Research Article

Calibration error, a neglected error source in the clinical laboratory quality control

Atilla Barna Vandra[1*]

[1]*Spitalul Clinic Județean de Urgență, Brașov, Romania (retired)*

Article Info

[*]***Corresponding Author:***
Atilla Barna Vandra
E-mail: vandraattila@gmail.com
Spitalul Clinic Județean de Urgență, Brașov, Romania (retired)

Keywords

Bias, calibration error, calibration graph, incorrigible bias, quality control, random error, repeatability conditions, reproducibility within laboratory conditions, systematic error

Abstract

"Calibration" conveys the connotative meaning of "correction." Therefore, calibration is frequently viewed as "perfect," but it is a measurement, and no measurement is error-free. This study aims to uncover the sources of calibration errors, to estimate their size, and assess their consequences in quality control.

The analytical bias is the difference between the working (determined) graph and the ideal graph (how the reagents behave).

The source of the calibration random component is the random error committed in the calibration. The primary source of the systematic component is the reference material value error, which cannot be reduced to the nominal value error. Even if the avoidable human errors are neglected, the reconstitution errors, including two volume measurements, are inherent.

The random component was estimated by making five calibrations in repeatability conditions and calculating the coefficient of variation of the slope factors. The total calibration error was estimated by comparing the slope factors of new calibrations using the same reagent and calibrator lots (one-year data).

The results confirmed the presumptions: the calibration error is bigger than the coefficient of variation measured in repeatability conditions. Smaller biases are incorrigible by calibration, and quality control rules must be designed to prevent them from being detected.

Using the σ parameter in the QC graphs would result in too frequent alarms. Westgard proportionally increased the decision limits by overestimating σ with the standard deviation measured in reproducibility within laboratory conditions. A more accurate solution is to increase all decision limits to account for the incorrigible bias and design the QC graphs with the standard deviation measured in repeatability.

Introduction

Measurement errors can be committed in the pre-pre-analytical, pre-analytical, analytical, and post-analytical phases [1]. Analytical errors hold a special status in the error budget. According to JO. Westgard:

"Quality design in a laboratory must begin with analytical quality because it is the essential quality characteristic of any laboratory test; unless analytical quality can be achieved, none of the other characteristics matter [2]."

The analytical errors encompass method, matrix, and true analytical or measurement errors, including calculation and rounding errors. Because the control materials used in the internal QC are not certified reference materials, method and matrix errors can be measured only in EQA.

Only true analytical (measurement) errors can be quantified in the internal QC and mathematically described by the Gauss equation [3]. Only in this case can we differentiate between the systematic (SE) and the random (RE) error components. Both parameters of the equation are ideal values. Therefore, estimators are used. The bias (B≈SE) is the estimator of the SE, while the standard deviation SD≈σ is the estimator of the mean RE.

An accurate classification is complete, avoids redundant categories, and uses accurately defined categories to prevent misinterpretations. Unfortunately, the classifications of the bias found in the literature do not have all these properties. For example, WP Oosterhuis et al. mention four sources of bias (SE): (1) laboratory, (2) reagent, (3) instrument, and (4) operator bias [4]. The "ladder of errors" of M. Thompson shows four rungs of the laboratory errors: (1) laboratory, (2) method and measurement system, (3) day-to-day variation ("run bias"), and (4) repeatability [5], [6]. This paper focuses on normal laboratory conditions, including well-trained personnel, an air-conditioned laboratory, proper maintenance, and the absence of analyzer failures.

A detailed analysis of the error sources reveals that the environmental influence is mostly pre-analytical and insignificant; the "laboratory errors" are a sum of pre-analytical, human, environmental, instrumental, and non-instrumental errors. Human (operator) errors are mostly pre-analytical, and if not, they indirectly influence instrumental and non-instrumental errors. Therefore, all of the aforementioned sources are redundant in the list. Only two error sources must be considered: the instrumental and the non-instrumental errors. (Calculation errors have similar properties and, therefore, can be categorized as instrumental errors.)

The instrumental errors are linked to the construction of the analyzer and its changeable parts. Assuming correct maintenance, the latter has a negligible influence. In the meantime, the construction of the analyzer remains constant and is non-specific. The inherent inconstant functionality of the instrument (e.g., sampling or reading error) is the source of a constant random error; however, it may also cause constant biases. In the absence of failures, the mean random error is constant!

The non-instrumental errors (linked to the reagents and calibration graphs) are variable and specific to each measurement.

The inherent reagent instability is the primary source of the bias variability. It causes a quasi-linear, predictable drift (gradual variation) in the systematic error component, masked by the random error on the QC graphs. Random variation in reagent properties contradicts the fundamental laws of chemistry. (An unpredictable change is still possible in the case of reagent impurification.) It cannot be the source of random error, only of inherently variable bias. The primary role of a QC is to identify and correct these bias variations.

The reagents and electrodes are not constructive parts of the instrument. They are specific, and their property significantly influences the results. Producers only guarantee trustworthy results after recalibration within the validity term [7].

The bias caused by reagent instability cannot increase indefinitely. It is corrected by (re)calibrations, resulting in cyclical, sawtooth-like graphs of the daily biases, usually masked by random errors. Surprisingly, CLIA requirements do not mention the reagent instability among the error sources:

"The control procedures must detect immediate errors that occur due to test system failure, adverse environmental conditions, and operator performance [8]."

