



Bias in Laboratory Medicine: The Dark Side of the Moon

Abdurrahman Coskun , M.D.

Department of Medical Biochemistry, School of Medicine, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey

Physicians increasingly use laboratory-produced information for disease diagnosis, patient monitoring, treatment planning, and evaluations of treatment effectiveness. Bias is the systematic deviation of laboratory test results from the actual value, which can cause misdiagnosis or misestimation of disease prognosis and increase healthcare costs. Properly estimating and treating bias can help to reduce laboratory errors, improve patient safety, and considerably reduce healthcare costs. A bias that is statistically and medically significant should be eliminated or corrected. In this review, the theoretical aspects of bias based on metrological, statistical, laboratory, and biological variation principles are discussed. These principles are then applied to laboratory and diagnostic medicine for practical use from clinical perspectives.

Key Words: Bias, Confidence interval, Diagnostic error, Quality control, Total quality management, Uncertainty

Received: January 21, 2023

Revision received: April 15, 2023

Accepted: August 4, 2023

Corresponding author:

Abdurrahman Coskun, M.D.

Department of Medical Biochemistry,
School of Medicine, Acibadem Mehmet Ali
Aydinlar University, Kayisdagi cad. No 32,
Atasehir, Istanbul 34752, Turkey
E-mail: coskun2002@gmail.com



© Korean Society for Laboratory Medicine

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Physicians increasingly use patients' laboratory test results for disease diagnosis, patient monitoring, treatment planning, and the evaluation of treatment effectiveness [1-4]. Laboratory test values do not represent exact data but rather vary within certain confidence limits due to systematic or random variation [1,5,6]. Bias is the systematic deviation of laboratory test results from the actual value. A significant bias in measurement results can cause misdiagnosis or misestimation of disease prognosis and increased healthcare costs [7-9]. Some causes of bias have been presented previously [10-12]. Although bias has been extensively discussed within the last decades, it has rarely been properly addressed, representing the "dark side of the moon," particularly in the field of laboratory medicine. To handle bias properly, the terminology, pre-analytical and analytical conditions, and statistical techniques used to evaluate bias must be standardized [13-19]. Notably a "purist" approach in which "ev-

erything is expected to be perfect" is not a pragmatic method for solving common laboratory medicine problems. Laboratory resources should not be wasted on correcting insignificant and minor differences that do not affect clinical decisions.

Biological and non-biological samples have distinct properties. Non-biological samples are affected by pre-analytical or analytical variations, whereas biological samples such as whole blood, plasma, and urine are affected by both pre-analytical and analytical variations as well as by biological variation (BV) [1]. This is particularly evident in sequential sampling. Within the human body, analytes fluctuate around (homeostatic) set points, which is known as the within-subject BV [20]. Measurement results of patient samples vary over time due to BV, even if pre-analytical and analytical variations are negligible [1, 20]. Deviations that cannot be tolerated in industrial measurements can be tolerated in medical laboratories. According to Albert Einstein, "everything should be made as simple as possible, but not simpler" [21]. The practical aspect should be as simple as possible but

not at the expense of the theoretical background of the concepts under study.

In this review, bias is evaluated from metrological and statistical, laboratory, and clinical perspectives. The theoretical aspects of bias based on metrological and BV principles are summarized, and these principles are applied to laboratory and diagnostic medicine for practical use.

METROLOGICAL AND STATISTICAL PERSPECTIVES OF BIAS

Terminology

The terms bias, trueness, and systematic error are interrelated [13, 22]. According to Vocabulary International Metrology (VIM) edition 3, measurement bias is the “estimate of a systematic measurement error” (2.18) [13]. Measurement trueness is defined as the “closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value” (2.14) and “is inversely related to systematic measurement error, but is not related to random measurement error” (Note 2). Instrumental bias is defined as the “average of replicate indications minus a reference quantity value” (4.20).

Based on these definitions, estimating bias requires two main components: (1) a reference quantity or assigned value and (2) a replicate measurement of the quantity (Fig. 1). If one of these two components is unknown or has not been properly determined, bias cannot be estimated correctly.

Mathematically, bias can be calculated using the following equation:

$$\text{Bias}(A) = O(A) - E(A) \quad (1)$$

where $O(A)$ and $E(A)$ are observed (measured) and expected values of analyte A, respectively. In practice, $O(A)$ and $E(A)$ correspond to the mean of repeated measurements and reference data, respectively.

Estimated bias is not a precise value. Each measurement result has a systematic and random component, and the mean of repeated measurements has a degree of variation depending on the probability selected.

Types of bias

Measurement accuracy varies across different concentrations, and the linearity of measurement methods is lost in the region near the limit of quantitation and upper measurement limits [23, 24]. Measured bias can be constant or proportional. In constant bias, the difference between the target and measured values is constant, whereas in proportional bias, the difference between the target and measured values is proportional to the amount of the measurand (i.e., it is a function of the measurand concentration) (Fig. 2) [25-28]. The bias between two methods can be evaluated using a Bland-Altman graph, which is a powerful graphic tool for evaluating the agreement between two methods, particularly when it is correctly interpreted and based on an adequate sample size [29-31]. Passing-Bablok regression analysis can also be used to evaluate the presence of constant and proportional bias between two methods (Fig. 2) [32].

The regression equation for two methods can be written as follows:

$$y = ax + b \quad (2)$$

where a is the slope and b is the intercept.

If $y = x$ (i.e., $a = 1$ and $b = 0$), it can be considered that there is no significant bias between two methods or instruments. Note

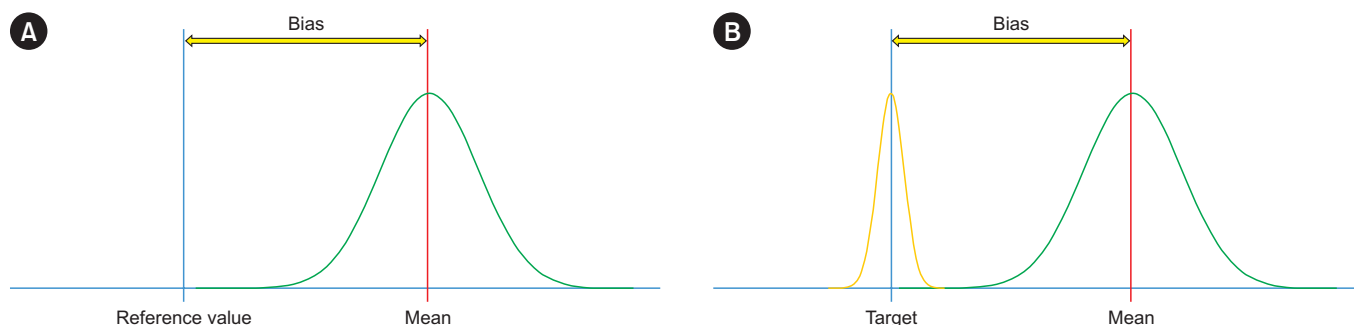


Fig. 1. Bias is the difference between the target/reference value and the mean value of repeated measurements of the sample. (A) The estimation of bias requires two main components: (1) reference quantity or assigned value and (2) replicate measurements of the quantity. (B) If the reference quantity value is not available, an assigned value can be used to estimate the bias.

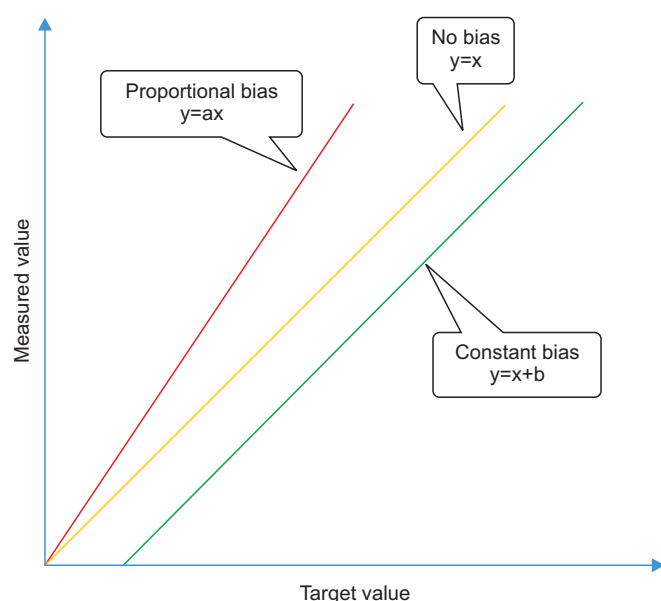


Fig. 2. Constant and proportional bias. If $a \neq 1$ and $b \neq 0$, the significance of a and b should be evaluated using the 95% confidence intervals of a and b .

that in the case of $a \neq 1$ and $b \neq 0$, the significance of a and b should be evaluated using the 95% confidence intervals (CIs) of the slope and intercept. If the 95% CI of a includes 1, it can be concluded that there is no significant proportional bias between two methods. Similarly, if the 95% CI of b includes 0, it can be concluded that there is no significant constant bias between two methods (Fig. 2). Details for detecting proportional and constant bias have been presented previously [33-35].

