LEARNING GUIDE

SIX SIGMA-BASED QUALITY CONTROL

CORELABORATORY.ABBOTT

Abbott

ACKNOWLEDGEMENTS

DR. JAMES O. WESTGARD

Dr. James O. Westgard is president of Westgard QC, Inc., a small business providing education and training for laboratory quality management. He is an emeritus professor in the Department of Pathology and Laboratory Medicine at the University of Wisconsin Medical School. He spent many years as faculty director of Quality Management Services for the Clinical Laboratories at the University of Wisconsin Hospital and Clinics.

A native of North Dakota, Dr. Westgard obtained a bachelor's degree in chemistry from Concordia College in Moorhead, Minnesota, and master's and doctorate degrees in analytical chemistry from the University of Wisconsin, Madison. He began his career at the University of Wisconsin as a clinical chemist in 1968. His early interest was in the development of method evaluation protocols, and he served as the first chairman of the Evaluation Protocols Area Committee at the Clinical and Laboratory Standards Institute (CLSI). His interest in quality control began in 1976 and 1977 when he was on sabbatical leave at Uppsala University in Sweden working with Professor Carl-Henric deVerdier and Drs. Torgny Groth and Torsten Aronsson. This work led to the multirule control procedure, internationally known as "Westgard Rules."

STEN WESTGARD, M.S.

Sten Westgard, M.S., is the Director of Client Services and Technology for Westgard QC. For more than 20 years, Mr. Westgard has managed the Westgard website, course portal and blog, creating and administering online training, as well as editing and writing hundreds of reports, essays and applications on quality control, method validation, Six Sigma risk management and other lab management topics. He has edited and contributed to numerous books on quality, including *Basic QC Practices, Basic Method Validation, Six Sigma QC Design and Control, Six Sigma Risk Analysis, CLIA Final Rules, Assuring the Right Quality Right, The Poor Lab's Guide to the Regulations* and *Nothing but the Truth about Quality*. Mr. Westgard is also an adjunct faculty member of the Mayo Clinic School of Health Sciences in Rochester, Minnesota.

HOW TO USE THIS LEARNING GUIDE

This guide is organized into six chapters and a glossary. Each chapter includes a "What's the Point" summary and references for additional reading. The glossary includes an extensive list of key terms found in each chapter, followed by a list of common abbreviations. Readers are encouraged to visit the Westgard QC website (www.westgard.com) for additional information.

CONTENTS

FOREWORD

It is standard practice in the clinical lab to test quality control (QC) samples (typically called "controls") at least daily and, preferably, multiple times per day. Controls are necessary to ensure analytical systems are of sufficient quality and "fit for purpose" because the performance of even the best assays can change over time. Statistical quality control (SQC) practices have long been established for this purpose.

However, once QC practices are established, there's a tendency to accept them without question and to continue to follow routine algorithms without periodic critical re-examination of their appropriateness. A one-size-fits-all approach, such as testing two controls once a day, is common, but it's a minimal practice. Such a simple QC rule is easy to follow, but it ignores the fact that not all assays are of equal analytical quality. An optimal QC program recognizes the need for customized QC rules for assays based on their inherent variability and establishes practical rules to minimize false rejection of acceptable patient test results and false acceptance of unacceptable results.

Dr. James Westgard has devoted his career to developing QC best practices and assessing analytical quality through tools such as the Sigma-metric. This learning guide presents his concepts, based on real working conditions in routine clinical labs. Clinical labs have used the Westgard Rules for years, and labs routinely apply the Six Sigma metrics approach today. But, as Dr. Westgard himself notes, each lab must assess its performance and apply the QC algorithms best suited to it. This requires a lab to set quality targets for every analyte and measure each assay's bias and imprecision. With these basic data, the lab can calculate the Sigma-metric and select appropriate QC rules based on analytical quality. This learning guide provides labs with sufficient basic information to create a practical, useable QC plan specific for their facilities. It also describes current risk-based approaches to QC.

It's important to note that SQC is necessary for optimal lab practice and patient care, but it is not sufficient on its own. SQC addresses variability in the analytical phase, but errors can also occur in the pre- and post-analytical phases, as well. In addition, external quality assessment/proficiency testing (EQA/PT) programs are an essential and independent means to assess a QC program's effectiveness.

Dave Armbruster, Ph.D., DABCC, FACB
Director, Clinical Chemistry, Abbott Diagnostics

CHAPTER 1

MANAGING ANALYTICAL QUALITY

INTRODUCTION

Statistical quality control (SQC) is an essential lab practice to ensure reported test results achieve the quality required for intended medical use. Analytical quality demands continue to increase as physicians and patients depend increasingly on test results for optimal diagnosis and treatment. Although modern highly automated systems provide many function checks to ensure proper operations to produce accurate results, no analytical system is perfectly stable. Labs need SQC to provide a final independent check to detect performance changes potentially causing medically important errors.

SQC's long history started with Levey and Jennings in the 1950s[1]. Today, labs still use the classic Levey-Jennings control chart, though they've updated the decision criteria and often employ Westgard Rules[2]. Current practice is to optimize SQC rules for individual assays based on their inherent quality (bias and precision) and the accuracy required for their intended clinical use.

The quality required for intended use is defined as allowable total error (TEa). The observed precision– (SD or %CV) and the observed bias are used to calculate the Sigma-metric, that is:

Sigma-metric = (TEa– Bias)/CV,

where all values are either in concentration units or percentages. Assays with high Sigma-metrics require minimal SQC, and assays with low Sigma-metrics require more extensive SQC rules.

"Do the right SQC right" is the objective for good laboratory practice. "Do the right SQC" means selecting the appropriate control rules and number of controls to detect medically important errors. QC tools available include:

- Sigma-metric SQC selection tool[3]
- Charts of operating specifications[4,5]
- Westgard Sigma Rules[6]

With proper selection and design, SQC is a powerful technique to monitor performance and ensure the quality of test results meets the defined clinical needs.

"Do SQC right" means:

- Selecting controls at appropriate concentrations
- Determining assay precision
- Calculating the right control limits
- Testing controls at the right times
- Interpreting control results correctly
- Taking appropriate actions based on the control results
- Documenting those actions

SQC will not achieve optimal performance unless it is properly implemented.

Figure 1-1 shows lab staff SQC responsibilities. Managers or technical specialists are responsible for:

- Establishing the SQC procedure by designing the SQC rules
- Selecting controls
- Determining the QC means and SDs from control measurements
- Calculating control limits
- Preparing control charts or setting the QC software parameters used

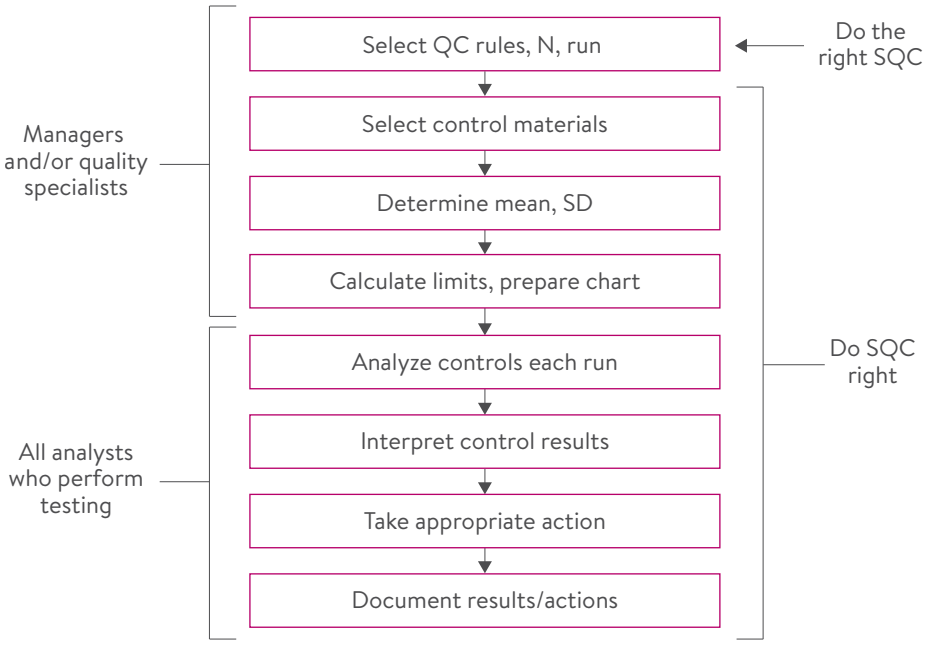

Figure 1-1: Laboratory process for "doing the right SQC right."

All analysts implement SQC systematically, following the standard operating procedures. This includes:

- Analyzing controls at the times and intervals prescribed
- Interpreting control results
- Taking appropriate corrective action
- Documenting all control results and actions

This learning guide's focus is to help labs establish, maintain and perform good SQC practice, but SQC is only one part of a lab's quality management system (QMS). A lab's QMS integrates all SQC technical and management requirements, as described in international quality management guidelines and national lab regulations. It includes many critical factors that contribute to achieving quality test results. Therefore, whether a laboratorian's responsibilities involve all or only some of the steps listed above, all lab personnel must understand the overall SQC process, as well as the larger QMS.

REGULATORY AND ACCREDITATION REQUIREMENTS

Quality control practices must adhere to regulatory and accreditation requirements. In the U.S., the Clinical Laboratory Improvement Amendments (CLIA)[7] provides the minimum regulatory requirements, which include three options:

- Implement control procedures that monitor the accuracy and precision of the complete analytic process, which includes establishing the number, type and frequency of testing control materials; ensuring control procedures detect immediate errors that occur due to test system failure, adverse environmental conditions and operator performance; and monitor, over time, the accuracy and precision of test performance that may be influenced by changes in test system performance, environmental conditions and variance of operation performance.
- At least once each day, analyze or examine patient specimens, using the following controls:
  - For each quantitative procedure, include two control materials of different concentrations.
  - For each qualitative procedure, include a negative and positive control material.
- Perform control procedures that provide *equivalent quality testing*, as specified in Appendix C of the State Operations Manual. As of January 2016, this option is described as an Individualized Quality Control Plan (IQCP), which consists of three components: a risk assessment, a QC plan and a quality assessment program.

In comparison, ISO 15189 provides the global standard of practice for accreditation[8] by stating:

> ***"The laboratory shall design quality control procedures that verify the attainment of the intended quality of results."***

This requires defining the intended quality, the quality goals or the requirement to be achieved. Defining the quality goals is the starting point for managing quality in a lab.

QUALITY GOALS

Quality at a minimum is "conformance to requirements" and at a maximum is "demonstration of competency." Lack of quality is measured by defects (that is, test results that exceed allowable error limits for the intended medical use). Quality goals are defined as the allowable total error (ATE in the terminology preferred by FDA; TEa in historical terminology), such as the criteria applied in proficiency testing (PT) or external quality assessment (EQA) surveys. CLIA sets performance criteria for some 70 to 80 "regulated" tests.

Example:

- Glucose should be correct within ± 8% of the target value (TV) or within ± 6 mg/dL at 60 mg/dL and lower.
- Cholesterol should be correct within ± 10% of the TV.

Criteria for more assays are defined by other EQA/PT programs. The College of American Pathologists (CAP) provides PT surveys for all CLIA-regulated tests, plus many others.

Example:

- HbA1c must be accurate within ± 6% of the TV.

Other quality goals are based on clinical outcome studies, the expected biologic variation, opinions of expert groups, surveys of physician use and interpretation of test values. The most extensive recommendations are found in the Ricos biological variability goals database, originally published in 1999[9] and updated every two or three years on the Westgard website[10].

Biological variability goal-setting models described by Fraser and Petersen[11] are used to define the allowable analytical CV (CV_a), the allowable analytical bias ($Bias_a$) and the allowable biologic total error (TEa_b), as follows:

$$\mathbf{CV_a = 0.5 \times CV_i}$$
$$\mathbf{Bias_a = 0.25 \times (CV_i^2 + CV_g^2)^{1/2}}$$
$$\mathbf{TEa_b = Bias_a + 1.65CV_a,}$$

where CV_i is the intra-individual variation and CV_g is the between-individual variation.

The CV_i is used to set CV_a for monitoring individual patients, and CV_i and CV_g are used to set $Bias_a$ for diagnostic classifications versus reference intervals. Combining the two sets is a desirable goal for TEa and, thus, PT/EQA goals[11].

The criteria above are sometimes described as *desirable* in a three-level model that includes *optimal* (more demanding) and *minimal* (less demanding) criteria[12]. In the equation, optimal criteria are based on multipliers of 0.25 for CV_a and 0.125 for $Bias_a$ while minimal criteria are based on multipliers of 0.75 for CV_a and 0.375 for $Bias_a$.

The lab medical director is responsible for defining assay quality goals, which drive the lab's QMS. (For a more detailed discussion of quality goals, see reference 13.)

QUALITY MANAGEMENT SYSTEM (QMS)

The Plan-Do-Check-Act (PDCA) cycle described by W. Edwards Deming provides the fundamental building block for developing, implementing and operating a QMS. Deming assigned management the responsibility for maintaining the balance among the many parts of a production operation and for applying the PDCA cycle to make objective data-driven decisions[14].

PDCA is the "scientific method" for experimentation. Plan an experiment, Do the experiment, Check the data and then Act on that data. Acting on the data often leads to a new experiment and better data and decisions.

PDCA is fundamental for quality improvement in labs, providing a continuous cycle to solve problems and improve quality.

Burnett's book describes how ISO 15189 management and technical requirements fit the PDCA cycle[15]. As shown in **Figure 1-2**, Burnett organizes the quality management process under the headings "Organization and Management," "Resource Management," "Examination Processes," and "Evaluation and Improvement." Management requirements are identified with the number 4 and technical requirements with 5. This process focus is critical for understanding how all the different requirements work together to provide an effective QMS.

Figure 1-2: Burnett's PDCA process model for ISO 15189 QMS.

For analytical quality management, the lab can also implement Six Sigma concepts, metrics and tools as a PDCA cycle, as shown by the Six Sigma quality management system (6σ QMS) in **Figure 1-3**[6,16].

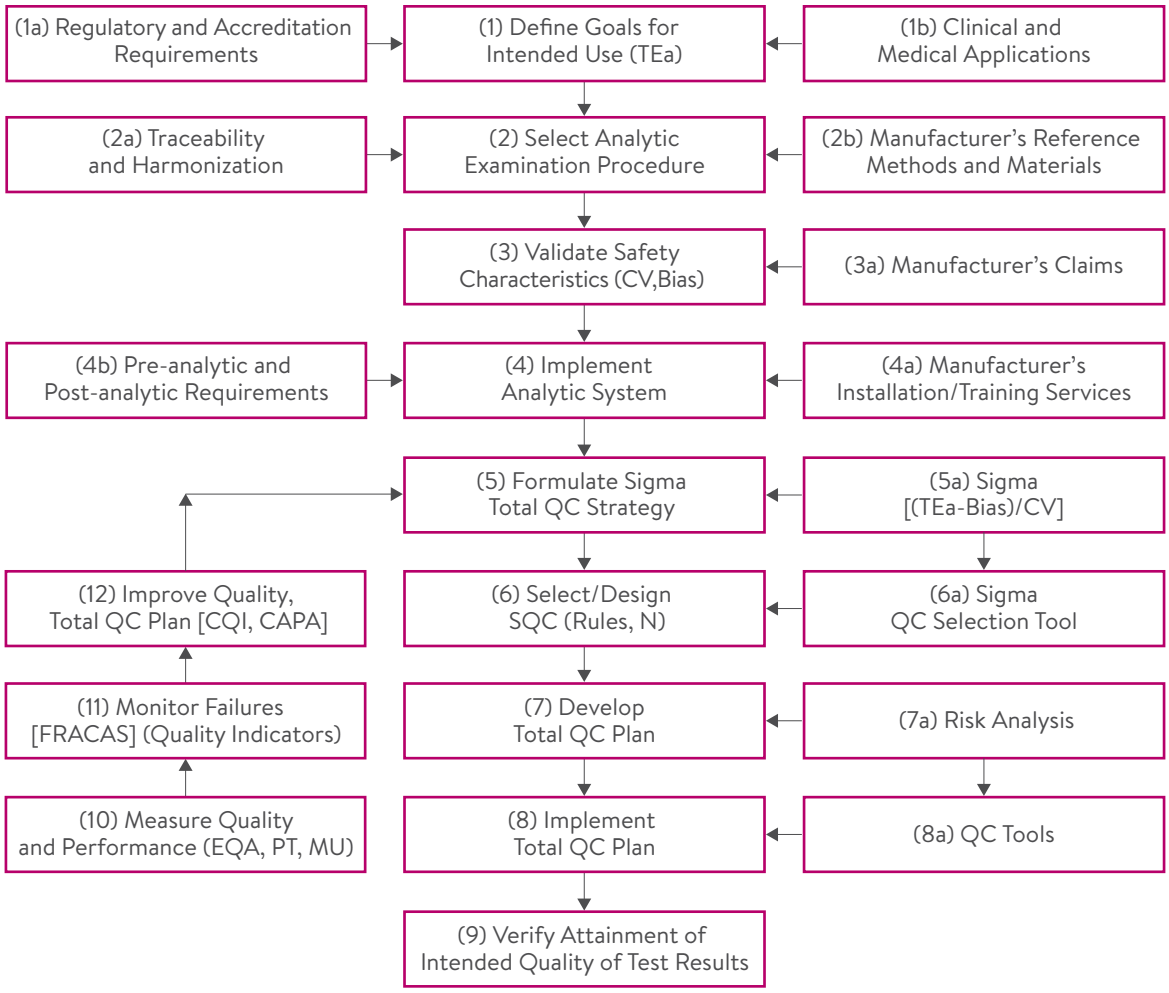

Figure 1-3: Six Sigma Quality Management System.

- **Plan (Steps 1–2):** Define quality goals as allowable total error (TEa). TEa guides selection of the analytic measurement procedure, or examination procedure in ISO terminology.
- **Do (Steps 3–4):** Validate safety characteristics (e.g., precision, bias, reportable range, interferences) using experimental studies and statistical data analysis. Using method performance data and the definition of TEa, calculate a Sigma-metric [Sigma-metric = (TEa – |Bias|)/CV]. Assuming the Sigma-metric indicates acceptable performance (that is, greater than 3), preferably at least 4 and, better yet, 5 or 6, proceed to implement the analytical method. Implementation requires establishing standard operating procedures (SOPs), maintenance schedules, QC procedures, etc., as well as training analysts to understand and follow the SOPs.

- **Check (Steps 5–9):** Knowledge of the Sigma quality drives the check stage, starting with formulation of a total QC strategy encompassing both statistical and non-statistical control mechanisms. The SQC procedure optimizes the control rules and numbers of controls to detect medically important errors. Design a total QC plan to integrate SQC with other control mechanisms that are needed to monitor specific failure modes that may occur with a particular analytic method or instrument system. New risk-based thinking and risk assessment tools are useful for identifying additional controls, particularly for the pre-analytic and post-analytic parts of the total testing process (TTP). Implementation of the total QC plan uses available QC tools and information technology. The outcome is an effective QC process to "verify the attainment of the intended quality of test results," as required by ISO 15189.
- **Act (Steps 10–12):** Finally, monitor the quality of the testing process over time to characterize performance, identify failures and improve the QC plan (go back to Step 5) or to overhaul the entire testing process (go back to Step 1).

WHAT'S THE POINT?

This is the context in which medical labs should practice SQC. SQC is essential but only one part of the QMS. Application of SQC assumes the method has been carefully evaluated and meets the requirements for intended use. The Total QC strategy is adopted based on the known Sigma quality of the testing process and use of the right control rules and right number of controls. Additional pre-analytic and post-analytic controls are implemented as part of the Total QC Plan. Additional controls are used to monitor critical failure modes of the particular testing process. Quality is monitored with EQA/PT surveys and other quality indicators, and problems are identified, corrected and prevented.

References

1. Levey S, Jennings ER. The use of control charts in the clinical laboratory. *American Journal of Clinical Pathology* 1050;20:1059-1066.

2. Westgard JO, Barry PL, Hunt MR, Groth T. A multirule Shewhart chart for quality control in clinical chemistry. *Clinical Chemistry* 1981;27:493-501.

3. C24A3. Statistical Quality Control for Quantitative Measurement Procedures. Wayne, Pennsylvania, Clinical and Laboratory Standards Institute, 2006.

4. Westgard JO. Charts of operational process specifications (OPSpecs Charts) for assessing the precision, accuracy, and quality control needed to satisfy proficiency testing performance criteria. *Clinical Chemistry* 1992;38:1226-1233.

5. Westgard JO, Hyltoft Petersen P, Wiebe D. Laboratory process specifications for assuring quality in the U.S. National Cholesterol Education Program. *Clinical Chemistry* 1991;37:656-661.

6. Westgard JO, Westgard SA. *Basic Quality Management Systems: Essentials for quality management in the medical laboratory.* Madison, Wisconsin, Westgard QC Inc. 2014.

7. U.S. Centers for Medicare & Medicaid Services (CMS). Medicare, Medicaid, and CLIA programs: Laboratory requirements relating to quality systems and certain personnel qualifications. Final Rule. *Federal Register* Jan 24, 2003;16:3650-3714.

8. ISO 15189. Medical laboratories – Requirements for quality and competence. Geneva: ISO, 2012.

9. Ricos C, Alverez V, Cava F, et al. Current databases on biological variation: pros, cons, and progress. *Scandinavian Journal of Clinical Laboratory Investigation* 1999;59:491-500.

10. Online Ricos Database, www.westgard.com/biodatabase1.htm.

11. Fraser CG, Hyltoft Petersen P. Quality goals in external quality assessment are best based on biology. *Scandinavian Journal of Clinical Laboratory Investigation* 1993;53(Suppl 212):8-9.

12. Fraser CG. *Biological Variation: From Principles to Practice.* Washington DC, AACC Press, 2001.

13. Westgard S. *Quantitating quality: Best practices for estimating the Sigma-metric*. White paper available from Abbott Diagnostics.

14. Nillson OJ. *The Essential Deming: Leadership principles from the father of quality, W. Edwards Deming.* McGraw-Hill, 2013.

15. Burnett D. *A Practical Guide to ISO 15189 in Laboratory Medicine.* London, ACB Ventures Publications, 2013.

16. Westgard JO, Westgard SA. Quality control review: Implementing a scientifically based quality control system. *Annual of Clinical Biochemistry* 2016;53:32-50.

CHAPTER 2

DETERMINING QUALITY ON THE SIGMA SCALE

INTRODUCTION

The initial steps for implementing a Six Sigma quality management system (6ΣQMS) are to define the quality goal/requirement for intended use, select an analytical measurement procedure and determine method performance from lab data.

Figure 2-1 illustrates the initial steps to calculate a Sigma-metric from the quality goal in the form of an allowable total error (ATE or TEa) and the accuracy (bias) and precision (SD or %CV) observed for the method. The metric reflects quality on the Sigma scale and provides guidance for selecting the appropriate SQC procedure (that is, the control rules and number of controls needed), which can be easily identified with a variety of SQC planning tools. Some of those tools are illustrated in the next chapter.

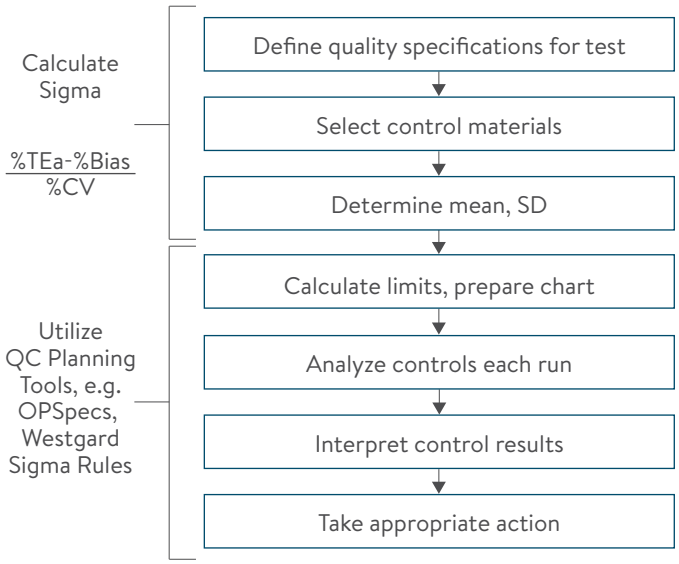

Figure 2-1: Laboratory process for "doing the right SQC right."