The word "calibration" conveys the connotative meaning of "correction." Therefore, calibration is frequently viewed as a "perfect" process, subject to only the calibrator's nominal value uncertainty (specified in the traceability documents); however, it is a measurement, and no measurement is error-free. Many authors share the myth of error-free calibration [9]. According to JO Westgard and T.L. Groth [10]:

"In the simulation procedure, the calibration process is assumed to be ideal and to contribute no errors in addition to the primary ones being introduced."

Despite this assumption, they state later:

"…these primary errors could represent calibration errors or other sources of errors…"

This myth persists. Despite several authors mentioning calibration errors [4], [5], [11], [12] some specifying even mosaic pieces of the sources (e.g., calibrator preparation [6]), others focusing on human errors [13] or calibration procedures [14]), a detailed analysis and a quantitative evaluation of the inherent calibration errors have not been done yet.

This study aims to fill the gap by focusing on the inherent errors in calibration, evaluating their source and magnitude.

Materials and methods

The size of the calibration random error was estimated in two ways.

1. Five calibrations were performed under repeatability conditions, with a 5-minute interval between each, to determine the calibration parameters, with a focus on the slope factor

Fcal. Based on the proportionality between the Fcal and the concentration, the calibration random error was estimated by calculating the variability of the slope factor Fcal, expressed as a coefficient of variation ($CV_{F_{cal}}$). The $CV_{F_{cal}}$ was compared with the CV_r, which was determined during the method adequacy testing a few months prior. Measurements were made on a Cobas 6000 analyzer using Roche reagents and reference materials. The method neglected the influence of the error in the null-point absorption (A_0) and the variability of the reconstitution error. (The same reference material was used.)

2. To estimate the variability of the F_{cal} in real-life situations, constant conditions must be maintained. This includes using the same reagent lotcalibrator lot and considering only new calibrations. All conditions were fulfilled, with sufficient data in the case of urea on a Cobas pro analyzer (The producer recommends recalibrating each reagent bottle.) Recalibrations of used reagents, different reagent lots, and calibrator lots would introduce supplementary errors, overestimating the calibration errors. If recalibrations are also included, and lots are mixed, the estimated CV_{cal} becomes similar to CV_{RW}.

The method also neglected the influence of A_0 variability. However, it accounted for the reconstitution error. (More reconstituted calibrator bottles were used.)

Results

Because of the proportionality between the F_{cal} values and concentration, equals the CV calculated from the variable bias values (CV_{VCSE}). (An SD or CV can be calculated from any variable data set, not only from normally distributed random errors. However, the SD (CV) can be used as the correct estimator of the σ parameter of the Gauss equation (the measure of the mean pure RE) only under constant conditions. Under variable conditions (a variable bias), the SD (CV) is the measure of all variable components (RE + VCSE) [7]. For this reason, $CV_{F_{cal}} \approx CV_{VCSE}$, and equals the mean B% caused by the mean variations.

The values for the estimated $CV_{F_{cal}}$ compared with the CVr obtained in the method validation for four analytes are presented in Table 1.

Table 1: Comparison between the measured $CV_{F_{cal}} \approx CV_{VCSE}$ and CV_r.

Analyte	AST/GOT	Creatinine	Glucose	Urea
$CV_{F_{cal}} \approx CV_{VCSE}$	2.85%	0.98%	1.25%	0.99%
CVr	1.64%	0.96%	0.84%	1.13%

A Cochran test for equality of two variances did not find significant differences between $CV_{F_{cal}}$ and CV_r values, except for AST/GOT, where $CV_{F_{cal}}$ is larger than CV_r).

The $CV_{F_{cal}}$ calculated from the F_{cal} values obtained only in new calibrations for the same reagent lot and the same calibrator lot for urea are shown in Table 2 (one-year database). As in the first case, a Cochran's F-test did not find significant differences in comparison with the CV_r:

Table 2: $CV_{F_{cal}}$ for urea calculated from new calibration data, CV_r=1.04%.

Reagent lot	629883	612739	597097	583201	583201
Calibrator (RM) lot	539941	539941	539941	539843	410065
$CV_{F_{cal}} \approx CV_{VCSE}$	1.29	1.39	1.33	0.8	1.1

As a comparison, the mean value for $CV_{F_{cal}}$ calculated from all F_{cal} values (including recalibrations) was 2.5%, similar to CV_{RW} (2.6% and 2.7% on the two control levels).

Discussion

"Calibration is a process of testing and adjusting an instrument or test system to achieve a correlation between the measurement response (measurement signal) and the concentration or amount of substance measured by the test procedure [14]."

A calibration graph is a mathematical relationship between absorption and concentration. In biochemistry, this relationship is usually linear within the measuring range. (The case of non-linear calibration graphs will not be discussed.)

Calibration, like all measurements, has both random and systematic components.

