Performance Standards for Quality Control Systems

Learning objectives:

1. Identify the steps to design an evidence-based QC system

2. Set QC limits and frequency based on the sigma value of each assay

3. Use the concept of total error to evaluate method performance

Abstract:

Standard statistical QC has been around since its introduction in 1950 by Levey and Jennings. Today, while most laboratories use standard statistical QC processes, with or without Westgard multi-rules to evaluate quality, few incorporate performance standards in their QC systems. With performance standards, the quality level of each test is determined on a sigma scale. This guides the selection of optimal QC frequency and rules to monitor the analytical system. QC rules are selected based on the sigma score and are specific to each test. While a single rule (1_{3s}) is sufficient to efficiently monitor tests with sigma quality at or above 6, a more stringent rule ($1_{3s}/2_{2s}/R_{4s}/4_{1s}$) is required for those performing at four sigma. The comparison of analytical total error (TE) to the total error allowable (TEa) helps ensure that the system is operating within the defined quality specifications; therefore accurate patient results are produced.

Abbreviations: ΔSEc- Critical systematic error, DPM- Defects per million, EQA- External quality assurance programs, IQC- Internal quality control, QC- Quality control, SD- Standard deviation, TE-Total error, TEa- Total error allowable.

Index terms: Performance standards, quality goals, evidence-based quality control, critical systematic error.

Introduction

  Laboratory testing plays a tremendous role in the diagnosis, treatment, monitoring, and management of patients' conditions. The accuracy of laboratory results is critical for the delivery of quality healthcare.[1] Laboratories ensure the accuracy of their test results through the implementation of an internal quality control (IQC) system and through the participation of external quality assurance programs (EQA). A properly designed quality control (QC) system should effectively detect errors in the analytical system while minimizing false rejection.[2,3]

  The standard QC practice in most laboratories is a statistical QC system in which the control limits are set at $\pm$ two standard deviation (SD) from the mean.[3] With the improvement of technology and methodology over the years, analytical systems have shown better accuracy and precision, and small changes in analyzers' performance may not be easily identifiable by standard QC processes.[2,3] In order to optimize quality control processes, clinical laboratories need to implement evidence-based quality control. Evidence-based quality control requires quality goals against which the analytical system performance will be compared.[3,4] Quality goals, also known as performance standards, are the acceptable limits for both random and systematic analytical errors and are used to determine if analytical methods are producing clinically acceptable results.[4,5] Performance standards are expressed as total error allowable (TEa). With the incorporation of quality goals, not only can standard QC systems detect an error in the operating system, but they can also safely determine if the error will lead to unacceptable patient results.[3]

Analytical Systems

The laboratory testing process can be divided into three phases (pre-analytical, analytical, and post-analytical) and errors can occur at any stage.[1] Errors associated with analytical systems are monitored through the implementation of an internal quality control (IQC) system and are the main focus of this paper. Laboratories also participate in external quality assurance (EQA) or proficiency testing (PT) programs to ensure that they are meeting the quality requirements set by regulatory bodies.

Analytical systems are subject to systematic and random errors. Bias is an estimate of systematic error and standard deviation (SD) is the measure of a random error. When a stable quality control material is measured for a period of time, it creates a population of data points which, when plotted, will display a normal or Gaussian distribution. With a single data population, it can be predicted that approximately 68% of the data, or area under the curve, will fall within $\pm$ 1SD of the mean, 95% between $\pm$ 2SD, and 99.7% within $\pm$ 3SD. Gaussian statistics and predictions are the key to understanding statistical quality control in clinical laboratory settings.[3,4,5]

History of Quality Control

Quality control (QC) has been utilized in medical laboratories for several decades and has evolved over the years with advances in technology. Before the introduction of statistical QC by Levey and Jennings, laboratory tests were performed manually in batches where quality control samples were run along with patients' samples. Statistical QC arose with the introduction of the Levey-Jennings chart in 1950, and quality control samples were run in duplicate with the range of acceptability set at $\pm$ 3 standard deviation (SD) from the statistical mean.[4,5]