Measurement of bias

Practically, bias measurement requires the availability of reference values and the mean of repeated measurements (Fig. 1A). The reference quantity value can be determined using certified reference materials (CRMs) or fresh patient samples measured using reference methods [36, 37]. If the reference quantity value is not available, an assigned value can be used to estimate the bias (Fig. 1B). Bias should not be estimated by simply subtracting the mean of the measured value from the reference or assigned value. Additionally, the significance of the bias should be evaluated and confirmed. Subtracting a single measurement result from a reference or assigned value does not yield bias; this is a common error made in medical laboratories, particularly when calculating the sigma metric (SM) [38] of the measurement procedure. The characteristics of bias depend on the measurement procedure and the duration of data collection

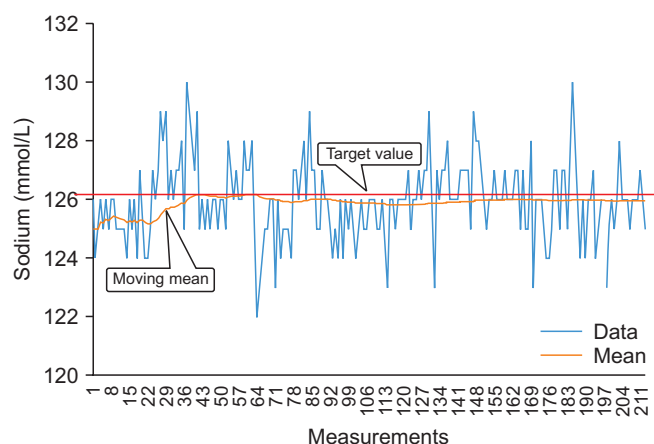


Fig. 3. Characteristics of bias change over time. Data collected for sodium under intermediate precision or reproducibility conditions contain both random and systematic (bias) variations.

for bias estimation; therefore, measurement conditions have a significant influence on bias and its significance [12, 39].

Measurement conditions

Metrologically, three bias measurement conditions [39] can be defined, which are discussed as follows:

Repeatability conditions

For repeatability conditions, (1) the measurement procedure, instrument, operating conditions, operator, and location (laboratory) must be the same, and (2) the repeated measurements must be completed within a short period (no longer than one day) and in a single run.

Repeated measurements under repeatability conditions yield the smallest random variation, and if a bias exists, it can be easily detected.

Intermediate precision conditions

Intermediate precision conditions are referred to as the variation in a measurand analyzed in a single laboratory over several months using different instruments, operating conditions, operators, reagents, and calibrators. Repeated measurements under *intermediate precision conditions* show higher random variation than those under repeatability conditions, and if a bias exists, it may be difficult to detect (Fig. 3).

Reproducibility conditions

In addition to repeatability and intermediate precision, reproducibility conditions also include the total variation of different laboratories. The variation of repeated measurements under repro-

ducibility conditions includes all types of variations originating from different sources, such as measurement procedures, instruments, operating conditions, operators, and locations (laboratories) over several months. Among repeated measurements, those conducted under reproducibility conditions demonstrate the highest degree of random variation compared to those conducted under intermediate precision conditions, and if bias exists, it may be difficult to detect.

Significance of bias

Since bias is defined as the difference between a target value and the mean of repeated measurements (Fig. 1), the significance of a calculated bias should be evaluated before further calculations [40, 41]. The significance of bias can be evaluated using *t*-test. Alternatively, while it may not be statistically accurate in some instances, the significance of bias can be evaluated using the 95% CI in a very practical context. This evaluation is more visual in nature, as opposed to a strict statistical assessment. If the 95% CI of the mean of repeated measurement results and the target value overlap, bias is not considered to be significant, whereas if there is no overlap bias is considered to be significant (see Supplemental Method for an explanation and Supplemental Tables S1, S2 for practical examples). Note that since bias and imprecision are related, the imprecision of the method significantly impacts the significance of the bias [14].

CLINICAL LABORATORY PERSPECTIVE OF BIAS

In clinical laboratories, bias should be evaluated using fresh patient samples or commutable samples. The use of commutable samples in clinical laboratories has been reviewed previously [42-44] (see also the “Commutability and bias” section below). As the analytical responses of fresh patient samples and commutable samples are similar [45], commutable samples can represent fresh patient samples in performance evaluations of measurement procedures. A pragmatic procedure including analytical performance of the instruments, sample types, measurement procedures, data collection period, and statistical techniques is required to handle bias in clinical laboratories.

Analytical performance specifications (APSs)

APSs are a set of criteria that specify the quality required for the analytical performance of measurement procedures to deliver laboratory test results that achieve the best possible health outcomes for patients without causing harm [46]. In daily practice, the analytical performance of measurement systems is evalu-

ated by calculating the systematic and random variations, namely, bias and imprecision. In addition to bias and imprecision, total allowable error (TEa) has been accepted as a component of APS over the last four decades and has been used for various purposes. However, TEa has limitations, including a lack of definition in the VIM and a lack of fit in metrology [13]. The standard equation of TEa is as follows:

$$\text{TEa} = \text{Bias} + 1.65 \text{ CV} \quad (3)$$

In the linear combination of bias and the CV, only one side of the CV (normal distribution) is included in the calculation; therefore, the appropriate multiplier for a 95% probability is 1.65. The CV represents the imprecision of the measurement procedure.

According to the Guide to the Expression of Uncertainty in Measurement (GUM) [47], bias should be corrected and known bias should not be included when calculating APSs and other indicators. Because of the limitations mentioned above, TEa should not be used in laboratory medicine to represent a tolerance limit and/or measurement uncertainty (MU). In metrology, total error (not TEa) corresponds to accuracy. Accuracy is the combination of bias and imprecision, which can be used to evaluate the error of a single measurement result. Therefore, accuracy is used in External Quality Assessment Scheme (EQAS) programs. In clinical laboratories, TEa has been incorrectly used or recommended instead of the tolerance limit (TL) [38], MU [48], or other reliable indicators. The TL or tolerance interval contains a specified proportion of units from the sampled process or population; the detailed calculation method is presented previously [49]. Although TEa explains many phenomena in laboratory medicine, in reality, it cannot solve any problems and has no place in metrology. Therefore, it is not considered a part of APSs in this review.

To prevent misdiagnosis, acceptable limits for bias should be determined for the measurand measurement results reported to patients.

Models for deriving acceptable bias

The acceptable limits or TLs can be determined based on various factors, including customer requirements, clinical needs, established guidelines, and statistical methodologies such as the Taguchi loss function [50, 51]. Despite intensive efforts, the acceptable limits for laboratory analytes measured in biological samples are not well-defined. Two international meetings were organized to define the criteria for APSs in medical laboratories [52-55].

Stockholm and Milan consensuses

The first conference on global analytical quality specifications was held in Stockholm in 1999. According to the Stockholm consensus, APSs are based on five hierarchical criteria [54] with the highest-ranking criterion given the highest priority. If it is not feasible to apply the first criterion, then the second criterion should be utilized, and so on, in descending order of priority [1, 53].

Although the Stockholm consensus was aimed at defining acceptance criteria for APS based on medical needs, it did not have the expected effect in laboratory medicine for 15 years. In 2014, the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) organized a strategic conference in Milan to revise the Stockholm consensus, which was named the Milan consensus [55]. The APS criteria were revised and simplified based on clinical and laboratory requirements, and technological achievability was considered a critical parameter.

In both the Stockholm and Milan consensuses, the first criterion in the hierarchy is based on the effect of analytical performance on clinical outcomes and the second criterion is based on the components of the measurands' BV. In the Milan consensus, the third and fourth criteria of the Stockholm consensus are excluded and the APS criteria are simplified and limited to three criteria. The last criterion (the fifth in the Stockholm consensus and the third in the Milan consensus) is based on the state of the art of the measurement, i.e., the highest level of analytical performance technically achievable.