METHOD VALIDATION

Validating safety characteristics is important for risk-based thinking. In ISO terminology, safety characteristics for medical devices are reportable range, precision, trueness or bias, detection limit, interference, and recovery — also commonly called performance characteristics. The key ISO guideline for risk management of medical devices[1] emphasizes design for safety:

> ***IVD medical devices have performance characteristics that determine the accuracy of examination results. Failure to meet the performance characteristics required for a specific medical use could result in a hazardous situation that should be evaluated for risk to the patients.***

Manufacturers address safety as part of design and validation of test systems. If precision and bias are not acceptable, manufacturers redesign the system until they achieve the required performance. The performance is documented as claims that the FDA reviews as part of the 510(k) approval to market a new test system.

CLIA requires labs to validate performance of new test systems and verify that they achieve a manufacturer's claims[2]:

§493.1253 Establishment and verification of performance specifications.

(a) Applicability. Laboratories are not required to verify or establish performance claims for any test system used by the laboratory before April 24, 2003.

(b) (1) Verification of performance specifications. Each laboratory that introduces an unmodified, FDA-cleared or approved test system must do the following before reporting patient test results.

(i) Demonstrate that it can obtain performance specifications comparable to those established by the manufacturer for the following characteristics: accuracy, precision, reportable range of the test system.

(ii) Verify that the manufacturer's reference intervals (normal values) are appropriate for the laboratory's patient population.

DATA FOR RISK ASSESSMENT

U.S. CLIA requirements for a QC program call for in-house data to conduct a risk assessment of the testing process[3].

To conduct a risk assessment, the laboratory must identify the sources of potential failures and errors for a testing process, and evaluate the frequency and impact of those failures and sources of error.

In-house data, established by the laboratory in its own environment and by its own personnel, must be included to demonstrate that the stability of the test system supports the number and frequency of QC documented in the QCP. Data from verification or establishment of performance specifications and historical (existing) QC data can be included. Published data or data from manufacturers (e.g., package inserts) may be taken into consideration, but may not be used as the sole criteria for decision-making.

Specific recommendations are to use data from performance verification/validation studies and existing QC records. Those data are supposed *to demonstrate that the stability of the test system supports the number and frequency of the QC documented in the QCP*.

In-house data can be used to determine test system quality on the Sigma scale (the Sigma-metric) and if observed bias and precision are suitable for clinical use. Sigma is inherently risk-based and predicts the expected number of defective test results for a test system in terms of precision and bias and the quality required for intended use of the test.[4] Sigma SQC planning tools guide the selection of appropriate SQC procedures. In-house data are used to right-size SQC procedures for the quality required for the intended use of the test.

SIGMA AS AN INDICATOR OF RISK

Six Sigma quality management assesses the quality of any process on the Sigma scale. Sigma provides a measure of observed quality relative to the quality required. In manufacturing, the quality required for intended use is defined as a tolerance specification. The quality produced typically shows variation around a target or ideal specification.

Part A of **Figure 2-2** applies the Sigma model to the clinical lab, with tolerance limits replacing allowable error requirement for the intended clinical use (ATE or TEa). Precision is represented by the SD or %CV, characterizing the width of the distribution. The effect of bias is shown by the location of the distribution relative to the target or true value. Bias shifts the distribution toward one of the tolerance specifications, thus reducing the amount of variation allowable.

The goal for world-class quality is process variation (i.e., test performance) that fits well within the tolerance specifications.

Example:
± 6 SD for an assay, as shown in part B of **Figure 2-2**.

Six Sigma quality ensures essentially no errors exceed the defined quality requirement.

In most industries, minimum acceptable quality is defined as three Sigma, shown in part C of **Figure 2-2**. For three Sigma, the tolerance limits are completely consumed by 3 SD of variation, and even under optimal operating conditions, a few defects are produced. Any change in process performance (e.g., decreased precision or increased bias) causes an increased risk of producing poor-quality test results. A Six Sigma process is considered world class, but it's possible to attain a Sigma value >6, or less than 3.4 defects per million opportunities, with exceptional precision and/or minimal bias.

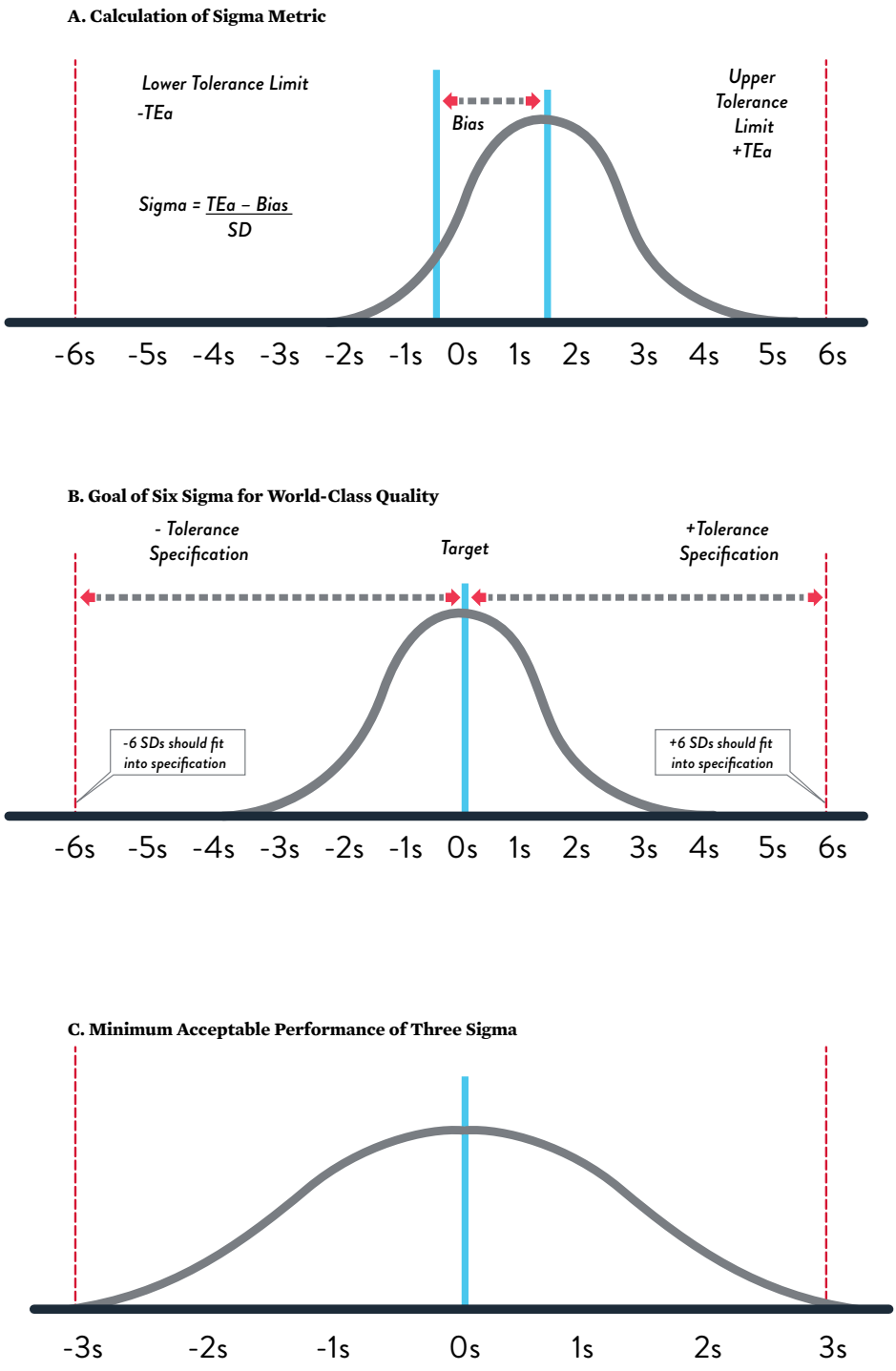

Figure 2-2: (A) Illustration of calculation of Sigma-metric from the allowable total error (ATE or TEa), inaccuracy (Bias), and imprecision (SD). (B) Comparison of Six Sigma goal for world class quality with (C) minimum acceptable quality of Three Sigma.

CALCULATION OF SIGMA METRIC

The Sigma-metric is calculated using the equation:

$$\textbf{Sigma-metric} = (\textbf{TEa} - |\textbf{Bias}|)/\textbf{SD}$$

where TEa is the allowable total error, bias is the systematic error (inaccuracy) and is treated as an absolute value (|Bias|), and SD is the random error (imprecision), with all terms expressed in concentration units.

Percentage units may also be used, as in the equation below:

$$\textbf{Sigma-metric} = (\%\textbf{TEa} - |\%\textbf{Bias}|)/\%\textbf{CV}$$

TEa

TEa may be defined by the criteria for acceptable performance for EQA/PT surveys.

Example:
The U.S. CLIA criterion for glucose is "Target Value ± 6 mg/dL or ± 10% (whichever is greater)."

The larger of the two limits should be used, depending on the target value (TV) or concentration of the PT survey material.

If the TV is:	Acceptable performance is...	Which is a range of...
50 mg/dL	50 ± 6 mg/dL	44 to 56 mg/dL
125 mg/dL	125 mg/dL ± 10% or 125 ± 12.5 mg/dL	112.5 to 137.5 mg/dL

Example:
CLIA provides a list of acceptable performance limits for 70–80 tests. These are the regulated analytes for which PT performance is assessed by five samples per survey and three surveys per year. Non-regulated analytes also require PT but may only be assessed by two surveys per year, with as few as two samples per survey. Non-regulated analyte acceptance limits may be based on various goal-setting models (e.g., intended clinical use, biologic variation and expert group recommendations). HbA1c is a non-regulated analyte for which the College of American Pathologists (CAP) and the National Glycohemoglobin Standardization Program (NGSP) have established an TEa of +/- 6.0% in 2014.

Traditionally, EQA/PT programs used peer group grading - that is, a lab's performance is compared to the performance of all labs using the same analytical method (i.e., analyzer, reagent, methodology). Peer group performance is acceptable if a lab's results agree with the mean value of its peer group within the established acceptance limits. Thus, satisfactory performance is relative to the peer group. Accuracy-based EQA/PT programs are becoming more prevalent. Accuracy-based grading compares a lab's performance to a predetermined target value established by an accepted gold standard reference method. Accuracy-based grading is absolute because it's based on the best available estimate of scientific truth, using a reference method.

BIAS

Inaccuracy, trueness or systematic error is determined during method validation studies from a comparison of methods experiment. Labs perform these experiments to verify a manufacturer's claim after installation of new test systems. After initial validation, labs are required to monitor bias using EQA/PT samples with target values established by reference methods, the mean of a survey group or the mean of a survey peer group. Results are generally reported as the deviation from the target and expressed as a multiple of the observed group variation (that is, a z-value that describes the deviation from the target as a multiple of the group standard deviation). For calculation of a Sigma-metric, %Bias is calculated as the observed bias divided by the target value, multiplied by 100.

$$\%Bias = (Bias/TV) \times 100$$

An important note about bias: Initially when determining Sigma quality, it may be difficult to obtain a good bias estimate. It is permissible to assume bias is zero and calculate Sigma simply as the ratio of TEa/SD or %TEa/%CV. This calculation yields a Sigma-metric that is too high (i.e., an optimistic estimate of quality). Nonetheless, if Sigma is low (<3 when bias is assumed to be zero), it's sufficient to indicate the new test system is high risk! If Sigma is >3, it is still important to get a better estimate of bias for a more reliable determination of Sigma.

SD

Imprecision (random error) is determined from a replication experiment during method validation studies or SQC data collected during routine operation. Labs perform replication experiments to verify precision and then monitor ongoing performance from SQC data collected under conditions of routine operation. %CV is calculated as the observed SD divided by the mean and then multiplied by 100.

$$\%CV = (SD/Mean) \times 100$$

Example Calculations for HbA1c

Given the importance of HbA1c for diagnosing and managing diabetes, the global agreement on quality requirements, the availability of accuracy-based EQA/PT programs, and the widespread application of methods in central labs, as well as point-of-care settings, HbA1c provides a good example for Sigma calculations.

These examples illustrate the demanding performance required of current analytical methods for the test quality needed for clinical use of HbA1c.

TEa	Bias	CV	Sigma
6.0%	0.0%	1.0%	(6.0–0.0)/1.0 or 6.0
6.0%	1.0%	1.0%	(6.0–1.0)/1.0 or 5.0
6.0%	0.0%	1.5%	(6.0–0.0)/1.5 or 4.0
6.0%	1.5%	1.5%	(6.0–1.5)/1.5 or 3.0

GRAPHICAL DETERMINATION OF SIGMA

The method decision chart is a plot of allowable bias versus allowable precision constructed once an TEa quality goal is defined. TEa specifies the size of the error budget, consisting of both random and systematic errors (precision and bias). Historically, the requirement for acceptable performance was bias plus 2 SD, still a common way of calculating the total analytic error (TAE). The TEa requirement became more demanding as assay performance improved. The criterion for Six Sigma quality requires bias plus 6 SD to fit within the TEa.

The multiplier of the SD is the Sigma of interest, and a graphical tool can be constructed to show performance limits for two, three, four, five and Six Sigma, as shown by the method decision chart in **Figure 2-3.** This tool allows plotting an operating point, where the y-coordinate represents the observed bias and the x-coordinate represents the observed precision. This point represents the Sigma quality of any method.

Example:
The operating point in **Figure 2-3** represents a bias of 2.0% and a CV of 1.0%. It falls on the line representing four Sigma quality, which agrees with the calculated Sigma-metric: [(6 –2)/1 = 4].

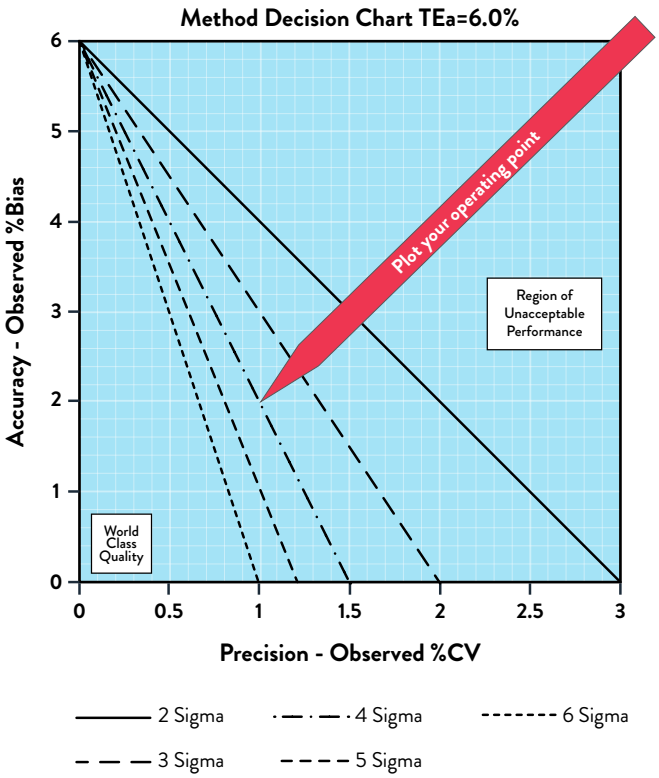

Figure 2-3: Method Decision chart for TEa equal to 6.0%. The observed %Bias is plotted on the y-axis vs the observed %CV on the x-axis. The different diagonal lines, from top to bottom, represent 2-Sigma, 3-Sigma, 4-Sigma, 5-Sigma, and 6-Sigma quality. The operating point represents an examination procedure having an observed bias of 2.0% and an observed precision of 1.0%.

To construct a Method Decision chart for TEa of 6.0%:

1. Scale the y-axis from 0% to TEa, or 6.0%. Label this axis "observed inaccuracy" in units of %Bias.
2. Scale the x-axis from 0% to half of TEa, which is 3.0%. Label this axis "observed imprecision" in units of %CV.
3. Draw lines for Sigma quality by determining the y-intercept and x-intercept as described below:

TEa	Bias	CV	Sigma
10%	Bias Plus 2s	TEa or 6.0%	6.0% ÷ 2, or 3.0%
	Bias Plus 3s	TEa or 6.0%	6.0% ÷ 3, or 2.0%
	Bias Plus 4s	TEa or 6.0%	6.0% ÷ 4, or 1.5%
	Bias Plus 5s	TEa or 6.0%	6.0% ÷ 5, or 1.2%
	Bias Plus 2s	TEa or 6.0%	6.0% ÷ 6, or 1.0%

It is also possible to construct a normalized method decision chart that can be used for any specified TEa. This is done by scaling the y-axis from 0 to 100 and the x-axis from 0 to 50, calculating the x- and y-intercepts as above, and drawing the lines for Sigma. To apply the normalized chart, it is necessary to express the observed bias and SD or CV as percentages of the TEa. For the HbA1c example above, the y-coordinate would be 2/6, or 33%, and the x-coordinate 1/6, or 17%. That method is plotted as point A on the normalized chart shown in **Figure 2-4.** Method B represents a central lab glucose method for which TEa is 10%, and method C represents a point-of-care glucose meter for which TEa is 20%. The advantage of the normalized chart is that many different methods can be shown on the same chart.

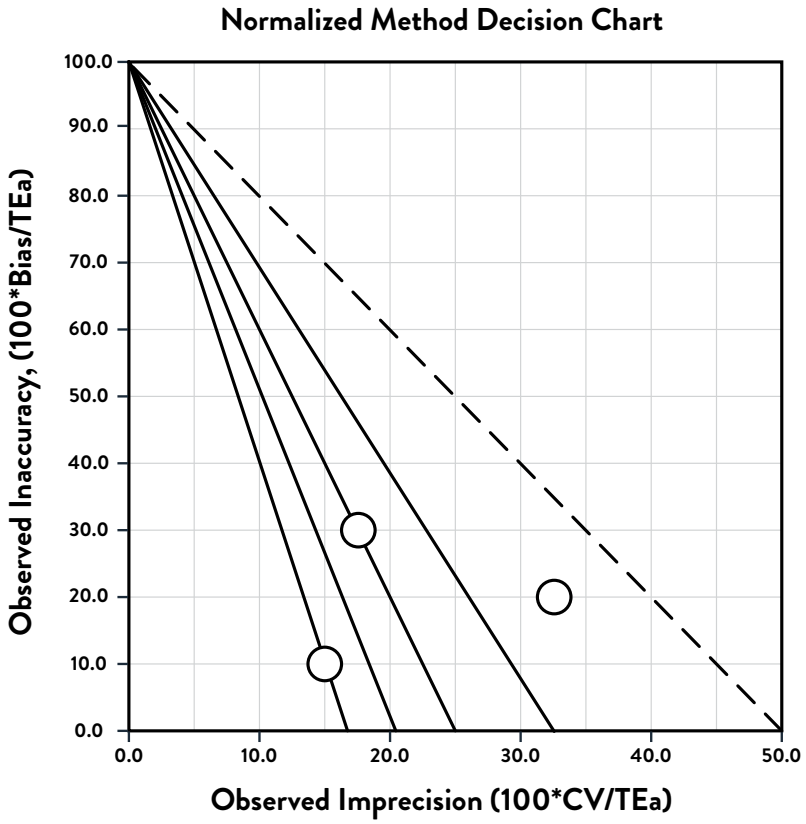

Figure 2-4: "Normalized" Method Decision chart where observed inaccuracy is calculated as 100*Bias/TEa and observed imprecision is calculated as 100*CV/TEa, when original parameters are all in units of %. Example A is the same HbA1c method as shown in **Figure 2-3**. Example B is a laboratory glucose examination procedure where TEa is 10% and C a point of care glucose meter where TEa is 20%.

For a more complete discussion of Six Sigma concepts, metrics and application tools, see reference 5.

DETERMINATION OF SIGMA-METRIC FROM RESULTS OF METHOD VALIDATION STUDIES

Typically, results from replication experiments are represented by the mean, SD and %CV from 20 or more replicates. If replicates are performed within one run or one day, they reflect "within-run" or "within-day" precision. Replicates analyzed over many days (e.g., $\geq$ 20 days) are preferred and reflect "between-day" or "total" precision. The SDs or %CVs for short-term precision are typically smaller than those for long-term precision.

Comparison of method results are presented by plotting the new test system results on the y-axis versus the comparative method results on the x-axis. The data are subjected to regression equation analysis to describe results as the equation for a straight line:

$$\mathbf{y = ax + b,}$$

where a is the slope and b is the y-intercept.

To determine bias at an important medical decision level, Xc:

1. Calculate Yc (aXc + b).
2. Subtract Yc – Xc to estimate bias.
3. Calculate %Bias as Bias(100)/Xc.

REAL-WORLD EXAMPLE 1

A published report on the performance of HbA1c devices provides the following information[6]:

- Precision [Table 1, reference 5]: CV = 1.9% at 6.5%Hb; CV = 3.2% at 8.9%Hb
- Accuracy [Table 2, lot #1, vs. Tina-Quant, reference 5]: y = 1.04x – 0.35

To determine Sigma:

Define TEa at a critical medical decision level (Xc): TEa is defined as 6.0% by the U.S. NGSP and CAP PT. The cut-point for diagnosis of diabetes is 6.5%Hb. Therefore, TEa = +/- 6.0% at 6.5%Hb.

1. Select the appropriate estimate of precision: Precision at 6.5%Hb is represented by the lower control material (i.e., CV = 1.9% at 6.5%Hb.) Note: It's not always that easy because the chosen control materials may not align exactly with the Xc of interest, in which case, it may be necessary to interpolate between the stated performance claims.
2. Calculate bias at Xc:
    a. Yc = (1.04 x 6.5) – 0.35 = 6.76 – 0.35 = 6.41
    b. Bias = Yc – Xc = 6.41 – 6.50 = -0.09
    c. |Bias| = 0.09%Hb
    d. |%Bias| = (0.09 x 100)/6.5 = 1.4%
3. Calculate Sigma:
    a. Sigma = (%TEa – |%Bias|)/%CV
        i. Sigma = (6.0% – 1.4%)/1.9% = 4.6/1.9 = 2.4

REAL-WORLD EXAMPLE 2

This same report[6] provides the following information for a second test system:

- Precision [Table 1, reference 5]: CV = 2.1% at 4.7%Hb; CV = 1.2% at 8.0%Hb; CV = 1.1% at 10.9%Hb
- Accuracy [Table 2, lot #1, vs. Premier, reference 5]: y = 1.08x – 0.41

To determine Sigma:

1. Define TEa at a critical medical decision level (Xc): TEa = +/- 6.0% at 6.5%Hb.
2. Select the appropriate estimate of precision: Precision at 6.5%Hb is probably best represented by the middle control material at 8.0%Hb (i.e., CV = 1.1%). Here, judgment is important for interpreting the results of the experimental studies.
3. Calculate bias at Xc:
   a. Yc = (1.08 x 6.5) – 0.41 = 7.02 – 0.41 = 6.61
   b. Bias = Yc – Xc = 6.61 – 6.50 = 0.11
   c. |Bias| = 0.11%Hb
   d. |%Bias| = (0.11 x 100)/6.5 = 1.69%
4. Calculate Sigma:
   a. Sigma = (%TEa – |%Bias|)/%CV
      i. Sigma = (6.0% – 1.69%)/1.1% = 4.31/1.2 = 3.6

REAL-WORLD EXAMPLE 3

Another report[7] in the same issue of *Clinical Chemistry* examined performance of HbA1c methods used in central lab testing. This study used the NGSP lab to obtain comparison results by an official U.S. reference method:

- Precision [Table 1, reference 6]: CV = 1.66% at 5.24%Hb; CV = 1.33% at 7.9%Hb
- Accuracy [Table 1 vs. NGSP, reference 6]: y = 0.998x + 0.016

1. Define TEa at a critical medical decision level (Xc): TEa = +/- 6.0% at 6.5%Hb.
2. Select the appropriate estimate of precision: Precision at 6.5%Hb is probably best represented by taking the average of the CVs because their means bracket the critical decision level of 6.5%Hb. The average of 1.66% and 1.33% is 1.50%. Again, here's an example where judgment is important for interpreting the results.
3. Calculate bias at Xc:
   a. Yc = (0.998 x 6.5) + 0.016 = 6.487 + 0.016 = 6.503
   b. Bias = Yc – Xc = 6.503 – 6.50 = 0.003
   c. |Bias| = 0.003%Hb
   d. |%Bias| = (0.003 x 100)/6.5 = 0.05%
4. Calculate Sigma:
   a. Sigma = (%TEa – |%Bias|)/%CV
      i. Sigma = (6.0% – 0.05%)/1.5% = 5.95/1.5 = 3.97

DETERMINING SIGMA-METRIC FROM PT AND SQC DATA

Results from PT surveys may be compared to the target value to determine the difference in the observed results. Such differences can be expressed in concentration units, in percentage units or in the form of a z-value that describes a multiple of the group SD or %CV. It is useful to calculate the differences in concentration units as a percentage of the target value and then average those differences to obtain an estimate of bias. There usually will be two to five samples in a CAP survey in the United States. Regulated tests require three survey events per year, with five samples per event; non-regulated tests (all others except waived tests) require two survey events per year, with a minimum of two samples per event.