With the introduction of automation in 1960, QC samples were reduced to a single measurement and the QC limits were set at ± 2SD.[4,6] Statistical QC became the basis of QC for quantitative tests in clinical laboratories.[6] This QC acceptability criteria worked well with the single test auto-analyzer until multi-channel analyzers made their way into clinical laboratories. There is a 5% false rejection rate when the ± 2SD rule is used for one analyte. For two levels of control the probability of false rejection increases to 9.5%.[6]

With multi-test systems, several different tests can be performed simultaneously on one analyzer. With simultaneous measurements, the rate of false rejection increases due to a multiplier effect.[6] In order to maximize error detection and minimize the false rejection rate, Dr. Westgard developed a set of rules, known as Westgard rules[7], in which QC limits were set at ± 3 SD and ± 2SD violation was used as a warning.[4,6]

Today, most laboratories still use the standard statistical QC with or without Westgard multi-rules. Statistical QC design is based on the mean and standard deviation (SD) mostly calculated from 20 QC data points when a new QC material is started.[3] This standard QC system sets limits at ± 2SD from the mean to assess quality.[2,3] The question arises: to what level of quality are we controlling?

Performance Standards

Expressed as total error allowable (TEa), quality goals, also known as performance standards, are the acceptable limits for both random and systematic errors and are used to determine if analytical methods are producing clinically acceptable results.[3,4,5] While selecting quality goals, laboratories should ensure that they are based on scientific evidence and are attainable by the analytical system.

Various quality requirements have been identified over the years leading to several sources of performance standards. In 1999, a conference was held in Stockholm to reach a consensus on how to set quality goals in laboratory medicine. The conference, sponsored by the International Union of Pure and Applied Chemistry (IUPAC), the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), and the World Health Organization (WHO), included participants from 27 countries.[7] The consensus statement laid out the hierarchy of models from the highest to the lowest. Five models were described and based on:

1. Clinical requirements or clinical outcomes

2. Biological variation within and between individuals

3. Professional recommendations such as those seen in professional expert publications

4. Quality specifications set by regulatory bodies (e.g. CLIA), and proficiency testing organizer (e.g. CAP)

5. State of the art performance.

These five models were later reduced to three after the Stockholm consensus was revised in 2014 in Milan. The first and second models were kept intact, while the remaining three were combined under the umbrella of state of the art performance.[8]

Evidence-based Approach to Quality Control

Unlike traditional QC design, evidence-based QC requires the laboratory to set quality goals for each analyte tested. Expressed as total error allowable (TEa), quality goals or performance standards, are the acceptable limits for both random and systematic errors and are used to determine if analytical methods are producing clinically acceptable results.[3]

Method performance is assessed by evaluating the analytical total error (TE), which is the combined effect of bias and imprecision.[3,9] By comparing the TE to TEa, the laboratory can determine if their analytical system is operating within their defined quality specifications. In this model of design, QC limits are determined based on the analytical system capability in terms of sigma metric (σ). Another useful quality indicator of method performance is critical systematic error (ΔSEc) which is the number of SD that the mean can shift before 5% of results fall outside a defined quality specification (TEa).[3,4,10,]

True/target value

The analyte true value, used to monitor a method's accuracy, is the best available estimate of the analyte value and is preferably based on a peer group method mean.[11] True value should not be confused with the laboratory observed mean. If a laboratory doesn't participate in an inter-laboratory comparison program, the manufacturer's target value may be used. When those two options are not available, the laboratory's own historical data may be used.[3,4]

