

Quality Control

How Labs Can Apply Six Sigma Principles To Quality Control Planning

BY JAMES O. WESTGARD, PHD

Since the publication of the Centers for Medicare and Medicaid Services' (CMS) Clinical Laboratory Improvement Amendments (CLIA) final rules in January 2003, many in the clinical laboratory community have questioned the amount of quality control (QC) needed in labs today. Based on the improved performance of new analytical systems, some laboratorians might be misled into thinking that less QC is required for such devices. However, in order to determine the appropriate amount of QC, the question that needs to be answered is: How does an instrument's analytical performance relate quantitatively to QC?

The emergence of Six Sigma Quality Management as a laboratory quality management tool provides the framework for addressing the issues surrounding what has become known as equivalent QC (EQC). While much has been written about the complexities of Six Sigma, laboratorians can actually use a simple calculation to determine a sigma metric that characterizes the level of quality required for a test and the precision and accuracy observed for the measurement procedure. This metric can then be related to the rejection characteristics of QC procedures in order to select the appropriate control rules and the number of control measurements for an individual lab test.

This article will both describe the relationship of Six Sigma to QC, as well as provide a practical QC planning tool that will allow laboratorians to apply this approach in their own labs.

Six Sigma Methodology

The first step to setting quality standards is defining tolerance limits, or quality requirements, that both describe good quality and identify poor quality or defective products

For analytic processes, the tolerance limits can be defined in the form of allowable total error (TE_a), such as those stated in the CLIA requirements for acceptable performance in proficiency testing events. For

observed for the measurement procedure. Figure 1A illustrates this calculation for a measurement procedure where TE_a is 10%, bias_{obs} is 1.0%, and s_{obs} is 2.0%, which gives a sigma metric of 4.5. This calculated sigma metric is predictive of process quality if and only if the measurement procedure remains stable, implying that the measurement procedure itself must be properly controlled, otherwise errors would not be detected.

With today's highly automated analytical systems, laboratorians are mainly concerned with systematic errors because random errors are well controlled by the operation of the system. An important point to remember, however, is that changes in calibrators, reagents, parts, or people may introduce systematic errors.

Figure 1B shows the size of a systematic error that must be detected in order to limit the risk of producing an erroneous result to 5%. This critical systematic error (ΔSE_{crit}) can be calculated using the equation $\Delta SE_{crit} = [(TE_a - bias_{obs})/s_{obs}] - 1.65$, where 1.65 is the z-value corresponding to an area of 0.05 in the tail exceeding the quality requirement. The important point here is that the size of the error that must be detected depends directly on the expression $[(TE_a - bias_{obs})/s_{obs}]$, which is the sigma metric for the measurement procedure. In simple terms, a higher sigma metric means that the systematic error that must be detected is larger, and therefore it is easier for the QC procedure to do its job. On the other hand, a lower sigma metric means that smaller systematic errors must be detected, and it is more difficult for the error to be detected by the QC procedure.

Performance Characteristics of QC Procedures

QC can be thought of as an error detector, much like a smoke detector in a fire alarm system. As with any alarm system, two conditions are of interest. The first characteristic has to do with false alarms or false rejections, and the second deals with true alarms or error detection.

or services. For pre-analytic and post-analytic test processes, the outcome can be inspected to count the defects, expressed in terms of defects per million (DPM), and this result can then be converted to a process sigma using a standard statistical table, commonly referred to as the area in the tail of a normal distribution. The goal when establishing quality standards is to achieve six-sigma quality, which has become known as the world class level of quality. However, performance at the three-sigma level is the minimum acceptable quality for a production process.

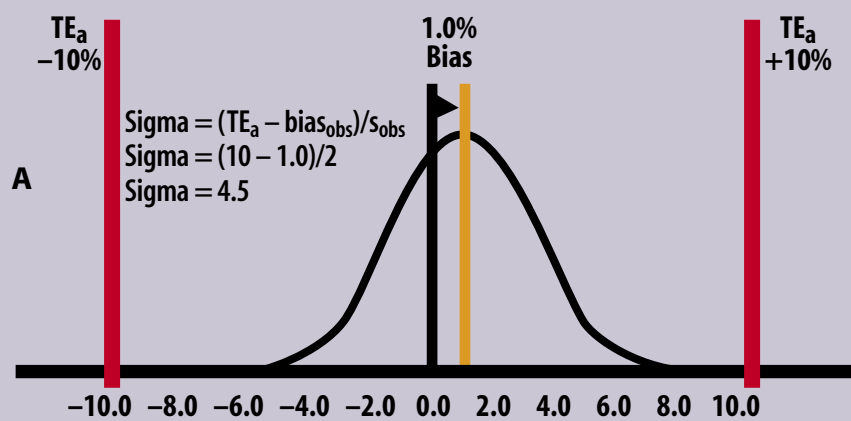
U.S. labs, these requirements are given in the *Federal Register* for some 70–80 tests. In other countries, External Quality Assessment (EQA) programs provide similar information. Alternatively, an allowable biologic total error can be calculated from the extensive databank available on biologic goals.