The few PT samples available (only two to five) is a limitation leading to a large uncertainty in the bias estimate. The minimum number of samples for a comparison of methods experiment is usually 20, and often 40 or more samples are included. Because of the low number of survey samples, it's generally good practice to calculate Sigma both with and without bias.

REAL-WORLD EXAMPLE 4

Most U.S. labs analyze two levels of controls per day to comply with the CLIA QC regulations. Typically, 20 to 30 control observations are available each month. Data are summarized monthly by calculating the mean, SD and %CV. CAP provides an HbA1c survey widely used in the U.S. and allows monitoring of the nearly 30 different test systems certified by the NGSP. More than 3,000 labs participate in the CAP survey, with assay peer groups ranging from 20 to 300 labs. Typically, three samples are provided for each survey event, and there are only two survey events per year because HbA1c is not a regulated test. Target values are assigned from analysis by reference methods.

- Precision: Routine SQC for two levels of control yielded an SD of 0.105%Hb at 5.58%Hb (1.9%CV) and an SD of 0.155%Hb at 9.58%Hb (1.6%CV).
- Accuracy: The 2014 CAP GH2 survey event A included three samples for HbA1c (GH2-01 = 6.49%Hb, GH2-02 = 6.97%Hb and GH2-03 = 9.65%Hb). The lab results were 6.7, 7.3 and 9.9%Hb, respectively.

To determine Sigma:

1. Define TEa at a critical medical decision level (Xc): TEa is +/- 6.0%.
2. Select the appropriate estimate of precision: The CV for the controls brackets the critical Xc of 6.5%Hb; thus, the value should be between 1.9% and 1.6%. Interpolating between the controls, a CV of 1.75% is a good estimate.
3. Calculate bias from the differences between the lab results and the CA-assigned reference values:
   a. Calculate the differences between the lab results and the assigned values.
      i. 6.80 – 6.49 = 0.21%Hb or 3.24% [(0.21 x 100)/6.49]
      ii. 7.30 – 6.97 = 0.33%Hb or 4.73% [(0.33 x 100)/6.97]
      iii. 9.90 – 9.65 = 0.25%Hb or 2.59% [(0.25 x 100)/9.60]
   b. Average the differences to estimate bias.
      i. (3.24% + 4.73% + 2.59%)/3 = 3.52%

4. Calculate Sigma both with and without bias:

   a. Sigma = (%TEa – |%Bias|)/%CV

      i. Sigma = (6.0% – 3.52%)/1.75% = 1.42

   b. Sigma = %TEa/%CV

      i. Sigma = 6.0/1.75 = 3.43

REAL-WORLD EXAMPLE 5

The same lab in example 4 analyzed the CAP survey samples for GH2, event B, in 2014, whose samples had assigned values of 6.58, 8.39 and 5.65%Hb, respectively. The lab results were 6.7, 8.5 and 5.6%Hb, respectively.

1. TEa is +/- 6.0%.

2. The long-term CVs were again 1.9% and 1.6% at means of 5.58 and 9.58%Hb. Interpolating between the controls, a CV of 1.75% is a good estimate.

3. Bias is determined from the differences between the lab results and the CAP-assigned values:

   a. Calculate the differences between the lab results and the assigned values.

      i. 6.70 – 6.58 = 0.12%Hb or 1.82% [(0.12 x 100)/6.58]

      ii. 8.50 – 8.39 = 0.11%Hb or 1.31% [(0.11 x 100)/8.39]

      iii. 5.60 – 5.65 = -0.05%Hb or -0.88% [(-0.05 x 100)/5.65]

   b. Average the differences to estimate bias.

      i. (1.82% + 1.31% – 0.88%)/3 = 0.75%

4. Calculate Sigma both with and without bias:

   a. Sigma = (%TEa – |%Bias|)/%CV

      i. Sigma = (6.0% – 0.75%)/1.75% = 3.00

   b. Sigma = %TEa/%CV

      i. Sigma = 6.0/1.75 = 3.43

Note the difference between the estimates of bias (3.52% vs. 0.75%) and Sigma in examples 4 and 5 (1.42 vs. 3.00). Those differences likely represent the limitation of having only three samples analyzed by single measurements in the lab for comparison. The results would be more reliable with more survey samples or multiple measurements on each sample, but that's not permitted by CLIA.

IMPORTANCE OF DETERMINING QUALITY ON THE SIGMA SCALE

The examples above suggest the Sigma quality for current HbA1c methods does not yet approach the goal of Six Sigma for world-class quality. Keep in mind, HbA1c is one of the most standardized tests in the world. A global IFCC lab network supports reference methods and materials, and several national lab networks, such as NGSP, certify the equivalence of virtually all test systems being marketed in the U.S.

The two studies [6,7] cited above were published in 2014. The data from the study of POC HbA1c test systems[6] report Sigmas that range from 0.44 to 4.23, with three of the seven methods demonstrating quality > three Sigma. Those results are summarized on the method decision chart in **Figure 2-5**, and most lab personnel will find this graphical summary to be much more understandable than the statistical tables in the paper. In the second paper, the data from the study of central lab HbA1c test systems[7] proves only one in six test systems provides quality > three Sigma.

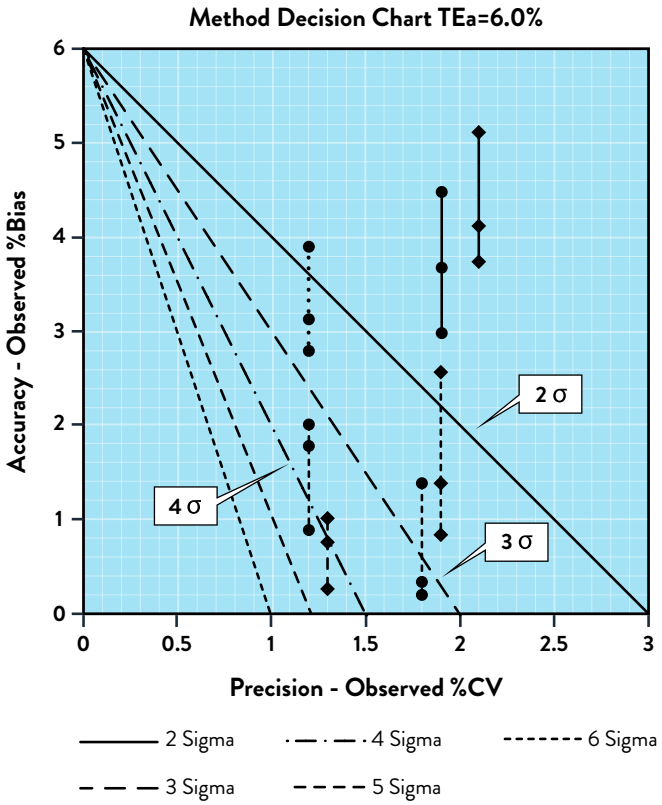

Figure 2-5: Summary of performance data for 7 test methods, each compared with 3 reference methods. Method Decision chart prepared for TEa=6.0%.

Additional data from the 2014 CAP HbA1c survey provide further evidence of the low Sigma quality of many test systems. **Figure 2-6** shows the performance of current U.S. test systems by plotting the bias on the y-axis and SD on the x-axis for each method subgroup. Note that this is simply a two-sided method decision chart adapted for use with PT and EQA data[8]. The inner diagonal >-shaped line represents three Sigma quality, and the outer line represents two Sigma quality. Only one method subgroup achieves three Sigma quality, six achieve between three and two Sigma, and 19 achieve less than two Sigma. This shows the importance of PT and EQA surveys for evaluating the comparability of results from different method subgroups (in this case, 26 different methods approved by the FDA and certified as equivalent by NGSP).

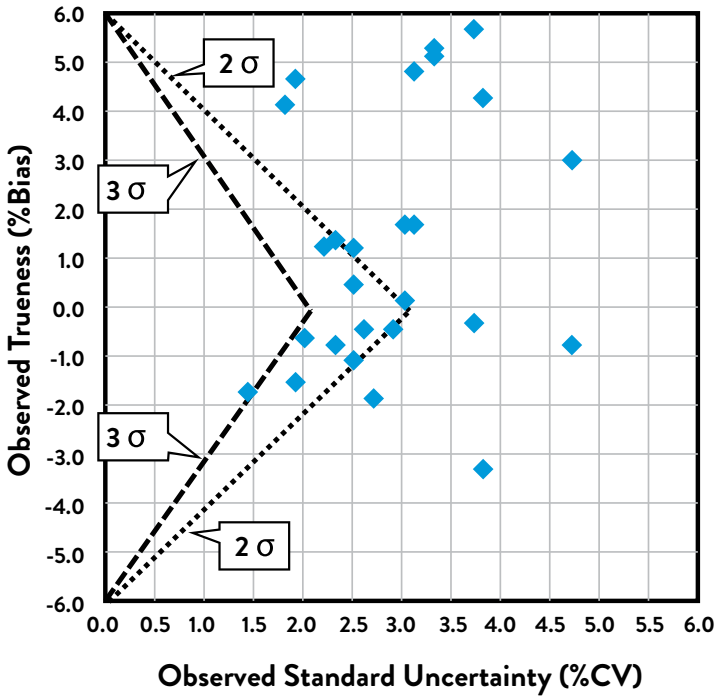

Figure 2-6: Sigma Proficiency Assessment Chart for 2014 College of American Pathologists (CAP) survey results for HbA1c GH2-01 sample with concentration of 6.49%Hb. TEa=6.0%. Each point represents the observed trueness (%Bias, y-axis) and the observed Standard Uncertainty (%CV, x-axis) for one of 26 examination subgroups. Results represent a total of 3,187 laboratories.

WHAT'S THE POINT?

It is absolutely critical to determine quality on the Sigma scale to assess the risk of any test or test system. Validation of safety characteristics is the most important first step in risk assessment. It is essential to use in-house validation and QC data to perform a risk assessment. Determination of Sigma quality is the single best predictor of the risk. Sigma is also a useful predictor of the QC needed to minimize the risk of poor-quality test results.

References

1. ISO 14971:2007. Medical devices – Application of risk management to medical devices. ISO, Geneva, 2007.

2. U.S. Department of Health and Human Services. Medicare, Medicaid, and CLIA Programs: Laboratory requirements relating to quality systems and certain personnel qualifications. Final Rule. *Federal Register* Jan 24 2003;16:3640-3714. See also CMS website www.cms.hhs.gov/clia/ and the CMS State Operations Manual, Appendix C, Regulations and Interpretive Guidelines for Laboratories and Laboratory Services, www.wms.hhs.gov/CLIA/03_Interpretive_Guidelines_for_Laboratories.asp.

3. CMS Memo of August 16, 2013: Individualized Quality Control Plan (IQCP): A new Quality Control (QC) option. www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/Downloads/IQCPbenefits.pdf.

4. Westgard S. Prioritizing risk analysis quality control plans based on Sigma-metrics. In *Quality Control in the Age of Risk Analysis,* eds. Westgard JO, Westgard S. *Clinics in Laboratory Medicine* 2013;33(Number 1):41-53.

5. Westgard S. *Six Sigma-metric analysis for analytical testing processes*. White paper available from Abbott Diagnostics.

6. Lenters-Westra E, Slingerland RJ. Three of 7 hemoglobin A1c point-of-care instruments do not meet generally accepted analytical performance criteria. *Clinical Chemistry* 2014;60:1062-1072.

7. Woodworth A, Korpi-Steiner N, Miller JJ, Rao LV, Yundt-Pacheco J, Kuchipudi L, Parvin CA, Rhea JM, Molinaro R. Utilization of assay performance characteristics to estimate Hemoglobin A1c result reliability. *Clinical Chemistry* 2014;60:1073-1079.

8. Westgard JO, Westgard SA. Assessing quality on the Sigma scale from proficiency testing and external quality assessment surveys. *Clinical Chemistry and Laboratory Medicine* 2015;53:1531-1535.

CHAPTER 3

SELECTING THE RIGHT SQC PROCEDURE

INTRODUCTION

Selecting an SQC procedure begins with defining the quality required for intended use, evaluating the performance (precision and bias) of the assay (method, examination procedure), and determining quality on the Sigma scale. The previous chapter focused on these initial steps and the determination of the Sigma-metric. This chapter describes the selection of an appropriate SQC procedure.

The objective of SQC is to achieve a high level of error detection and a low level of false rejections with the simplest control rules and the least number of controls. Four different QC planning tools are described:

- Sigma-metric SQC selection tool
- Chart of operating specifications
- Normalized chart of operating specifications
- Westgard Sigma Rules

Each tool has advantages and disadvantages related to simplicity of use and ease of understanding, but all are based on the performance characteristics of SQC and provide similar, if not identical, results.

PERFORMANCE OF SQC PROCEDURES

SQC is an error detector, and its response depends on the size of the error. It is similar to a smoke alarm. A small fire may not set off the alarm, but as the size of the fire increases, the probability the alarm will go off also increases. False alarms cause building evacuations when there isn't a fire. True alarms and false alarms are performance characteristics of any detector, including the SQC error detector.

Figure 3-1 describes the typical smoke detector response. The chance that the alarm will go off is on the y-axis, and the size of the fire is on the x-axis. As the fire gets larger, the probability of an alarm increases. There is a small probability of an alarm even when there is no fire, as shown by the y-intercept. That's the chance of a false alarm.

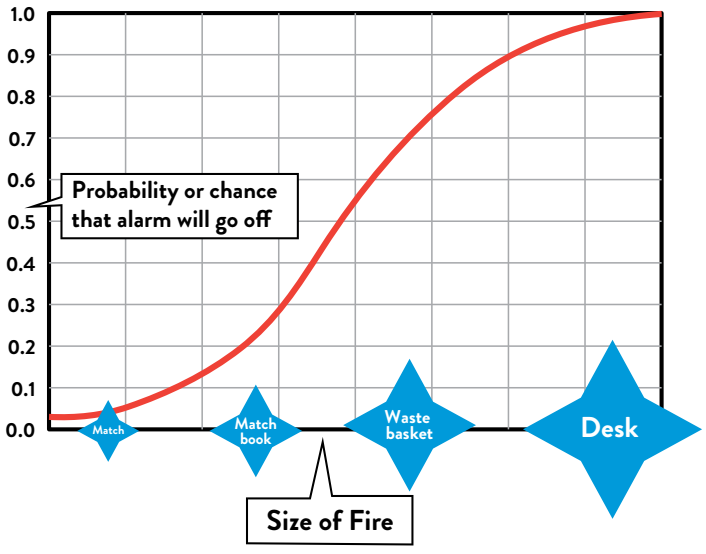

Figure 3-1: Typical response curve for a detector: The level of false alarms is shown by the y-intercept of the response curve; chance for true alarms depend on the size of the fire.

For SQC detectors, response curves are based on statistical theory or from computer simulation studies. **Figure 3-2** shows five response curves for control procedures, all having two controls per run but using different control rules. This is a power function graph, where each line is a power curve showing the probability of rejection on the y-axis versus the size of the error on the x-axis for a particular SQC procedure[1]. The probability of rejection varies from 0.0, when there is never a rejection, to 1.0, when there will always be a rejection. The probability of false rejection (P_{fr}) is determined from the y-intercept of a power curve. The probability for error detection (P_{ed}) is determined by identifying the size of an error on the x-axis, drawing a vertical line, locating the intersection with the power curve and reading the probability from the y-axis.

Example:
For the power curve second from the bottom, P_{fr} is essentially 0.0. If the systematic error to be detected is 2.5 on the x-axis, as shown by the vertical line, the intersection with the power curve indicates P_{ed} is approximately 0.53, meaning there is a 53% chance of detecting a systematic shift equivalent to 2.5 times the SD of the method. For comparison, the top power curve provides a P_{ed} of 0.90, which is much better, but it suffers from a high P_{fr} of nearly 0.10 or 10%. The high false rejections compromise the use of that SQC procedure because analysts won't know whether or not an observed rejection is a true or a false alarm.

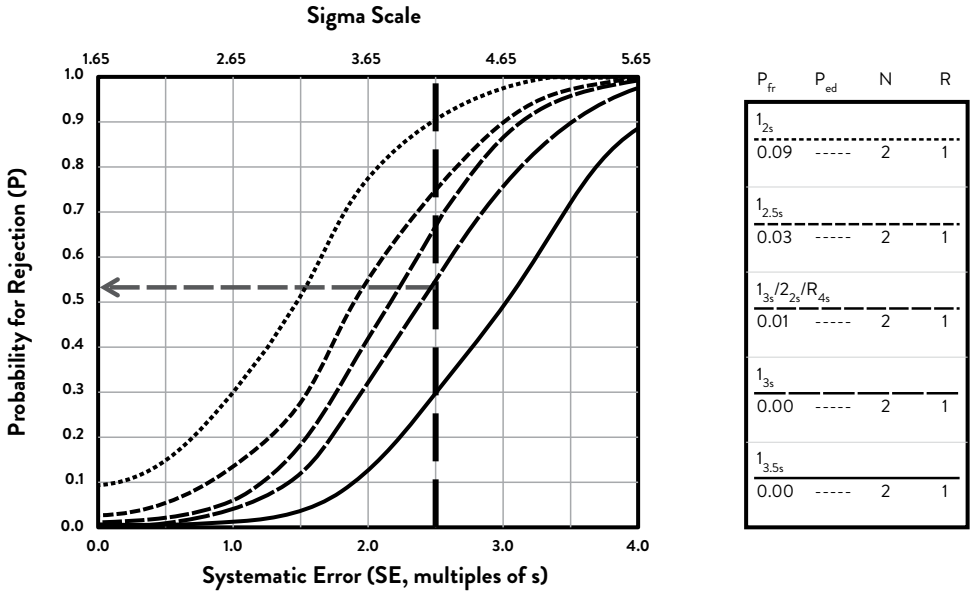

Figure 3-2: Power function graph showing probability of error detection on the y-axis versus size of systematic error (lower x-axis) and Sigma quality (upper x-axis). Different power curves represent different SQC procedures whose control rules are identified in the key at the right. Lines top-to-bottom represent control rules and number of control measurements (N) top-to-bottom in key.

SIZE OF A MEDICALLY IMPORTANT ERROR

The critical systematic error (ΔSE_{crit}) that needs to be detected by SQC is calculated from the quality required for intended use and the observed precision and bias, as follows:

$$\mathbf{\Delta SE_{crit} = [(TEa - Bias)/SD] - 1.65,}$$

where TEa is the allowable total error, bias represents inaccuracy, and SD is the imprecision.

Note the Sigma-metric can be substituted for the expression [(TEa – Bias)/SD]:

$$\mathbf{\Delta SE_{crit} = Sigma - 1.65}$$

The Sigma-metric indicates of the size of the medically important systematic error, and the equation can be rearranged:

$$\mathbf{Sigma = \Delta SE_{crit} + 1.65}$$

This means the x-axis of a power function graph can be rescaled in terms of Sigma by adding 1.65 to the value of the systematic error, as shown by the x-axis at the top of **Figure 3-2**.

DEFINITIONS OF CONTROL RULES

The key at the right in **Figure 3-2** identifies different SQC procedures (control rules), the number of controls (N) and the number of runs (R) over which the control rules are applied. Control rules are abbreviated in the form A_L and defined as follows:

- $\mathbf{1_{2s}}$ – The control rule commonly used with a Levey-Jennings chart, with control limits set as the mean ± 2s (s = SD). This rule is sometimes used as a rejection rule, with problems due to false rejections (5% for N = 1, 10% for N = 2). In multirule SQC, it can be used as a warning rule to trigger careful inspection of the control data by other rejection rules.
- $\mathbf{1_{3s}}$ – Reject when one control measurement exceeds the mean ± 3s.
- $\mathbf{1_{2.5s}}$ – Reject when one control measurement exceeds the mean ± 2.5s control limits.
- $\mathbf{1_{3.5s}}$ – Reject when one control measurement exceeds the mean ± 3.5s control limits.
- $\mathbf{2_{2s}}$ – Reject when two consecutive control measurements exceed the same mean + 2s control limit or the same mean – 2s control limit.
- **2 of 3**$_{2s}$ – Reject when two out of three control measurements exceed the same mean + 2s or mean – 2s control limit.

- **R_{4s}** – Reject when one control measurement in a group exceeds the mean + 2s control limit and another exceeds the mean – 2s control limit. (Note: This rule is best applied within a single run.)
- **3_{1s}** – Reject when three consecutive control measurements exceed the same mean + 1s or the same mean – 1s control limit.
- **4_{1s}** – Reject when four consecutive control measurements exceed the same mean + 1s or the same mean – 1s control limit.
- **6x** – Reject when six consecutive control measurements fall on one side of the mean.
- **8x** – Reject when eight consecutive control measurements fall on one side of the mean.
- **9x** – Reject when nine consecutive control measurements fall on one side of the mean.
- **10x** – Reject when 10 consecutive control measurements fall on one side of the mean.

Note that the SQC rule in **Figure 3-2** with the high false rejection (top line, 1_{2s} with N = 2) corresponds to Levey-Jennings chart limits set at the mean ± 2 SD, whereas the other SQC rule (second line from the bottom) has a very low $P_{fr,}$ but also a lower P_{ed} (1_{3s} with N = 2), and corresponds to Levey-Jennings chart limits set at the mean ± 3 SD. Comparison of performance shows the practical difficulty in selecting SQC rules: There is a tradeoff between error detection and false rejection. Narrow control limits lead to higher error detection but also higher false rejection. Wide control limits provide low false rejection but also lower error detection.

A good compromise is multirule SQC rules that increase error detection by applying several control rules, each chosen to have a low P_{fr}.

Example:
The middle curve in **Figure 3-2** combines the $1_{3s}/2_{2s}/R_{4s}$ control rules with N = 2 and provides higher error detection than the 1_{3s} with the same number of controls. Typically, a multirule SQC rule is constructed with some rules sensitive to systematic error (2_{2s}, 3_{1s}, 4_{1s}, 6_x, 8_x) and some rules sensitive to random error (1_{3s}, R_{4s}). Rules using a single value outside a wide limit respond to increased SD (random error). Rules using a series of consecutive values exceeding the same limit are sensitive to shifts in distribution (systematic error). The closer the limit line, the more consecutive observations are needed to maintain a low P_{fr}.