Total error

Once the true value is determined, the analytical mean bias can be calculated. The bias represents the systematic error and is the measure of the difference between the laboratory observed mean and the true value. Random error, also known as imprecision, is expressed as standard deviation and is the measure of dispersion around the mean. The total variation of a test result from the true value combines the effects of bias and random error and is defined as total error (TE).[3,4] Figure 1 below shows the visual representation of total error.[9,18]
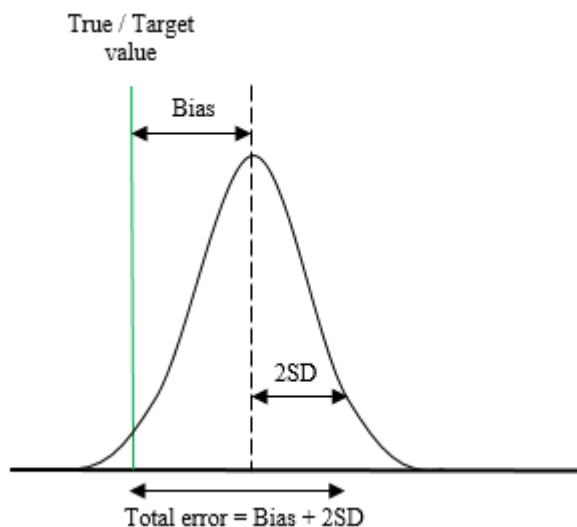
Figure 1. Illustration of the concept of total error which is the combination of systematic error (bias) and random error (2SD).

The following formula is used to calculate the analytical total error:

TE = Bias +2SD

Method performance is evaluated by comparing TE to the total error allowable.

Total error allowable (TEa)

TEa is the tolerance limit, the maximum acceptable variation from the true value of an assay.[3,4] As pointed out earlier, there are several sources of TEa limits. In the US, the Clinical Laboratory Improvement Amendments (CLIA) set limits for acceptable method performance. For example, the CLIA limit for chloride is the target value ± 5%. The TEa in this case is ± 5% of the true value of the analytes.[3,4] This criterion is also used by the College of American Pathologists (CAP). Regulatory agencies from different countries may have different limits. In Australia, the Royal College of Pathologists of Australasia (RCPA) set the limit for acceptable analytical performance for chloride at 3mmol/L when levels are below or equal to 100 mmol/L and at ± 3% if levels are above 100 mmol/L.[13] Chloride's quality specification is more stringent when based on biological variation and is set at ±1.5% for desirable performance.[14]

Sigma metric

By definition, the sigma metric is the number of SD that the closest tolerance limit is from the mean of the assay.[3,4] The concept of six sigma is widespread in industry quality management and requires defined tolerance limits in order to evaluate quality in terms of a sigma metric.[3,4,15,16,17] In industry, a sigma value ranges from 1 to 6 and relates to the number of defects per million (DPM). A system with one sigma produces 697,700 DPM, while a six sigma process generates 3.4 DPM. Six sigma is regarded as world class quality and three sigma (66,807 DPM) is the minimum recommended for routine production.[4,11,12] In the laboratory, total allowable error (TEa) represents the tolerance limits and is used to calculate an assay performance on a sigma scale (Figure 2).[18] Sigma metric is calculated as follows: Sigma metric $(\sigma) = (TEa - Bias)/SD$
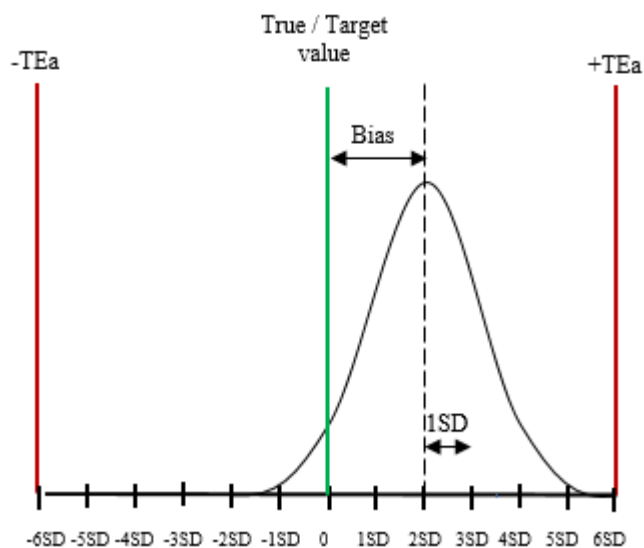
Figure 2: Sigma metric. The graph shows how six sigma tolerance limits apply to QC in a clinical laboratory. The sigma value of this essay is 4 as the mean is at 4SD from the upper tolerance limit.

