIES

VERIFIKACIJA KVANTITATIVNIH ANALITIČKIH METODA U MEDICINSKIM LABORATORIJAMA

Muhammad T. Abdel Ghafar¹, Muhammad I. El-Masry²

¹Departments of Clinical Pathology, Faculty of Medicine, Tanta University, Egypt
²Departments of Clinical Pathology, Faculty of Medicine, Kafr El-sheikh University, Egypt

Summary

Background: Globally, all medical laboratories seeking accreditation should meet international quality standards to perform certain specific tests. Quality management program provides disciplines targeted to ensure that quality standards have been implemented by a laboratory in order to generate correct results. The hallmark of the accreditation process is method verification and quality assurance. Before introducing a new method in your laboratory, it is important to assess certain performance characteristics that reflect the concept of method verification.

Methods: In this review, we illustrated how to verify the performance characteristics of a new method according to the recent guidelines. It includes an assessment of precision, trueness, analytical sensitivity, detection limits, analytical specificity, interference, measuring range, linearity, and measurement uncertainty.

Conclusions: Although the presence of several updated guidelines used to determine the performance characteristics of new methods in clinical chemistry laboratories, the real practice raised several concerns with the application of these guidelines which in need for further consideration in the upcoming updates of these guidelines.

Keywords: method verification, accreditation, quality, errors

Kratak sadržaj

Uvod: Sve medicinske laboratorije koje traže akreditaciju treba da ispunjavaju međunarodne standarde kvaliteta kako bi mogle da sprovode određene specifične testove. Program upravljanja kvalitetom pruža jasne smernice koje imaju za cilj da se obezbedi da laboratorija primenjuje standarde kvaliteta u cilju tačnosti rezultata. Osnova procesa akreditacije je verifikacija metoda i obezbeđivanje kvaliteta. Važno je da pre uvođenja nove metode u vašu laboratoriju procenite određene karakteristike performansi koje odražavaju koncept metode verifikacije.

Metode: U ovom pregledu smo ilustrovali kako sprovesti proveru karakteristika performansi nove metode prema nedavnim smernicama. To uključuje procenu preciznosti, istinitosti, analitičke osetljivosti, granice detekcije, analitičku specifičnost, interferencije, opseg merenja, linearnost i mernu nesigurnost.

Zaključak: Iako postoji nekoliko ažuriranih smernica koje se koriste za utvrđivanje karakteristika performansi novih metoda u laboratorijama za kliničku hemiju, stvarna praksa ukazuje na nekoliko problema u primeni ovih smernica koje je potrebno dodatno razmotriti u narednim verzijama ovih smernica.

Ključne reči: verifikacija metoda, akreditacija, kvalitet, greške

Address for correspondence:

Muhammad T. Abdel Ghafar
Department of Clinical Pathology, Faculty of Medicine
Aljaysh St, Medical Campus, Postal code: 31511
Tanta University, Egypt
Tel: +2 01008580811, Fax Number: 2+ 040 3282344
e-mail: mohammedtarek5514@yahoo.com

List of abbreviations: ILAC, International Laboratory Accreditation Cooperation; VIM3, International Vocabulary of Metrology 3; CLIA 88, Clinical Laboratory Improvement Amendments of 1988; SD, Standard deviation; CV, Coefficient of variation; TEa, Total error allowable; BNP, B-natriuretic peptide; PSA, Prostate-specific antigen; LD, Lactate dehydrogenase; TV, Target value; TSH, Thyroid-stimulating hormone; HIL, haemolysis/icterus/lipaemia; AMR, Analytical measurement range; CRR, Clinical reportable range; LOB, Limit of blank; LOD, Limit of detection; LOQ, Limit of quantification.

Introduction

The hallmark of the health care process in a country is the correct diagnosis, risk factor assessment, effective prophylactic and curative handling of the diseases. The pathologists and laboratory personals contribute mainly to diagnosis, effective treatment and follow up of patients. To achieve this role, efficient implementation of the quality system should be assigned in every laboratory seeking accreditation (1). Accreditation regarding the ISO/IEC 17025 for testing and calibration laboratories (2) and ISO 15189 for medical laboratories (2) considered any institution or program meets the standards of quality set to perform certain specific tests presented forth by a worldwide scale of accreditation bodies including International Laboratory Accreditation Cooperation (ILAC). It spans the managerial and technical capabilities of a laboratory. The hallmark of the process of accreditation is method validation, verification and quality assurance (3).

Quality is a maintained process aiming for the appropriate performance of the tests from the first time and every time. Quality assurance system is the whole framework incorporating the inside and outside laboratory activities with appropriate laboratory practice and enhanced management skills to ensure that the correct assay performed on a correct sample obtained from right subjects at the correct well equipped place, generating perfect result interpreted precisely based on correct reference data. Implementation of quality concepts in the medical laboratories requires the presence of a quality management program targeted to ensure the reliability of the results generated by the laboratory (4).

Quality control is the process concerned with the control of performance errors in the analytical testing phase and verification of test results. Quality control can be assigned as an internal control that is performed by laboratories offering the day to day basis working quality assurance or external control performed consequently by many laboratories, and their results are statistically compared and evaluated for proficiency testing (4).

Discussion

Validation versus Verification concepts

According to International Vocabulary of Metrology 3 (VIM3), verification is defined as 'provision of objective evidence that a given item fulfils specified requirements' (5) whereas validation is 'verification, where the specified requirements are adequate for the intended use' (5). In other words, validation is establishing the performance of a new diagnostic tool which is a manufacturer concern. However, verification is a process to determine performance characteristics before a test system is utilized for patient testing which is laboratory or user concern.

Once the method (reagents, procedure and the measurement instrument) has been manufactured by a company, a proper method validation emerged and its results should be provided to the user. In this situation, a laboratory method validation is not required and method verification is more convenient (6).

There have been several publications discussing the guidelines of method validation and verification adopted by national and international organizations, regulated and used by ISO 17025, ISO 15189 or by the Clinical Laboratory Improvement Amendments of 1988 (CLIA 88) (7) either generally (7–15) or specifically in the analytical chemistry (16, 17), toxicology (18), chemical pathology (19), food and drugs industry (20, 21).

Errors in measurements

The main purpose of method validation and verification is error assessment – what is the scope of possible errors within your laboratory assay results, and to what extent this degree of errors could affect clinical interpretations and, consequently, patient care.