WESTGARD MULTIRULE SQC PROCEDURE

A multirule control procedure employing a series of five rules ($1_{3s}/2_{2s}/R_{4s}/4_{1s}/10_x$) is commonly known as Westgard Rules and is widely used in labs today. Westgard Rules use a control chart with limits drawn at the mean ± 1SD, mean ± 2 SD and mean ± 3 SD[2]. **Figure 3-3** describes the logic for the traditional Westgard multirule SQC, including an initial 1_{2s} warning rule, followed by five different rejection rules[3]. This multirule procedure was introduced in the 1980s, when QC charting was done manually. For that reason, the 1_{2s} warning rule was included to trigger inspection by the full set of rules. The warning rule is not needed when rule checking is done easily and quickly by SQC software.

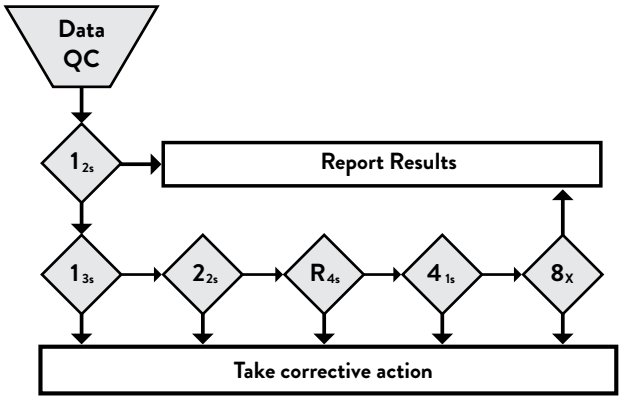

Figure 3-3: Logic diagram for application of Westgard multirule SQC procedure.

Westgard Rules have been applied broadly for many methods, and the multirule concept provides a flexible set of rules that can be tailored for desired error detection while maintaining relatively few false rejections[4]. Adding rules to the basic 1_{3s} rule (Levey-Jennings chart with 3 SD control limits) increases error detection. Increasing the number of controls also increases both error detection and false rejection. Selecting an SQC procedure is a matter of balancing the number of rules and control measurements, based on the expected probabilities for error detection and false rejection. Fortunately, several QC planning tools simplify and support the selection process[5].

SIGMA SQC SELECTION TOOL

A power function graph with a Sigma scale is a Sigma SQC selection tool. As shown in **Figure 3-4**, the power curves allow comparison of the performance of single and multirule SQC rules with Ns from 2 to 8. The desirable error detection is often set at P_{ed} = 0.90 (90% chance). The desirable false rejection is generally set at P_{fr} = 0.05 or less (5% chance or less), satisfied by all but one of the SQC rules shown here.

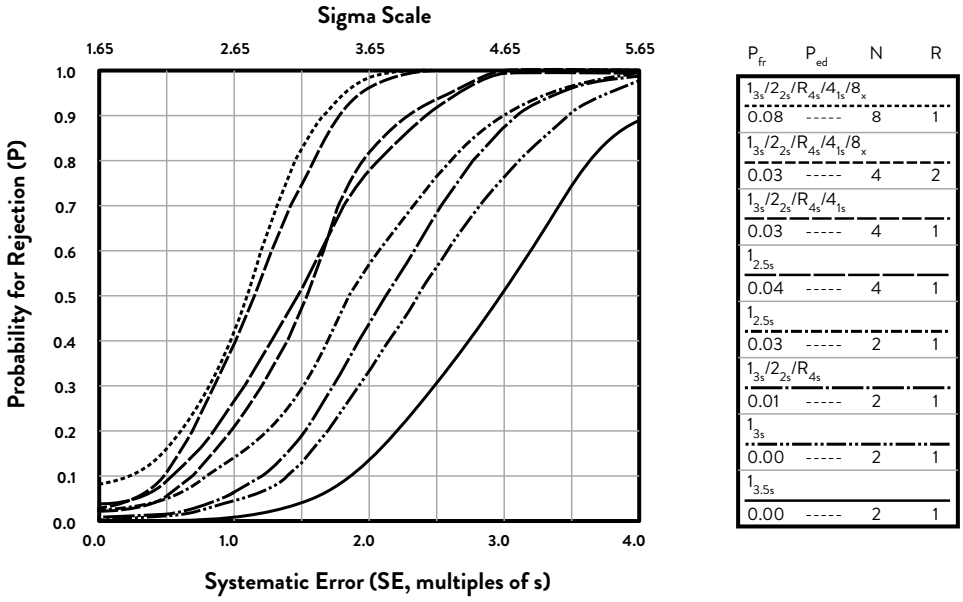

Figure 3-4: Sigma SQC Selection Tool. Probability of error detection on the y-axis vs size of systematic error (lower x-axis) and Sigma quality (upper x-axis). Different power curves represent different SQC procedures whose control rules are identified in the key at the right. Lines top-to-bottom represent control rules and number of control measurements (N) top-to-bottom in key.

To select an appropriate SQC rule, draw a vertical line corresponding to the Sigma-metric of the test (x-axis, top scale). To identify appropriate control rules and the number of controls, inspect the graph and compare the error detection at the points where the vertical line intersects the power curves.

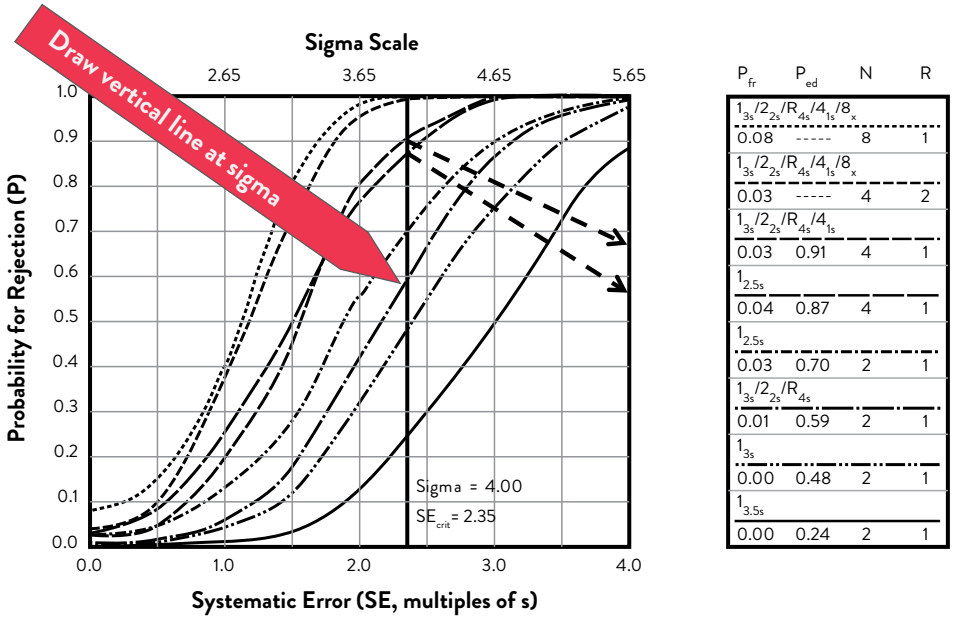

Figure 3-5: Example application of Sigma SQC Selection Tool for test with HbA1c method where TEa=6.0%, Bias=2.0%, and CV=1.0%, or 4.0 Sigma quality [Sigma = (TEa-Bias)/CV]. Appropriate SQC procedures would be a 13s/22s/R4s/41s multirule with N=4 or a 12.5s single-rule with N=4.

Figure 3-5 illustrates a test of four Sigma quality. The key shows the P_{fr} and P_{ed} values for all SQC procedures. Note those with Ns of 4 and higher provide the appropriate error detection. An N of 4 refers to the total number of controls (e.g., two measurements on each of two controls at different concentrations, one measurement on each of four controls, or even four measurements on one control). There is no need to go beyond the $1_{3s}/2_{2s}/R_{4s}/4_{1s}$ rules and an N of 4 because it provides a P_{ed} of 0.91 and a P_{fr} of 0.03. A $1_{2.5s}$ single rule procedure with N = 4 provides a P_{ed} of 0.87 and a P_{fr} of 0.04. The performance of the two SQC rules is essentially equivalent. The choice between them depends on which is easier to implement depending on the SQC software and the training and skills of the analysts. A Sigma of four is the level of quality at which it is essential to implement multirule SQC.

CHARTS OF OPERATING SPECIFICATIONS

Another tool for selecting SQC procedures is the operating specifications chart, or OPSpecs chart[6]. **Figure 3-6** uses the same format as the method decision chart described earlier. The OPSpecs chart is for a stated quality requirement. In this example, TEa = 6.0%, per the label at the top of the chart. The top label specifies this chart is for SQC rules, providing a 90% detection of systematic errors. Like the method decision chart, the y-axis shows the allowable bias and the x-axis shows the allowable precision. The difference is the lines represent the allowable limits of bias and precision for different SQC rules. To use an OPSpecs chart, plot an operating point where the y-coordinate equals bias and the x-coordinate equals precision. The lines above the operating point identify SQC rules appropriate for the observed performance of the method.

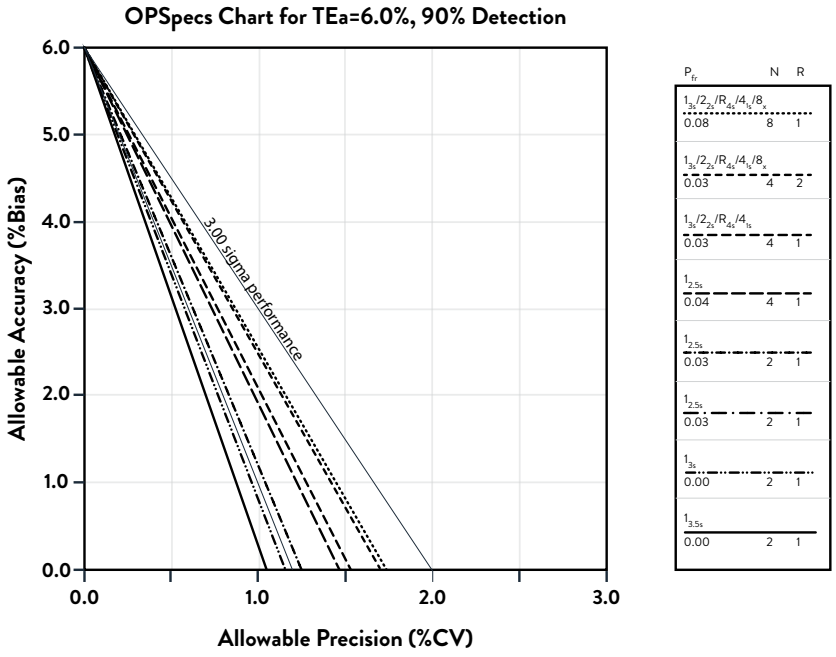

Figure 3-6: Chart of Operating Specifications for TEa=6.0% (with 90% error detection) showing allowable bias on y-axis versus allowable precision on x-axis for different SQC procedures whose rules and number of control measurements (N) and number of run (R) are shown in the key at the right. Lines top-to-bottom corresponds to SQC procedures top-to-bottom in key.

Example:
Figure 3-7 shows an HbA1c method with TEa = 6.0%, bias = 2.0% and precision = 1.0%. The lines closest to the operating point are identified in the key at the right as a $1_{3s}/2_{2s}/R_{4s}/4_{1s}$ multirule procedure with an N of 4 and a $1_{2.5s}$ single-rule with an N of 4. Both provide 90% error detection, and false rejection probabilities are 0.03 and 0.04, respectively.

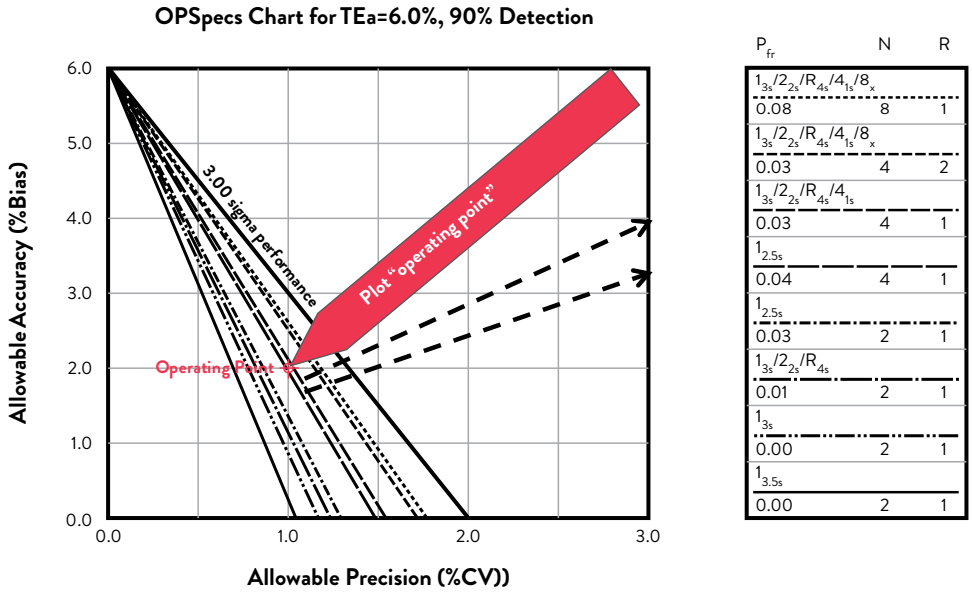

Figure 3-7: Example application of a Chart of Operating Specifications for TEa=6.0% (with 90% error detection) for HbA1c method having a bias of 2.0% and CV of 1.0%. Appropriate SQC procedures would be a 13s/22s/R4s/41s multirule with N=4 or a 12.5s single-rule with N=4.

It is not just a coincidence that the OPSpecs chart provides the same answers as the Sigma SQC selection tool. The OPSpecs chart is constructed by rearranging the equation for the calculation of the critical systematic error, as follows:

$$\Delta SE_{crit} = [(TEa - Bias)/SD] - 1.65$$

$$\Delta SE_{crit} + 1.65 = [(TEa - Bias)/SD]$$

$$(\Delta SE_{crit} + 1.65)SD = TEa - Bias$$

$$Bias = TEa - (\Delta SE_{crit} + 1.65)SD$$

This is the equation for a straight line when bias is plotted on the y-axis and SD on the x-axis. The line has a y-intercept of TEa and a slope of (ΔSE_{crit} + 1.65). The slope of the line depends on the size of the systematic error. By specifying Ped = 0.90, the power curves for different SQC rules determine the size of the error detected, and that value determines the slope of the line. In other words, the OPSpecs chart uses the error detection capability from the power curve for each of the SQC rules.

The OPSpecs chart's advantage is its similarity to the method decision chart. In fact, the OPSpecs chart includes a three Sigma line to show the relationship between Sigma performance and the performance of the SQC rules. Methods need to achieve better than three Sigma performance for a cost-effective SQC procedure.

A limitation of the OPSpecs chart is the difficulty of preparation, due to the need to scale the chart for each TEa requirement and calculate the slopes of the lines. This can be readily accomplished by a specialized computer program, but the need to prepare new OPSpecs charts for each of the multiple quality requirements (TEa) is a practical problem.

NORMALIZED OPSPECS CHART

A remedy is a normalized chart, for which the y-axis represents the ratio %Bias/%TEa, and the x-axis represents the ratio %CV/%TEa. This chart is shown in **Figure 3-8**, scaled from 0 to 100 on the x-axis and 0 to 50 on the y-axis. To use this chart, the operating point is determined as follows:

y-coordinate = (%Bias/%TEa) x 100

x-coordinate = (%CV/%TEa) x 100

For example, for the HbA1c method in **Figure 3-7**, the operating point for a normalized OPSpecs chart is:

y-coordinate = (2.0%/6.0%) x 100 = 33.3%

x-coordinate = (1.0%/6.0%) x 100 = 16.7%

Like the normalized method decision chart, a single OPSpecs chart can be used for multiple tests with different TEas. Normalized charts are convenient with a multitest analytical system for which the performance of several tests is evaluated and appropriate SQC rules are selected.

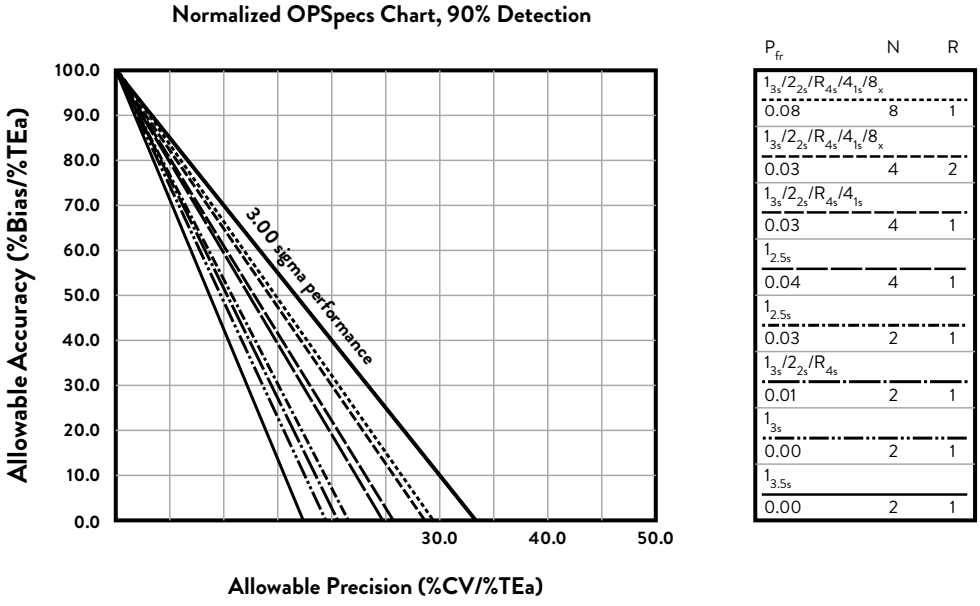

Figure 3-8: Example application of a chart of operating specifications for TEa=6.0% (with 90% error detection) for HbA1c method having a bias of 2.0% and CV of 1.0%. Appropriate SQC procedures would be a 13s/22s/R4s/41s multirule with N=4 or a 12.5s single-rule with N=4.

WESTGARD SIGMA RULES

We have long advocated customizing SQC, such as the Westgard Rules, to account for the required quality for the intended use of a test and the precision and bias for a method[7]. Over time, we have developed a variety of QC planning tools to select the SQC that is right for the intended clinical use and method performance. We continue to look for faster, simpler tools to help labs select the right SQC for their own applications.

In our recent book, *Basic Quality Management Systems*[8], we introduced a new tool that is quicker and easier to use than previous ones: Westgard Sigma Rules™ (to distinguish this approach from the original Westgard Rules). **Figure 3-9** shows the Westgard Sigma Rules for two controls.

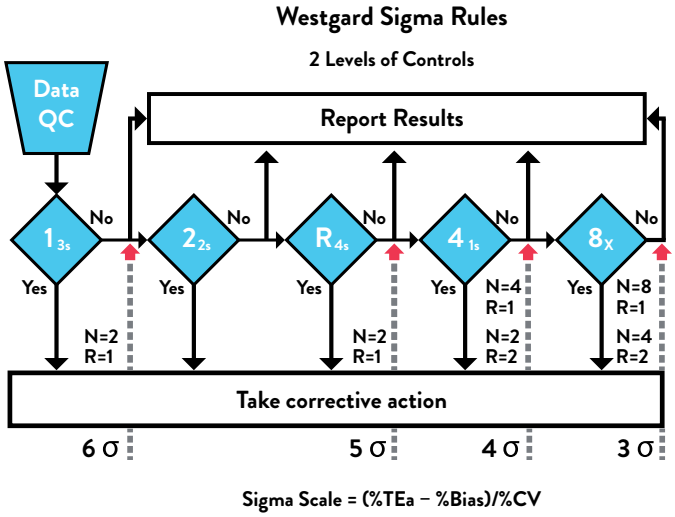

Figure 3-9: Westgard Sigma Rules for 2 levels of controls. Note Sigma scale at the bottom of the diagram. To apply, determine Sigma-metric, locate on the Sigma scale, identify rules above and to the left, find N and R above the Sigma value.

At first glance, it looks just like the Westgard Rules diagram, except there is no 2 SD warning rule, an important distinction. The most important change is the Sigma scale at the bottom of the diagram that provides guidance for which rules should be applied based on the Sigma quality.

Here's how it works. The dashed vertical lines on the Sigma scale show the rules that should be applied based on the Sigma quality. **Figure 3-10** shows examples for Six Sigma (A), five Sigma (B) and four Sigma (C). Locate the Sigma value on the scale at the bottom; then look up and select the control rules to the left. Identify the number of controls (N) and number of runs (R) from the notation immediately to the left above the Sigma value.

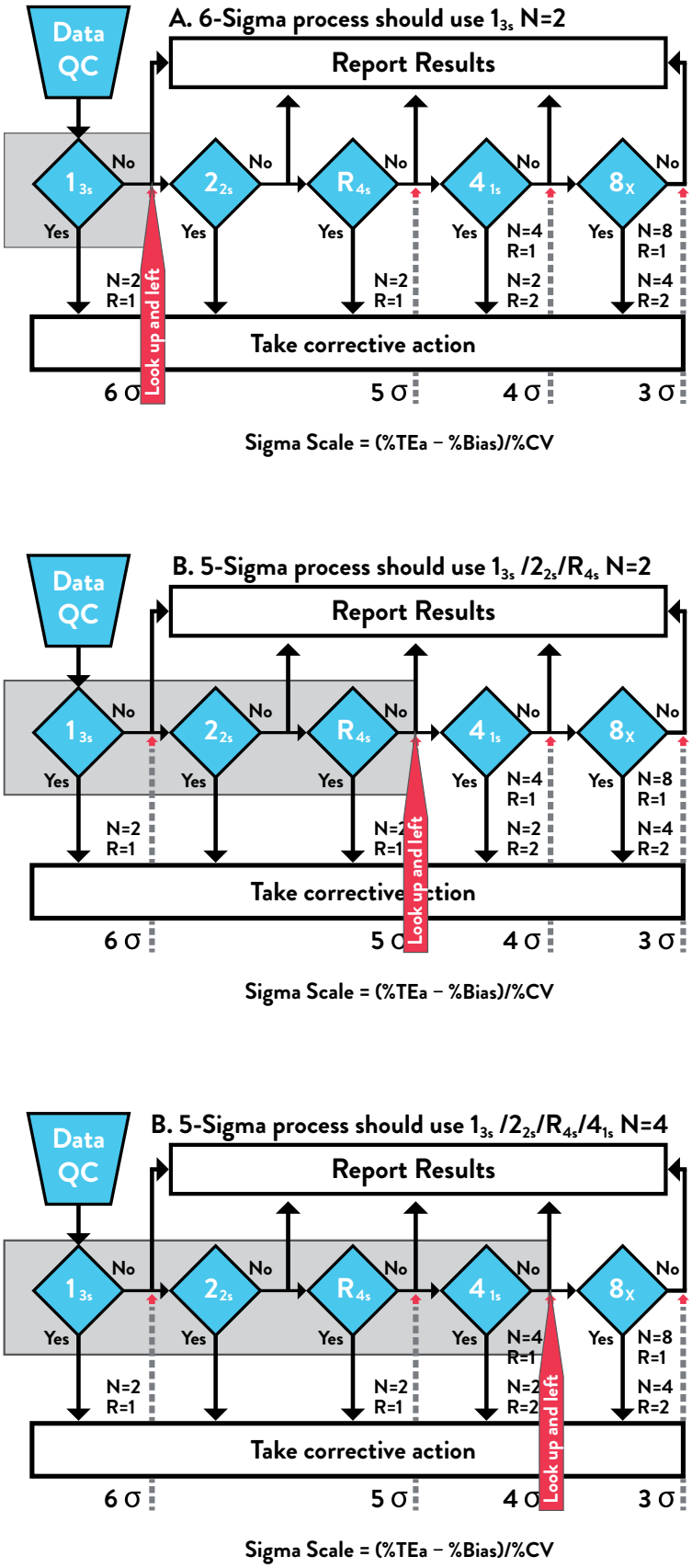

Figure 3-10: Example applications of Westgard Sigma Rules for HbA1c methods having (A) 6-Sigma, (B) 5-Sigma, and (C) 4-Sigma quality.