Since 1999, APSs of laboratory analytes based on clinical decision limits have been difficult to define. A single laboratory test can be used for numerous clinical purposes, resulting in various associated clinical decision limits. Although the first criterion is excellent, it is not widely used in practice. No single model can be applied to all measurands, and therefore, applying different models is a pragmatic approach to estimate the APSs of different measurands. Alternatively, APSs can be based on a compromise between different models [56]. Selecting the most appropriate model for a measurand can be challenging. Ceriotti, *et al.* [57] proposed a simple and pragmatic workflow to select the most appropriate models for various measurands.

Because of the nature of laboratory tests, the first criterion of both the Stockholm and Milan consensuses is not widely applied in practice, and therefore, the APSs of laboratory analytes are mainly based on the second criterion, BV.

Acceptable bias derived from components of BV of the measurands Unlike clinical outcomes, BV in an analyte can be easily estimated. The EFLM BV Working Group has greatly contributed to

laboratory test BV. In the last decade, the group has developed checklists [58] and standards [59] for BV studies, measured the BV of numerous analytes [60] using a strict protocol [61], performed meta-analyses of BV data of numerous analytes published in the literature [62-69], and launched a BV database for most laboratory tests [70]. The database is dynamic and is updated when a new paper on BV is published.

The question remains as to how to develop a model to estimate acceptable limits for bias based on BV. There is a model for the acceptable limits of imprecision based on BV. This model is based on the contribution of analytical variation to the total variation, which is calculated using the following equation:

$$CV_T^2 = CV_A^2 + CV_I^2 \quad (4)$$

A triple model for performance evaluation has been proposed as follows [1]:

Desirable performance is defined as $CV_A < 0.50 CV_I$. Here, the maximum contribution of analytical variation to total variation is 12%.

Optimum performance is defined as $CV_A < 0.25 CV_I$. Here, the maximum contribution of analytical variation to total variation is 3%.

Minimum performance is defined as $CV_A < 0.75 CV_I$. Here, the maximum contribution of analytical variation to total variation is 25%.

The acceptable limit for imprecision can be modeled based on the contribution of the analytical variation to total variation; however, this method cannot be used for bias. Because bias is a linear parameter, the reference interval (RI) can be used to model the limits of acceptable bias [71-74].

Physicians generally use conventional RIs for clinical decisions. If a patient's laboratory results are within the RI, they are accepted as normal; otherwise, they are considered abnormal. Therefore, bias has various effects on disease misdiagnosis [71]. A positive bias in laboratory test results will increase the percentage of RIs outside the upper limits (ULs) and decrease the percentage outside the lower limits (LLs). A negative bias will have the opposite effect. Laboratory test results within RIs are considered normally distributed. The geometrical shape of the normal distribution is not rectangular [75, 76]; therefore, the effect of bias on the UL and LL is not symmetric. Using normal distribution mathematics, we can calculate the percentage of individuals outside the RI when bias exists. This can be used to model the acceptable limit for bias based on BV as with imprecision [77].

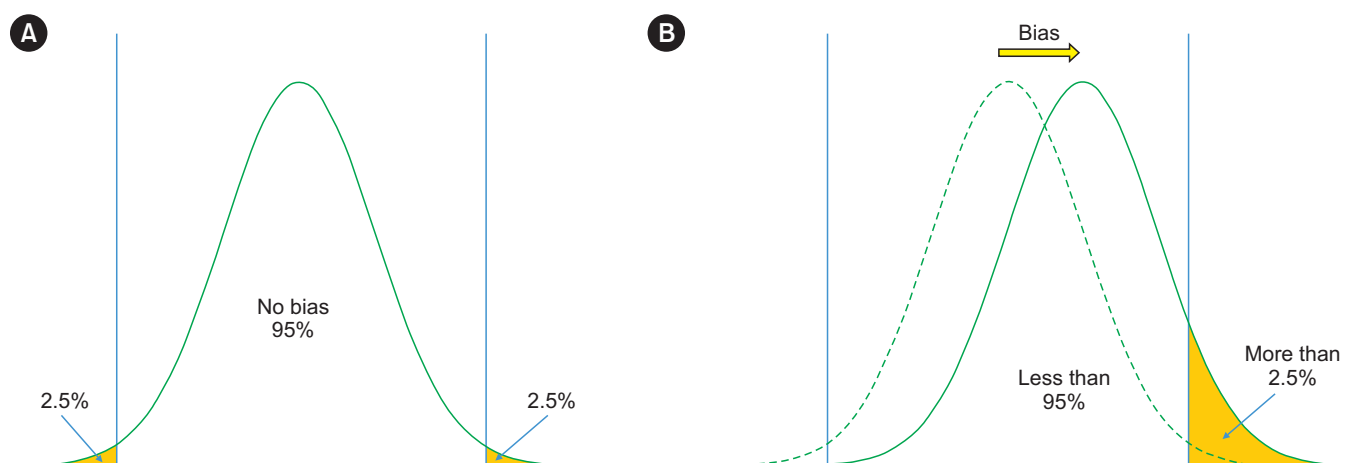


Fig. 4. Effect of bias on population values inside and outside the reference interval. Given the geometric shape of the normal distribution curve, an increase in bias results in an exponential shift of the population from within the reference intervals to beyond them. (A) When bias = 0, 5% of the population is situated outside the reference intervals. (B) When bias > 0, the proportion of the population located outside the reference intervals exceeds 5%.

The conventional population-based RI comprises both between- and within-subject BV, so that acceptable bias can be calculated according to their Gaussian combination. The model for acceptable bias is based on the acceptable number of people outside the RI when bias exists. Details of the model to derive the acceptable bias from BV data have been reported previously [1]. As shown in Fig. 4A, the LL and UL of the RI are set to cover 95% of the population values. If the measurement procedure has a positive bias, the curve will shift to the right (Fig. 4B); > 2.5% of the population will have values higher than the UL and < 2.5% of the population values will be outside the LL. Because of the bell shape of the curve, the increase in population values outside the UL will be higher than the decrease in population values outside the LL. The change in population values inside and outside of the RI caused by increasing bias is presented in Fig. 4B.

Mathematically, the area under the curve (AUC) can be used to calculate the population values inside and outside of the RI as follows:

$$AUC = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]} d(x) \quad (5)$$

where σ is the standard deviation, μ is the mean, and x is the variable.

From Equation 5, the AUC within the RI can be calculated using Equation 6:

$$AUC_{RI-inside} = \int_{LL}^{UL} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]} d(x) \quad (6)$$

From Equations 5 and 6, the population values outside the RI can be calculated according to Equation 7:

$$AUC_{RI-outside} = \int_{-\infty}^{LL} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]} d(x) + \int_{UL}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]} d(x) \quad (7)$$

Equations 5–7 are very complex and cannot be used in daily practice. Instead, z-transformation and a z table can be used to calculate the AUC and population values inside and outside the RI. A practical method is presented in Fig. 5. When bias exists, this graph can be used to easily estimate the population values outside and inside the RI.

Based on the normal distribution, a triple model for bias has been suggested (see Supplemental Table S3):

$$CV_B^2 = CV_i^2 + CV_g^2 \quad (8)$$

where CV_B is the total BV.

Desirable performance is defined as $B_A < 0.250CV_B$. Here, 0.8% of the additional population will be outside the conventional RI and 5.8% will be outside the RIs due to bias.

Optimum performance is defined as $B_A < 0.125CV_B$. Here, 0.1% of the additional population will be outside the conventional RI and 5.1% will be outside the RI due to bias.

Minimum performance is defined as $B_A < 0.375CV_B$. Here,

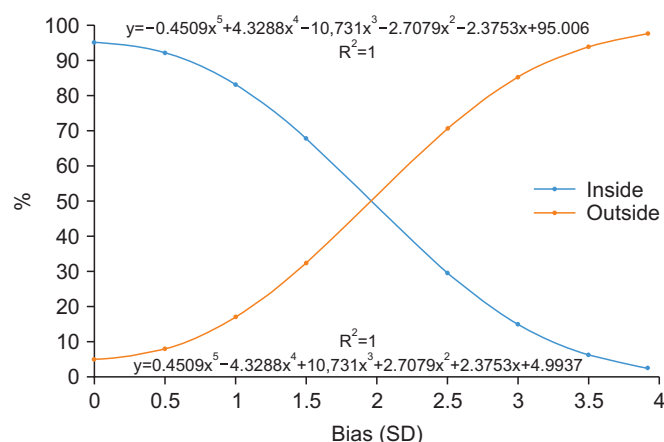


Fig. 5. Bias and area under the curve inside and outside reference intervals. For a given bias value, this diagram can be used to easily estimate the population values inside and outside reference intervals.

1.7% of the additional population will be outside the conventional RI and 6.7% will be outside the RI due to bias.