Random Error

Random error is a type of measurement errors arising from the repeated assay. Hence it is considered a sort of imprecision issue and determined by the standard deviation (SD) and the coefficient of variation (CV) of test values (Figure 1). It is characterized by wide random dispersion of control values around the mean and exceeding both the upper and lower control limits (22). It reflects problems affecting measuring techniques as noises, sample preparation as improper pipetting, sample reagent mixing, incubation timing, and reaction environment as improper

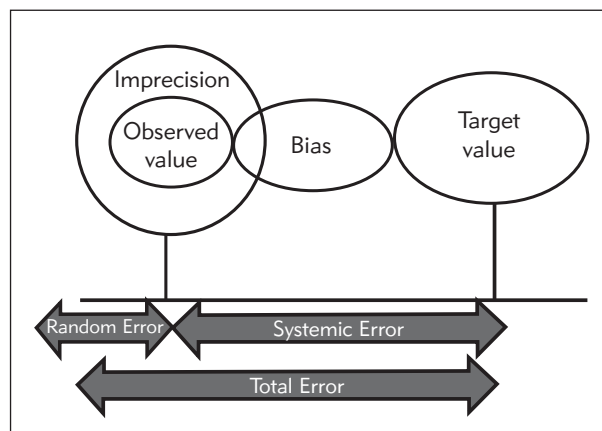


Figure 1 Total analytical errors in medical laboratories. Note: the total error is the sum of both random and systematic errors which in turn represent the bias and imprecision values.

Table I Equations used for estimation of different verification parameters.

Parameter	Equation no	Equation	Remarks
1	Random error	1	$S_{y/x} = \sqrt{\frac{\sum (y_i - Y_i)^2}{(n-2)}}$ Where: $S_{y/x}$ =Standard error of estimate $y_i - Y_i$ =The distance of each y-value from the regression line n =number of y-values
2	Systematic error	2	$Y = a + bX$ $a = \frac{(\sum y)(\sum x^2) - (\sum y)(\sum xy)}{(n(\sum x^2) - (\sum x)^2)}$ $b = \frac{n(\sum x - \sum y)(\sum y)}{(n(\sum x^2) - (\sum x)^2)}$ Where: Y =Reference method values X =Test method values a =y-intercept b =Slope of the regression line
3	Interference	3	$\text{Bias \%} = \frac{(\text{concentration with interference} - \text{concentration without interference})}{(\text{concentration without interference} \times 100)} \times 100$
4	Trueness	4	$\text{Verification interval} = X \pm 2.821 \sqrt{S_x^2 + S_a^2}$ Where: X =mean of tested reference material S_x =standard deviation of tested reference material S_a =Uncertainty of assigned reference material (Manufacturer SD of IQC, Uncertainty of PT sample, Uncertainty of calibrator) 2.821 is the 99 per cent point of, t , of the t-distribution with 9 degrees (2n-1) degrees of freedom
5	Precision	5A	$S_r = \sqrt{\frac{\sum (X_{di} - \bar{X}_d)^2}{D(n-1)}}$ Where: Σ =summation S_r =repeatability D =total days number (5), n =total replicates number per day (3), d_i =replicates result per day (3 replicates) x_d =average of all results for day (d).
		5B	$S_b = \sqrt{\frac{\sum (X_d - \bar{X})^2}{D-1}}$ where: X_d =average of all results for day d \bar{X} =average of all results.
		5C	$S_t = \sqrt{\frac{n-1}{n} (S_r^2 + S_b^2)}$ Where: S_t =Total within lab precision n = number of replicates per run (three).
6	Detection limits	6A	LOB= Mean blank+1.645 * SD blank
		6B	LOD= Mean blank+3.3 * SD blank
		6C	LOQ= Mean blank+10 * SD blank
		6D	LOD=3.3 σ /Slope
		6E	LOQ=10 σ /Slope
			Where: σ =the standard deviation of the response at low concentrations Slope=the slope of the calibration curve.
7	Error index	7	$\text{error index} = (x-y)/TEa$ Where: TEa (total allowable error)
8	Uncertainty	8A	$U_s = \frac{(SD)L1^2 + (SD)L2^2}{2}^{1/2}$ Where: U_s =Standard uncertainty (SD) L1 and (SD) L2= the average SD of each control level, respectively, for the past 6 months.
		8B	$U_B = \text{Test Result} - \text{Reference value}$ Where: U_B =Bias uncertainty
		8C	$U_c = [(U_s)^2 + (u_B)^2]^{1/2}$ $U_c = \left[\frac{U_s^2 + U_B^2}{2} \right]^{1/2}$ Where: U_c =the combined standard uncertainty
		8D	$U = U_c \times 1.96$ Where: U =the expanded uncertainty of the method 1.96=coverage factor