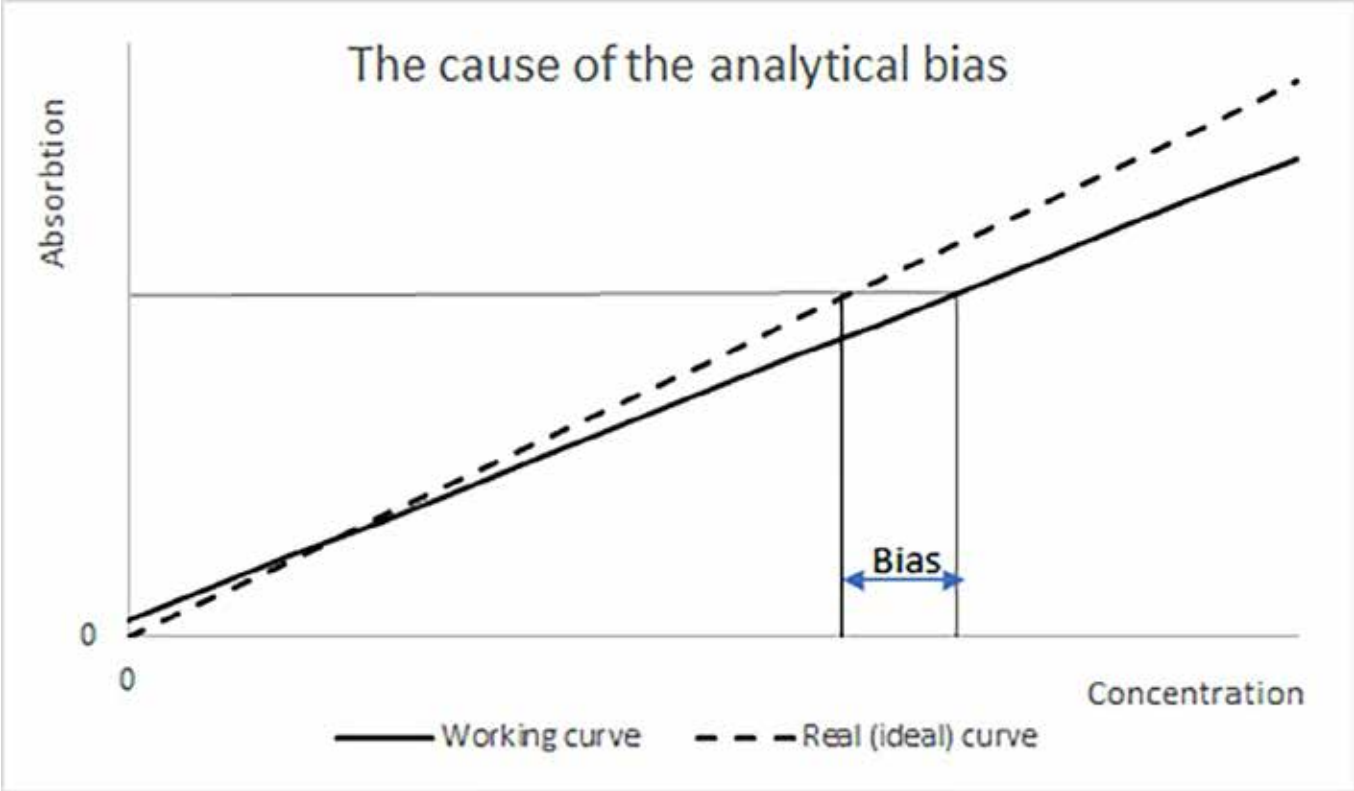
The SE (bias) variation (the VCSE) has two primary sources: reagent instability and calibration errors. Both phenomena are linked to the calibration graphs.

When calibrating, two parameters are determined: the null-point absorption, A_0, and the slope factor F_{cal}. The determined parameters differ from the ideal values. Therefore, two different calibration graphs exist, and both are variable. One is the determined (working) graph used in calculations, which remains constant between calibrations but changes randomly within each calibration. The consequence is a randomly variable systematic error, an alternation between predictable (constant) periods and random variations. The other is the ideal graph, reflecting how the reagent behaves. It changes gradually depending on the reagent stability, causing quasi-linear bias variations. Combining the two phenomena produces sawtooth-like graphs of the daily bias, which are masked on the QC graphs by the random error, allowing only significant shifts to be observable [7]. According to E. Theodorsson, B.

Magnusson, and I. Leito:
"Within a given day, the small deviations of the calibration graph from an 'ideal calibration graph' affect all the samples in a systematic way [6]."

The analytical bias is the difference between the result obtained on the determined graph (the measured value) and the ideal graph (the real value) [14] (Figure 1).

Figure 1: The cause of the analytical bias is the error of the calibration graph.



The error of A_0 introduces a constant bias, while the error of the slope factor F_{cal} causes a bias proportional to the concentration. While the determined graph remains constant until the next calibration, the ideal one depends on the reagent quality and varies gradually over time. Therefore, different measured values with gradually increasing biases (in absolute values) are obtained on the following days [7], [13], [14].

We must distinguish between calibration errors and calibration graph errors. The calibration errors, which cause a bias in the first run after calibration, have a single source: the measurement error committed during the calibration process. The calibration graph errors, which are the permanent causes of analytical bias, have two sources: calibration errors and reagent degradation.