A predictive measure of analytical quality on the sigma scale can be calculated from the simple equation, sigma metric = $[(TE_a - bias_{obs})/s_{obs}]$, where TE_a represents the tolerance interval in the form of an allowable total error, and bias_{obs} and s_{obs} represent the accuracy and precision, respectively, that are

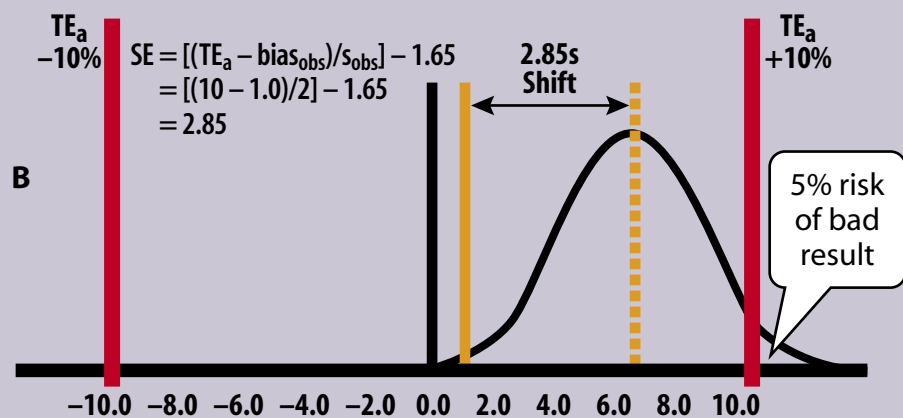


Figure 1

Calculation of Method Sigma



Calculation of Critical Systematic Error



False alarms or false rejections. Obviously, keeping false alarms to a minimum is critical, otherwise, people waste time responding to them. Of even greater concern, however, is the likelihood that people will become conditioned to false alarms and will change the way they respond to all alarms. For example, think about what you would do if the fire alarm went off right now—you'd probably follow the procedure and leave the building. But what would you do if the alarm went off again after you returned? Are you still going to leave the building, or has the experience with a false alarm modified your behavior?

Now, consider the normal behavior in many laboratories where two standard deviations define the control limits and the false rejection rate is approximately 10% when two controls are analyzed in a run. The response to an out-of-control signal is generally to repeat the controls again and again as many times as is needed to make it right. This type of response to a QC alarm can totally corrupt the lab's QC process, because when true alarms occur, the actual problems never get fixed.

True alarms or error detection. It should come as no surprise to laboratorians that QC procedures don't automatically detect all errors. Again, using the smoke detector analogy, the chance for detection depends on the size of the fire. For example, if you light a single match, the fire alarm probably would not go off. But if you were to throw that match into a waste basket full of paper, then the fire is much more likely to set off the fire alarm. The larger the fire, the greater the chance it will be detected. The same is true for analytical errors.

The error detection capabilities of a QC procedure can be described by its probabilities of rejecting runs with different sizes of errors. This capability is also known as statistical power, and this response function can be called a power curve or a power function. Such information is available in clinical chemistry textbooks for most commonly used QC procedures.

To illustrate this point, Figure 2 shows a power function graph that describes the

performance of two different QC procedures. The y-axis shows the probability for rejection, and the x-axis shows the size of the systematic error. The higher curve represents a multi-rule procedure with a total of four control measurements per run ($N = 4QC$), which would commonly be achieved by analyzing two levels of control in duplicate. The lower curve represents a single-rule procedure with a total of two control measurements per run ($N = 2QC$), which would commonly be achieved by performing single measurements on each of two different materials to comply with the CLIA requirement to analyze a minimum of two levels of control per day. As expected, the figure shows that more errors are detected with the higher number of control measurements. In the example method, which has 4.5 sigma quality, or a critical systematic error of 2.85s, the probability for error detection (P_{ed}) is 0.70 (or a 70% chance) for the $N = 2QC$ procedure and 0.98 (or a 98% chance) for the $N = 4QC$ procedure.