Short- and long-term biases

In routine practice, clinical laboratories use different consumables, QC samples, calibrators, and reagents with different lot numbers. The accuracy of measurement results is generally monitored using QC materials, and if the measurement result of QC materials is not within acceptable limits, the instruments/measurement system is calibrated. Calibration may correct a shift from the target value; consequently, any measurement system that is frequently calibrated may show bias patterns around the mean value (Fig. 3). Therefore, the characteristic of bias changes over time [12], and the bias pattern estimated from repeated measurement results obtained under repeatability conditions is expected to differ from that estimated from data collected under intermediate precision or reproducibility conditions.

A reliable target or consensus value that can be obtained from the EQAS peer group is required to estimate bias [43, 78-80].

Bias and external quality assessment schemes

Although it is recommended to estimate bias using CRM and reference methods from a metrological perspective, this is not practically achievable in clinical laboratories. Modern clinical laboratories analyze thousands of measurands in different sample types; therefore, procedures to assess the quality and performance of measurement procedures should be practical and cost-effective, rather than purist and theoretical. Numerous laboratories use assigned values from EQAS to estimate bias. How-

ever, this does not represent the real bias that is determined using CRMs and reference methods. Bias calculated from EQAS data is performance bias. Since performance of a laboratory is evaluated using data from other laboratories, bias in the laboratory reflects the position of the laboratory within the peer group. The assigned value or mean of the peer group does not reflect the actual value of the analyte measured using CRMs and a reference method. EQAS programs are not aimed at estimating bias and EQAS samples are generally not commutable [42-44]. Therefore, bias estimated using EQAS data may not represent the actual bias in patients' laboratory test results.

Commutability and bias

According to the VIM, commutability of a reference material is defined as the "property of a reference material, demonstrated by the closeness of agreement between the relation among the measurement results for a stated quantity in this material, obtained according to two given measurement procedures, and the relationship obtained among the measurement results for other specified materials" [13]. In other words, for commutability, the analytical response of tested materials obtained from measurement procedures should be the same as that of patient samples [45]. Therefore, commutable materials can represent fresh patient samples for method comparison [81-83].

In clinical laboratories, human samples (whole blood, serum, plasma, urine, and other body fluids and samples) are analyzed. Therefore, the samples used to evaluate quality indicators must represent human samples. However, in practice, commercial QC samples are used to evaluate quality indicators. As commercial QC samples are used for a long period, they must be stable. To increase the stability of QC materials, lipids are removed and the samples are lyophilized, resulting in a matrix that differs from that of the fresh patient samples. Although fresh patient samples are commutable, they are unstable and cannot be used for long periods. Because of the lack of commutability, commercial QC samples do not represent the patient samples, and therefore, the bias and imprecision estimated from commercial QC and fresh patient samples may be different. Ideally, the reference materials and/or commercial QC samples should be commutable with patient samples. The commutability of samples can be estimated according to CLSI guidelines [84,85]. An estimate of the bias observed between reference and routine methods is required to evaluate sample commutability [86]. Consequently, commutability can be estimated using correctly designed bias experiments and bias can be estimated correctly using commutable materials [87-89].

Bias and MU

As MU is an inseparable part of all types of measurements, accurate calculation of MU has long been a research focus in metrology [90-92].

Various methods for MU calculations have been suggested; however, a global consensus has not been reached, particularly for use in daily practice. Numerous parameters, including instruments, reagents, methods, laboratory environments, and technical staff, affect the MU of analytes. The number of parameters and their contribution to MU vary depending on the analytes and laboratory. To overcome this problem, two major methods to calculate the MU of analytes have been proposed: bottom-up (Type A) and top-down (Type B) methods [93-96].

In the bottom-up method, all possible sources of MU are included in the calculation of the total MU [92]. This method is generally applied to newly developed methods, in-house methods, and measurement procedures that have multiple components. However, this method is time-consuming and requires a detailed road map analysis before MU calculations. It may not be possible to determine all possible sources of MU, particularly in automated measurement systems. In the top-down method, QC data collected in a laboratory, such as internal QC or EQAS data, are used to calculate the total MU [97,98]. This method is more practical and pragmatic for calculating the MU of analytes in medical laboratories, particularly for auto-analyzers.

MU is applied in nearly all industrial sectors but not effectively in medical laboratories. Unlike other calculations, the MU calculation has not been standardized in medical laboratories. Although International Organization for Standardization (ISO) guidelines 15189 [99] recommend the MU calculation for each analyte in medical laboratories, there is no explanation of how to make these calculations. Some guidelines recommend using the bottom-up method to calculate MU, whereas ISO guidelines to calculate the MU of analytes in medical laboratories were not available until 2019 with the release of guidelines ISO/TS 20914:2019 [100], recommending the inclusion of three major parameters to calculate MU: precision, bias, and calibration uncertainty. Different approaches are suggested depending on the availability of MU components as follows.

If all components (i.e., imprecision, bias, and calibration uncertainty) are available, the following equation can be used to estimate the MU of analytes.

$$U = \sqrt{U_{\text{cal}}^2 + U_{\text{Rw}}^2 + U_{\text{Bias}}^2} \quad (9)$$

If bias or calibration uncertainty is not available, these compo-

nents can be excluded from the MU equation as follows:

$$U = \sqrt{U_{\text{cal}}^2 + U_{\text{Rw}}^2} \quad (10)$$

$$U = \sqrt{U_{\text{Rw}}^2} \quad (11)$$

Although the guidelines suggest the inclusion of long-term imprecision, the long-term data collected under intermediate precision conditions or reproducibility conditions also include bias. As mentioned above when discussing short- and long-term biases, for frequently calibrated instruments, the long-term bias becomes a random variation. The graph of data collected under reproducibility conditions (Fig. 3) shows that bias cannot be evaluated as a separate parameter, and the imprecision calculated from the dataset collected under reproducibility conditions covers all known variations, including bias. The data shown in Fig. 3 contains all measurement results collected from an instrument that was frequently calibrated. Based on the EQAS evaluation, instrument performance was acceptable in comparison with that in the peer group. There was no reason to expand the variation in these data by including additional parameters.

However, the guidelines do not provide a strict framework for these parameters. In ISO/TS 20914:2019 [100], imprecision is calculated from the internal QC data, bias is calculated from EQAS data, and calibration uncertainty is obtained from the manufacturers.

Details on how to obtain these parameters are unclear because the imprecision of the measurement procedure can be calculated from the data collected under repeatability, intermediate precision, or reproducibility conditions [39]. The imprecision is expected to be the lowest for data collected under repeatability conditions and the highest for data collected under reproducibility conditions. Bias can also be calculated using CRMs and reference methods or EQAS data; however, the significance of bias should be evaluated before further calculations.

Laboratory data are not exact and have various degrees of variation depending on several factors, including methods and samples, resulting in differences in numerical data, which may be significant or insignificant [101]. As bias is the difference between reference data and the mean of repeated measurements, the significance of bias must be addressed before using a bias in further calculations. Using bias without evaluation of its significance in MU calculation can artificially increase the total MU.

MU should include the most influential factors affecting patients' test results rather than numerous insignificant compo-

nents. Estimating MU from data collected under reproducibility conditions is a practical method for medical laboratories [39]. Because bias is a component of data collected under reproducibility conditions, it should not be included in the MU calculation as a separate parameter.

Another issue is the treatment of bias in the MU calculation. As shown in Equation 9, bias is generally included in MU calculations as a quadratic parameter similar to imprecision. In mathematical terms, it is not valid to sum a variance with a linear parameter; therefore, only the variances of variables can be added together [102].

In conclusion, (1) the inclusion of bias in MU calculation, particularly if the imprecision is calculated from data collected under reproducibility conditions, artificially increases the total MU; (2) it is mathematically incorrect to treat bias as variance; and (3) the significance of bias should be considered before further calculations [39, 40, 103].

Bias and Six Sigma

Six Sigma is a widely accepted standard methodology for total quality management [104] in the new millennium. The performance of processes can be evaluated objectively using the Sigma scale [105, 106]. If a process has a level of 6 sigma, it produces only 3.4 defects per one million opportunities (DPMO), which can be considered the gold standard [107]. The SM of a process can be calculated using equation 12:

$$SM = \frac{TL}{2 \times SD} \quad (12)$$

where TL is the tolerance limit (from the upper to lower limit) of the process and SD is the standard deviation of the process.

In the 1980s, Bill Smith and engineers at Motorola Inc. developed the Six Sigma methodology. Because the SM is considered

the number of SDs between the mean and the UL/LL of the process, the mean of the process can be centered, which is the same as the target of the process. In practice, the situation is different, and a shift can be observed between the mean and the target of the process. Based on long-term observation, this shift is approximately 1.5 SDs (Fig. 6) [108].