It's possible to quickly assess the SQC appropriate for different Sigma quality levels:

- Six Sigma quality requires only a single control rule, 1_{3s}, with two controls of different concentration in each run. The notation N = 2, R = 1 indicates two controls are needed in a single run.
- Five Sigma quality requires three rules, $1_{3s}/2_{2s}/R_{4s}$, with two controls in each run (N = 2, R = 1).
- Four Sigma quality requires addition of a fourth rule and the $1_{3s}/2_{2s}/R_{4s}/4_{1s}$ multirule, preferably with four controls in each run (N = 4, R = 1) or, alternatively, two controls in each of two runs (N = 2, R = 2), using the 4_{1s} rule to inspect the control rules across both runs. This second option suggests dividing a day's work into two runs and monitoring each with two controls.
- Less than four Sigma quality requires a multirule procedure that includes the 8x rule, which can be implemented with four control measurements in each of two runs (N = 4, R = 2) or, alternatively, with two control measurements in each of four runs (N = 2, N = 4). The first option suggests dividing a day's work into two runs with four control measurements per run, whereas the second option suggests dividing a day's work into four runs and monitoring each with two controls.

A similar diagram shown in **Figure 3-11** describes Westgard Sigma Rules for three levels of controls.

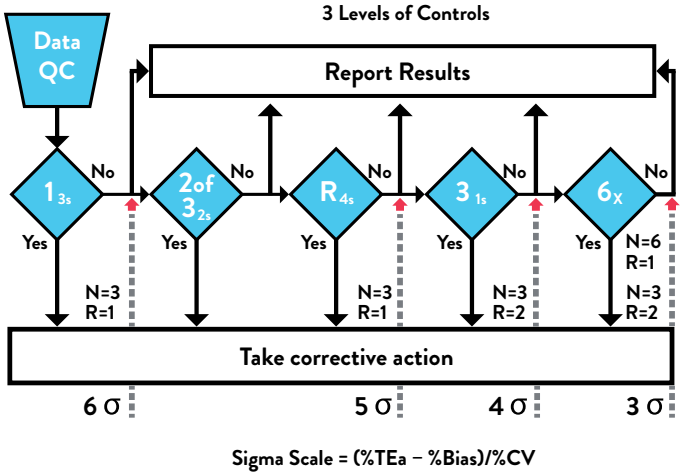

Figure 3-11: Westgard Sigma Rules for 3 levels of controls. Note Sigma scale at the bottom of the diagram. To apply, determine Sigma-metric, locate on the Sigma scale, identify rules above and to the left, find N and R above the Sigma value.

- Six Sigma quality requires only a 1_{3s} rule and one measurement on each of three levels of controls.
- Five Sigma quality requires adding the $2of3_{2s}$ and R_{4s} rules for use with one measurement on each of three levels of controls.
- Four Sigma quality requires adding a 3_{1s} rule for use with one measurement on each of three controls.
- Less than four Sigma quality requires multirule SWQC, including the 6x rule, and a doubling of controls to a total of six, suggesting that three levels of controls be analyzed in duplicate in one run (N = 6, R = 1) or the day's work divided into two runs with three control measurements per run (N = 3, R = 2). If a 9x rule is substituted for the 6x rule, then a day's work could be divided into three runs with three controls per run (N = 3, R = 3).

SQC IN THE REAL WORLD

Evaluating quality on the Sigma scale, many highly automated systems provide a majority of tests with Five to Six Sigma quality. For those with Six Sigma quality, use of a Levey-Jennings QC chart, with control limits set as the mean ± 3 SD and analysis of two controls per run, should provide reliable detection of medically important errors. Two controls of different concentrations or two measurements on one control can be used. For those with five Sigma quality, a simple multirule such as $1_{3s}/R_{4s}/2_{2s}$, with one measurement on each of two controls of different concentrations, should be adequate. Such systems typically include a few tests of lower quality, which require more QC, with the addition of $4_{1s,}$ possibly the 8x rules and doubling the number of controls to a total N of 4.

In point-of-care applications, many test devices do not demonstrate high quality on the Sigma scale; therefore, the SQC required may be very demanding. As a specific example, HbA1c devices evaluated by Lenters and Slingerland[9] demonstrated Sigmas ranging from about 0.0 to 4.5. (See discussion in the previous chapter.) All seven devices were NGSP-certified and classified by the FDA as waived tests, meaning operators need no formal lab training. (They only need to follow manufacturer's QC directions, and devices are not subject to proficiency testing.) There is clearly a need for rigorous QC with many of these devices. Based on the Sigma quality, the best SQC would be the multirule procedure $1_{3s}/2_{2s}/R_{4s}/4_{1s}$ with four controls, but some may require even more QC.

WHAT'S THE POINT?

Labs need to determine the Sigma quality of tests and use SQC to manage testing properly. SQC rules and the number of controls should be optimized for quality and efficiency. It is very easy to determine the right SQC by using Westgard Sigma Rules. The hard part is defining how good a test needs to be for its intended clinical use (i.e., TEa), determining the precision (SD, CV) from a replication experiment or routine SQC data, and determining accuracy (bias) from a comparison of method experiments or PT results. If Sigma is known, the Westgard Sigma Rules make it easy to select the right control rules and number of controls.

References

1. Westgard JO, Groth T, Aronsson T, Falk H, de Verdier C-H. Performance characteristics of rules for internal quality control: Probabilities for false rejection and error detection. *Clinical Chemistry* 1977;23:1857-1867.

2. Westgard JO, Groth T. Power functions for statistical control rules. *Clinical Chemistry* 1979;27:1536-1545.

3. Westgard JO, Barry PL, Hunt MR, Groth T. A multirule Shewhart chart for quality control in clinical chemistry. *Clinical Chemistry* 1981;27:493-501.

4. Koch DD, Oryall JJ, Quam EF, Feldbruegge DH, Dowd DE, Barry PL, Westgard JO. Selection of medically useful quality control procedures for individual tests done in a multitest analytical system. *Clinical Chemistry* 1990;36:230-233.

5. C24A3. Statistical Quality Control for Quantitative Measurement Procedures. Wayne, Pennsylvania, Clinical and Laboratory Standards Institute, 2006.

6. Westgard JO. Charts of operational process specifications (OPSpecs Charts) for assessing the precision, accuracy, and quality control needed to satisfy proficiency testing performance criteria. *Clinical Chemistry* 1992;38:1226-1233.

7. Westgard JO, Hyltoft Petersen P, Wiebe D. Laboratory process specifications for assuring quality in the U.S. National Cholesterol Education Program. *Clinical Chemistry* 1991;37:656-661.

8. Westgard JO, Westgard SA. Basic Quality Management Systems. Chapter 12. *Designing SQC procedures.* Madison, Wisconsin, Westgard QC, 2014.

9. Lenters-Westra E, Slingerland RJ. Three of 7 hemoglobin A1c point-of-care instruments do not meet generally accepted analytical performance criteria. *Clinical Chemistry* 2014;60:1062-1072.

CHAPTER 4

IMPLEMENTING SQC RIGHT

INTRODUCTION

"Doing the right SQC right" describes two important strategies for effective SQC. The first "right" applies to the design of the SQC procedure. It ensures the appropriate control rules and number of controls are based on the quality required and the observed precision and bias, so the SQC detects medically important errors. The second "right" applies to the implementation of the selected SQC, which includes using the appropriate control materials, determining the means and SDs, calculating the appropriate control limits, correctly interpreting control results, taking necessary corrective actions, and documenting the SQC activities. These practices are critical for ensuring the selected SQC behaves as expected.

The previous chapter described how to select the right SQC and provided practical tools for doing so. This chapter provides more background on how to implement and apply SQC properly, so the expected performance is achieved in practice. The basic SQC practice is to analyze a control repeatedly to establish the expected range of results[1]. This is similar to the replication experiment used to verify or validate precision, and the results may be used to calculate the mean and SD used for a control chart. A control chart displays control results over time to identify deviations and changes in method performance. Control limits are drawn on the chart for the range of variation expected and to identify unexpected and unusual patterns of results.

Figure 4-1 provides an overview of the SQC process and identifies who is responsible for each step. SQC is a shared responsibility between lab managers and analysts. Management personnel are responsible for establishing the SQC strategy as part of the SQC. This includes specifying the control rules (including how control limits are calculated), the number of levels of controls, the number of measurements for each level of control, the location of the controls in an analytic run, and the frequency of analysis of the controls. Thereafter, individual analysts or operators are responsible for following the specified SQC strategy. In large labs, quality specialists may exercise the management responsibilities, but all analysts should be involved, so quality control is part of their responsibilities.

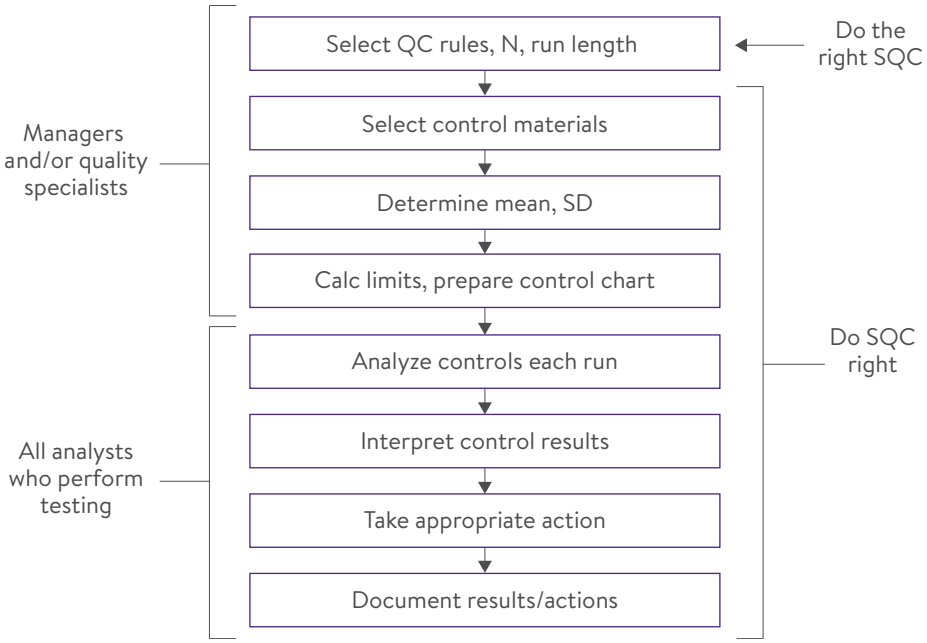

Figure 4-1: Laboratory process for implementing SQC.

SELECTION OF CONTROL MATERIALS

Controls should behave like patient samples. Originally, labs prepared their own patient pools for controls. Today, labs purchase controls from manufacturers who specialize in their production. The matrix of commercial controls ideally should be the same as patient samples, but there are practical limitations, due to additives and processing necessary to ensure long-term stability. For example, lyophilized controls are certainly different from fresh liquid patient samples. Likewise, liquid controls contain additives, making them different from fresh patient samples. These differences may lead to matrix effects, whereby different methods obtain different test results on the same control. That is not necessarily a big problem when the controls are used for an individual method and the mean and SD are established from repetitive testing with that method. However, commutability is a serious issue for controls with assigned values for different methods and analytic systems, such as samples used in EQA/PT to evaluate lab performance. There are precision controls, which may have assigned values that are peer-group-specific, and accuracy controls, where target values should be applicable for all methods.

Stability is an important issue. Ideally, controls should be stable for a year or longer to minimize potential lot-to-lot variability. Vial-to-vial variability should be minimal, so observed differences reflect primarily the analytic variability. Liquid controls minimize vial-to-vial variability but may introduce problems due to viscosity. Controls generally achieve the desired stability for much of clinical chemistry, but they have limited stability in other areas, such as hematology controls for cell counts.

Target values or mean concentrations should be close to medical decision levels (MDLs). However, multianalyte controls that have 20 or 40 or more analytes will likely not hit the MDLs for all of them. For critical tests, such as HbA1c, special controls intended specifically for a given test may be needed. General practice in clinical chemistry is to analyze two controls at different concentrations, whereas three controls at different concentrations are often analyzed for blood gas measurements, immunoassays and hematology testing.

DETERMINATION OF MEAN AND SD

Both assayed and unassayed control materials are available, and it is generally recommended that a lab establish its own mean and SD using a minimum of 20 values over a period of 20 days, each value from a different vial of control material. The mean is calculated as follows:

$$\textbf{Mean} = \sum x_i/n,$$

where the individual control observations are summed and then divided by the number of measurements to determine the mean or average.

The SD is calculated as follows:

$$\textbf{SD} = [\sum(x_i\text{-mean})^2/(n\text{-}1)]^{1/2}$$

Longer-term mean and SD estimates are based on cumulative data obtained over several months to account for the effects of changes in reagent lots, calibrations, operator variability and environmental conditions. Cumulative control limits may be based on three to six months of control data to provide reliable estimates of process variation.

For estimating a cumulative SD, it may be convenient to use the following form of the equation:

$$\textbf{SD} = \{[n\sum x_i^2 - (\sum x_i)^2]/[n(n\text{-}1)]\}^{1/2}$$

CALCULATING CONTROL LIMITS

Control limits should be calculated from the mean and SD determined in the lab by the method operating under stable conditions. The use of bottle values or values from assayed controls is not recommended, except as a stop-gap when introducing a new lot of controls that has not been analyzed in parallel with the old lot. Likewise, use of group mean and SD values from peer-comparison programs is not recommended. Such practices can widen control limits and reduce the false rejections with two SD control limits. The preferred practice is to select the right SQC rules and avoid the use of two SD control limits, followed by use of the mean and SD determined in the lab to calculate the control limits.

SQC software may allow use of assayed values, user-assigned values, monthly values, moving interval values or cumulative values. Management personnel are responsible for setting the control limits, so the control rules will provide the desired performance. This may require a detailed understanding of the SQC software to implement control limits reflecting the lab's own performance limits.

PREPARING A CONTROL CHART

The standard quality control chart used in medical labs is the Levey-Jennings chart, introduced in 1950[2] and modified for use with individual control values by Henry and Segalove in 1952[3]. Individual controls are plotted on the y-axis versus time on the x-axis. Control limits are drawn on the chart to interpret the results. They are typically calculated as the mean plus or minus a certain multiple of the SD, commonly the mean ± 3 SD, the mean ± 2 SD and sometimes the mean ± 1 SD. It is expected that 99.7% (almost all) control results fall within the mean ± 3 SD limits, whereas about 95% are expected within the mean ± 2 SD limits and 67% within the mean ± 1 SD. It is very unexpected for a control to exceed a 3 SD limit (only 0.03%) but only somewhat unexpected for a control to be outside a 2 SD limit (about 5.0%, or 1 out of 20). When there are two controls per run, as required by U.S. CLIA regulations, the false alarms effectively double, with about 10% false rejections, or one out of 10 runs, with at least one control exceeding a 2 SD limit.

With the use of ± 2 SD limits, it is hard to distinguish true alarms from false alarms. Because of the many false alarms, labs may respond to an out-of-control situation by just repeating controls again and again, until they are finally in. Thus, labs may do the wrong QC wrong, causing analysts and operators to get very frustrated by continuous control problems. That's why it is so important to select the right SQC upfront and implement the SQC properly.

To construct the Levey-Jennings control chart shown in **Figure 4-2:**

1. Determine the mean and SD for the control material.
2. Scale the y-axis in concentration units from the mean – 4 SD to the mean + 4 SD, and label it "Control Result."
3. Scale the x-axis in time (typically a month) or consecutive run numbers, and label it accordingly.
4. Draw a solid line at the mean.
5. Draw control limits as the mean ± 3 SD, mean ± 2 SD and mean ± 1 SD.

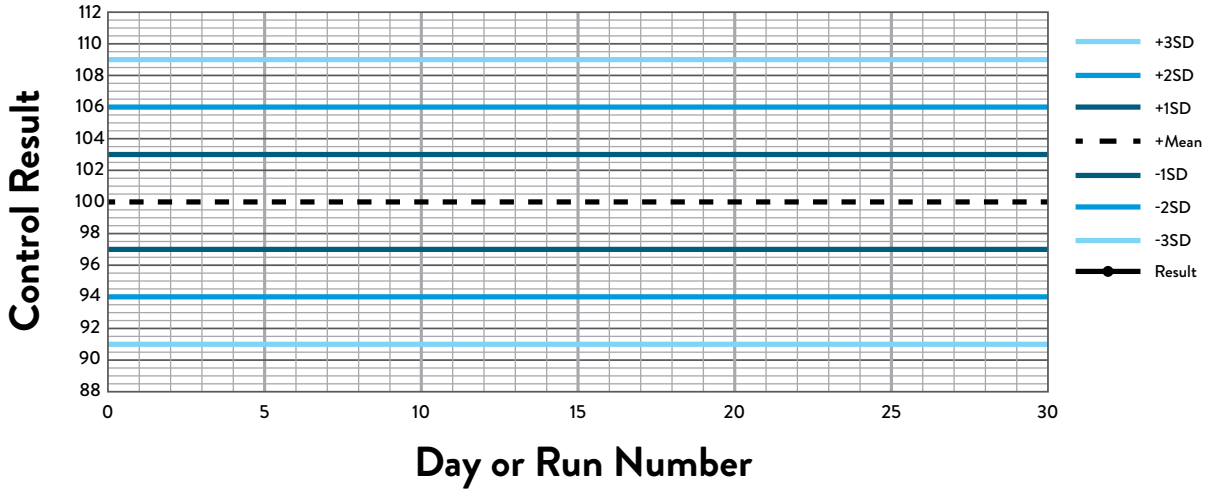

Figure 4-2: Preparation of a Levey-Jennings control chart with control limits set as the mean ± 3SD, mean ± 2SD, and mean ± 1SD for control material having mean of 100 and SD of 3.

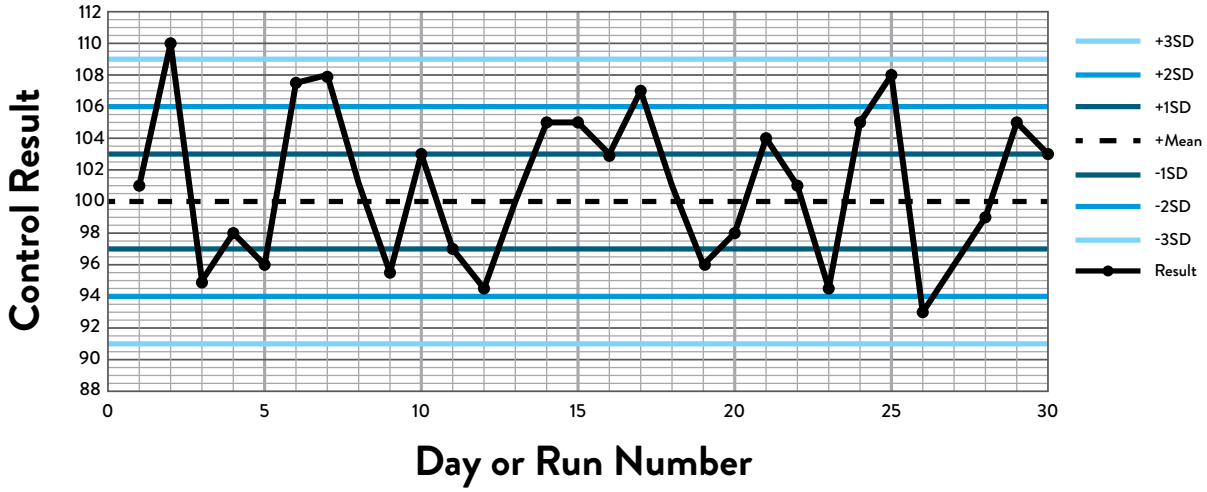

Figure 4-3: Example Levey-Jennings control chart with control limits set as the mean ± 3SD, mean ± 2SD, and mean ± 1SD for control material having mean of 100 and SD of 3.

Figure 4-3 shows the chart with controls plotted. The general practice is to plot each control immediately to examine the control data. Drawing straight lines from point to point visualizes the changes and patterns in the controls. The practical difficulty with SQC is a signal-to-noise problem (that is, detecting a change in performance in the midst of the random error, due to imprecision under stable operation). Additional instability somehow needs to be identified in the presence of that random error or analytical noise.

REVIEWING CONTROL RULES

Specific control rules describe the particular patterns likely to identify changes in performance. Common control rules were defined in the previous chapter, but **Figure 4-4** provides a quick graphical review of the rules used in the Westgard multirule procedure[4] or Westgard Sigma Rules[5] when two controls are analyzed per run. Each rule violation is shown by the points at the end of each graph in **Figure 4-4**, which illustrates violations of the 1_{3s}, 2_{2s}, 4_{1s}, R_{4s}, 8_{x} and 10_{x} rules, respectively.

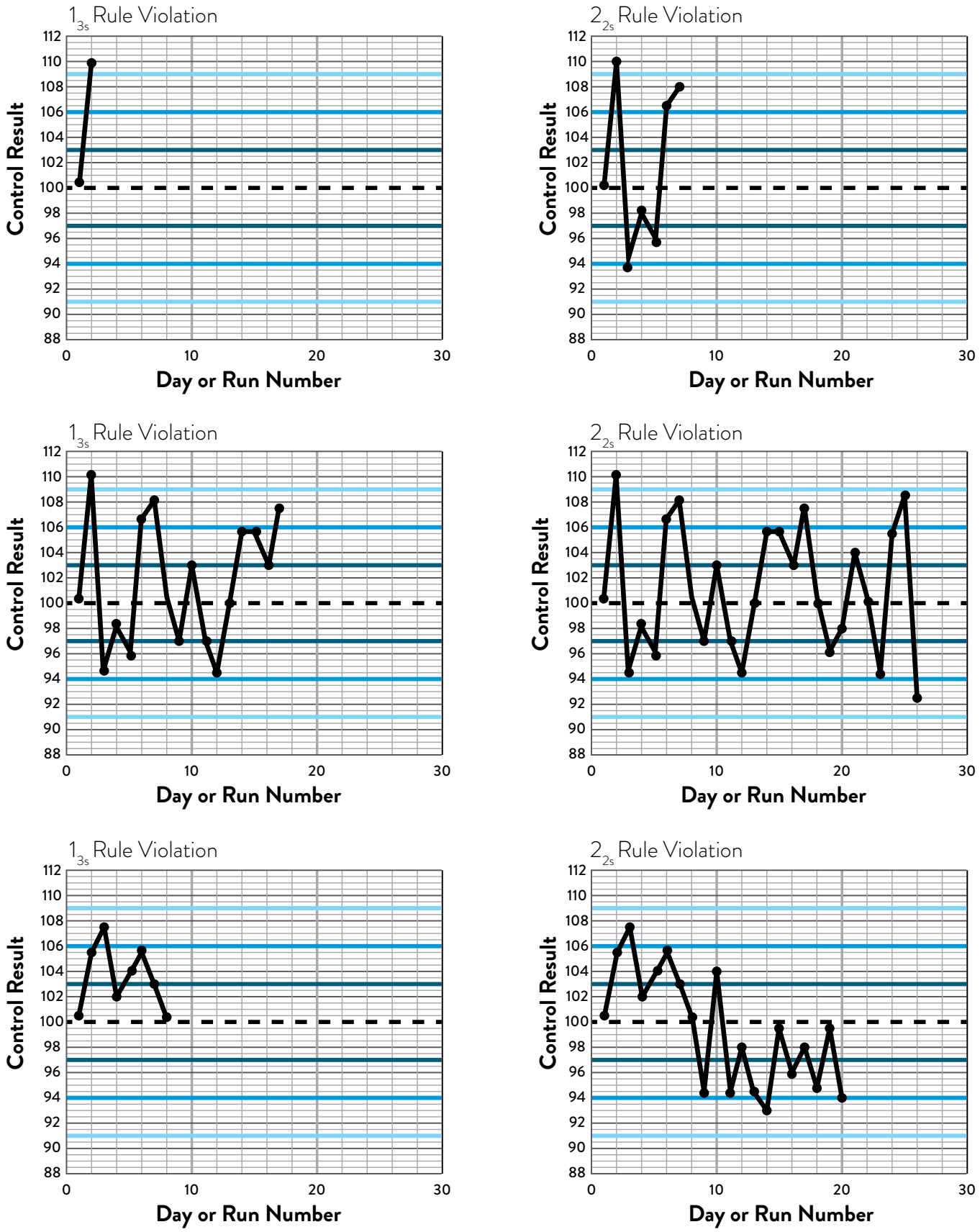

Figure 4-4: Review of control rules commonly used with Westgard multirule SQC procedure with 2 levels of controls. Control charts prepared for a mean of 100 and SD of 3.