Calibration is a measurement, subject to analytical errors comprising systematic and random components. The source of the calibration random errors is the measurement errors in the calibration process. The measurement of the calibrator is subject to the same random errors as a patient or control sample, comprising both systematic and random components. Theoretically, the systematic components are eliminated or at least reduced when the absorption variations ($\Delta A=A-A_0$) are calculated, but the random ones are not. The SD when measuring the calibrator is also sr (calibrations are made in repeatability conditions), and the CV is CV_r. In linear calibrations, four measurements are made: two to determine A_0 and the other two measurements are used to measure the calibrator and calculate the slope factor F_{cal}.

Due to the proportionality between concentrations and absorptions, the coefficient of variation (CV) of the absorptions will be equal to the CV of the concentrations.

Calculating the average absorption of two calibrator measurements will not eliminate the RE; it only reduces the CV $\sqrt{2}$ times. A similar RE is committed when measuring A_0. The slope factor F_{cal} is determined as a difference, and the CV of a difference is $\sqrt{2}$ times bigger than the CV of a single measurement. Therefore, the RE committed in the slope factor F_{cal} determination is $CV_{F_{cal}}$~CV_r. Due to the proportionality between the concentrations and Fcal, the probable calibration-induced bias is B% ≈ CV_r. (In 32% of cases, even bigger). In absolute values, the probable bias induced by the calibration RE is ≈1s_r.

The calibration random error induces a constant bias until the next calibration, when the new calibration random error replaces its value.
The primary source of systematic calibration errors is the error in the reference material (calibrator) value. The uncertainty of the reference material's nominal value is well-known. It is specified in the traceability certificates. However, the fact that different calibrator values are specified for different methods raises some doubts about the credibility of these uncertainties. Some of these uncertainty intervals are disjoint (have no common values). This study does not aim to analyze and explain the phenomenon but only to draw attention to the fact that nominal value errors may be more significant than previously thought.
The reference material value errors cannot be reduced to the nominal value errors [6].

According to M. Thompson:
"Systematic errors in calibration, such as might be brought about by the incorrect preparation of a stock solution, or lack of fit in an estimated calibration function, contribute to the higher-level biases [4]."
T. Badrick [15], referring to the Tietz Textbook of Clinical Chemistry [16], underlines:
"The act of reconstitution can introduce an error far greater than the inherent error of the rest of the analytical process."
Avoidable human errors may also contribute to reconstitution errors, but in most cases, these errors are inherent. The reference materials, after their concentration (activity) is determined by the producers, undergo a lengthy process. They are distributed into bottles (a first volume measurement), lyophilized, and then deposited. They are transported to the laboratory, reconstituted (a second volume measurement), and subsequently frozen. After thawing, they are homogenized, acclimatized, and only then measured. Their value is different from the nominal value specified by the producers. (Reference material value error). Even if the nominal value errors and human errors (stability, homogenization, temperature errors, and avoidable human errors in reconstitution) are neglected, the two volume measurements still induce inherent errors in the reconstitution.
The precision of the actual pipettes, expressed as a CV, is ±0.2-0.3%, and the accuracy is 0.5-0.6%. The expanded uncertainty is twice as big, and there are two volume measurements. Furthermore, the reconstitution of the control materials is subject to the same errors; therefore, the observable reconstitution error may be even bigger than CV_r.
Consequently, the estimated average total calibration error is 1-2 CV_r. According to a quintessential principle valid in all exact sciences, an error (bias) cannot be corrected if the uncertainty of the value is greater than the error itself, nor if the uncertainty (mean error) of the corrective action is greater than the corrected error. For example, a bias cannot be corrected as GUM recommends if its uncertainty is bigger: B<Ubias [17]. Additionally, we cannot eliminate a smaller bias than the average calibration error. In both cases, there is a risk of increasing the bias.
The estimation of the random component of the calibration error was verified by five consecutive calibrations of four analytes (Table 1). All four cases confirmed the assumption that the random component of the calibration error and the induced bias (B%) is at least 1 CV_r. Because the error of the A_0 was not considered, it is an under-evaluated estimation.
Because the total (systematic and random) calibration error can be estimated only in the case of new calibrations (assuming that all bottles from the same lot are identical and recalibrations cannot be used), a sufficient number of data points, allowing for an accurate estimation, were available for urea. The data in Table 2 once again confirm that the average calibration error is at least 1 CV_r.
The assumed $B_{incorrigible}$>1 CV_r is sustained by the fact that, despite efforts to correct the biases, 1 CV_r is a typical value for the mean bias calculated from monthly QC data, thereby sustaining its incorrigibility.
Due to the calibration error, there exist incorrigible biases. The question arises: Why must the QC rules detect them?
According to Theodorsson et al.:
"There is no point in trying to eliminate or correct a small and clinically unimportant bias since both elimination and correction need resources and may increase the measurement uncertainty [6]."
Other authors also share similar ideas [18].
It is harder to define what "small and clinically unimportant bias" means. If an incorrigible bias (B<CV_r) is not insignificant, it indicates an ineffective quality control system. Usually, this is not the case. 1 CVr is a minimal value for the insignificant bias. Other possibilities for estimating it are: $B_{insignificant}$=TEa/7 or a quarter of the within-subject biological variation ($B_{insignificant}$=CVI /4).
The most frustrating moments in the clinical laboratory biochemistry department occur when a QC alarm is triggered, and despite recalibration, no improvement is achieved; the situation may even worsen. In the meantime, the phones are ringing because of the delayed results. Usually, incorrigible biases can be discovered behind these phenomena.
The necessity to avoid alarms in the cases of incorrigible biases highlights the weaknesses of the actual QC system based on Westgard rules and the false assumption of a normal distribution of long-term control results. Correct decisions based on the Westgard rules (consistent with the calculations of JO Westgard and T. Groth [10]) are only possible if the Levey-Jennings graphs are designed with the σ parameter of the Gauss equation (or with its trustworthy estimator). Unfortunately, a correctly applied, Westgard-rules-based QC system is dysfunctional. The former statement seems doubtful because almost everybody uses the Westgard rules. The highlight is on "correctly applied." A rhetorical question: "Does anybody design the Levey-Jennings graphs with the σ parameter?"