In Equation 12, bias is not directly included in the SM calculation. However, the Six Sigma methodology does not neglect bias but rather treats it correctly. If bias is detected, it should be eliminated; including bias even if it can be eliminated is not pragmatic. However, if the system does not provide real-time monitoring (as is the case in most medical laboratories), we cannot be certain that bias does not exist. In daily practice, bias is the dark side of the moon. To overcome this problem, a 1.5 SD bias is included in all calculations related to the SM, and a table of conversion of DPMO to SM and vice versa is prepared accordingly. Therefore, 6 sigma corresponds to 3.4 DPMO. However, if we neglect bias, it corresponds to 0.002 DPMO.

In medical laboratories, the process performance is calculated using a modified equation proposed by Westgard:

$$SM = \frac{TL - \text{Bias}}{SD} \quad (13)$$

Equation 13 differs from Equation 12 in that it includes bias. This method has two main disadvantages. First, incorporating bias in the equation is mathematically incorrect, and the SM obtained from Equation 13 dramatically underestimates the process performance [103, 109, 110]. Second, 1.5 SD bias is included in the DPMO tables and the SM calculated using Equation 13 can significantly underestimate process performance. Because bias is included twice in the calculations (once in Equation 13 and once in the 1.5 SD bias inclusion), the performance of numerous medical instruments and laboratory tests has been

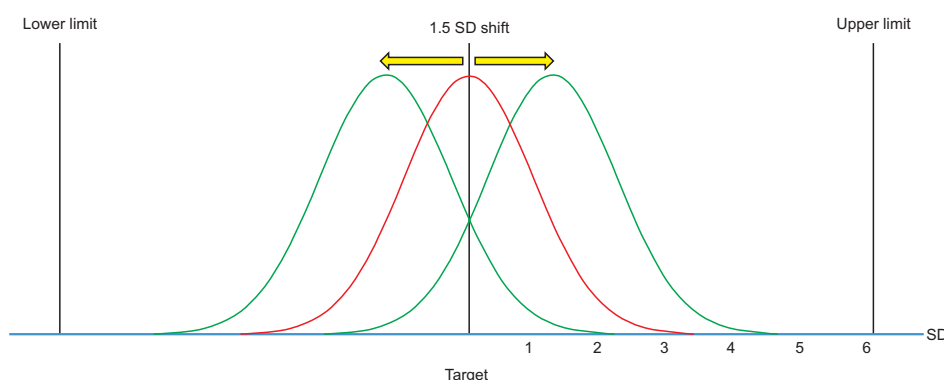


Fig. 6. Sigma metrics is the number of SDs located between the target and upper/lower limits. 1.5 SD shift is considered the standard bias.

calculated as 3–4 SM [111–114]. This implies that the quality of medical laboratory instruments is lower than that of industrial instruments, which is not true. The low-quality level calculated for medical laboratory instruments is due to the incorrect equation and is unrealistic. In reality, medical laboratory analyzers are high-technology products of the same quality as industrial analyzers.

In statistics, various distributions, such as normal, *t*, and chi-square distributions, are used in different situations [75,76]. The mathematics of the SM are based on the normal distribution [115], which is geometrically bell-shaped and mathematically asymptotic to the X-axis (Fig. 6). A shift of the normal distribution curve to the right or left will change the AUC within the TL; however, this change will not be linearly proportional to the shift [103, 109, 110]. The relationship between bias and the AUC can be calculated using the normal distribution equation (Equation 5). However, Equation 5 is very complex and cannot be used in daily practice. Instead, it is practical to use standard tables that show how performance changes with bias [107].

The second important point is that the bias included in the calculation rarely reflects the real bias. In medical laboratories, bias is calculated from EQAS data, and its significance requires confirmation.

Correction of bias

Before initiating the correction procedure, it is essential to evaluate the significance of a bias and confirm its existence. Correcting statistically insignificant or clinically unimportant bias would be a waste of time and money [12, 116]. For a significant bias, a root-cause analysis should be conducted, and if the cause is unknown, correction is not recommended. In this case, bias should be accepted and the bias of the analyte should be considered in all reported information. If a bias is significant and clinically important, it can be eliminated by modifying the methods. If elimination is not possible, a correction procedure should be initiated.

DIAGNOSTIC PERSPECTIVE OF BIAS

Diagnostic accuracy is directly related to the clinical performance characteristics of the measurands. Sensitivity, specificity, positive and negative predictive values, likelihood ratios, and ROC curves are used to describe the relationship between test results and diagnostic accuracy [117–120].

The sensitivity of a test reflects the fraction of patients with a specific disease correctly predicted by the test and can be calcu-

lated using the following equation:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (14)$$

where TP represents the true positives (patients with a disease with a correct diagnosis based on the test result) and FN represents the false negatives (patients with a disease with an incorrect diagnosis based on the test result).

In contrast to sensitivity, the specificity of a test reflects the fraction of individuals without a specific disease correctly predicted by the test, which can be calculated using the following equation:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (15)$$

where TN represents the true negatives (individuals without diseases who are correctly predicted by the test result) and FP represents the false positives (individuals without diseases who are incorrectly predicted by the test result).

Sensitivity and specificity are key components of method performance, and the correct estimation of both metrics is affected by bias (Fig. 7) [121,122].

Predictive values (positive and negative) are functions of sensitivity, specificity, and the disease prevalence and can be formulated as follows:

$$\text{PV}^- = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (16)$$

$$\text{PV}^+ = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (17)$$

The predictive value of a negative test result (PV^-) is the fraction of healthy individuals with negative test results, whereas the predictive value of a positive test result (PV^+) is the fraction of patients with a disease and positive test results.

The odds ratio shows the prevalence of a disease in a population and is expressed as the ratio of the probability of the presence of the disease to the probability of its absence, as follows:

$$\text{Odds ratio} = \frac{\text{Probability of the presence of a specific disease}}{1 - \text{Probability of the presence of a specific disease}} \quad (18)$$

Significant bias will decrease the diagnostic accuracy of laboratory tests (Fig. 8).

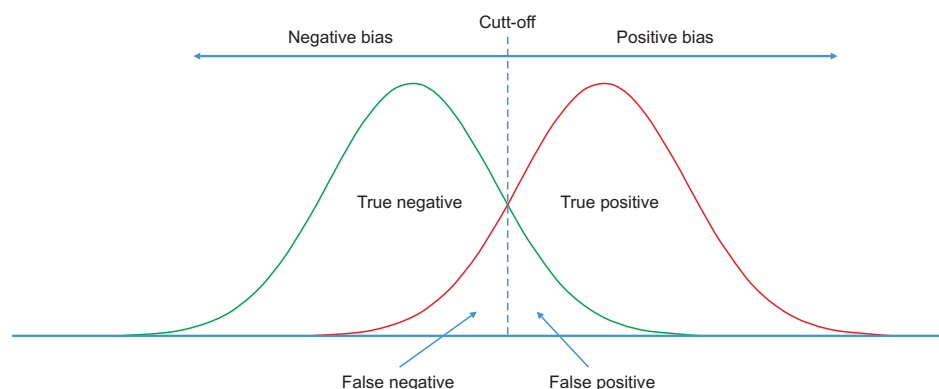


Fig. 7. Effect of bias on true and false positive and negatives.

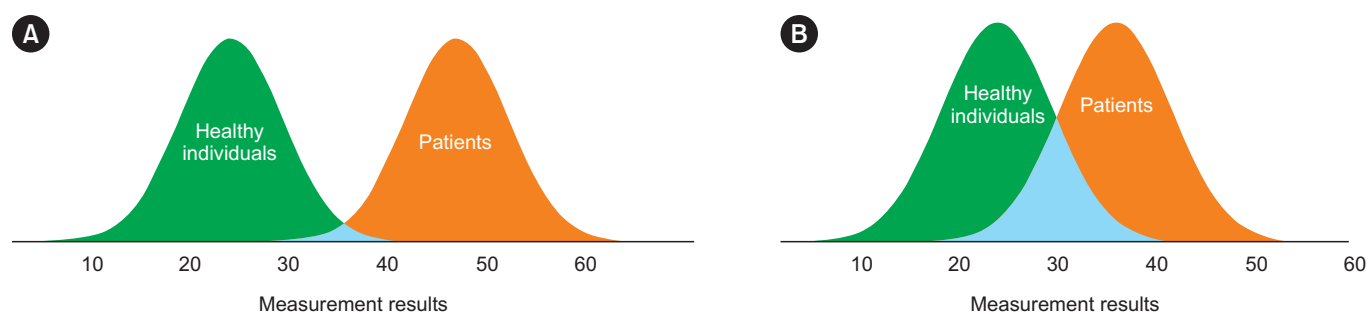


Fig. 8. Bias and the diagnostic accuracy of laboratory tests. Positive or negative bias dramatically affects the diagnostic accuracy of laboratory tests. (A) Acceptable bias does not have a significant negative effect on the diagnostic accuracy of laboratory tests. (B) Increasing bias can lead to misdiagnosis of diseases (blue area) and can dramatically impair the diagnostic power of laboratory tests.