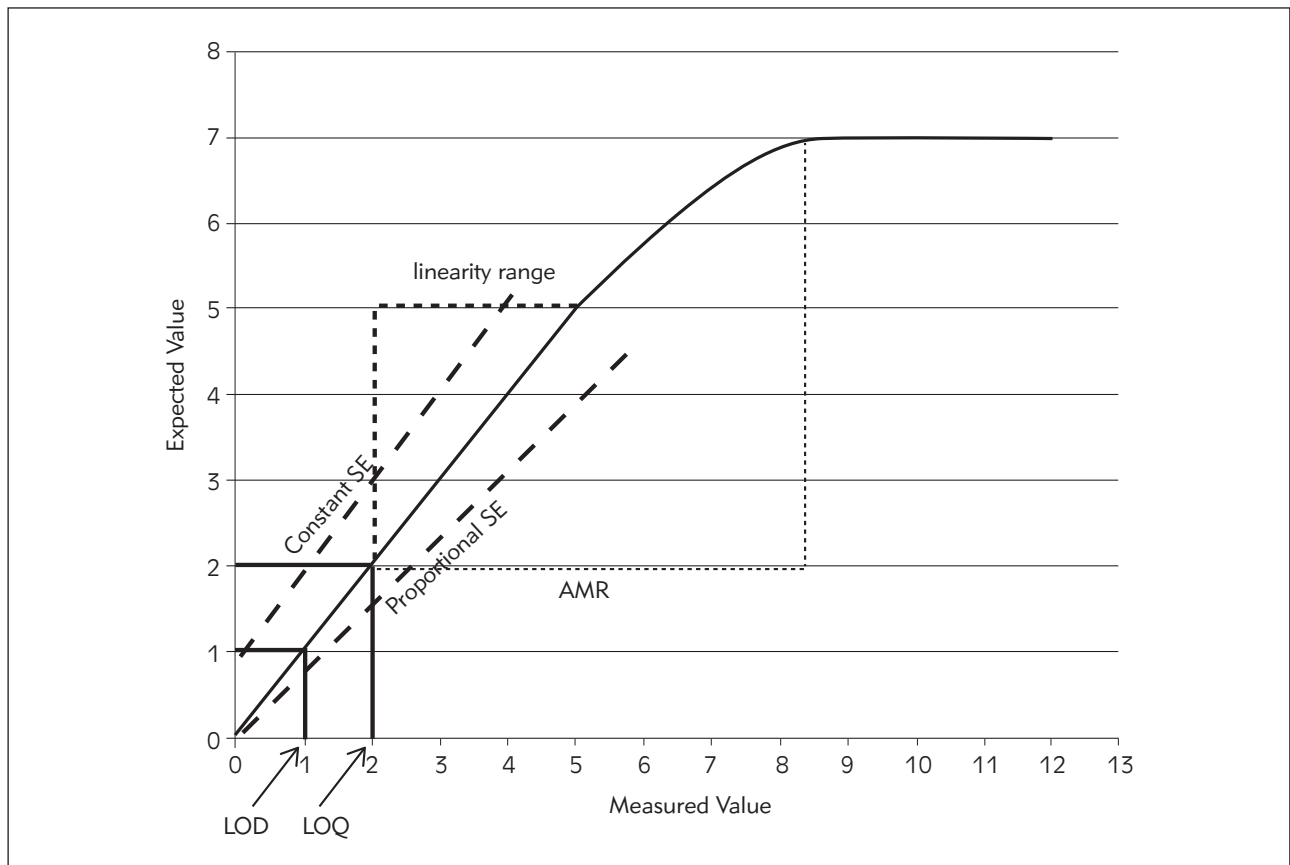


Figure 2 Point to point linear curve of the measured and expected values of a laboratory assay results of the serially diluted sample. Note: it represents the limit of detection (LOD), the limit of quantitation (LOQ), linearity range, analytical measurement range (AMR), constant and proportional systematic errors (SE).

temperature stability (23). Random error is calculated as the standard error of estimate (Sy/x) which is the SD of the points about the regression line. Sy/x represents the square root of the squared distance of results from the regression line divided by their numbers as in equation 1 (Table I). The higher the Sy/x , the wider is the scatter and the higher is the amount of random error (24).

Systematic Error

Systematic error reflects the inaccuracy problem in which the control observations are shifted in one direction of the mean and may exceeding one of the upper or lower limits. It is related mainly to the calibration problems such as impure, unstable calibration materials or improper standards preparation and inadequate calibration. In contrast to random errors, systematic errors could be evaded via correction of their causes (22). Systematic errors could be proportional or constant (Figure 2). Systematic error is detected by linear regression analysis with y-intercept of the linear regression curve indicates the constant error while the slope indicates the proportional error as in equation 2 (Table I) (24).

Total Error Allowable (TEa)

TEa is the total error permitted by CLIA 88, based on the medical requirements, the available analytical methods and compatibility with proficiency testing (PT) expectations. The CLIA 88 have published allowable errors for a wide range of clinical and laboratory tests (25). The original CLIA list of regulated analytes was based on medical practice in the past and missed several tests such as HbA1c or PSA that are used frequently at present. Recently, new proposed rules have been developed to expand the list of regulated analytes and define new criteria for acceptance limits to reflect the technological advancements and changes in the use of laboratory tests. The recent document proposed by Westgard JO and Westgard S, 2019 includes some tests deletions and many additions based on the PT availability, test significance, and feasibility of implementation. Several tests were added such as B-natriuretic peptide (BNP), hemoglobin A1c (HbA1c), prostate-specific antigen (PSA), troponin, C-reactive protein (highly sensitive), while other tests are deleted such as lactate dehydrogenase (LD) isoenzymes, quinidine, primidone, and procainamide. Moreover, the acceptance limits of other tests are changed from the target value (TV) \pm SD to

TV \pm % such as alpha-fetoprotein, complement, immunoglobulins, human chorionic gonadotropin, thyroid-stimulating hormone (TSH) and others (26).

The allowable errors limits assigned by CLIA denote the maximum error allowable by federally mandated proficiency testing. These performance characteristics are currently used to investigate the acceptability of clinical chemistry analyzer performance (27, 28). The total analytical error involves the sum of the random and systematic errors generated by the laboratory (Figure 1). The target of a quality assurance system is to maintain the total Analytical Error released by the laboratory below the total allowable error. Hence, the method should be modified or even rejected if the potential error generated by it exceeds the error allowable and could lead to misinterpretation (29). It is possible to calculate the total errors using the precision and trueness protocols described in this literature by considering the confidence-interval criteria and single-value criteria in the evaluation of error (30, 31).

Verification Outline Parameters

Analytical Specificity: Interference studies

Analytical specificity is referred to the measurement capability of an analytical method to exclude any interference in general, either cross-reacting substances, matrix effects and other effects as heterophilic antibodies, and to detect only the measurand of interest (5). Analytical specificity is verified using interference studies. The hemolysis, icterus, and lipaemia abbreviated as (HIL) together with the anticoagulants and preservatives as stated by Young (32) and Glick and Ryder (33, 34) to be the common interferences. They assessed the effect degree of interference produced by each interferent concentration on the analyte detection level by each of the chemistry instruments presented as »interferographs«.

The effect of 'HIL' on the analyte concentration can be avoided by using certain analytical systems that minimize haemoglobin, triglyceride, and bilirubin interference (35). This can be achieved by using certain assay methods that are not liable to interference. Icterus, for instance, has been shown to have no significant effect on some phosphate assays (Ortho Vitros 250/950 and Hitachi 700/900 platforms) (36). Some automated systems have a »HIL« index. However, it is not immune to interference itself. Samples containing extremely high bilirubin concentrations or gross haemolysis may produce non-specific lipemic flags. Moreover, raised lipemic index results may occur in samples containing high concentrations of immunoglobulins (37). This interference may result in a false high HIL index which can be corrected by correlation of HIL index results with the sample results for bilirubin and triglyceride level. The available data with analyst experience play a role in iden-

tifying the interferences that are most likely to affect the measurement method (38).