If the QC graphs are designed with σ, according to the normal distribution tables, the probability of an $R_{1\text{-}2S}$ warning in the absence of bias is 4.56% [19]. Assuming 25 biochemistry parameters, two control levels, three runs per day, and 30 days per month (resulting in 4500 control measurements per month), this yields more than two false $R_{1\text{-}2S}$ warnings in each run, totaling ≈200 per month. A B=1 SD is a typical value in practice because it is at the limit of incorrigibility. Assuming only 11 (less than half of 25) parameters with B≈1 SD, an $R_{1\text{-}3S}$ alarm in each second control run is quasi-lawful. Less frequent $R_{1\text{-}2S}$ warnings and $R_{1\text{-}3S}$ rejections are evidence of the overestimation of the standard deviation (SD) used in the design of the QC graphs. Not σ, but a significantly bigger SD value is used in the design of the graphs.

Instead of frequent warnings and rejections, several "impossible" QC graphs are observed in practice. For example, graphs with no $R_{1\text{-}2S}$ warning in a month (probability 0.0224%), no $R_{1\text{-}3S}$ violation in a month if B>1SD, (Probability <1.57%) no $R_{1\text{-}3S}$ violation in a month if the reagent is unstable (visible drift), more $R_{4\text{-}1S}$ violations without $R_{1\text{-}2S}$ warnings, more frequent $R_{4\text{-}1S}$ and R_{10X} violations, than $R_{1\text{-}3S}$ or $R_{2\text{-}2S}$ (and examples may continue). All these examples are evidence of overestimations of the "SD" used in the design of the QC graphs.

The Gauss equation is only valid, and the predictions are correct if the conditions are constant [3]. Constant conditions assume a constant mean and bias, which cannot be maintained in the long term in a clinical laboratory. Data sets collected when the bias is different relate to each other as two sets of length measurements, some expressed in meters and others in yards [7].

The distribution of all data will be bimodal, with two distinct peaks rather than a normal (Gaussian) distribution.

More authors questioned the normal distribution of the long-term control results [20],[21],[22]. According to AB Vandra, the non-Gaussian distribution is caused by SE (bias) variability [7].

The σ parameter of the Gauss equation can be estimated correctly only in constant (repeatability) conditions. The SD measured in variable conditions is not a measure of the true RE but of all variable components, including the RE and the bias variations [7]. An SD can be calculated from any variable data set, not only from normally distributed ones. Several authors (using different definitions and notations) concluded that the link between the SD measured in repeatability conditions (s_r) and the SD measured in reproducibility within laboratory conditions (s_{RW}) is the bias variations [7],[23],[24],[25]. The true RE and its correct estimator can be determined only under constant conditions of repeatability $\sigma \approx s_r$.

According to another quintessential principle valid in all exact sciences, a parameter must be determined under the same conditions in which it is used. QC decisions are made in repeatability conditions, consistent only with s_r, not with s_{RW}. The former principle suggests the same: to design the QC graphs with s_r, not with s_{RW}, as Westgard recommends [26].

As shown previously, a correctly applied Westgard-rules-based QC system (with QC graphs designed with σ) is dysfunctional due to excessive false alarms. The "solution" proposed by JO Westgard is to overestimate σ with s_{RW}, which is only apparently efficient. The system becomes functional, but such a "correction" is inaccurate and questionable.

J.O. Westgard and T. L. Groth [10] acknowledged:

"The calculations based on computer simulations behind the power function graphs are made assuming "within-run (repeatability) SD", while the graphs are designed with "total SD".

And also, that: "…the error level that may occur is likely to be greater than that predicted."

However, they acknowledged the inconsistency between calculations and decisions, they preferred the larger limits to maintain the functionality of the QC system but using different SD in calculations (s_r) and decisions (s_{RW}) is a serious error source.

The usual $\frac{s_{RW}}{s_r}$ ratio is 1.25-2; however, it may be even bigger in some cases. For an approximate comparison, let us accept an $\frac{s_{RW}}{s_r}$=1.5 ratio. It means that by designing the QC graphs with s_{RW}, instead of the $R_{1\text{-}3S}$ rule, we de facto apply the $R_{1\text{-}4.5S}$ rule, rather than the $R_{2\text{-}2S}$ rule, the $R_{2\text{-}3S}$ rule, rather than the $R_{4\text{-}1S}$ rule, and the $R_{4\text{-}1.5S}$ rule. Additionally, instead of the $R_{1\text{-}2S}$ warning, we use the $R_{1\text{-}3S}$ "warning." Using s_{RW} instead of σ, we make a proportional increase in the decision limits. Following Westgard's recommendations strictly, we partially avoid alarms in cases of incorrigible biases. The probability of no warnings in a month becomes 61%, a plausible value to observe the phenomenon.