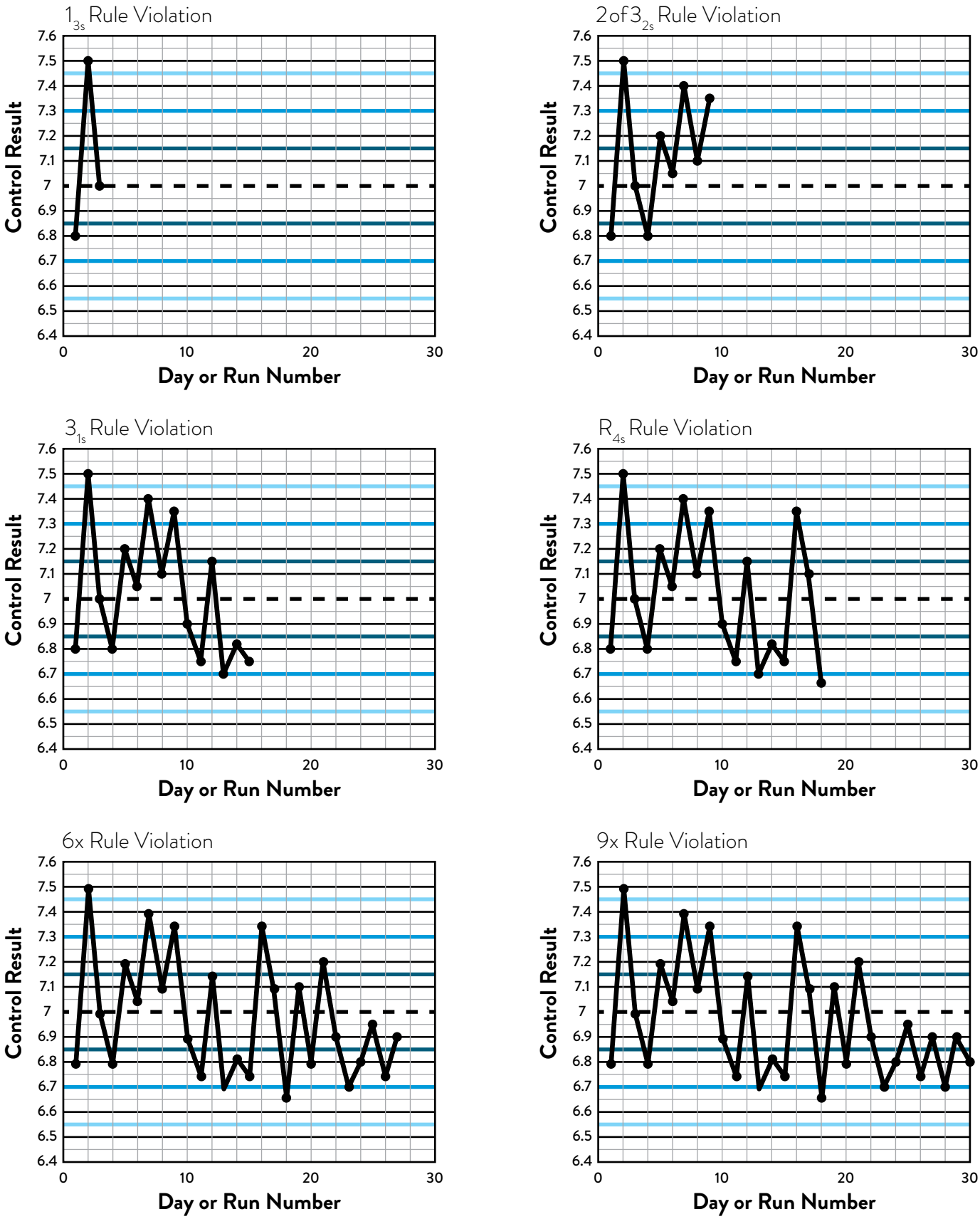

Figure 4-5: Review of control rules commonly used with Westgard multirule SQC procedure with 3 levels of controls. Control charts prepared for a mean of 7.0 and SD of 0.15.

Figure 4-5 provides a graphical review of the 1_{3s}, 2 of 3_{2s}, R_{4s}, 3_{1s}, 6_{x} and 9_{x} rules commonly used when three controls are tested per run. This control chart is prepared for a mean of 7.0 and an SD of 0.15, suitable for monitoring HbA1c. Note the 2 of 3_{2s} rule does not require consecutive measurements but, instead, two out of the three controls in that run. Likewise, the R_{4s} applies to the highest and lowest of the three measurements in a run.

These rule violations are illustrated for a single control, but keep in mind, common practice is to analyze two or three controls in a run and apply the rules across controls (i.e., across control charts).

ANALYZING CONTROLS IN EACH RUN

Determining how often to test controls is still an issue requiring experience and judgment. Guidance is provided by the CLSI C24A3 document[6]:

> ***"For purposes of quality control, the laboratory must consider the stability of the analytical testing process, its susceptibility to problems that may occur, and the risk associated with an undetected error."***
>
> ***"An analytical run is an interval (i.e., a period of time or series of measurements) within which the accuracy and precision of the measuring system is expected to be stable; between which events may occur causing the measurement process to be more susceptible (i.e., greater risk) to errors that are important to detect."***

Stability, susceptibility and risk are the critical factors. What is the interval, period of time or number of patient specimens for which the measuring system is stable? What events might introduce instability? What are the causes of system failures? What is the risk of such failures?

Events are actually the key to making practical sense of this guidance. Consider frequency of QC to be event-driven, as discussed by Parvin[7,8]. There are *expected events*, such as the daily setup of an analyzer, a change of reagents, a new lot of calibrators, replacement of an instrument component, preventive maintenance, and possibly the change of analysts or operators. There are also *unexpected events*, such as reagent deterioration; instrument drift; failure of an instrument component; and change of environmental conditions, such as temperature or humidity. Controls evaluate the effects of these events. For expected events, controls can be scheduled. For unexpected events, controls must be analyzed periodically to ensure changes don't cause medically important errors.

Risk analysis helps labs identify events or failure modes[9]. Likewise, risk analysis prioritizes the importance of these events and failure modes and how they can be detected. The QC plan should schedule controls both for the times of known events and to provide periodic monitoring of unexpected events. As a general SQC strategy, regulatory requirements set a minimum frequency of two controls of different concentrations per day. The lab should add controls for known events to assess the significance of any changes, plus additional periodic controls to monitor unexpected events during testing.

INTERPRETING CONTROL RESULTS

The lab's SOPs should define the SQC and provide directions for interpreting control results. Graphical displays of control data on Levey-Jennings charts are useful for visual assessment, but it is still necessary to define specific control rules to ensure systematic and uniform interpretation. Single rule SQC is preferred for manual applications and low-volume testing sites, but the methodology employed may require a greater number of controls and possibly multirule interpretation of the controls. Use of 2.5s control limits or a $1_{2.5s}$ control rule provides about the same error detection as multirule with the same N, but as N increases, the false rejections are somewhat higher for the $1_{2.5s}$ control rule.

APPROPRIATE ACTION

The lab's SOPs should describe the appropriate actions when control results are in control or out of control. CLSI C24A3 recommends that the common practice of repeating controls should be avoided, emphasizing the importance of selecting the right SQC to minimize false rejections and maximize error detection. Instead of repeating controls, labs should investigate the problem, identify its cause and take corrective actions. The test performance should then be re-evaluated, any questionable patient test results should be examined and, if necessary, the tests should be repeated.

DOCUMENTING RESULTS AND ACTIONS

Controls and the corresponding actions should be documented to provide an accurate history of process performance. Given the importance of events, all changes made to testing must also be documented. Changes should be reviewed whenever control problems occur. Changes that make the process susceptible to errors should be identified for preventive actions and should be monitored by event-driven SQC. The control record of a testing process is often the most valuable information for improving the QC plan.

WHAT'S THE POINT?

An SQC procedure will not perform *as needed* for patient care unless the lab establishes the right control rules, the right number of controls, and the right run length or right frequency of SQC in the SOP.

An SQC procedure will not perform *as expected* for lab operation unless the right mean and SD are used, the right control limits are calculated, the right interpretation is made of the control data, and the right actions are taken.

Doing the "right SQC right" is not easy, but it is essential in any QC plan to provide a safety net for catching medically important errors.

References

1. Westgard JO. *Basic QC Practices*, 3rd ed. Madison, Wisconsin, Westgard QC, 2011.
2. Levey S, Jennings ER. The use of control charts in the clinical laboratory. *American Journal of Clinical Pathology* 1950;20:1059-1066.
3. Henry RJ, Segalove M. The running of standards in clinical chemistry and the use of the control chart. *Journal of Clinical Pathology* 1952;27:493-501.
4. Westgard JO, Barry PL, Hunt MR, Groth T. A multirule Shewhart chart for quality control in clinical chemistry. *Clinical Chemistry* 1981;27:493-501.
5. Westgard JO, Westgard SA. *Basic Quality Management Systems*. Madison, Wisconsin, Westgard QC, 2014.
6. CLSI C24A3. *Statistical Quality Control for Quantitative Measurement Procedures*. Wayne, Pennsylvania, Clinical and Laboratory Standards Institute, 2006.
7. Parvin CA. Quality-control (QC) performance measures and the QC planning process. *Clinical Chemistry* 1997;43:602-607.
8. Parvin CA. Assessing the impact of the frequency of Quality Control testing on the quality of reported patient results. *Clinical Chemistry* 2008;54:2049-2054.
9. CLSI EP23A. *Laboratory Quality Control Based on Risk Management*. Clinical and Laboratory Standards Institute, Wayne, Pennsylvania, 2011.

CHAPTER 5

DEVELOPING A QC PLAN INCLUDING RISK ASSESSMENT

INTRODUCTION

Beginning in January 2016, U.S. medical labs had the option of implementing a risk-based QC plan to comply with U.S. CLIA regulations. This new option is called an Individualized Quality Control Plan, or IQCP. The Centers for Medicare and Medicaid Services (CMS) described an IQCP as consisting of three parts:

- A risk assessment that identifies critical error sources
- A QC plan that assembles practices, resources and procedures to control the quality of a particular total testing process (TTP)
- A quality assessment program that monitors the IQCP[1]

The initial introduction of risk-based QC plans may be targeted at U.S. labs, and there appears to be an increasing interest in risk management, due to the new 2015 edition of ISO 9001[2] and its emphasis on risk-based thinking[3].

> ***"Explicit in the new standard is the requirement that some minimal risk management be integrated into an organization's quality system. The writers deliberately created the term 'risk-based thinking' to encompass the varying, acceptable degrees in which organizations may choose to manage risk."***

ISO 9001 is the basic guideline for quality management for all types of businesses and organizations, so risk-based thinking will be widely supported and a variety of risk management tools will be used more commonly. Adoption of risk-based thinking is evident in the recent 2015 update on QC practices from the Hong Kong Association of Medical Laboratories, which includes a new section on "QC Practices and Risk Management"[4]. In the introduction of the new edition, the editor comments, "[T]he focus in this revision is entirely on risk-based QC management and its practical applications to IQC." Risk-based thinking should help labs optimize the detection of process failures, prevent problems and improve quality.

IDEA OF A QC PLAN

The advantage of a QC plan is that it expands QC coverage to include the pre-analytic and post-analytic phases of the TTP, as well as permits a wide variety of controls to be applied in the analytic phase. Some advocates hope SQC procedures can be replaced by other specific risk controls, particularly for point-of-care applications. However, a major disadvantage is that risk assessment is a complicated process and few labs have any experience with it.

CLSI provides guidance on the use of risk management for developing QC plans in the EP23A guideline[5], developed with the support of CMS and IVD manufacturers. CLSI promotes risk-based QC as the right QC that can be customized for particular measurement procedures and lab conditions (that is, the QC that is right for an individual lab).

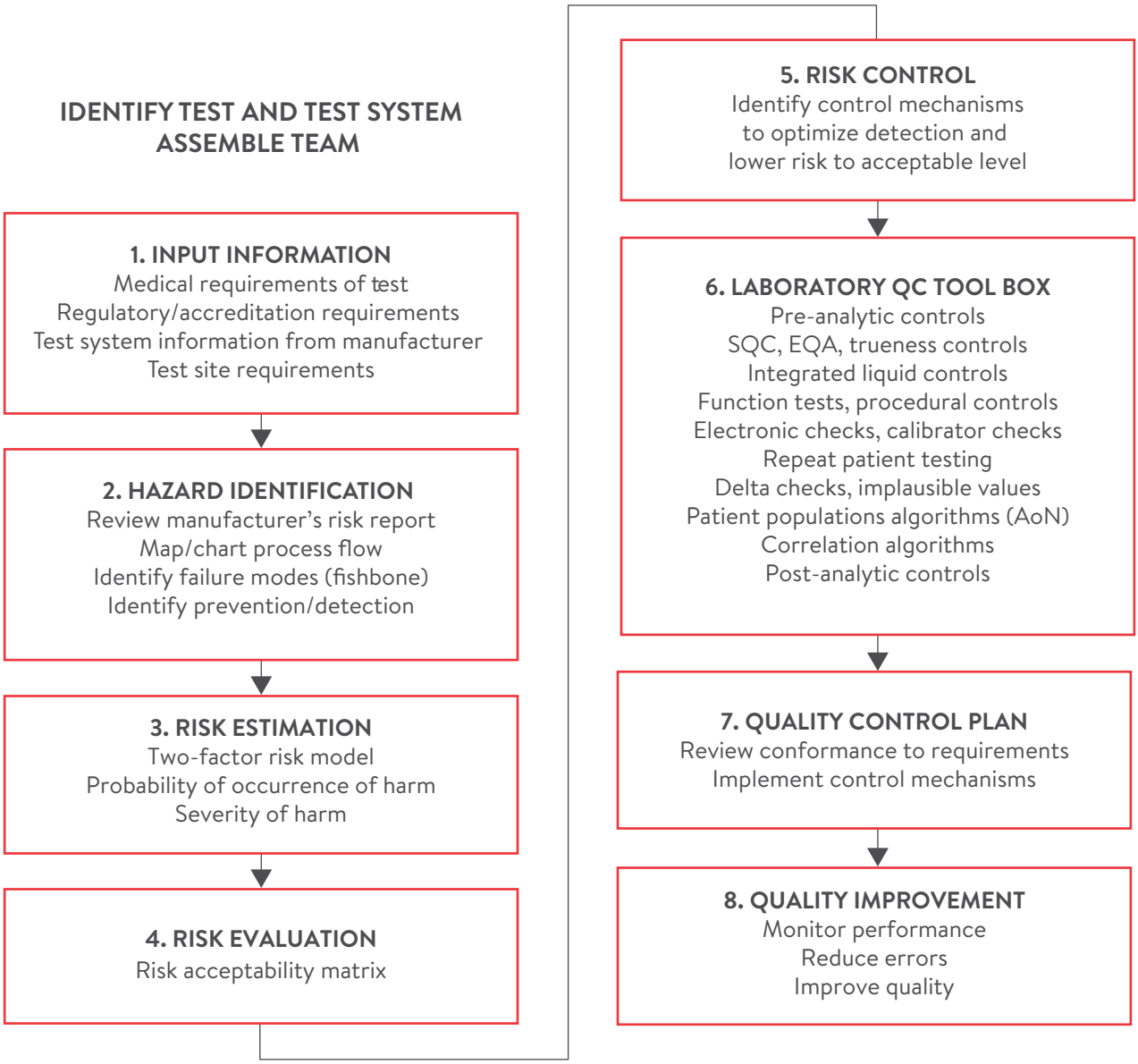

Figure 5-1: CLSI EP23A Process for developing a Quality Control Plan (QCP) based on risk management.

The CLSI process for developing a risk-based QC plan is outlined in **Figure 5-1**. Typically, a lab assembles a team to review the information about a test and analytic system, identify hazards, estimate risk, evaluate risk, identify risk controls, assemble those controls in a QC plan, and monitor their performance to ensure the quality of the QC plan itself. Developing a risk-based QC plan is a complicated process, and labs face a steep learning curve. Furthermore, risk estimation and evaluation are qualitative and subjective, which makes it difficult to assess whether or not a QC plan can "verify the attainment of the intended quality of results," which is the objective of QC, according to ISO 15189[6].

For these reasons, it's recommended to start with a total QC plan, including an SQC procedure selected on the basis of the quality required for intended use and the precision and bias observed. Because this approach includes right-sizing SQC to ensure detection of medically important errors, compliance with the CLIA QC requirements is achieved by analysis of control materials and does not depend on a formal risk assessment. Nonetheless, a total QC plan provides an approach that can also include risk-based controls; thus, it provides a natural foundation for integrating the concepts and ideas for risk management.

CONTROLLING THE TOTAL TESTING PROCESS

Figure 5-2 illustrates the TTP, its three phases and several steps for each phase. The pre-analytic phase begins with a test order that drives the identification and preparation of the patient for specimen collection and identification of the specimen for subsequent processing, concluding with evaluating the quality of the sample and its suitability for analysis. The analytic phase may also begin by evaluating the quality of the sample, preparing the analyzer (reagents, calibration, etc.), analyzing samples, monitoring performance, reviewing QC and releasing the test results. The post-analytic phase may include reviewing the test results, immediate notification of critical values, preparing a test report (including safety information: reference ranges, specimen conditions, etc.), storing samples for future use, monitoring turnaround time, and documenting process performance.

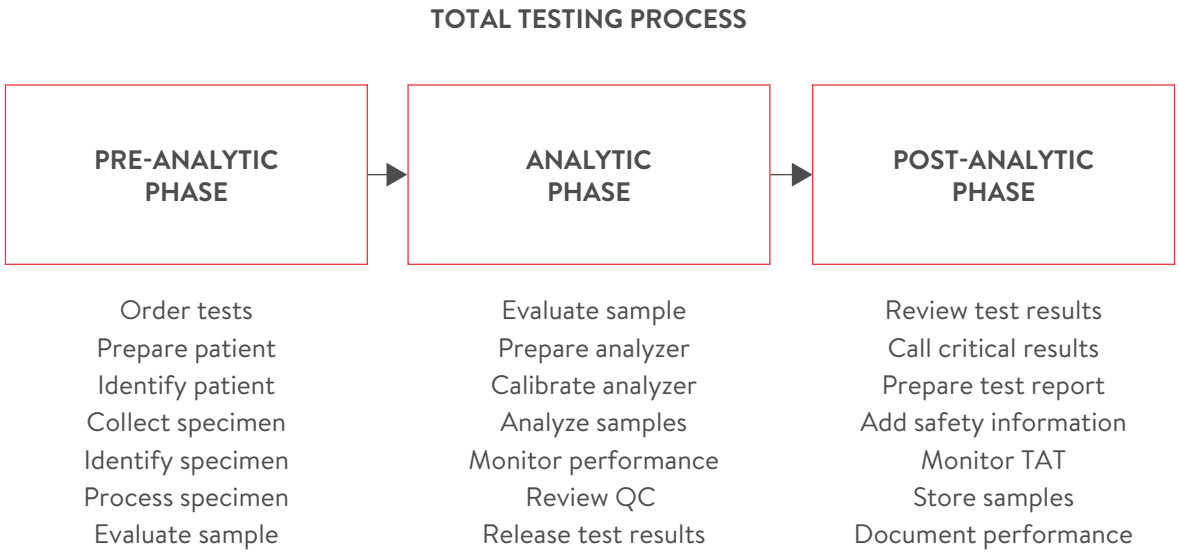

Figure 5-2: Top-down flowchart of the Total Testing Process.

Historically, many labs have developed separate quality systems for the pre-analytic, analytic and post-analytic phases. CLIA also recommends this practice, which is especially appropriate in labs where different people are involved in the different phases of the TTP. The criticism of statistical QC has been that it only monitors the analytic phase of the testing process, and it is important to add pre-analytic and post-analytic controls to right-sized SQC procedures to control the TTP.

The idea of a total QC plan is to provide a complete description of the critical control mechanisms for a particular test system and the TTP. This is a good practice, especially in small labs and point-of-care testing, because the same personnel are involved in all phases of the TTP. In fact, it might be good practice to limit IQCPs to these applications, where it is difficult to implement right-sized SQC procedures.

CMS/CDC GUIDANCE

Specific IQCP guidance ("Developing an IQCP: A Step-by-Step Guide") was issued by CMS and CDC in mid-2015[7] and is available free for download from a CDC website. This guidance strangely does not follow the CLSI EP23A recommendations and provides a much simpler process. In effect, it is hazard identification, without any actual assessment of risks. Identifying hazards is an important step in risk management. CMS provides specific guidance about the factors that should be reviewed by the lab, as shown in **Figure 5-3**. Specific information related to the specimen, environment, reagent, test system and testing personnel should be examined to identify hazards. This worksheet is modeled after the CMS/CDC worksheet that addresses three questions:

1. What are the possible sources of errors?

2. Can the identified sources of errors be reduced?

3. How can the identified sources of errors be reduced?

Although this is called *risk assessment* in the CMS/CDC guidance, it is confusing because risk assessment should involve determining the probability of occurrence of a failure mode, severity of harm and detection capabilities of controls. That is the complicated part of risk-based QC, but the CMS/CDC guidance only asks whether or not the lab can do something about an identified hazard without any actual assessment of risk and prioritization of failure modes.

The CMS/CDC guidance then jumps directly to the quality control plan, which is organized in a worksheet or table with the following headings:

- Type of quality control
- Frequency
- Criteria for acceptability

This is the basic outline for a QC plan: select control mechanisms, specify their frequency and define criteria for acceptability.

Finally, CMS/CDC provides a worksheet for documenting a quality assessment program that includes the following headings:

- QA activity to monitor
- Frequency
- Assessment of QA activity (variation from policy?)
- Corrective action (when indicated)

Following the CMS/CDC guidance will certainly be acceptable in lab inspections in the U.S., but labs should also optimize their SQC practices and make SQC an essential part of any QC plan. SQC provides a basic safety net to catch many of the errors occurring in the analytic phase. Addition of controls to monitor pre-analytic and post-analytic factors provides a total QC plan to which risk-based controls may be added to monitor specific failure modes.

HAZARDS CHECKLIST (BASED ON CMS/CDC RISK ASSESSMENT WORKSHEET)

Test, Test System	
Project Analyst or Group, Date	
Sigma Quality = (TEa-Bias)/SD	

	What are our possible sources of errors?	Can our identified sources of errors be reduced?	How can they be reduced?
SPECIMEN			
Patient preparation			
Collection			
Labeling			
Storage, preservation, stability			
Transportation			
Acceptability and rejection			
Referral (to other labs)			
ENVIRONMENT			
Temperature			
Airflow/ventilation			
Light intensity			
Noise and vibration			
Humidity			
Altitude			
Dust			
Water			
Utilities (elect stability)			
Adequate space			
REAGENT			
Shipping/receiving			
Storage condition requirements			
Expiration data			
Preparation			
TEST SYSTEM			
Inadequate sampling			
Clot detection			
Interference detection			
Hemolysis			
Lipemia			
Icterus			
Turbidity			
Calibration			
Mechanical/electronic failures			
Optics			
Pipettes, pipettors			
Barcode readers			
System controls and function checks			
Procedural, electronic controls			
Liquid controls			
Temperature controls			
Software/hardware			
Data transmission to LIS			
Result reporting			
TESTING PERSONNEL			
Training			
Competency			
Education, experience qualification			
Adequate staffing			

Figure 5-3: Hazards checklist for identifying potential error sources or failure modes.