The interference study is done by adding the tested interferent in a maximum concentration to the patient sample. With generating an effect, its concentration is lowered gradually till reached the concentration at which estimated analyte results considered valid. The bias % could be calculated using equation 3 (Table I). This concentration cutoff of the interferent is thus determined and acceptable results could be generated if interferences are at levels below this (39). Samples with interference in some methods for example, in the diagnosis of thyroid diseases due to the presence of heterophilic antibodies should be tested by other methods for estimation of the degree of interference (6).

Analytical Sensitivity: Calibration curve, Detection limit

The sensitivity of a measurement system is the 'quotient of the change in an indication of a measuring system and the corresponding change in the value of a quantity being measured' (5). In other words, analytical sensitivity is the possibility of detection of a low concentration and or change in concentration of an analyte in a biological specimen. Concisely, sensitivity is the function of detection limit studies and the slope of the calibration (6).

Reportable Range (Analytical Measurement Range)

The reportable range is defined by CLIA (40) as »the span of test result values over which the laboratory can establish or verify the accuracy of the instrument or test system measurement response.« However, CAP defines the reportable range in the context of two distinct concepts; the analytical measurement range (AMR) and the reportable clinical range (CRR). AMR is the »range of analyte values that a method can directly measure on the specimen without any dilution, concentration, or other pretreatments not part of the usual assay process« (41). The AMR is determined by the manufacturers, and should not exceed the linearity range as demonstrated between the estimated and actual analyte concentrations (Figure 2).

Verification of AMR is performed using matrix-appropriate materials which could be calibrators or commercial linearity standards via the linearity experiment (41). They should be of low, mid, and high concentration covering the activity spectrum of the claimed AMR by the manufacturer or spanning from 50 to 150% of the target analyte concentration. Each of the samples with distinct concentrations should be assayed at least in duplicate to eliminate the imprecision effect. The measured values (on the X-axis) are

displayed versus the assigned values (on the Y-axis) in point to point line graph for each analyte with the assessment of the slope, intercept and correlation coefficient data (42). The reportable range should be within the manufacturer's AMR claims. The manufacturer's detection limit can be verified if the concentration of the highest calibrators or other used materials within the per cent TEa (10% to 15%) of the AMR upper limit and the lowest calibrator is within the minimum detectable level or approximating TEa per cent of the lower limit of the claimed AMR (43).

A clinically reportable range (CRR) is another proposed concept of the CLIA reportable range by CAP. It is similar to AMR but assumed to refer the wide range of analyte results, including those obtained with dilution or concentration of a patient sample. It should be applied for samples with analyte concentrations exceeding the AMR (44). Each laboratory should verify the maximal amount of dilution is allowed to cover this range without affecting the assay accuracy and exceeding manufacturer's protocols for dilution. With implying this allowed dilution, samples giving an inaccurate assay result should be reported as exceeding the upper CRR limits.

Accuracy and Trueness (Bias)

Accuracy is defined as the closeness of the agreement between the test result and accepted true value whereas, trueness is »closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value« (5). It can be evaluated as measurement bias which is the quantitative »estimate of a systematic error« (5).

Several factors found to share in the production of measurement bias. One of these factors is the presence of analytical interferents as haemolysis, lipaemia and icterus or cross-reaction as heterophilic antibodies in immunoassays. Others involve the improper calibration matrix or faulty preparation. Also, poor sample processing, transport and preservation may have a role in measurement bias (6).

Measurement bias can be determined via comparing the sample results estimated by a certain method with certified reference materials values purchased from companies or organizations of high metrological competence or participating in external quality control program that compare with consensus mean of estimated control values among the different laboratories using the same method or implying the interference study or assessing the recovery of the measurand in spiked natural samples (6).

The trueness can be assessed using two levels of quality control materials or certified reference materials as calibrators, or the proficiency testing material. If proficiency testing material used, the consensus mean and SD are used only if obtained from at least

ten peer participants. Each analyte is measured in three replicates for 5 days with the calculation of mean and SD of the 15 measurements. Calculation of verification interval for the bias was performed using equation 4 (*Table I*). If the verification interval includes the consensus peer group mean, trueness is accepted (45).

Precision (Replication study)

The precision is a hallmark of the verification process. It is defined as the closeness of agreement between the measurands results or test values provided by multiple estimations in replicates under particular conditions which could be similar or changed (5). These conditions could be the same method, by the same operator, on the same instrument, within the same laboratory, using the same reagent materials within a short period which represents the process entitled repeatability or over an extended period, but may include other conditions involving changes as new calibrations, calibrators, operators entitled as intermediate precision. Whereas, the previous process options but with different operators, on different instruments carried on different laboratories but with the same test reagents and the same test method was considered as reproducibility (46).

Accordingly, precision could be viewed at three levels; repeatability, intermediate precision and reproducibility. Or, precision could be regarded in accordance to the time component within the laboratory as repeatability (within run precision – intra-assay), between run precision (inter-assay), within-day precision, and between-day precision.

All over the entire precision estimation process, a single lot of reagents and calibrators should be used as using different lots could increase the observed variability. The material used for precision estimation should be stable, frozen pools of two or more concentration spanning the medical decision levels or measuring range of the instrument. Quality control material or pooled serum of minimum two concentration levels are the ideal materials for precision estimation (47).

For each of the provided levels of quality control materials or pooled sera, three replicates were measured per day (within run-repeatability-intra-assay), for five days (in between runs-intermediate precision). All of these replication results are collected and analyzed in the way to obtain the mean, standard deviation (SD) and coefficient of variations (CV) for each parameter level (45). The within run repeatability standard deviation (S_r) is calculated using the following equation 5A whereas inter-assay precision standard deviation (S_b) is calculated using equation 5B and the total intra-laboratory precision standard deviation is calculated by equation 5C (*Table I*) (45).

The standard deviations and CVs obtained for each analyte as estimated repeatability or within-laboratory precision should be statistically compared to the manufacturer's claim using analysis of variance (ANOVA) test of significance. The estimated within-laboratory or repeatability standard deviation should be less than the manufacturer's claim and hence demonstrated as precision consistent with the claim (45).