While the within-run rules ("$R_{1\text{-}3S}$" and "$R_{2\text{-}2S}$") are sensitive to the proportional overestimation of σ. They efficiently avoid alarms in the case of incorrigible biases. In contrast, "$R_{4\text{-}1S}$" is less sensitive, and R_{10X} is insensitive, resulting in a relative increase in efficiency of the cross-run rules, along with a reduced but not insignificant number of false alarms. The Westgard recommendations are only practical for higher sigma metric values.

A more efficient solution is to increase all decision limits not proportionally, but with a constant value, the estimated Bincorrigible. (E.g., using the $3s_r + B_{incorrigible}$ limit instead of the $3s_{RW}$). This increase in the decision limits must be reflected in the name of the rule, consistent with the calculations. To state that you apply "$R_{1\text{-}3S}$" and de facto apply $R_{1\text{-}4.5S}$ is misleading and a source of errors.

When comparing the monthly s_{RW} values, the inaccuracy of the overestimation of σ by s_{RW} becomes more evident. A value in a month may double or halve in the next month [27]. On the last day of the month, calculating the new s_{RW} value and redesigning the graphs, we will become conscious that we missed several alarms, or vice versa; we have made several unnecessary calibrations.

A real-life example of cholesterol is suggestive. For an easier and more accurate comparison between the control levels, the results are expressed as a percentage. TEa = 8%, CV_r = 0.98%, and the average CV_{RW} = 2.11% (the average of both control levels, measured in the previous month). Calculated with CV_r = 0.98% and negligible bias in EQA, Nσ=8.16. Assuming $B\%_{incorrigible} \approx CV_r$=0.98%, the R_{1-4S} limit is 3.92%. All control data were between -2.05% and +3.91%, with a mean of +1.38%. No alarm was observed. According to Westgard's recommendation, Nσ=3.82; therefore, "all Westgard rules" are recommended. Decision limits: R_{1-3S}: 6.33%, R_{2-2S}: 4.22%, R_{4-1S}: 2.11%. Consistent with a mean of 1.38%, several R_{10X} alarms could be observed if applied, but no R_{1-2S} warning. (The actual recommendations on westgard.com do not use the R1-2S warning in multi-rules [28].) At the end of the month, the recalculated CVRW was 1.41%, and the Nσ = 5.67. According to Westgard's recommendations, only the R_{1-3S} and R_{2-2S} rules were necessary to apply, with the following decision limits: 4.23% for R_{1-3S} and 2.82% for R_{2-2S}. If applied to the same dataset, more R_{1-2S} warnings would occur, but no R_{1-3S} or R_{2-2S} rule violations would be observed. In the meantime, the violations of the unapplied R_{4-1S} or R_{10X} rules suggest the existence of (insignificant) biases. Due to the variability of the s_{RW}, applying Westgard's recommendation may raise doubts about the accuracy of decisions.

In contrast to s_{RW}, the s_r values (determined from control values obtained under constant conditions) are less variable, making the decisions more trustworthy.

Conclusions

The Westgard-rules-based QC system is based on an erroneously assumed normal distribution of the long-term control data. The laws of the normal distribution can be applied only to data collected under constant, repeatability conditions. Because of the internal QC decisions are made under repeatability conditions. Therefore, according to a principle valid in all exact sciences, the standard deviation (SD) used in the design of the QC charts must be s_r.

Calibration is a measurement, and, like all measurements, it has both random and systematic error components. The primary source of the random component is the random error committed in the measurement, which is at least 1 CV_r. The systematic component is linked to the reference material value error, which cannot be reduced to the uncertainty of the nominal value specified in the traceability certificates. A more significant source is the reconstitution error, which includes two volume measurements. The overall calibration error was estimated to be 1-2 CV_r. The estimation was experimentally verified through repeated calibrations (under repeatability conditions) and by comparing the calibration factors obtained in new calibrations using the same reagent and reference material lots. Both confirmed the estimations.

According to another quintessential principle valid in all exact sciences, calibration cannot correct smaller errors than the average calibration error. Therefore, such minimal biases are incorrigible, and alarms must be avoided in those cases. The actual Westgard-rules-based QC system neglects the importance of incorrigible biases. The decisions based on the Westgard rules are only correct if the QC graphs are designed with a correct estimator of the σ parameter of the Gauss equation, which is s_r. Calculations based on the laws of the normal distribution predict that the Westgard rules, based on s_r, are a guaranteed fail, causing hundreds of false alarms. The fact that these alarms are not observed suggests an overestimation of the σ parameter.