Assuming a laboratory test has a bias = 3, SD = 2, and the TEa set at 13, the sigma metric for the assay is (13-3)/2 = 5. If the same test has a larger bias = 4 and SD = 3, the sigma value

will be (13-4)/3 = 3. There a positive correlation between the sigma score and a test performance. The higher the sigma metric, the better the test performance.

Setting rules based on sigma value

Standard QC systems use the same QC rules (mostly ± 2SD) for all analytes regardless of the level of performance, whereas the rules in evidence-based QC systems are based on each analyte's performance expressed in the sigma metric.[10] Westgard Sigma rules allow laboratories to select QC limits and the number of control measurements needed to achieve optimal error detection (≥ 90%) and keep false rejection below 5%.[6,10,19,20]

Table 1 shows QC limits and frequency based on the assay's capability. Highly capable tests with sigma scores at or above six don't require stringent QC rules as they can be controlled easily and efficiently.[4,9] Two levels of control run once every 24 hours, with limits set at 1_{3s}, are enough to control the quality of the results. The probability of false rejection is 0%, meaning any error detected is significant and should be addressed. At five sigma, the $1_{3s}/2_{2s}/R_{4s}$ QC rule is needed to bring about error detection above 90% and false rejection below 5%.[5,20]

Table 1: QC rules based on sigma metric for two control measurements per run

Sigma	QC rules	Number of run
≥ 6	1_{3s}	1
5	$1_{3s}/2_{2s}/R_{4s}$	1
4	$1_{3s}/2_{2s}/R_{4s}/4_{1s}$	2
<4	$1_{3s}/2_{2s}/R_{4s}/4_{1s}/8_{x}$	4

Tests performing at four sigma require additional rules ($1_{3s}/2_{2s}/R_{4s}/4_{1s}$) and increased QC frequency (two levels of control run twice a day or 4 levels run once a day).

The maximum QC rule ($1_{3s}/2_{2s}/R_{4s}/4_{1s}/8_{x}$) and higher QC frequency is recommended for performance below four sigma. The appropriate QC frequency is four times in 24 hours when

two levels of control are used, and twice a day when four levels of control are measured per run.[5,20]

Critical systematic error (ΔSEc)

ΔSEc is used to monitor method performance over time. As defined previously, ΔSEc measures how far the mean can shift before 5% of the results exceed a TEa limit.[3,4] ΔSEc is calculated as:

ΔSEc = (TEa – Bias)/SD -1.65 or ΔSEc = sigma metric -1.65

1.65 represents the z-score associated with 5% risk of producing erroneous results. It corresponds to 95% confidence limits.[3,4] For example, a test method with a quality level at six sigma is associated with a ΔSEc of 4.35 (6-1.65 = 4.35), therefore, the mean can shift 4.35SD before 5% of results exceed TEa. On the other hand, a test method with 1.65 sigma value is already producing 5% erroneous results as ΔSEc is 0 (1.65-1.65 = 0). The probability of producing clinically misleading patient results or unacceptable proficiency tests will exceed 5% with any increase in bias or imprecision.

Conclusion

Even though multiple analytes are run on the same analyzer platform, they do not always perform at the same level. Quality varies between tests and from one analytical system to another, therefore it may not be appropriate to set the same QC rules for different analytes. Setting quality goals allows laboratories to design a more efficient QC system and select error detection limits specific to the quality level (sigma score) of each analyte. Such systems can be objectively monitored for improvement. The calculated monthly ΔSEc tracks the performance of the analytical system relative to the quality goals (TEa) set. The ΔSEc decreases as the total error produced moves closer to the error limits set (TEa). A ΔSEc of 0 indicates that the

probability of producing clinically misleading patient results or unacceptable proficiency tests is greater than 5%.[7]

References

1. Hammerling JA. A review of medical errors in laboratory diagnostics and where we are today. Laboratory medicine 2012; 42(2): 41-44.