Detection Limits

It is impossible for any reagent to detect zero concentrations of the analyte regardless of the reported measuring range of some reagents to have in their reagent package. It is necessary to present enough analyte concentration to be measured as an analytical signal not as »analytical noise,« that produced in the absence of analyte (16). The lowest quantity of the analyte that could be reliably measured by the analytical methods is referred as the detection limit of the reagent. However, detection limits exist in three levels expressed as Limit of blank (LOB), Limit of Detection (LOD), and Limit of Quantification (LOQ) (Figure 2) (48).

Limit of Blank (LOB) is determined by EP17 as if repeated measures of the reagent blank devoid of any analyte were performed, and the highest reachable concentration confirmed (49). The limit of blank should be determined if the repeated blank assay revealed different results. LoB is determined by repeated measurements of a blank sample and calculation of the mean and the standard deviation (SD) of the obtained results using equation 6A (Table 1) (50).

Limit of Detection (LOD) is the minimal quantity of the analyte in a sample which produces excess signals than those generated by the blank but not necessarily expected as a true value. Limit of Quantification (LOQ) is the minimal concentration by which the analyte could be reliably measured with particular acceptable bias and imprecision value, commonly CV=20%. It is also defined without bias inclusion as »Functional sensitivity« (34, 51).

Several methods have emerged for estimation of detection and quantitation limits. These methods include visual definition, the signal-to-noise ratio (with 2 or 3 times for LOD and 10 times for LOQ), the standard deviation of the blank and the calibration line at low concentrations (11).

Implying the standard deviation of the blank method, the blank is measured ten or more times with calculation of the mean and SD of the blank results. The limits are consequently calculated using equation 6B, 6C (Table 1) (52). However, if the blank had no background noise, method implying the standard deviation of the response and the slope could be performed using a standard curve to determine LOD and LOQ (48). Five concentration of calibrators at very low values close to zero are measured six or

more times. The y-intercepts of regression lines represent the standard deviation of the response. Hence, the detection limits may be expressed following equations 6D, 6E (Table 1) (52).

Verification of reference intervals

Reference interval is the most important element in the analysis process as it is the value by which the clinician could interpret the patient laboratory results. Most of the laboratories pay little attention to establishing the reference intervals or even verifying the reference interval provided by the manufacturers.

Reference interval typically represents the central 95% interval of the observed analyte values among the healthy subjects. The role of laboratories is just verifying the reference intervals claimed by the manufacturer and »transferring« them to the laboratory. »Establishment« of reference intervals is another issue.

The reference interval could be verified by the laboratory by collecting samples from 20 reference individuals, selected from the studied population, and comparing the obtained test results to those of the manufacturer (53). The selection of the reference individual should obey certain exclusion criteria as (history of diseases, surgery, smoking, drugs, and conception) and partitioning criteria including age, gender, race, etc. (54). Written informed consent should be signed by all of the reference subjects and the pre-analytical precautions (e.g., fasting state, physical exercise, medication), with specimen collection variables as timing, tourniquet application, tube selection, and sample processing as centrifugation, transport and preservation should be assigned (55).

If no more than two of the 20 tested subjects' values (or 10% of the test results) fall outside the manufacturers claimed interval, this reference interval is considered verified. If more than 2 values exceed the limits, the claimed interval could not be adopted and more data collection is emerged (53).

Establishment of reference interval has emerged if verification of the claimed interval has been not assigned. The non-parametric approach is the most commonly used method for establishing reference intervals due to its ease and applicable for any distribution nature (56). In this method, 120 samples obtained from reference subjects and ranked by concentration order one simply puts the values obtained from reference individuals in rank order by concentration with rank 3 is the 2.5th percentile; rank 118 is the 97.5th percentile and ranks 1 and 7 define the 90% confidence interval of the 2.5th percentile, and ranks 114 and 120 define the 90% confidence interval of the 97.5th percentile. The reference interval occupies the central 95% of the distribution and the 90% confidence limits on both endpoints (56).

Comparability experiments

Comparability is the agreement between results obtained for an analyte using different measurement procedures (different methods or different instruments). As stated by Westgard et al. (31) and CLIA (57), it could be assessed using 40 to 100 samples assayed on both methods under examination (two field methods), or between one tested method and a reference method or on both instruments on the same day over 8 to 20 days (preferably within 4 hours), with specimens spanning the clinical range and representing a diversity of pathologic conditions. Daily analysis of two to five patient specimens should be followed for at least 8 days if 40 specimens are compared and for 20 days if 100 specimens are compared in replication studies (57).

Different methods could be used for the assessment of the proper comparability as Spearman coefficient of Correlation, Paired test for difference, linear regression as Deming regression or Passing-Bablok regression and Bland-Altman analysis.

Either a linearity regression analysis for the result obtained from the 2 instruments or methods and calculate the correlation coefficient »r« or error-index plot using equation 7 (Table I) can be used to assess the acceptability of comparability. The test-method results (y-axis) are displayed versus the comparative method (x-axis) if the two methods correlate perfectly, the data pairs plotted as concentrations values from the reference method (x) versus the evaluation method (y) will produce a straight line, with a slope of 1.0, a y-intercept of 0, and a correlation coefficient (*r*) of 1 (58). The results are considered to be comparable if no more than 10% of results' error-index exceed +1 or are less than -1 or the correlation coefficient »r« more than or equal 0.95 (59–61).

Uncertainty

It is a parameter reported with the test results characterizes the probability interval concerning the true value around the laboratory result (62). Hence, uncertainty offers a quantitative determination of the confidence range and the expected variability in a laboratory result when the test is performed on different instances. Consequently, both imprecision (SD) and inaccuracy (bias) are taken into account in the measurement uncertainty estimate. With negligible or corrected bias, the measurement uncertainty can be estimated using only the imprecision times a coverage factor (magnitude of the factor is based on the confidence level assigned to the result distribution). Thus, measurement uncertainty is adopted in standard deviation (SD) units, the coefficient of variation (% CV), confidence intervals (CI's) or ranges (R's) (63).

Two interchangeable approaches were employed for uncertainty estimation; the so-called »bot-

tom-up« and »top-down« approach. The »bottom-up« approach as proposed by the Guide to the Expression of Uncertainty of Measurement (GUM) identify, quantify and incorporate each source or origin of measurement uncertainty in the assay process into a final estimate entitled expanded combined uncertainty using statistical measures (64, 65). However, it is a complicated diverse approach applicable mainly for metrology institutions and accredited reference laboratories and challenging to be assigned by the medical laboratories (66). Instead, the »top-down« approach can be easily applied by clinical laboratories using intra-laboratory (for imprecision) and inter-laboratory (for bias estimation) quality control data (67–69).