SELECTING CONTROL MECHANISMS

The basic mechanisms needed to monitor the TTP should consider the following pre-analytic, analytic and post-analytic controls:

PRE-ANALYTIC CONTROLS

- **Patient identification:** Quality testing begins by ordering the right test on the right patient. The identification of a patient must be checked carefully and repeatedly by each service provider.
- **Collection and processing of specimens:** The right type of specimens must be correctly obtained using the right specimen collection devices.
- **Labeling of specimens and samples:** The specimens and processed samples must be correctly labeled to identify the patient source.
- **Sample requirements:** The type of sample, volume, and presence or absence of possible interfering conditions (e.g., hemolysis) should be assessed as a control on the pre-analytic phase to address needs for new specimens as early as possible.

ANALYTIC CONTROLS

- **Reagent acceptability:** Reagents should be stored under conditions specified by the manufacturer and used within the shelf-life date on the reagent label.
- **Operational acceptability:** Properly trained operators should follow SOPs for preparing reagents, calibrators, controls and samples, and specific checklists for qualifying the test system for routine operation.
- **Instrument and environmental conditions:** Specific functions of the test system are often monitored by internal mechanisms, and inadequate conditions are identified by error messages or flags. Environmental conditions (e.g., temperature) may need to be monitored separately if not identified by the instrument flags.
- **Sample acceptability:** The samples should be checked visually for hemolysis, lipemia and icterus by the analyst or by measurement of sample indices by the test system.
- **Calibration and trueness:** The correctness of calibration should be checked periodically with use of reference materials. Trueness controls with assigned values are available for some measurands and can be used to verify calibration.
- **Statistical QC:** Stable control materials should be measured, along with patient samples, to evaluate the performance of the test system, environmental conditions and system operators.
- **QC review:** Before results are released, the SQC measurements and error messages are evaluated by operator review or, under certain conditions, by automated review.
- **Test result review:** Patient test results should be reviewed to identify any inconsistencies or questionable results. Possible control mechanisms include limit checks, critical value checks, delta checks, cross-check algorithms and population algorithms (such as average of normals or average of patients).

POST-ANALYTIC CONTROLS

- **Immediate notification of critical values:** Test results that represent critical values for patient management need to be transmitted to the physician as soon as possible, often by an alternative communication, such as telephone, email or text mail.
- **Interpretative guidance and safety information:** Along with the test result, information on reference intervals, possible sample interferences and interpretive guidance should be provided to ensure safe use of the test results.
- **Test report delivery:** Appropriate test reports should be delivered, as needed, to facilitate the care of patients.
- **Turnaround time:** Throughout the process, from specimen collection to reporting of test results, samples should be tracked for location and time to identify delays throughout the TTP.
- **Customer complaints:** All feedback from physicians, nurses and patients should be documented to identify problems throughout the TTP. Corrective and preventive actions should be documented, and plans should be made for improvements.

To review controls, **Figure 5-4** provides a summary of available control mechanisms based on the QC toolbox recommended in CLSI EP23A[6]. This list is useful to audit existing controls and currently documented policies and procedures. With this information, the lab should be able to identify controls to be implemented.

Control of Mechanisms	Frequency	Criteria for Acceptance
PRE-ANALYTIC CONTROLS		
Physician test order	Every patient	Readable, match with sample
Patient identification	Every patient	Correct ID, match with sample
Specimen labeling	Every specimen	Correct ID, match request and sample
Specimen processing	Every specimen	Proper container, time
Sample inspection	Every sample	No visible hemolysis, lipemia
ANALYTIC OPERATOR CONTROLS		
Standard Operating Procedure	Yearly SOP review	Up to date, signed by director
Operator training	Every operator	Demonstrated proficiecy
Operator checklists	Daily	Supervisor review
System maintenance	Manufacturer Schedule	On time, on schedule
Operator competency	Yearly	PT, supervisor review
ANALYTIC TEST SYSTEM CONTROLS		
Reagent storage & expiration	Every run	Within outdate period
Sample acceptability	Every sample	Visual inspection
Electronic checks	Manufacturer	Manufacturers specs
Function tests	Manufacturer	Manufacturers specs
Process tests	Manufacturer	Manufacturers specs
Calibration checks	Manufacturer/Regulations	Within TEa limits
Statistical QC	Startup + Monitor	Westgard Sigma Rules
Trueness controls	Periodic	TV ± limits of uncertainty
Proficiency testing	3 times per year	CLIA criteria for acceptance
ANALYTIC TEST REVIEW CONTROLS		
Limit checks	Every sample	Limits defined per test
Implausible values	Every sample	Limits defined per test
Repeat patient tests	Daily	Limits defined per test
Delta checks	Each sample	Limits defined per test
Correlation algorithms	Each sample	Limits defined per test
Patient population algorithms	Each run	Limits defined per test
POST-ANALYTIC CONTROLS		
Review test results	Every run	Range criteria per test
Confirm/call critical values	Each test result	Critical value criteria
Interpretive and safety info.	Each report	Codes on report
Report delivery schedule	Each report	Clinic criteria
Turnaround time	Each sample	TAT stats, routine
Customer feedback	Each complaint	Supervisor review

Figure 5-4: Available control mechanisms along with example specifications for application.

PRIORITIZING ANALYTIC CONTROLS

Pre-analytic controls, operator controls and post-analytic controls all have a high priority regardless of the Sigma quality of a testing process. Analytic controls, however, can be prioritized in relation to Sigma when SQC procedures have been right-sized to optimize the detection of medically important errors, as shown in **Figure 5-5**[8].

The priority for control review is low when Sigma is high (> 5.5) and high when Sigma is low (< 3.5). That means controls with low priority may not need to be included for high Sigma quality test systems, therefore simplifying the total QC plan because SQC can be relied on to detect medically important errors. Low Sigma quality test systems, on the other hand, will require more intensive and extensive control mechanisms, which may be difficult to implement in certain lab settings, such as point-of-care applications. In principle, point-of-care test devices should perform at a Sigma quality of 5.5 or better to ensure simple control mechanisms will reliably detect medically important errors.

Control Mechanisms	Sigma >5.5	Sigma 3.5-5.5	Sigma <3.5
ANALYTIC OPERATOR CONTROLS			
Standard Operating Procedure	High	High	High
Operator training	High	High	High
Operator checklists	High	High	High
System maintenance	High	High	High
Operator competency	High	High	High
ANALYTIC TEST SYSTEM CONTROLS			
Reagent storage and expiration	Low	Medium	High
Sample acceptability	High	High	High
Electronic checks	Low	Medium	High
Function tests	Low	Medium	High
Process tests	Low	Medium	High
Calibration checks	Low	Medium	High
Statistical QC	High	High	High
Trueness controls	Low	Low	Low
Proficiency testing	Regulatory	Regulatory	Regulatory
ANALYTIC TEST REVIEW CONTROLS			
Limit checks	High	High	High
Implausible values	High	High	High
Repeat patient tests	Low	Medium	High
Delta checks	Low	Medium	High
Correlation algorithms	Low	Medium	High
Patient population algorithms	Low	Medium	High

Figure 5-5: Priority of analytic controls on basis of the Sigma quality of the test system.

FREQUENCY OF CONTROLS

Frequency is often difficult to specify, but some guidance is available from manufacturers for system maintenance, calibration, test system checks, SQC, etc. Regulatory and accreditation requirements set maximum times for other controls, such as calibration checks, SQC and PT. The Sigma quality of the test system also can be used to specify higher frequency controls for lower Sigma testing processes.

For SQC, the Westgard Sigma Rules provide initial guidance in terms of number of runs per day or per shift as a starting point for establishing the frequency of SQC. Other factors are also important in determining when controls should be analyzed. It is useful to identify events or changes that occur in the testing process that should be qualified by analysis of new controls. There are both expected events and unexpected events. The first refers to known, scheduled or observed changes that occur at specific times; the latter refers to unexpected changes that might occur anytime. An important strategy is to schedule controls for all expected events, such as changes in reagent lots, calibrator lots, system maintenance, replacement of parts, changes in environmental conditions, and even possibly changes in analysts or operators. For unexpected events, it is important to periodically monitor the testing process to limit the possible exposure to unknown changes that may affect the quality of patient test results.

The SQC design for expected events should include the right control rules and right number of controls to detect medically important errors (i.e., right-sized Westgard Sigma Rules). For unexpected events, a monitoring design should be employed that uses a single control rule, such as 1_{3s} or $1_{2.5s}$, and spaces control measurements throughout the analytic run. One practical consideration is the number of patient samples that would need to be retested if a run were determined to be out of control. Cost of the repeat testing should be weighed against the cost of analyzing periodic controls throughout the run.

As an overall strategy, start with the regulatory requirements that mandate a minimum of two control levels per day. Add controls for expected events to evaluate the changes in the testing process. Add controls for unexpected events to monitor the process during routine operation to minimize the risk and costs of unexpected changes. Finally, add risk-based controls to monitor specific failure modes of the particular test and test system.

How to deal with expected and unexpected events is actually the realm of risk management. Here's where risk management can be helpful in developing a total QC plan. What changes or failures might occur? What is the probability of such failures happening? What is the severity of harm from such failures? What control mechanisms can be implemented to detect such failures? What frequency of controls is needed to monitor that failure mode? What corrective actions or, preferably, preventive actions might be taken to reduce harm?

ASSEMBLING A TOTAL QC PLAN

In the approach recommended here, the SQC procedure should have already been right-sized by the selection of appropriate Westgard Sigma Rules, in which case, the lab will be in compliance with the CLIA requirements to analyze at least two controls and to detect medically important errors. Relevant pre-analytic and post-analytic controls can be added without the need for a formal risk assessment. The strategy is to satisfy the CLIA QC requirement without performing a risk assessment but to still ensure quality testing by inclusion of the most critical pre-analytic, analytic and post-analytic controls. Those controls should be identified on the basis of experience, with help from others involved in the TTP, guidance based on the Sigma quality of the test system, and guidance from the manufacturer's instructions for use.

An example total QC plan is shown in **Figure 5-6**. The minimum pre-analytic control is to examine samples for acceptability for analysis (sample type, volume, potential interferences). Operator controls are essential, such as proper training and following SOPs, checklists and maintenance schedules. SQC is the essential analytic control and should be supplemented by limit and implausible value checks. Proficiency testing is important for long-term monitoring of the analytical quality of the testing process. Post-analytic controls should include identification of critical values for immediate notification and monitoring of turnaround time.

Control of Mechanisms	Frequency	Criteria for Acceptance
PRE-ANALYTIC CONTROLS		
Patient identification	Every patient	Correct ID
Specimen labeling	Every specimen	Correct name on label
Sample inspection	Every sample	No visible hemolysis or lipemia
ANALYTIC OPERATOR CONTROLS		
Standard Operating Procedure	Yearly SOP review	Signed by technical supervisor
Operator training	Every operator	Proficiency by supervisor
Operator checklists	Daily	Supervisor review
System maintenance	Manufacturer schedule	Supervisor review
Operator competency	Yearly	Proficiency assessment
ANALYTIC TEST SYSTEM CONTROLS		
Sample acceptability	Every sample	Instrument indices and volume limits
Calibration checks	Manufacturer/Reg.	Controls within limits
Statistical QC	Startup + monitor	Controls within limits
Proficiency testing	3 times per year	Acceptable scores
ANALYTIC TEST REVIEW CONTROLS		
Limit checks	Every sample	Instrument working range check
Implausible values	Every sample	Critical limits
POST-ANALYTIC CONTROLS		
Confirm/call critical values	Each test result	Critical limits
Turnaround time	Each sample	60 minutes for stats, 3 hours other
Customer feedback	Each complaint	Supervisor review

Figure 5-6: Available control mechanisms along with example specifications for application.

WHAT'S THE POINT?

Medical labs can adopt the concept of a Total QC plan without the need to perform a formal risk assessment, adding pre-analytic and post-analytic controls to right-sized SQC procedures to monitor the TTP. Right-sized SQC is the key building block in a Total QC plan and can be relied on to detect many of the possible failure modes in the analytic process. Other controls can be added based on knowledge and experience, but there is no requirement to perform a formal risk assessment. Instead, the determination of Sigma quality provides a general assessment of risk that guides both the selection of SQC procedures and the addition of other controls[8]. A Total QC plan provides a good starting point for improving QC practices and a logical building block for developing risk-based QC plans. Formal risk assessment is described in reference 9, and its integration with Six Sigma concepts is described in great detail in reference 10.

References

1. CMS Memo of August 16, 2013: Individualized Quality Control Plan (IQCP): A New Quality Control (QC) Option. www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/Downloads/IQCPbenefits.pdf.

2. ISO 9001:2015. Quality Management Systems. International Organization of Standards, Geneva.

3. Robitaille D. Consider risk. *Quality Progress* 2015 (September), 24-25.

4. Pang R, ed. *A Practical Guide to Internal Quality Control (IQC) for Quantitative Tests in Medical Laboratories: Version 2.0*. Hong Kong Association of Medical Laboratories Limited, February 2015.

5. CLSI EP23A. *Laboratory Quality Control Based on Risk Management*. Clinical and Laboratory Standards Institute, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania, 2011.

6. ISO 15189:2012. *Medical laboratories – Requirements for quality and competence*. International Organization of Standards, Geneva.

7. CDC, CMS, U.S. Department of Health and Human Services. Developing an IQCP: A Step-by-Step Guide. www.cdc.gov/CLIA/Resources/IQCP/

8. Westgard S. Prioritizing Risk Analysis Quality Control Plans based on Sigma-Metrics. In Quality Control in the Age of Risk Management. Eds. JO Westgard, SA Westgard. *Clinics in Laboratory Medicine* 2013;33(1):41-53.

9. Westgard JO, Westgard SA. *Basic QC Practices*, 4th ed. Madison, Wisconsin, Westgard QC, 2016.

10. Westgard JO. *Six Sigma Risk Analysis: Designing Analytic QC Plans for the Medical Laboratory*. Madison, Wisconsin, Westgard QC, 2011.

CHAPTER 6

MONITORING QUALITY AND PERFORMANCE

INTRODUCTION

Recalling the Deming PDCA cycle and the final steps in the 6ΣQMS discussed in Chapter 1 (please review **Figure 1-3**), the Act part of the cycle involves the following steps:

- Measure quality and performance (EQA, PT, MU)
- Monitor failures (quality indicators)
- Improve quality of the QC plan and/or the testing process (continuous quality improvement, or CQI)

The purpose is to ensure the SQC works effectively in routine operation over time to identify failure modes to be corrected or prevented and to improve the QC plan itself or, if necessary, to start over and update the QC plan to ensure the necessary quality by using a new analytic system and updating the SQC and total QC plan.

Also, remember individualized QC plans require three components: a risk assessment, a QC plan and a quality assessment (QA) program.[1]

> ***CMS has approved a procedure that permits laboratories to develop and customize quality control procedures in their healthcare setting. This procedure is termed individualized quality control plan (IQCP). An IQCP comprises three parts: a risk assessment (RA), a quality control plan (QCP), and a quality assessment (QA) plan. The RA is the identification and evaluation of potential failures and errors in a testing process. A QCP is a laboratory's standard operating procedure that describes the practices, resources, and procedures to control the quality of a particular test process. The QA is the laboratory's policy for the ongoing monitoring of the effectiveness of its IQCP.***

QUALITY ASSESSMENT (QA)

QA is the ongoing assessment of the testing process quality through review of performance measures and indicators to identify problems and make improvements. CMS recommends this monitoring include the analysts, specimens, reagent test system and environmental conditions. Recommended review documents include specimen-rejection logs, QC records and corrective actions, maintenance records and preventive actions, patient test results, turnaround time records, personnel competency assessments, and proficiency testing scores.

> ***When the laboratory discovers a testing process failure, the laboratory must conduct an investigation to identify the cause of the failure and its impact on patient care, and make appropriate modifications to its QCP, as applicable. The investigation must include documentation of all corrections, corresponding corrective actions for all patients affected by the testing process failure, and evaluation of the effectiveness of the corrective action(s). The laboratory must implement the correction(s) and corresponding corrective action(s) necessary to resolve the failure and reduce the risk of recurrence of the failure in the future. If necessary, the laboratory must update the risk assessment with the new information and modify the QCP as needed.***

The targeted failure modes should be monitored to measure the failure frequency, review the corrective actions and identify preventive actions leading to improvements. The initial risk assessment ranked the frequency of failures based on the team members' judgments. Now, it should be possible to determine failure frequency and the effectiveness of the controls based on real lab data.

CLSI EP23A describes the QA activity as "post-implementation monitoring of the quality control plan"[2] and provides the following guidance:

> ***"[A] laboratory should establish a review system for monitoring quality benchmarks, or the effectiveness of the QCP over time. One quality benchmark that could be monitored is the frequency of a specific error over time to ensure that the QCP effectively reduces the frequency of error occurrence. Unacceptable performance will trigger an investigation to identify the root cause and potentially trigger appropriate modifications to the QCP ...."***

This guidance focuses on the targeted failure modes from the initial risk assessment. Those potential error sources should be monitored to estimate the actual frequency of error occurrence, identify and correct the root causes, and improve the testing process and/or the QCP.

QUALITY INDICATORS

Here is a general list of the kinds of data that should be collected to monitor the failures of a lab testing process[3]:

- **Specimen conditions and sample acceptability**
  - Incorrect identification/specimen labeling problems and number of hemolyzed samples, clotted samples, samples with inadequate volume and samples redrawn
- **Test system failures**
  - Number of runs rejected and error flags observed
    - Types of failures, such as reagent, calibration or control degradation, hardware failure, software failure, inadequate maintenance, operator errors and adverse environmental conditions
    - Corrective and preventive actions
  - Number of patient error conditions
    - Delta-check errors, correlation-check errors, reportable range problems and panic values called
  - Instrument flags and error messages
    - Type of flag, number of occurrences and corrective actions
- **Test system performance**
  - Precision observed from SQC or repeat patient controls
  - Bias observed determined from method comparison, PT/EQA surveys and/or peer comparison programs

- **Test reports**
  - Turnaround time (TAT), average and 95% limit of TAT distribution
- **Customer complaints**
  - Department/source, specimen/sample problems, TAT, analytic quality and other service conditions

A specific list should be developed for the test and test system. The effectiveness of these indicators depends on thorough data collection and the capabilities for accessing and analyzing those data. Periodic reports should be reviewed by the lab director and manager. The lab director should identify and prioritize quality issues that need resolution and improvement.

EXAMPLE QA PLAN

As part of an IQCP, the QA plan may be described as shown in **Figure 6-1**, in which the quality indicator is identified in the first column and its implementation is described in the second column.

QUALITY INDICATOR	MONTHLY PERFORMANCE SUMMARY
Workload	Count total number of tests performed
Samples rejected	Count samples rejected
	Number due to hemolysis
Processing time	Measure time from specimen collection to analysis
	Measure analysis time
	Measure total time from specimen to report
Test system flags	Count total number of device alerts and error flags
Runs/tests rejected	Count number of runs rejected because of flags and controls
	Count number of patient samples rejected
Operator variability	Calculate SD of duplicates for RPT controls
Bias versus reference	Calculate bias versus comparative method
Bias from PT survey	Calculate bias each survey events
Turnaround time	Tabulate turnaround time measurements
	Calculate average turnaround time
	Determine approx. 95% upper limit
Customer feedback	Count number of complaints
	Summarize causes of complaints
	Summarize corrective actions

Figure 6-1: Example quality assessment program to monitor quality and performance of a laboratory testing process.

While it would be preferable to computerize the QA data collection, it may be necessary in small labs to develop a manual system. The system requires logs to track patient samples as they progress through the TTP, from specimen acquisition through reporting of patient test results. Logs should allow monitoring of processing time for the pre-analytic phase, the analytic phase and the total time to reporting. QC records must track all of the results from control mechanisms, such as the alerts and flags from the test system, measures of operator and test system variability from a repeat patient test (RPT) control, and periodic assessments of bias by comparison to another method and from EQA/PT results. Corrective actions for control failures must be documented, and the number of patient test results affected should be part of that record. Customer feedback and complaints should be summarized, along with the corrective actions.

ERROR OR DEFECT RATES

Remember, the purpose of collecting these data is to determine how often failures occur. That's why there must be some measure of the workload to express the number of opportunities for failure. The actual number of failures is referenced to the total number of opportunities to estimate a defect rate.

Example:
If five samples out of 100 are found to be unacceptable due to hemolysis, the defect rate is 5%.

Another way to express defect rate is defects per million (DPM) or defects per million opportunities (DPMO), commonly used in the industry and in Six Sigma quality management. A 5% defect rate corresponds to 50,000 DPM. An advantage of the DPM figure is that it can be converted to Sigma-metrics by using standard tables available in any Six Sigma QM text.

Example:
50,000 DPM corresponds to 3.15 Sigma. While a 5% error rate doesn't sound so bad, 3.15 Sigma isn't very good and would be considered minimal acceptable quality for a manufacturing process.

When quality is expressed on the Sigma scale, it becomes clear that defect rates of 0.5% or better (5,000 DPM) need to be achieved to provide quality equivalent to airline baggage-handling errors (4.15 Sigma), which many have unhappily experienced. (See Chapter 14 in *Basic Quality Management Systems*[3] for a more complete discussion of monitoring nonconformities and converting percentage defects to quality on the Sigma scale.)

Defective results should also influence a lab's measure of turnaround time (TAT). Labs often cite average TAT in discussions with customers. However, when customers state they expect results within 60 minutes, they will expect all the results to be available in that time frame, not just half of them. A more realistic indicator of lab performance is to use the upper 95% limit of the observed distribution of TATs and compare that limit to the customer's requirement.

EXTERNAL QUALITY ASSESSMENT/PROFICIENCY TESTING (EQA/PT)

Participation in EQA/PT surveys is almost universally required for medical laboratories, with the exception of waived tests in U.S. labs. External samples are submitted to the lab from the EQA/PT program and analyzed by the lab, and the results are reported to the survey provider, who scores the results and documents the observed performance in a report. The survey report typically identifies a target value, determined either by reference method analysis, the observed survey group mean, or the method subgroups' or peer groups' means.

A lab is usually scored according to the TEa criteria defined by regulatory requirements (e.g., CLIA) or established by the survey program. In the U.S., acceptable performance for regulated analytes requires acceptable results within defined CLIA criteria for four out of five samples in a survey event for each regulated test (CLIA lists about 80 tests requiring PT). Non-regulated tests are often surveyed with two samples and two events per year and graded by the PT/EQA provider. (More details about U.S. CLIA PT regulations are found in reference 4.)

Survey results are often considered to be proprietary by the providers, although results must be available to regulatory agencies. The U.S. National Glycohemoglobin Standardization Program (NGSP) publicly discloses the CAP survey results for HbA1c[5] and provides an example of accuracy-based grading using a reference method target value. These data are very useful for assessing the comparability of results using the bias and standard deviation for each method subgroup. Those results can be used to assess quality on the Sigma scale[6], as in **Figure 2-6**.

For the individual lab, the most important information is the bias observed. However, given the small number of samples and the requirement that samples must be tested in the same way as patients (i.e., a single measurement), it is necessary to average the observed biases across samples to minimize the effects of random variation. Thus, bias is not known very well for a method in an individual lab. Problems with PT/EQA results often require further testing and investigation with other reference materials or methods. For some tests, certified reference materials (trueness controls) are available that have assigned values and documented uncertainty.