CONCLUSIONS

Bias is the systematic deviation of measurement results from the true value, and it has a significant effect on the information produced from laboratory medicine. However, bias is rarely handled correctly. While imprecision is estimated based on repeated measurements, bias is estimated based on both repeated measurements and a reference/target value. Additionally, the significance of bias should be evaluated and confirmed. In clinical laboratories, bias is the dark side of the moon, and its estimation should be based on appropriate experimental design, data collection, statistical evaluation, and commutable samples. Treating bias appropriately reduces laboratory errors, improves patient safety, and significantly reduces healthcare costs. Statistically significant and medically important biases should be eliminated or corrected. Medical laboratories should develop policies to eliminate the impact of bias on data reported to patients. Future studies are required to illuminate the dark side of the moon, i.e., to eliminate the negative impact of bias on medical decisions and healthcare costs.

SUPPLEMENTARY MATERIALS

Supplementary materials can be found via <https://doi.org/10.3343/alm.2024.44.1.6>

ACKNOWLEDGEMENTS

Not applicable.

AUTHOR CONTRIBUTIONS

Coskun A was involved in conducting the literature review; manuscript writing, editing, proofreading; and reference formatting.

CONFLICTS OF INTEREST

None declared.

RESEARCH FUNDING

None declared.

REFERENCES

- Fraser CG. Biological variation: from principles to practice. Washington, DC: AACC Press, 2001.
- Forsman RW. Why is the laboratory an afterthought for managed care organizations? *Clin Chem* 1996;42:813-6.
- McRae MP, Rajsri KS, Alcorn TM, McDevitt JT. Smart diagnostics: Combining artificial intelligence and in vitro diagnostics. *Sensors (Basel)* 2022;22:6355.
- Hicks AJ, Carwardine ZL, Hallworth MJ, Kilpatrick ES. Using clinical guidelines to assess the potential value of laboratory medicine in clinical decision-making. *Biochem Med (Zagreb)* 2021;31:010703.
- Dybkaer R. Result, error and uncertainty. *Scand J Clin Lab Invest* 1995;55:97-118.
- Frenkel R, Farrance I, Badrick T. Bias in analytical chemistry: A review of selected procedures for incorporating uncorrected bias into the expanded uncertainty of analytical measurements and a graphical method for evaluating the concordance of reference and test procedures. *Clin Chim Acta* 2019;495:129-38.
- Nielsen AA, Petersen PH, Green A, Christensen C, Christensen H, Brandslund I. Changing from glucose to HbA1c for diabetes diagnosis: Predictive values of one test and importance of analytical bias and imprecision. *Clin Chem Lab Med* 2014;52:1069-77.
- Weykamp C, John G, Gillery P, English E, Ji L, Lenters-Westra E, et al. Investigation of 2 models to set and evaluate quality targets for hb a1c: Biological variation and sigma-metrics. *Clin Chem* 2015;61:752-9.
- Mower WR. Evaluating bias and variability in diagnostic test reports. *Ann Emerg Med* 1999;33:85-91.
- Schmidt RL and Factor RE. Understanding sources of bias in diagnostic accuracy studies. *Arch Pathol Lab Med* 2013;137:558-65.
- Haeckel R, Gurr E, Hoff T; on behalf of the working group Guide Limits of the German Society of Clinical Chemistry and Laboratory Medicine (DGKL). Bias, its minimization or circumvention to simplify internal quality assurance. *J Lab Med* 2016;40:263-70.
- Theodorsson E, Magnusson B, Leito I. Bias in clinical chemistry. *Bioanalysis* 2014;6: 2855-75.
- International Organization of Legal Metrology. International vocabulary of metrology-Basic and general concepts and associated terms (VIM) 3rd ed. https://www.oiml.org/en/files/pdf_v/v002-200-e07.pdf (Updated on Aug 2008).
- Eurachem. Treatment of an observed bias. <https://www.eurachem.org/index.php/publications/leaflets/bias-trt-01> (Updated on Oct 2022).
- CLSI. Measurement procedure comparison and bias estimation using patient samples. 3rd ed. CLSI EP09c. Wayne, PA: Clinical and Laboratory Standards Institute, 2018.
- Tate J and Panteghini M. Standardisation - The theory and the practice. *Clin Biochem Rev* 2007;28:93-6.
- Johnson R. Assessment of bias with emphasis on method comparison. *Clin Biochem Rev* 2008;29 (Suppl 1): S37-42.
- Millsap RE. Statistical approaches to measurement invariance. Milton Park: Routledge, 2012.
- Livesey JH, Ellis MJ, Evans MJ. Pre-analytical requirements. *Clin Biochem Rev* 2008;29(suppl 1):S11-5.
- Sandberg S, Carobene A, Bartlett B, Coskun A, Fernandez-Calle P, Jonker N, et al. Biological variation: Recent development and future challenges. *Clin Chem Lab Med* 2022;61:741-50.
- Wikipedia. Albert Einstein. https://en.wikiquote.org/wiki/Albert_Einstein (Updated on March 2023).
- Menditto A, Patriarca M, Magnusson B. Understanding the meaning of accuracy, trueness and precision. *Accredit Qual Assur* 2007;12:45-7.
- Jhang JS, Chang CC, Fink DJ, Kroll MH. Evaluation of linearity in the clinical laboratory. *Arch Pathol Lab Med* 2004;128:44-8.
- Jeong TD, Kim SK, Kim S, Lim CY, Chung JW. Comparison between polynomial regression and weighted least squares regression analysis for verification of analytical measurement range. *Clin Chem Lab Med* 2022;60:989-94.
- Hazra A and Gogtay N. Biostatistics series module 6: Correlation and linear regression. *Indian J Dermatol* 2016;61:593-601.
- Magari RT. Evaluating agreement between two analytical methods in clinical chemistry. *Clin Chem Lab Med* 2000;38:1021-5.
- Ludbrook J. Comparing methods of measurements. *Clin Exp Pharmacol Physiol* 1997;24:193-203.
- Martínez À, Del Río FJ, Riu J, Rius FX. Detecting proportional and constant bias in method comparison studies by using linear regression with errors in both axes. *Chemom Intell Lab Syst* 1999;49:179-93.
- Lu MJ, Zhong WH, Liu YX, Miao HZ, Li YC, Ji MH. Sample size for assessing agreement between two methods of measurement by Bland-Altman method. *Int J Biostat* 2016;12:20150039.
- Zaki R, Bulgiba A, Ismail NA. Testing the agreement of medical instruments: Overestimation of bias in the Bland-Altman analysis. *Prev Med* 2013;57:S80-2.
- Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)* 2015;25:141-51.
- Bilić-Zulle L. Comparison of methods: Passing and Bablok regression. *Biochem Med (Zagreb)* 2011;21:49-52.
- Ludbrook J. Statistical techniques for comparing measurers and methods of measurement: A critical review. *Clin Exp Pharmacol Physiol* 2002;29:527-36.
- Ludbrook J. Linear regression analysis for comparing two measurers or methods of measurement: But which regression? *Clin Exp Pharmacol Physiol* 2010;37:692-9.
- Martínez A, del Río FJ, Riu J, Rius FX. Detecting proportional and constant bias in method comparison studies by using linear regression with errors in both axes. *Chemometr Intell Lab* 1999;49:179-93.
- Theodorsson E. Reference materials and reference measuring systems. <https://cms.jctm.org/wp-content/uploads/2023/02/Reference-materials-and-reference-measuring-systems-2022-03-27.pdf> (Updated on March 2023).
- Bunk DM. Reference materials and reference measurement procedures: An overview from a National Metrology Institute. *Clin Biochem Rev* 2007;28:131-7.
- Westgard S, Bayat H, Westgard JO. Analytical Sigma metrics: A review of Six Sigma implementation tools for medical laboratories Special issue: Six Sigma metrics Review. *Biochem Med* 2018;28:20502.
- Coskun A, Theodorsson E, Oosterhuis WP, Sandberg S; European Federation of Clinical Chemistry and Laboratory Medicine Task and Finish Group on Practical Approach to Measurement Uncertainty. Measurement uncertainty for practical use. *Clin Chim Acta* 2022;531:352-60.
- Becker D, Christensen R, Currie L, Gills T, Hertz H, Klouda G, et al. Use of NIST Standard Reference Materials for decisions on performance of analytical chemical methods and laboratories. NIST Special Publication 829, Washington, DC: US Department of Commerce, 1992.
- Ioannidis JPA. Retiring statistical significance would give bias a free pass. *Nature* 2019;567:461.
- Badrick T, Punyalack W, Graham P. Commutability and traceability in EQA programs. *Clin Biochem* 2018;56:102-4.
- Miller WG. Specimen materials, target values and commutability for external quality assessment (proficiency testing) schemes. *Clin Chim Acta* 2003;327:25-37.