As JO Westgard and T Groth acknowledged, the calculations behind the Westgard-rules-based QC system are based on the assumption of "within-run (repeatability) SD" (s_r). In contrast, the graphs are designed with "total SD" (s_{RW}). JO Westgard et al. recommended a proportional but inaccurate increase in the decision limits, overestimating σ with s_{RW}. It is inaccurate for two reasons: (1) due to the variability of s_{RW}, which is not calculated from normally distributed data sets, and (2) s_{RW} is a parameter determined under variable conditions in the past, inconsistent with the QC decisions made under repeatability conditions in the present.

A more accurate correction is a constant increase in the decision limits by adding the estimated incorrigible bias to all decision limits and simultaneously using sr in the design of the QC graphs. In the case of moderate ratios (around 1.5), the decision limits for the within-run rules are similar to those proposed by Westgard et al. (for example, $3s_{RW} \approx 3s_r + B_{incorrigible}$). Significant differences can be observed in the case of bigger ratios and if the cross-run rules (R_{4-1S} and R_{10X}) are used. Correcting the estimator of the σ parameter (from s_{RW} to s_r) changes all calculations in the QC system, including the design of the QC graphs, Bcrit, and Nσ, and a complete reevaluation of the QC system becomes necessary.

A comparative analysis of the efficiency of the proposed rules with the Westgard rules exceeds the word count limits of this paper. An independent study is necessary.

Author statements

To this study, no other persons contributed except the author.

This study was neither funded nor sponsored.

The author did not use patient data in this study.

The author created all images and tables.

The author attests that this manuscript did not use generative artificial intelligence (AI) technology to generate figures, ideas, data, or other informational content. AI was used only for grammar correction and unintentional plagiarism detection. To assist with language correction, the author utilized the following Grammarly AI prompts: "Improve it" and "Find synonyms."

Abbreviations

QC	Quality control
RM	Reference material (calibrator)
TEa	Admitted total error
SE	Systematic error component
RE	Random error component
VCSE	Variable component of the systematic error
CCSE	Constant component of the systematic error
A	Absorption
A0	Null-point absorption, the intercept of the linear calibration graph
DA	Difference in absorption (=A-A0)
F_{cal}	Slope factor of the calibration graph in linear calibrations
σ	The sigma parameter of the Gauss equation, the estimator of the mean RE
SD	Standard deviation
s	Measured SD
CV	Coefficient of variation
CV_{Fcal}	Coefficient of variation calculated from the calibration slope factors
CV_{VCSE}	CV calculated from the variable daily bias values, or the VCSE
B	Bias expressed in absolute values
B%	Bias expressed in percent of the target value
$B_{incorrigible}$	Bias incorrigible by calibration
Nσ	sigma metrics
$_{r}$	Repeatability conditions of measurement
$_{RW}$	Reproducibility within laboratory conditions of measurement.

References

1. Plebani M. Errors in clinical laboratories or errors in laboratory medicine? Clin Chem Lab Med. 2006;44(6):750-759. doi: 10.1515/CCLM.2006.123. https://pubmed.ncbi.nlm.nih.gov/16729864/

2. Westgard, JO: Advanced Quality Management / Six Sigma, [Internet]. Westgard QC [cited 2025 Jan 25]. Available from: https://westgard.com/lessons/advanced-quality-management-six-sigma/lesson67.html

3. Gauss, CF. Bestimmung der Genauigkeit der Beobachtungen. Zeitschrift für Astronomie und Verwandte Wissenschaften.1816;1:187–197.

4. Oosterhuis WP, Bayat H, Armbruster D, Coskun A, Freeman KP., Kallner A, Koch D, Mackenzie F, Migliarino G, Orth M, Sandberg S, Sylte MS., Westgard S, and Theodorsson E. The use of error and uncertainty methods in the medical laboratory CCLM. 2018;56(2),209-219. DOI: 10.1515/cclm-2017-0341 https://doi.org/10.1515/cclm-2017-0341 Available from URL: https://www.degruyter.com/document/doi/10.1515/cclm-2017-0341/html#j_cclm-2017-0341_ref_030_w2aab3b7c59b1b6b1ab2b3c30Aa

5. Thompson, M. Towards a unified model of errors in analytical measurement, Analyst, 2000;125 (11),2020-2025, The Royal Society of Chemistry, DOI:10.1039/B00637M, https://doi.org/10.1039/B006376M Available from URL: https://pubs.rsc.org/en/content/articlelanding/2000/an/b006376m

6. Theodorsson E, Magnusson B, Leito I: Bias in Clinical Chemistry, Bioanalysis, 2014;6(21), 2855–2875. https://doi.org/10.4155/bio.14.249, published online Dec 8, 2014, Available from: https://www.future-science.com/doi/10.4155/bio.14.249

7. Vandra AB. Reevaluation of the Variable Component of the Systematic Error Calls for Paradigm Change in Clinical Laboratory Quality Control. 2023 May 28; Preprint, medRxiv 2023.05.24.23290382; doi: 10.1101/2023.05.24.23290382 available from: https://doi.org/10.1101/2023.05.24.23290382

8. CLIA Final Rule: "Equivalent" Quality Control Practices. [Internet] Westgard QC [Cited 2025 Jan 26] Available from: https://westgard.com/clia-a-quality/clia-final-rule/249-cliafinalrule7.html

9. Toacşe G. , Toacşe AM: Controlul de calitate şi validarea metodelor analitice cantitative pentru laboratoarele medicale Bucureşti: Editura Tehnică, 2010.