The measurement uncertainty is estimated reliably via the collection of 180 measures over six months of at least two levels of a single lot of stabilized control materials (70, 71). For new methods, a minimum of 30 replicate determinations of appropriate control is required to calculate the standard deviation (SD). If bias is significant or known, calculate the combined standard uncertainty using the following equations 8A, 8B, 8C, 8D (Table I) (63).

As the usage of quality control samples may yield overestimated uncertainty values due to interference by matrix effects in stabilized control materials. So, split-sample techniques using patient samples could be more accurate and close to the real uncertainty estimates (72).

Main concerns and future perspectives

Several guidelines have been applied to determine the performance characteristics of new methods in clinical chemistry laboratories. However, several questions have emerged with the application of these guidelines in real practice.

Assessment of method trueness as previously illustrated, is hindered because of the presence of a wide range of verification interval. It can be raised due to the presence of discrepant results on repeated measurement which confer a wide standard deviation that reflected by a wide verification interval. Thus the mean result of the measurand may fall within the verification interval due to the inappropriate wide difference of the measurand results on repeated measures not because of the closeness of measurand results to the true value which is the hallmark of trueness definition.

The latest version of CLSI (EP15) for precision assessment expresses the main defect in the assessment of only 3 samples in the run which confer some statistical bias. Even though this guideline has improved relative to the previous version (CLSI EP5 A2) which employed only duplicate measures for 20 days, however, statistical bias may still exist due to the small number of replicates in the run.

The main concern arises with the application of the previously described guidelines is inappropriate acceptance criteria for some verification parameters. Regarding the reference interval verification, the acceptance criteria need to be revised for the presence of no more than 10% outside the validated range as the distribution shape of the results should be also considered. The sample results should be distributed normally. Skewed or bimodal distribution should not be accepted as it confers the inappropriate assignment of selection, exclusion and partitioning criteria for the included subjects for reference interval verification experiment. The recommended percentage (no more than 10%) of results that are accepted to fall outside the range of the claimed reference interval should be distributed equally outside the upper and lower limit of the claimed reference interval. Moreover, to our best knowledge, no acceptance criteria have been proposed for measurement uncertainty up to now.

Because of the previously described concerns experienced with the assignment of the current guidelines for method verification, it will be necessary to address these concerns in the future version of the verification guidelines. We recommend employing a much greater number of replicates in each run in the precision experiment to overcome the statistical bias of low sample size. Also, the estimated standard error of the mean of the measurand can be used instead of the standard deviation in the calculation of verification interval for the method trueness experiment. Moreover, the acceptance criteria for measurement uncertainty of a method are the main question and needs an answer in the future guidelines. Employment of CLIA 88 as acceptance criteria for measurement uncertainty of a method may be valuable with

the expanded uncertainty of a measurand should not exceed the accepted standard deviation proposed by CLIA 88 for this measurand.

Conclusion

Several global workgroups exert much effort in establishing protocols and guidelines to rule the process of verification in medical laboratories. Method verification is the main step in the process of enhancing the quality of laboratory results and a cornerstone in accreditation. Despite the presence of several updated guidelines used to determine the performance characteristics of new methods in clinical chemistry laboratories. However, real practice raised several concerns with the application of these guidelines which need further consideration in the upcoming updates of these guidelines.

Acknowledgements. We would like to express our special appreciation for Prof Amal El-Bendary for her intimate support.

Contributors. MTAG researched literature and conceived the study. MTAG and MIE wrote the first draft of the manuscript. MTAG reviewed and edited the manuscript and approved the final version of the manuscript.

Funding. No funding was received.

Ethical approval. Not applicable.

Conflict of interest statement

The authors stated that they have no conflicts of interest regarding the publication of this article.

References

- Handoo A, Sood SK. Clinical laboratory accreditation in India. *Clin Lab Med* 2012; 32(2): 281–92.
- International Organization for Standardization: ISO/IEC 17025 general requirements for the competence of testing and calibration laboratories, 2005.
- Burnett D, Blair C, Haeney MR, Jeffcoate SL, Scott KW, Williams DL. Clinical pathology accreditation: standards for the medical laboratory. *J Clin Pathol* 2002; 55(10): 729–33.
- Rahman T. Quality Assurance (QA) in Laboratory Testing. *AKMMC J* 2011; 2(2): 3–5.
- JCGM. International vocabulary of metrology – Basic and general concepts and associated terms (VIM) 2012, 3rd. Available from: www.bipm.org/en/publications/guides/vim.html.
- Theodorsson E. Validation and verification of measurement methods in clinical chemistry. *Bioanalysis* 2012; 4(3): 305–20.
- Clinical laboratory improvement amendments (1988). www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/IVDRegulatoryAssistance/ucm124105.htm
- Haeckel R. Verification, validation and evaluation of analytical procedures in laboratory medicine. *Clin Chem Lab Med* 2004; 42(1): 111–2.
- Milinković N, Ignjatović S, Šumarac Z, Majkić-Singh N. Uncertainty of measurement in laboratory medicine. *J Med Biochem* 2018; 37: 279–88.
- van Amsterdam P, Arnold M, Bansal S, Fast D, Garofolo F, Lowes S, et al. Building the Global Bioanalysis Consortium – working towards a functional globally acceptable and harmonized guideline on bioanalytical method validation. *Bioanalysis* 2010; 2(11): 1801–3.
- van Amsterdam P, Lausecker B, Luedtke S, Timmerman P, Brudny-Kloepfel M. Towards harmonized regulations for bioanalysis: moving forward! *Bioanalysis* 2010; 2(4): 689–91.