44. Braga F and Panteghini M. Commutability of reference and control materials: An essential factor for assuring the quality of measurements in Laboratory Medicine. *Clin Chem Lab Med* 2019;57:967-73.
45. Greg Miller W, Greenberg N, Budd J, Delatour V, IFCC Working Group on Commutability in Metrological Traceability. The evolving role of commutability in metrological traceability. *Clin Chim Acta* 2021;514:84-9.
46. Horvath AR, Bossuyt PMM, Sandberg S, John AS, Monaghan PJ, Verhaegen-Kamerbeek WDJ, et al. Setting analytical performance specifications based on outcome studies - is it possible? *Clin Chem Lab Med* 2015;53:841-8.
47. JCGM. Evaluation of measurement data-Guide to the expression of uncertainty in measurement. https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf (Updated on Sept 2008).
48. Westgard JO. Error methods are more practical, but uncertainty methods may still be preferred. *Clin Chem* 2018;64:636-8.
49. Hahn GJ, Meeker WQ, Escobar LA. Statistical intervals: A guide for practitioners and researchers. Hoboken: Wiley, 2016: 651.
50. Rao RS, Kumar CG, Prakasham RS, Hobbs PJ. The Taguchi methodology as a statistical tool for biotechnological applications: A critical appraisal. *Biotechnol J* 2008;3:510-23.
51. Kiran DR. Quality loss function. In: Total quality management: Key concepts and case studies. Oxford: Butterworth-Heinemann, 2017: 439-45.
52. Panteghini M and Sandberg S. Defining analytical performance specifications 15 years after the Stockholm conference. *Clin Chem Lab Med* 2015;53:829-32.
53. Fraser CG. The 1999 Stockholm Consensus Conference on quality specifications in laboratory medicine. *Clin Chem Lab Med* 2015;53: 837-40.
54. Kallner A, McQueen M, Heuck C. The Stockholm Consensus Conference on Quality Specifications in Laboratory Medicine, 25-26 April 1999. *Scand J Clin Lab Invest* 1999;59:475.
55. Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G, Oosterhuis W, 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; 53:833-5.
56. Haeckel R, Wosniok W, Kratochvila J, Carobene A. A pragmatic proposal for permissible limits in external quality assessment schemes with a compromise between biological variation and the state of the art. *Clin Chem Lab Med* 2012;50:833-9.
57. Ceriotti F, Fernandez-Calle P, Klee GG, Nordin G, Sandberg S, Streichert T, et al. Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM Strategic Conference. *Clin Chem Lab Med* 2017;55:189-94.
58. Bartlett WA, Braga F, Carobene A, Coskun A, Prusa R, Fernandez-Calle P, et al. A checklist for critical appraisal of studies of biological variation. *Clin Chem Lab Med* 2015;53:879-85.
59. Aarsand AK, Røraas T, Fernandez-Calle P, Ricos C, Díaz-Garzón J, Jonker N, et al. The Biological Variation Data Critical Appraisal Checklist: A standard for evaluating studies on biological variation. *Clin Chem* 2018;64:501-14.
60. Carobene A, Aarsand AK, Bartlett WA, Coskun A, Diaz-Garzon J, Fernandez-Calle P, et al. The European Biological Variation Study (EuBIVAS): A summary report. *Clin Chem Lab Med* 2021;60:505-17.
61. Carobene A, Strollo M, Jonker N, Barla G, Bartlett WA, Sandberg S, et al. Sample collections from healthy volunteers for biological variation estimates' update: A new project undertaken by the Working Group on Biological Variation established by the European Federation of Clinical Chemistry and Laboratory Medicine. *Clin Chem Lab Med* 2016;54: 1599-608.
62. Diaz-Garzón J, Fernández-Calle P, Minchinela J, Aarsand AK, Bartlett WA, Aslan B, et al. Biological variation data for lipid cardiovascular risk assessment biomarkers. A systematic review applying the biological variation data critical appraisal checklist (BIVAC). *Clin Chim Acta* 2019; 495:467-75.
63. Marques-Garcia F, Boned B, González-Lao E, Braga F, Carobene A, Coskun A, et al. Critical review and meta-analysis of biological variation estimates for tumor markers. *Clin Chem Lab Med* 2022;60:494-504.
64. Coskun A, Braga F, Carobene A, Tejedor Ganduxe X, Aarsand AK, Fernández-Calle P, et al. Systematic review and meta-analysis of within-subject and between-subject biological variation estimates of 20 haematological parameters. *Clin Chem Lab Med* 2019;58:25-32.
65. Coskun A, Aarsand AK, Braga F, Carobene A, Díaz-Garzón J, Fernandez-Calle P, et al. Systematic review and meta-analysis of within-subject and between-subject biological variation estimates of serum zinc, copper and selenium. *Clin Chem Lab Med* 2021;60:479-82.
66. Diaz-Garzon J, Fernandez-Calle P, Sandberg S, Özçürümmez M, Bartlett WA, Coskun A, et al. Biological variation of cardiac troponins in health and disease: A systematic review and meta-analysis. *Clin Chem* 2021; 67:256-64.
67. González-Lao E, Corte Z, Simón M, Ricós C, Coskun A, Braga F, et al. Systematic review of the biological variation data for diabetes related analytes. *Clin Chim Acta* 2019;488:61-7.
68. Fernández-Calle P, Díaz-Garzón J, Bartlett W, Sandberg S, Braga F, Beatriz B, et al. Biological variation estimates of thyroid related measurands - meta-analysis of BIVAC compliant studies. *Clin Chem Lab Med* 2021;60:483-93.
69. Jonker N, Aslan B, Boned B, Marqués-García F, Ricós C, Alvarez V, et al. Critical appraisal and meta-analysis of biological variation estimates for kidney related analytes. *Clin Chem Lab Med* 2020;60:469-78.
70. Aarsand AK, Fernandez-Calle P, Webster C, Coskun A, GonzalesLao E, Diaz-Garzon J, et al. EFLM Biological Variation Database. <https://biologicalvariation.eu/> (Updated on July 2023).
71. Petersen PH, De Verdier CH, Groth T, Fraser CG, Blaabjerg O, Hørdler M. The influence of analytical bias on diagnostic misclassifications. *Clin Chim Acta* 1997;260: 189-206.
72. Hyltoft Petersen P, Lund F, Fraser CG, Sandberg S, Sölétormos G. Valid analytical performance specifications for combined analytical bias and imprecision for the use of common reference intervals. *Ann Clin Biochem* 2018;55:612-5.
73. Ricós C, Doménech MV, Perich C. Analytical quality specifications for common reference intervals. *Clin Chem Lab Med* 2004;42:858-62.
74. Gowans EMS, Hyltoft Petersen P, Blaabjerg O, Hørdler M. Analytical goals for the acceptance of common reference intervals for laboratories throughout a geographical area. *Scand J Clin Lab Invest* 1988;48: 757-64.
75. Krishnamoorthy K. Handbook of statistical distributions with applications. 2nd ed. New York: Chapman and Hall, 2016.
76. Coskun A and Oosterhuis WP. Statistical distributions commonly used in measurement uncertainty in laboratory medicine. *Biochem Med (Zagreb)* 2020;30:010101.
77. Petersen PH, Fraser CG, Jørgensen L, Brandslund I, Stahl M, Gowans EM, et al. Combination of analytical quality specifications based on biological within- and between-subject variation. *Ann Clin Biochem* 2002; 39:543-50.
78. Thelen MHM, Jansen RTP, Weykamp CW, Steigstra H, Meijer R, Cobbaert CM. Expressing analytical performance from multi-sample evaluation in laboratory EQA. *Clin Chem Lab Med* 2017;55:1509-16.
79. Stockl D and Reinauer H. Candidate reference methods for determin-