10. Westgard JO, Groth TL. Power functions for statistical control rules. Clin Chem. 1979;25(6): 863-869. PMID: 445821. DOI: 10.1093/clinchem/25.6.863 Available from URL: https://www.researchgate.net/publication/22696098_Power_Functions_for_Statistical_Control_Rules

11. Magnusson B, Näykki T, Hovind H, Krysell M, Sahlin E: Handbook for calculation of measurement uncertainty in environmental laboratories Nordtest, NT TR 537 – Edition 4, p 10, Available from: http://www.nordtest.info/wp/2017/11/29/handbook-for-calculation-of-measurement-uncertainty-in-environmental-laboratories-nt-tr-537-edition-4/

12. Cooper G. Resolving QC failures. Medical Laboratory Observer online. [Internet] Jul 20, 2016, [cited 2025 Jan 28] Available from URL: https://www.mlo-online.com/management/qa-qc/article/13008813/resolving-qc-failures

13. Seamaty Technology: Common Errors and Pitfalls in Clinical Chemistry Analyzer Testing [Internet] Seamaty Technology, Apr 17, 2023 [cited 2025 Jan 27] Available from https://en.seamaty.com/index.php?s=/sys/547.html

14. Sonntag O, Loh TP. Calibration - an under-appreciated component in the analytical process of the medical laboratories. Adv Lab Med. 2023;5(2):148–152. doi: 10.1515/almed-2023-0127. PMID: 38939195; PMCID: PMC11206179. Available from URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC11206179/

15. T. Badrick: The quality control system. Clin Biochem Rev. 2008;Suppl 1: S67-S70. Available from URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2556587/

16. Burtis CA, Ashwood ER (editors). Tietz Textbook of Clinical Chemistry. 2. Philadelphia, USA: W.B. Saunders; 1994.

17. White GH. Basics of estimating measurement uncertainty.

Clin Biochem Rev. 2008;29 Suppl 1(1): S53-S60. Available from URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2556585/
18. Coskun A. Bias in Laboratory Medicine: The Dark Side of the Moon. Ann Lab Med. 2024;44(1) 6–20. doi: 10.3343/alm.2024.44.1.6. Epub 2023 Sep 4. PMID: 37665281; PMCID: PMC10485854. Available from URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC10485854/Bias in Laboratory Medicine: The Dark Side of the Moon - PMC
19. NIST. Engineering Statistics Handbook. 1.3.6.7.1. Cumulative Distribution Function of the Standard Normal Distribution [Internet] NIST [cited 2025 Jan 29] Available from URL: https://www.itl.nist.gov/div898/handbook/eda/section3/eda3671.htm
20. Haeckel R, Wosniok W. Observed, unknown distributions of clinical chemical quantities should be considered to be log-normal: a proposal. Clin Chem Lab Med. 2010;48(10):1393–1396. doi: 10.1515/CCLM.2010.273.
21. Badrick T. Biological variation: Understanding why it is so important? Pract Lab Med. 2021 Jan 4;23:e00199. Doi: 10.1016/j.plabm.2020.e00199. Available from URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7809190/#bib136
22. Katayev A. & Fleming J. Past, present, and future of laboratory quality control: patient-based real-time quality control or when getting more quality at less cost is not wishful thinking —Journal of Laboratory And Precision Medicine, 2020(5). Doi:10.21037/jlpm-2019-qc-03 Available from URL: https://jlpm.amegroups.org/article/view/5718/html#B6
23. Haeckel R, Schneider B: Detection of drift effects before calculating the standard deviation as a measure of analytical imprecision. J Clin Chem Clin Biochem. 1983;21(8):491-497. doi: 10.1515/cclm.1983.21.8.491. PMID: 6631335 Available from URL: https://pubmed.ncbi.nlm.nih.gov/6631335/
24. Mackay M, Hegedus G, Badrick T. Assay Stability, the Missing Component of the Error Budget, Clinical Biochemistry, 2017;50(18): 1136-1144, ISSN 0009-9120, https://doi.org/10.1016/j.clinbiochem.2017.07.004.
25. Krouwer JS: Setting Performance Goals and Evaluating Total Analytical Error for Diagnostic Assays, Clinical Chemistry, 2002;48(6):919–927, https://doi.org/10.1093/clinchem/48.6.919
26.Westgard JO Basic QC Practices. QC - The Calculations. [Internet] Westgard QC, [cited 2025 Jan 30] Available from URL: http://westgard.com/lessons/basic-qc-practices-l/lesson14.html
27. Kumar BV, Mohan T: Sigma metrics as a tool for evaluating the performance of internal quality control in a clinical chemistry laboratory, Tables 3 and 4, J Lab Physicians. 2018;10(2):194–199.,doi:10.4103/JLP.JLP_102_17, available from URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5896188/
28. Westgard JO, Westgard S. Introducing Westgard Sigma Rules, [Internet]2014 Sep. [Cited 2025 Apr 2. Available from URL: https://westgard.com/lessons/westgard-rules/westgard-rules/westgard-sigma-rules.html
29. Grammarly. Grammarly [Internet]. Grammarly.com. 2025. Available from: https://app.grammarly.com/