2. Stackhouse JW. Into quality control. Cola's insights. 2013; Available at: http://www.cola.org/insights-newsletters/2013/April/insights.pdf.

3. Brooks ZC. Performance-driven quality control. Washington DC, USA: AACC Press; 2001.

4. Westgard JO. Six sigma quality design & control: desirable precision and quantitative measurement processes, 2nd ed. Madison, WI. Westgard QC Inc. 2006.

5. Westgard JO. Basic QC practice: training in statistical quality control for medical laboratories. 3rd ed. Madison, WI: Westgard QC, Inc.; 2010.

6. Westgard JO. Basic QC practices: Historical Perspective on Laboratory QC: Where we've been and where we're going! 2011; Available at: https://www.westgard.com/history-and-future-of-qc.htm.

7. Petersen PH, Fraser CG, Kallner A, Kenny D. Strategies to set global analytical quality specifications in laboratory medicine. Scandinavian Journal of Clinical and Laboratory Investigation 1999; 59: 473-585.

8. Special issue: 1st EFLM strategic conference defining analytical performance goals- 15 years after Stockholm conference. Clinical Chemistry and Laboratory Medicine 2015; 53: 829-958.

9. Westgard JO, Westgard SA. Total analytical error from concept to application. Clinical laboratory news. Available at: https://www.aacc.org/publications/cln/articles/2013/september/total-analytic-error.

10. Westgard JO, Westgard SA., Quality control review: implementing a scientifically based quality control system, Ann. Clin. Biohem. 53 (2016) 32–50.

11. Schoenmarkers CHH, Naus AJM, Vermeer HJ, et al., Practical application of sigma metrics QC procedures in clinical chemistry, Clin. Chem. Lab. Med. 2011; 49: 1837 – 1843.

12. Jassam N, Bosomworth M, Thompson D, Lindsay C, Barth JH. Managing quality in networked laboratories: a system for managing quality in networked laboratories – statistical control rules design. Available at: https://www.westgard.com/networked-lab-quality.htm.

13. RCPA QAP Allowable limits of performance (ALP). Available at http://www.rcpaqap.com.au/docs/2014/chempath/ALP.pdf .

14. Westgard JO. Desirable Specifications for Total Error, Imprecision, and Bias, derived from intra- and inter-individual biologic variation. Available at: https://www.westgard.com/biodatabase1.htm .

15. Bais R. Use of Capability Index to Improve Laboratory Analytical Performance. The Clinical Biochemist Reviews. 2008; 29 (Suppl 1):S27-S31.

16. Singh B, Goswami B, Gupta VK, Chawla R, Mallika V. Application of sigma metrics for the assessment of quality assurance in clinical biochemistry laboratory in India: A pilot study. Indian J Clin Biochem. 2011;26:131–5.

17. Gami B, Patel D, Chauhan K, Shah H, Haridas N. Sigma metrics as quality marker for analyzing electrolytes in laboratory. International Journal of Advanced Research. 2013; 7(1): 197-201.

18. Westgard JO, Westgard SA. Assessing quality on the sigma scale from proficiency testing and external quality assessment surveys, Clin. Chem. Lab. Med. 2015; 53:1531 – 1535.

19. Jones G, Calleja J, Chesher D, et al. Collective opinion paper on a 2013 AACB workshop of experts seeking harmonization of approaches to setting a laboratory quality control policy. Clinical biochemistry. 2015; 36(3):87-95.

20. Westgard JO, Westgard SA. Westgard Sigma Rules. 2014; Available at: https://www.westgard.com/westgard-sigma-rules.htm.