12. Stockl D, D'Hondt H, Thienpont LM. Method validation across the disciplines—critical investigation of major validation criteria and associated experimental protocols. *J Chromatogr B Analyt Technol Biomed Life Sci* 2009; 877(23): 2180–90.
13. Lippi G, Mattiuzzi C. Project management in laboratory medicine. *J Med Biochem* 2019; 38: 401–6.
14. European Medicines Agency concept paper/ recommendations on the need for a guideline on the validation of bioanalytical methods (2008). www.ema.europa.eu/pdfs/human/ewp/53130508en.pdf.
15. European Medicines Agency draft guideline on validation of bioanalytical methods (2009). www.ema.europa.eu/pdfs/human/ewp/19221709en.pdf.
16. Arslan DF, Karakoyun I, Basok IB, Aksit ZM, Celik E, Dogan K, Duman C. The effects of education and training given to phlebotomists for reducing preanalytical errors. *J Med Biochem* 2018; 37: 172–80.
17. Hartmann C, Smeyers-Verbeke J, Massart DL, McDowall RD. Validation of bioanalytical chromatographic methods. *J Pharm Biomed Anal* 1998; 17(2): 193–218.
18. Peters FT, Drummer OH, Musshoff F. Validation of new methods. *Forensic Sci Int* 2007; 165(2–3): 216–24.
19. Khatami Z HR, Sturgeon C, Kearney E, Breadon P, Kallner A. Measurement verification in the clinical laboratory: a guide to assessing analytical performance during the acceptance testing of methods (quantitative examination procedures) and/or analysers. The Scientific Committee of the Association for Clinical Biochemistry 2010. Available from: www.acb.org.uk/An%20Ver/Measurement%20verification_final_090608.pdf
20. DeSilva B, Smith W, Weiner R, Kelley M, Smolec J, Lee B, et al. Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharm Res* 2003; 20(11): 1885–900.
21. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* 2006; 24(8): 971–83.
22. Reichenbecher M, Einax JW. Types of errors in instrumental analysis. In: Reichenbecher M, Einax JW, editor. *Challenges in Analytical Quality Assurance* [Internet]. Springer-Verlag Berlin Heidelberg: Springer, 2011: 7–35.
23. Romero A, Gómez-Salgado J, Romero-Arana A, Antonio Gómez-Fernández J, Cobos A, Carmen Ramos M, Rosa Iglesias M. Preanalytical errors in emergency department samples: investigating error sources. *J Med Biochem* 2020; 39: 474–80.
24. McCudden CR, Rogers M, Erickson J, Erickson R, Willis MS. Method Evaluation and Quality Management. In: Bishop ML, Fody, Edward P, Schoeff LE, editor. *Clinical Chemistry: Techniques, principles, correlations*. 6th ed. Baltimore, MD 21201 Philadelphia, PA 19106: Lippincott Williams & Wilkins, 2010: 94–5.
25. Medicare, Medicaid and CLIA programs; regulations implementing the Clinical Laboratory Improvement Amendments of 1988 (CLIA) – HCFA. Final rule with comment period. *Fed Regist* 1992; 57(40): 7002–186.
26. Westgard JO, Westgard S. New CLIA Proposed Rules for Acceptance Limits for Proficiency Testing. Part I. Background and Summary/Comparison. *Fed Reg* 2019; 84: 1536–67.
27. Ehrmeyer SS, Laessig RH, Leinweber JE, Oryall JJ. 1990 Medicare/CLIA final rules for proficiency testing: minimum intralaboratory performance characteristics (CV and bias) needed to pass. *Clin Chem* 1990; 36(10): 1736–40.
28. Westgard JO, Seehafer JJ, Barry PL. European specifications for imprecision and inaccuracy compared with operating specifications that assure the quality required by US CLIA proficiency-testing criteria. *Clin Chem* 1994; 40(7 Pt 1): 1228–32.
29. McCudden CR, Rogers M, Erickson J, Erickson R, Willis MS. Method Evaluation and Quality Management. In: Bishop ML, Fody, Edward P, Schoeff, L.E., editor. *Clinical Chemistry: Techniques, principles, correlations*. 6th ed. Baltimore, MD 21201 Philadelphia, PA 19106: Lippincott Williams & Wilkins; 2010: 115–6.
30. Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. *Clin Chem* 1974; 20(7): 825–33.
31. Westgard JO, de Vos DJ, Hunt MR, Quam EF, Garber CC, Carey RN. Concepts and practices in the evaluation of clinical chemistry methods: IV. Decisions of acceptability. *Am J Med Technol* 1978; 44(7): 727–42.
32. Young DS. Effects of drugs on clinical laboratory tests. *Ann Clin Biochem* 1997; 34: 579–581.
33. Glick MR, Ryder KW. Analytical systems ranked by freedom from interferences. *Clin Chem* 1987; 33(8): 1453–8.
34. Glick MR, Ryder KW, Jackson SA. Graphical comparisons of interferences in clinical chemistry instrumentation. *Clin Chem* 1986; 32(3): 470–5.
35. Ryder KW, Glick MR. Erroneous laboratory results from hemolyzed, icteric, and lipemic specimens. *Clin Chem* 1993; 39(1): 175–6.
36. Farrell CJ, Carter AC. Serum indices: managing assay interference. *Ann Clin Biochem* 2016; 53(Pt 5): 527–38.
37. Fliser E, Jerkovic K, Vidovic T, Gorenjak M. Investigation of unusual high serum indices for lipemia in clear serum samples on siemens analysers dimension. *Biochem Med (Zagreb)* 2012; 22(3): 352–62.
38. Wu SL, Li W, Wells A, Dasgupta A. Digoxin-like and digitoxin-like immunoreactive substances in elderly people. Impact on therapeutic drug monitoring of digoxin and digitoxin concentrations. *Am J Clin Pathol* 2001; 115(4): 600–4.
39. Clinical and Laboratory Standards Institute (CLSI). Interference testing in clinical chemistry; Approved Guideline—Second Edition. CLSI document EP07-A2, (ISBN 1-56238-584-8), Clinical and Laboratory Standards Institute, 940 West Valley Road, Suite 100, Wayne, Pennsylvania 19087-1898 USA, 2005.
40. Centers for Disease Control and Prevention. CLIA law & regulations Web site. Available from: <http://wwwn.cdc>