- ing target values for cholesterol, creatinine, uric acid, and glucose in external quality assessment and internal accuracy control. I. Method setup. *Clin Chem* 1993;39:993-1000.
80. Uldall A, Blaabjerg O, Elfving S, Elg P, Gerhardt W, Holmberg H, et al. A programme for assigning target values for external quality assessment schemes in countries with no authorized reference laboratories. Annex. Experiences with deviating results on Ektachem 700 XR. *Scand J Clin Lab Invest Suppl* 1993;212:31-7.
 81. Miller WG, Myers GL, Rej R. Why commutability matters. *Clin Chem* 2006;52:553-4.
 82. Zegers I, Beetham R, Keller T, Sheldon J, Bullock D, MacKenzie F, et al. The importance of commutability of reference materials used as calibrators: The example of ceruloplasmin. *Clin Chem* 2013;59:1322-9.
 83. Miller WG and Myers GL. Commutability still matters. *Clin Chem* 2013; 59:1291-3.
 84. CLSI. Characterization and qualification of commutable reference materials for laboratory medicine. 1st ed. CLSI EP30AE. Wayne, PA: Clinical and Laboratory Standards Institute, 2010.
 85. CLSI. Evaluation of commutability of processed samples. 4th ed. CLSI EP14Ed4. Wayne, PA: Clinical and Laboratory Standards Institute, 2022.
 86. Korzun WJ, Nilsson G, Bachmann LM, Myers GL, Sakurabayashi I, Nakajima K, et al. Difference in bias approach for commutability assessment: Application to frozen pools of human serum measured by 8 direct methods for HDL and LDL cholesterol. *Clin Chem* 2015;61:1107-13.
 87. Budd JR, Weykamp C, Rej R, MacKenzie F, Ceriotti F, Greenberg N, et al. IFCC Working Group recommendations for assessing commutability Part 3: Using the calibration effectiveness of a reference material. *Clin Chem* 2018;64:465-74.
 88. Nilsson G, Budd JR, Greenberg N, Delatour V, Rej R, Panteghini M, et al. IFCC Working Group recommendations for assessing commutability Part 2: Using the difference in bias between a reference material and clinical samples. *Clin Chem* 2018;64:455-64.
 89. Miller WG, Schimmel H, Rej R, Greenberg N, Ceriotti F, Burns C, et al. IFCC Working Group recommendations for assessing commutability Part 1: General experimental design. *Clin Chem* 2018;64:447-54.
 90. Fuentes-Arderiu X. Bio-metrological uncertainty in clinical laboratory sciences. *EJIFCC* 2006;17:6-7.
 91. Farrance I and Frenkel R. Uncertainty of measurement: A review of the rules for calculating uncertainty components through functional relationships. *Clin Biochem Rev* 2012;33:49-75.
 92. Coskun A, Inal BB, Serdar M. Measurement uncertainty in laboratory medicine: The bridge between medical and industrial metrology. *Turk J Biochem* 2019;44:121-5.
 93. Lim YK, Kweon OJ, Lee MK, Kim B, Kim HR. Top-down and bottom-up approaches for the estimation of measurement uncertainty in coagulation assays. *Clin Chem Lab Med* 2020;58:1525-33.
 94. Martinello F, Snoj N, Skitek M, Jerin A. The top-down approach to measurement uncertainty: Which formula should we use in laboratory medicine? *Biochem Med (Zagreb)* 2020;30:020101.
 95. Burr T, Croft S, Favalli A, Krieger T, Weaver B. Bottom-up and top-down uncertainty quantification for measurements. *Chemom Intell Lab Syst* 2021;211:104224.
 96. 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;53:1025-32.
 97. Magnusson B, Ossowicki H, Rienitz O, Theodorsson E. Routine internal- and external-quality control data in clinical laboratories for estimating measurement and diagnostic uncertainty using GUM principles. *Scand J Clin Lab Invest* 2012;72:212-20.
 98. Ceriotti F. Deriving proper measurement uncertainty from Internal Quality Control data: An impossible mission? *Clin Biochem* 2018;57:37-40.
 99. International Organization for Standardization. ISO 15189:2012(en), Medical laboratories — Requirements for quality and competence. <https://www.iso.org/obp/ui/#iso:std:iso:15189:ed-3:v2:en> (Updated on 2012).
 100. International Organization for Standardization. ISO/TS 20914:2019(en), Medical laboratories — Practical guidance for the estimation of measurement uncertainty. <https://www.iso.org/obp/ui/fr/#iso:std:iso:ts:20914:ed-1:v1:en> (Updated on 2019).
 101. Coskun A, Serteser M, Karpuzoglu HF, Unsal I. How can we evaluate differences between serial measurements on the same sample? A new approach based on within-subject biological variation. *Clin Chem Lab Med* 2017;55:e44-6.
 102. Pishro-Nik H. Introduction to probability, statistics, and random processes. Amherst, MA: Kappa Research, LLC, 2014:732.
 103. Coskun A, Serteser M, Unsal I. The short story of the long-term Sigma metric: Shift cannot be treated as a linear parameter. *Clin Chem Lab Med* 2019;57:E211-3.
 104. Harry M and Schroeder R. Six Sigma, the breakthrough management strategy revolutionizing the world's top corporations. New York: Doubleday, 2005.
 105. Cao S and Qin X. Application of Sigma metrics in assessing the clinical performance of verified versus non-verified reagents for routine biochemical analytes. *Biochem Med (Zagreb)* 2018;28:020709.
 106. Keleş M. Evaluation of the clinical chemistry tests analytical performance with Sigma Metric by using different quality specifications - Comparison of analyser actual performance with manufacturer data. *Biochem Med (Zagreb)* 2022;32:010703.
 107. Coskun A, Oosterhuis WP, Serteser M, Unsal I. Sigma metric or defects per million opportunities (DPMO): The performance of clinical laboratories should be evaluated by the Sigma metrics at decimal level with DPMOs. *Clin Chem Lab Med* 2016;54:e217-9.
 108. Coskun A and Ialongo C. Six Sigma revisited: We need evidence to include a 1.5 SD shift in the extraanalytical phase of the total testing process. *Biochem Med (Zagreb)* 2020;30:010901.
 109. Coskun A, Serteser M, Unsal I. Sigma metric revisited: True known mistakes. *Biochem Med (Zagreb)* 2019;29:010902.
 110. Oosterhuis WP and Coskun A. Sigma metrics in laboratory medicine revisited: We are on the right road with the wrong map. *Biochem Med (Zagreb)* 2018;28:020503.
 111. Shaikh MS, Ali SA, Rashid A, Karim F, Moiz B. Performance evaluation of a coagulation laboratory using Sigma metrics. *Int J Health Care Qual Assur*. 2018;31: 600-8.
 112. Teshome M, Worede A, Asmelash D. Total clinical chemistry laboratory errors and evaluation of the analytical quality control using Sigma metric for routine clinical chemistry tests. *J Multidiscip Healthc* 2021;14: 125-36.
 113. Ozdemir S and Ucar F. Determination of Sigma metric based on various TEa sources for CBC parameters: The need for Sigma metrics harmonization. *J Lab Med* 2022;46:133-41.
 114. Kumar BV and Mohan T. Sigma metrics as a tool for evaluating the performance of internal quality control in a clinical chemistry laboratory. *J Lab Physicians* 2018;10: 194.
 115. Coskun A, Serteser M, Kilercik M, Aksungar F, Unsal I. A new approach to calculating the Sigma Metric in clinical laboratories. *Accredit Qual Assur* 2015;20:147-52.
 116. Magnusson B and Ellison SLR. Treatment of uncorrected measurement bias in uncertainty estimation for chemical measurements. *Anal Bio-*

- anal Chem 2008;390:201-13.
117. Ashwood ER and Bruns DE. Clinical utility of laboratory tests. In: Burtis CA, Ashwood ER, Bruns DE, eds. Tietz textbook of Clinical Chemistry and Molecular Diagnostics. 6th ed. Saint Louis:Elsevier, 2012: 49-60.
118. Chu K. An introduction to sensitivity, specificity, predictive values and likelihood ratios. Emerg Med 1999;11:175-81.
119. Trevethan R. Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice. Front Public Health 2017;5:307.
120. Shreffler J and Huecker MR. Diagnostic testing accuracy: Sensitivity, specificity, predictive values and likelihood ratios. StatPearls. <https://www.ncbi.nlm.nih.gov/books/NBK557491> (Updated on March 2023).
121. Leeflang MMG, Moons KGM, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: Mechanisms, magnitude, and solutions. Clin Chem 2008;54: 729-37.
122. Ringham BM, Alonzo TA, Grunwald GK, Glueck DH. Estimates of sensitivity and specificity can be biased when reporting the results of the second test in a screening trial conducted in series. BMC Med Res Methodol 2010;10:3.