MEASUREMENT UNCERTAINTY (MU)

Determination of MU is not required in U.S. labs under the CLIA regulations, but MU is required for labs accredited under the 2012 edition of ISO 15189[7]. Given that ISO 15189 is the global standard of practice for quality management in medical labs, U.S. labs should also consider how to implement a methodology that meets the ISO requirement:

> ***"The laboratory shall determine measurement uncertainty for each measurement procedure in the examination phases used to report measured quantity values on patients' samples. The laboratory shall define the performance requirements for the measurement uncertainty of each measurement procedure and regularly review estimates of measurement uncertainty."***

The practical estimation of MU in the lab comes down to calculating the SD from SQC data under intermediate precision conditions and then multiplying that SD by a factor of two to provide a conventional 95% confidence limit for a test result. The SD is known as the *standard measurement uncertainty,* the factor of two is called the *coverage factor*, and the 95% limit, or interval, is known as the *expanded measurement uncertainty*.

INTERMEDIATE PRECISION CONDITIONS

This implies control values within one lab but with changes between reagent lots, calibrator lots, operators, operating conditions, routine maintenance, periodic service, etc. The practical issue is the appropriate time period for collecting and analyzing SQC data. This time period depends on the particular operating conditions for an individual test or analyzer (e.g., how often runs are performed, how often operators change, how frequently maintenance is performed, how often reagent and calibrator lots change). Other factors to consider include the number of measurements needed to obtain a reliable SD estimate, frequency of SQC and the time over which control limit data are collected.

NUMBER OF MEASUREMENTS

The reliability of an SD estimate is characterized by the confidence limit of the estimate, which depends on the number of measurements. While the rule of thumb is a minimum of 20 control measurements to calculate an SD for control limits, many more are needed to obtain a reliable estimate of the SD.

Example:
Assuming a true standard deviation of 10 units, the 90% confidence interval will range from 7.4 to 15.9 when N = 20. That is, an SD as low as 7.4 could be observed, which is 26% low, or an SD as high as 15.9 could be observed, which is 59% high. For N = 100, the confidence interval is 9.0 to 11.3. That is, the reliability of the estimate of the SD is much better, within approximately 10% of the correct value. Therefore, at least 100 measurements are preferable for estimating MU.

SQC FREQUENCY

There is no standard practice for SQC frequency, but many labs, globally and in the U.S., tend to follow the CLIA guideline that a minimum of two levels of controls be analyzed per day. Of course, high-volume labs will often analyze many more controls per day. The new emerging practice of risk-based QC plans may lead to low frequency of SQC, particularly in point-of-care applications. Clearly, the practicality of estimating MU from SQC data will depend on having a sufficient number of control measurements to provide a reliable estimate of the SD. A reliable estimate of MU may not be obtainable for unit-use devices used in point-of-care applications, even though knowledge of the quality in these settings is critically important for patient treatment.

CUMULATIVE SDs

Given the difficulty of obtaining a reliable estimate of an SD, CLSI C24A3[8] recommends labs utilize several months of data to establish cumulative control limits.

Example:
If a lab analyzes two levels of controls per day, data for more than 100 days will be needed to provide reliable SDs at the two levels.

C24A3 recommends labs combine control data from six consecutive monthly periods, calculate a cumulative SD, and implement control limits based on that cumulative SD.

In summary, there is no specific guidance for how many control measurements are needed, but the estimate of the SD will be more reliable if at least 100 data points are used, which often requires SQC data from over several months. A period of six months is practical for many labs and matches the CLSI recommendation for establishing control limits from a cumulative SD obtained from six successive months of routine SQC data.

WHAT'S THE POINT?

Quality management is an ongoing, continuous process, as exemplified by Deming's Plan-Do-Check-Act cycle. "Doing the right SQC right" is a critical part of that process and is essential to verify the attainment of the intended quality of results, which is the ISO 15189 requirement for quality control. A right-sized SQC procedure should be part of any QC plan, including risk-based QC plans. Once a QC plan is implemented, it is essential to monitor the quality and performance of the testing process to identify failures and make improvements. EQA/PT programs provide important independent measures of quality and performance. Measurement uncertainty, a quality indicator required by ISO 15189, can be estimated from SQC results obtained over three to six months. Other quality indicators must be specified and implemented by the lab to ensure quality and patient safety.

References

1. CMS Memo of August 16, 2013: Individualized Quality Control Plan (IQCP): A New Quality Control (QC) Option. www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/Downloads/IQCPbenefits.pdf.

2. CLSI EP23A. *Laboratory Quality Control Based on Risk Management*. Clinical and Laboratory Standards Institute, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania, 2011.

3. Westgard JO, Westgard SA. *Basic Quality Management Systems*. Madison, Wisconsin, Westgard QC, 2014.

4. Ehrmeyer SS. *The New Poor Lab's Guide to the Regulations – 2015 edition*. Madison, Wisconsin, Westgard QC, 2015.

5. NGSP website, www.ngsp.org, accessed December 1, 2014.

6. Westgard JO, Westgard SA. Assessing quality on the Sigma scale from proficiency testing and external quality assessment surveys. *Clinical Chemistry and Laboratory Medicine* 2015;53:1531-1535.

7. ISO 15189:2012. *Medical laboratories – Requirements for quality and competence*. 3rd ed. International Organization for Standards, Geneva, Switzerland, 2012.

8. CLSI C24A3. *Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions*. Clinical and Laboratory Standards Institute, Wayne, Pennsylvania, 2006.

GLOSSARY OF TERMS, CONTROL RULES AND ABBREVIATIONS

These definitions, some official and others unofficial, are referenced to the following sources: International Standards Organization (ISO), Clinical and Laboratory Standards Institute (CLSI), U.S. Centers for Medicare and Medicaid Services (CMS), U.S. Centers for Disease Control and Prevention (CDC), and Westgard QC (WQC).

APPENDIX A: GLOSSARY OF TERMS

ACCURACY – Closeness of agreement between a test result and the accepted reference value (ISO 5725-1). Note: The term *accuracy*, when applied to a set of test results, involves a combination of random components (imprecision) and a common systematic error or bias component (ISO 5725-1).

ALLOWABLE TOTAL ERROR (ATE) – An analytical quality requirement that sets a limit for both the imprecision (random error) and bias (systematic error) that is tolerable in a single measurement or single test result. **Note: Also called total error allowable (TEa)** (CLSI EP21).

BIAS (OF MEASUREMENT) – Difference between the expectation of the test result or measurement results and a true value (ISO 3534-2). Note: Bias is an estimate of the systematic measurement error (JCGM 200:2012).

BIOLOGIC GOALS – Specifications for precision, bias and total error based on within subject biologic variation (CVi) and between subject biologic variation (CVg). Maximum allowable CV is typically calculated as 0.5 x CVi, maximum allowable bias as 0.25 x $(CVi^2 + CVg^2)^{1/2}$, and maximum biologic total error as 0.25 x $(CVi^2 + CVg^2)^{1/2}$ + 1.65 x 0.5 x CVi. (See the Ricos biologic database for tabulations of biologic variation and goals.)

COEFFICIENT OF VARIATION (CV) – For a control material, the standard deviation divided by the mean times 100 to express variation as a percentage.

COMMUTABLE – Interassay properties of a reference material, calibrator material or QC material that are comparable to those demonstrated by authentic clinical specimens. **Commutability of a material** is defined as the degree to which a material yields the same numerical relationships between results of measurements by a given set of measurement procedures, purporting to measure the same quantity as those between the expectations of the relationships obtained when the same procedures are applied to other relevant types of material (CLSI EP31A, ISO 15194).

CONTROL CHART – A graph that displays control results on the y-axis versus time or run number on the x-axis. The standard control chart in medical labs is called the Levey-Jennings chart, and examples are illustrated in the text of this guide.

CONTROL RULE – A decision criterion for interpreting control data and making a judgment on the control status of an analytical run. Symbolized by A_L, where A is the abbreviation for a particular statistic or the number of control measurements and L is the control limit. For example, 1_{3s} indicates that a run should be judged as out of control if one measurement exceeds a control limit set as the mean ± 3 SD (WQC).

CRITICAL SYSTEMATIC ERROR – The size of the systematic error that would cause a medically important error, as calculated from the allowable total error (TEa) and the observed precision (SD, CV) and accuracy (bias) of the method or measurement procedure (WQC).

DEFECT – A departure of a quality characteristic from its intended level or use that occurs with a severity sufficient to cause the product or service to **not** satisfy the intended use or customer requirement.

GLOSSARY OF TERMS

DEFECTS PER MILLION (DPM); DEFECTS PER MILLION OPPORTUNITIES (DPMO) – The number of defects per million units provided or per million opportunities of service.

DEMING'S PDCA CYCLE (PDCA) – Application of the scientific method to provide objective data-driven decisions through a Plan-Do-Check-Act process. Plan an experiment, Do the experiment, Check the results, and Act on that data. Select an examination procedure to satisfy requirements for intended use (Plan), implement the examination procedure and validate its performance (Do), monitor quality in routine production (Check), and identify problems and make improvements (Act) (WQC).

EXTERNAL QUALITY ASSESSMENT; EXTERNAL QUALITY ASSURANCE (EQA) – A surveillance activity where samples are submitted to the lab for testing, then their correctness is evaluated by the survey group. Program may be voluntary or required by regulations, in which case, it may be called proficiency testing (PT).

IMPRECISION – The random dispersion of a set of replicate measurements and/or values expressed quantitatively by a statistic, such as standard deviation or coefficient of variation (CLSI). IFCC has recommended that the mean value and number of replicates should also be stated and the experimental design described in such a way that other workers can repeat it. This is particularly important whenever a specific term is used to denote a particular type of imprecision, such as within-run, within-day, day-to-day, total or between labs.

INACCURACY – Numerical difference between the mean of a set of replicate measurements and the true value. This difference (positive or negative) may be expressed in the units in which the quantity is measured or as a percentage of the true value (IFCC). Commonly expressed as bias, which is the ISO measure of trueness.

INDIVIDUALIZED QUALITY CONTROL PLAN (IQCP) – An option for compliance with the quality control requirements in the U.S. CLIA regulations. An IQCP consists of a risk assessment to identify sources of error, a control plan to identify mechanisms for mitigating the risks of failure, and a quality assessment program to monitor performance and identify needs for improvement (U.S. CMS/CDC agencies).

MEAN – The arithmetic average of a set of values. A measure of central tendency of the distribution of a set of replicate results, often abbreviated by an x with a bar over it.

MEASURAND – Quantity intended to be measured (JCGM 200:2012, CLSI EP31A).

MEASUREMENT UNCERTAINTY (MU) – Non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used. Note 1: Measurement uncertainty includes components arising from systematic effects, such as components associated with corrections and the assigned quantity values of measurement standards, as well as the definitional uncertainty. Sometimes, estimated systematic effects are not corrected for, but instead, associated uncertainty components are incorporated. Note 2: The parameter may be, for example, a standard deviation (SD) called *standard measurement uncertainty* (or a specified multiple of it) of the half-width of an interval, having a stated coverage probability (CLSI C51A).

GLOSSARY OF TERMS

METHOD DECISION CHART – A graphical tool that describes the allowable bias on the y-axis versus the allowable precision (CV) for a defined allowable analytic total error (ATE, TEa). Lines define zones for Sigma quality, from Six Sigma to two Sigma. The observed bias and precision of an examination procedure can be plotted as an operating point to assess Sigma quality (WQC).

NUMBER OF CONTROL MEASUREMENTS (N) – Used here to indicate the total number of control measurements available for assessing the quality of an analytical run. These measurements may be replicates of one level or material, individual measurements of two or more materials, or replicate measurements of two or more materials. For example, N = 4 could represent four measurements of a single control material or two replicates of each of two different control materials (WQC).

OPERATING POINT – Used here to describe a point whose y-coordinate represents the bias of a method and whose x-coordinate represents the precision (SD, CV) that is plotted on a method decision chart or a chart of operating specifications (WQC).

OPSPECS CHART – A graphical tool that shows the bias (on the y-axis) and precision (on the x-axis) that are allowable for different SQC procedures, having a stated level of error detection and a defined quality requirement. The observed bias and precision (SD, CV) for an examination procedure can be plotted as an operating point to select appropriate control rules and numbers of control measurements (WQC).

POWER FUNCTION GRAPH – A plot of the probability for rejection versus the size of errors for an SQC procedure (that is, for specified control rules or decision criteria and the specified number of control measurements) (WQC).

PRECISION (MEASUREMENT) – Closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions (JCGM200:2012). Note: Measurement precision is usually expressed numerically by measures of imprecision, such as standard deviation, variance or coefficient of variation, under the specified conditions of measurement (JCGM 200:2012).

PROBABILITY FOR ERROR DETECTION (P_{ED}) – A performance characteristic of an SQC procedure that describes how often an analytical run will be rejected when the test results contain errors in addition to the inherent precision of the examination procedure. Ideally, P_{ed} should be 1.00 for errors that are medically important. In practice, a P_{ed} of 0.90, or 90% detection, is often used in selecting and designing SQC procedures (WQC).

PROBABILITY FOR FALSE REJECTION (P_{FR}) – A performance characteristic of an SQC procedure that describes how often an analytical run is rejected when there are no errors occurring, except for the inherent precision of the examination procedure. Ideally, P_{fr} should be 0.00. In practice, values less than 0.05 or 0.01 (5% or 1% false rejections) are used in selecting and designing SQC procedures (WQC).

PROFICIENCY TESTING (PT) – Used in the U.S. to describe a program of external quality assessment, whereby specimens are submitted to a lab for analysis, with the purpose of grading the performance of the lab for regulatory purposes.

GLOSSARY OF TERMS

QUALITY – Degree to which a set of inherent characteristics fulfills requirements (ISO/CLSI).

QUALITY CONTROL (QC) – Part of quality management, focused on fulfilling quality requirements. Note 1: In healthcare testing, the set of procedures designed to monitor the test method and the results to ensure appropriate test system performance. Note 2: The purpose of quality control is to ensure that all quality requirements are being met. Note 3: The set of mechanisms, processes and procedures designed to monitor the measuring system to ensure the results are reliable for the intended clinical use (ISO/CLSI).

QUALITY CONTROL PLAN (QC PLAN) – A document that describes the practices, resources and sequences of specified activities to control the quality of a particular measuring system or test process to ensure the requirements for its intended purpose are met (CLSI EP23A).

QUALITY GOAL – A general term that describes a requirement for quality. Other terms used are quality specifications and quality requirements. For analytical measurement processes, quality requirements are commonly defined in terms of an allowable total error, allowable bias or allowable standard deviation.

QUALITY INDICATOR – Measurement (metric) to monitor specific activities as part of the quality management system (CLSI GP35).

QUALITY MANAGEMENT – Coordinated activities to direct and control an organization with regard to quality. Note (CLSI GP29): Direction and control with regard to quality usually include establishment of the quality policy and quality objectives, quality planning, quality control, quality assurance, and quality improvement (ISO/CLSI).

QUALITY MANAGEMENT SYSTEM (QMS) – Management system to direct and control an organization with regard to quality. Note 1: Systematic and process-oriented efforts are essential to meet quality objectives. Note 2: For purposes of ISO 15189, the quality referred to in this definition relates to matters of both management and technical competence. Note 3: A quality management system typically includes the organizational structure, resources, processes and procedures needed to implement quality management. Note 4: These principles include the following categories: documents and records, organization, personnel, equipment, purchasing and inventory, process management, information management, nonconforming event management, assessment, continual improvement, customer focus, and facilities and safety (ISO 15189).

RICOS BIOLOGIC DATABASE – Refers to a tabulation of biologic variation and calculated biologic goals provided by Dr. Carmen Ricos and a group of Spanish clinical chemists. Originally published in *Scandinavian Journal of Clinical and Laboratory Investigation* 1999;59:491-500. (Updates are available at www.westgard.com.)

RISK – Combination of the probability of occurrence of harm and the severity of that harm (ISO/IEC Guide 51, CLSI EP23A).

GLOSSARY OF TERMS

RISK ANALYSIS – Systematic use of available information to identify hazards and to estimate the risk (ISO/IEC Guide 51). Note: Risk analysis includes examination of different sequences of events that can produce hazardous situations and harm (ISO 15189, CLSI EP23A).

RISK ASSESSMENT – Overall process comprising a risk analysis and a risk evaluation (ISO/IEC Guide 51, CLSI EP23A).

RISK ESTIMATION – Process used to assign values to the probability of occurrence of harm and the severity of that harm (ISO 14971, CLSI EP23A).

RISK EVALUATION – Process of comparing the estimated risk against given risk criteria to determine the acceptability of risk (ISO 14971, CLSI EP23A).

RISK MANAGEMENT – Systematic application of management policies, procedures and practices to the tasks of analyzing, evaluating, controlling and monitoring risk (ISO 14971, CLSI EP23A).

SIGMA-METRIC – Calculated here as (TEa – Bias)/SD, where all terms are in concentration units, or (%TEa – %Bias)/%CV, when all terms are percentages. TEa is the allowable analytic total error, bias is the observed systematic error, and SD (or CV) is the observed random error or precision of an examination procedure (WQC).

SIGMA PROFICIENCY ASSESSMENT CHART – A two-sided method decision chart used for EQA or PT results to assess quality on the Sigma scale (WQC).

SIGMA SQC SELECTION TOOL – A graphical display of the probabilities of rejection for different control rules and different numbers of control measurements on the y-axis versus the size of a medically important systematic error or the Sigma-metric of an examination procedure on the x-axis. Visual inspection allows selection of SQC procedures with the desired rejection characteristics (WQC, CLSI C24A3).

SIX SIGMA QUALITY MANAGEMENT SYSTEM (6ΣQMS) – Application of Six Sigma concepts, tools and metrics, and Deming's Plan-Do-Check-Act cycle to provide objective and quantitative management of the analytical quality of an examination process (WQC).

STANDARD DEVIATION (SD) – Statistic that describes the dispersion or spread of a set of measurements about the mean value of a Gaussian or normal distribution.

STATISTICAL QUALITY CONTROL (SQC) – Procedure that involves the analysis of stable materials and comparison of measurement results with the expected distribution of results under stable operating conditions. Control limits are typically calculated from the mean and standard deviation observed during an initial period of stable operation. Control rules and the number of control measurements should be selected to identify analytic runs that have medically significant errors. Control results are typically displayed graphically by plotting the observed control measurement sequentially versus time, run or day (WQC).

GLOSSARY OF TERMS

TOTAL ANALYTICAL ERROR (TAE) – Defines the interval that contains a specified proportion (usually 95% or 99%) of the distribution of analytical measurement differences between a measurement procedure operating in its stable in-control state and a comparative measurement procedure that is either a definitive reference method or one that is traceable to one (CLSI EP21). Also commonly abbreviated as TE.

TOTAL ERROR – Includes all random and systematic errors that can occur during the total testing process, as well as the combined effect of all precision and bias errors that can affect the accuracy of an analytical result. Note: Total error incorporates error sources from the pre-analytical, analytical and post-analytical phases of a measurement procedure (CLSI EP21).

TOTAL ERROR ALLOWABLE (TE_a) – See Allowable Total Error.

TOTAL QUALITY CONTROL PLAN – A quality control plan that is first optimized to provide a right-sized SQC procedure that will detect medically important systematic errors and then expanded to include pre-examination and post-examination controls to monitor the total examination process (WQC).

TOTAL QUALITY CONTROL STRATEGY – Used here to describe the balance between SQC, other controls and quality improvement appropriate for the Sigma quality of an examination procedure (WQC).

TOTAL TESTING PROCESS (TTP); TOTAL EXAMINATION PROCESS – Includes the pre-analytic, analytic or post-analytic phases of the testing process. In ISO language, these are referred to as pre-examination, examination and post-examination phases of the examination process.

TRACEABILITY – (Metrological) property of measurement results, where the results can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty (JCGM 200:2012).

TRUENESS (MEASUREMENT) – Closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value (JCGM 200:2012). Note: Trueness is expressed numerically using the observed bias (ISO/CLSI).

UNCERTAINTY OF MEASUREMENT – Non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used. Note: The parameter may be, for example, a standard deviation (SD) called standard measurement uncertainty (or a specified multiple of it) or the half-width of an interval having a stated coverage probability (CLSI C51).

VALIDATION – Confirmation, through the provision of objective evidence, that the requirements for a specific intended use of application have been fulfilled (ISO 15189).

VERIFICATION – Confirmation, through the provision of objective evidence, that specified requirements have been fulfilled (ISO 15189).

WESTGARD SIGMA RULES – Control rules and the total number of control measurements that are selected on the basis of a Sigma-metric that relates the quality defined by a total allowable analytic error (TEa, ATE) with the precision (SD or CV) and bias observed for an examination procedure (WQC).

APPENDIX B: CONTROL RULE DEFINITIONS (WQC)

1_{2S} – Refers to the control rule commonly used with a Levey-Jennings chart, where control limits are set as the mean ± 2s. This rule is sometimes used as a rejection rule, in which case, there often are problems with false rejections (5% for N = 1, 10% for N = 2). In multirule SQC procedures, this rule is used as a warning rule to trigger careful inspection of the control data by other rejection rules.

1_{3S} – Reject when one control measurement exceeds the mean ± 3s.

$1_{2.5S}$ – Reject when one control measurement exceeds the mean ± 2.5s control limits.

2_{2S} – Reject when two consecutive control measurements exceed the same mean + 2s control limit or the same mean – 2s control limit.

2 OF 3_{2S} – Reject when two out of three control measurements exceed the same mean + 2s or mean – 2s control limit.

R_{4S} – Reject when one control measurement in a group exceeds the mean + 2s control limit and another exceeds the mean – 2s control limit. (Note: This rule is best applied within a single rule.)

3_{1S} – Reject when three consecutive control measurements exceed the same mean + 1s or the same mean – 1s control limit.

4_{1S} – Reject when four consecutive control measurements exceed the same mean + 1s or the same mean – 1s control limit.

6X – Reject when six consecutive control measurements fall on one side of the mean.

8X – Reject when eight consecutive control measurements fall on one side of the mean.

9X – Reject when nine consecutive control measurements fall on one side of the mean.

10X – Reject when 10 consecutive control measurements fall on one side of the mean.

APPENDIX C: ABBREVIATIONS

6ΣQMS – Six Sigma quality management system

AON – Average of normals

ATE – Allowable total error (also TEa)

BIPM – International Bureau of Weights and Measures

CAP – College of American Pathologists

CDC – U.S. Centers for Disease Control

CIPM – International Committee of Weights and Measures

CLIA – U.S. Clinical Laboratory Improvement Amendments

CLSI – Clinical and Laboratory Standards Institute

CMS – U.S. Centers for Medicare and Medicaid Services

CV – Coefficient of variation

CVB – Overall or total biologic variation

CVG – Between-individual biologic coefficient of variation

CVI – Within-individual biologic coefficient of variation

DPM – Defects per million

DPMO – Defects per million opportunities

EFLM – European Federation for Laboratory Medicine

EQA – External quality assessment

FMEA – Failure modes and effects analysis

FRACAS – Failure reporting and corrective action system

GUM – Guide for Estimation of Uncertainty of Measurements

IQCP – Individualized quality control plan

ISO – International Standards Organization, or International Organization for Standardization (IOS)

JCTLM – Joint Committee for Traceability in Laboratory Medicine

MDL – Medical decision levels

MU – Measurement uncertainty

N – Total number of control measurements for assessment in SQC

NGSP – U.S. National Glycohemoglobin Standardization Program

OPSPECS – Operating specifications

PDCA – Plan-Do-Check-Act cycle or process

POC – Point of care

PT – Proficiency testing

QC – Quality control

QMS – Quality management system

R – Number of runs over which control rules are applied

RPT – Repeat patient test

SD – Standard deviation

SOP – Standard operating procedure

SQC – Statistical quality control

TAE – Total analytical error (also TE)

TEa – Total error allowable (also ATE)

TQC – Total quality control

TQM – Total quality management

TTP – Total testing process

CORELABORATORY.ABBOTT

Abbott