- gov/clia/Regulatory. Updated May 31, 2012. Accessed December 30, 2013.
41. College of American Pathologists, Commission on Laboratory Accreditation. Chemistry and Toxicology Checklist. Northfield, IL: College of American Pathologists; 2012.
 42. Clinical and Laboratory Standards Institute (CLSI). Evaluation of the Linearity of Quantitative Measurement Approved Guideline-Second Edition, CLSI document EP6-A (ISBN 1-56238-498-8) Clinical and Laboratory Standards Institute, 940 West Valley Road, Suite 100, Wayne, Pennsylvania 19098-1898 USA, 2005.
 43. Westgard JO. Online Validation Training, Westgard QC, Inc. www.westgard.com, Sections 9-Determining Reportable Range.
 44. Killeen AA, Long T, Souers R, Styer P, Ventura CB, Klee GG. Verifying performance characteristics of quantitative analytical systems: calibration verification, linearity, and analytical measurement range. *Arch Pathol Lab Med* 2014; 138(9): 1173–81.
 45. Clinical and Laboratory Standards Institute (CLSI). User Verification of Performance for Precision and Trueness; Approved Guideline-Second Edition. CLSI document EP15-A2, (ISBN 1-56238-574-7), Clinical and Laboratory Standards Institute, 940 West Valley Road, Suite 100, Wayne, Pennsylvania 19087–1898 USA, 2005.
 46. ISO. Accuracy (trueness and precision) of measurement methods and results – Part 1: General principles and definitions. ISO 5725-1. 1994. Available from: <https://www.iso.org/standard/11833.html>.
 47. User Verification of Performance for Precision and Trueness; Approved Guideline-Second Edition. CLSI document EP15-A2, (2005).
 48. Little T. Method Validation Essentials, Limit of Blank, Limit of Detection, and Limit of Quantitation. *BioPharm International* 2015; 28(4): 48–51.
 49. Clinical and Laboratory Standards Institute (CLSI). Protocols for Determination of Limits of Detection and Limits of Quantitation, Approved Guideline. CLSI document EP17. Clinical and Laboratory Standards Institute, 940 West Valley Road, Suite 100, Wayne, Pennsylvania 19087–1898 USA, 2004.
 50. Armbruster DA, Pry T. Limit of blank, limit of detection and limit of quantitation. *Clin Biochem Rev* 2008; 29 Suppl 1: S49–52.
 51. Hay ID, Bayer MF, Kaplan MM, Klee GG, Larsen PR, Spencer CA. American Thyroid Association assessment of current free thyroid hormone and thyrotropin measurements and guidelines for future clinical assays. The Committee on Nomenclature of the American Thyroid Association. *Clin Chem* 1991; 37(11): 2002–8.
 52. NCCLS, EP17-A, Protocol for Evaluation of Limits of Determination and Limits of Quantitation Approved Guideline, 2004.
 53. Clinical Laboratory and Standards Institute. »How to Define and Determine Reference Intervals in the Clinical Laboratory; Approved Guideline – Second Edition« CLSI document C28-A2. Wayne, PA: 2000; 24.
 54. PetitClerc C, Solberg H.E. Approved recommendation on the theory of reference values. Part 2. Selection of individuals for the production of reference values. *J Clin Chem Clin Biochem* 1987; 25: 639–44.
 55. Solberg HE, PetitClerc C. Approved recommendation (1988) on the theory of reference values. Part 3. Preparation of individuals and collection of specimens for the production of reference values. *Clin Chim Acta* 1988; 177(3): S3–11.
 56. Clinical Laboratory and Standards Institute. »How to Define and Determine Reference Intervals in the Clinical Laboratory; Approved Guideline – Second Edition« CLSI document C28-A2. Wayne, PA: 2000; 13–22.
 57. National Committee for Clinical Laboratory Standards (NCCLS). Approved guideline for method comparison and bias estimation using patient samples. Villanova, Pa.: NCCLS, 2002. (Document no. EP09-A2).
 58. Shahzad K, Kim DH, Kang MJ. Analytic evaluation of the beta-human chorionic gonadotropin assay on the Abbott IMx and Elecsys2010 for its use in doping control. *Clin Biochem* 2007; 40 (16–17): 1259–65.
 59. Cornbleet PJ, Gochman N. Incorrect least-squares regression coefficients in method-comparison analysis. *Clin Chem* 1979; 25(3): 432–8.
 60. Feldmann U, Schneider B, Klinkers H, Haecel R. A multivariate approach for the biometric comparison of analytical methods in clinical chemistry. *J Clin Chem Clin Biochem* 1981; 19(3): 121–37.
 61. Westgard JO. Precision and accuracy: concepts and assessment by method evaluation testing. *Crit Rev Clin Lab Sci* 1981; 13(4): 283–330.
 62. The American Association for Laboratory Accreditation (A2LA). Frederick (MD): The American Association for Laboratory Accreditation; c2002. P102 – A2LA Policy on Measurement Traceability. 2008.
 63. Moses G, Crawford L. Traceability and Measurement Uncertainty for Medical Laboratories. Quality Management Program – Laboratory Services Ontario Laboratory Accreditation 1500–393 University Avenue Toronto, Ontario, Canada M5g 1e6. 2011.
 64. Crawford L, Moses G. Traceability and uncertainty of measurement for medical laboratories–OLA's expectations. QMP–LS News, No. 118. Toronto (ON): QMP–LS QView. c2007. Available from: <https://home.qmpls.org/qview/MainForm.aspx?dl=402885>
 65. JCGM 100: 2008 Evaluation of measurement data – Guide to the expression of uncertainty in measurement.
 66. Lee JH, Choi JH, Youn JS, Cha YJ, Song W, Park AJ. Comparison between bottom-up and top-down approaches in the estimation of measurement uncertainty. *Clin Chem Lab Med* 2015; 53(7): 1025–32.
 67. National Association of Testing Authorities (NATA). Technical Note #33 – Guidelines for estimating and reporting measurement uncertainty of chemical test results. Australia; 2006.
 68. Magnusson B, Naykki T, Hovind H, Krysell M. Handbook for Calculation of Measurement Uncertainty in environmental laboratories. NORDTEST Report TR 537 – 2003–05.
 69. Ceriotti F, Brugnoli D, Mattioli S. How to define a significant deviation from the expected internal quality control result. *Clin Chem Lab Med* 2015; 53(6): 913–8.

70. Feinberg M, Boulanger B, Dewe W, Hubert P. New advances in method validation and measurement uncertainty aimed at improving the quality of chemical data. *Anal Bioanal Chem* 2004; 380(3): 502–14.
71. Patriarca M, Castelli M, Corsetti F, Menditto A. Estimate of uncertainty of measurement from a single-laboratory validation study: application to the determination of lead in blood. *Clin Chem* 2004; 50(8): 1396–405.
72. Norheim S. Computer support simplifying uncertainty estimation using patient samples 2008. Available from: <http://liu.diva-portal.org/smash/record.jsf?pid=diva2:417298>

Received: October 27, 2019

Accepted: January 03, 2020