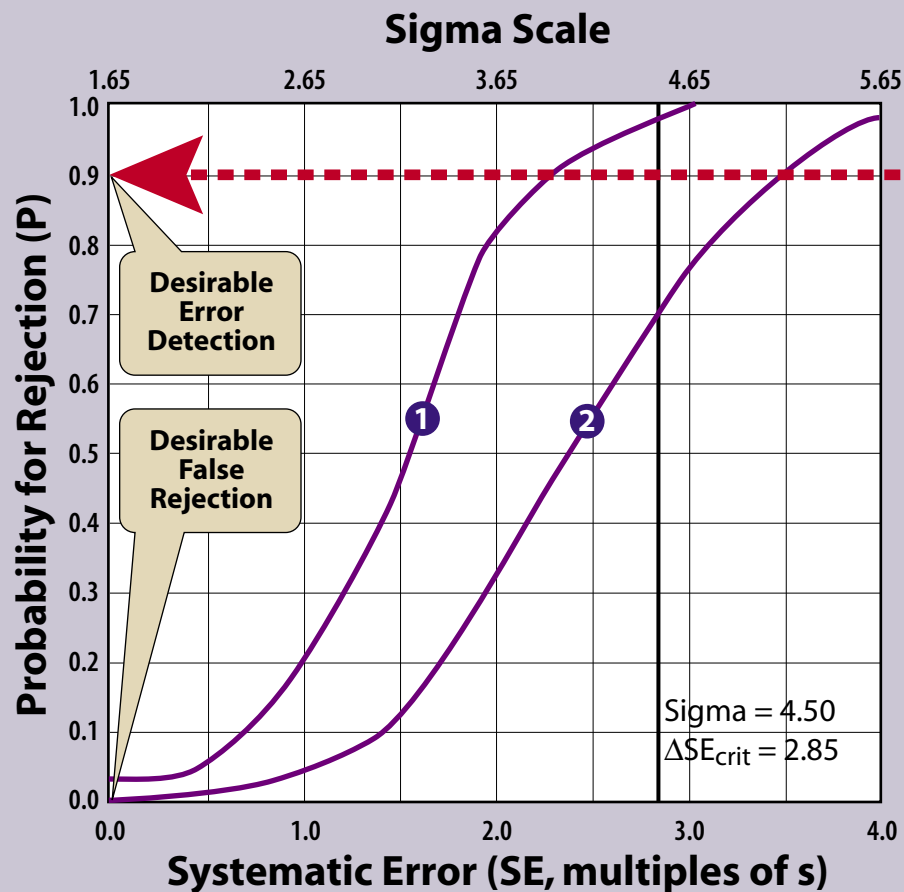
For labs, practical goals for the performance of a QC procedure could be set as P_{ed} of 0.90 (a 90% error detection rate) and a P_{fr} of ≤ 0.05 (a 5% false rejection rate, with lower being more desirable). The goal for error detection, as represented by the horizontal line in Figure 2, shows that a method with approximately 5.2 sigma could be adequately controlled by the $N = 2QC$ procedure and that a method with 4.0 sigma could be controlled by the $N = 4QC$ procedure. The probability of false rejections (P_{fr}) is shown by the y-intercepts of the power curves, which are 0.03, or 3%, for the $N = 4QC$ procedure and 0.00, or 0%, for the $N = 2QC$ procedure. In this case, both of the QC procedures provide acceptably low false rejections. Keep in mind that the 2QC procedure here uses 3s control limits, not 2s control limits, where the false rejections would approach 10%, which would clearly be too high.

A Simple QC Planning Tool

To help guide laboratorians in QC planning, Figure 3 illustrates eight different QC

Figure 2

Power Function Graph for QC Procedures



QC Procedure	QC Rule	P_{fr}	P_{ed}	N	R
1	$1_{3s}/2_{2s}/R_{4s}/4_{1s}$	0.03	0.98	4	1
2	1_{3s}	0.00	0.70	2	1

Probability for rejection is shown on the y-axis vs. the size of the systematic error on the x-axis (bottom scale) or the sigma-metric of the measurement procedure (top scale). The vertical line represents the specific method of interest in this example. The key above identifies the QC procedure. P_{fr} =probability for false rejection; N=total number of control measurements; R=number of runs to which the QC procedure is applied.

procedures. QC procedure 1 is a multi-rule QC procedure with $N = 6$, which could be achieved by analyzing three levels of controls, each in duplicate. QC procedure 2 is a multi-rule QC procedure with $N = 4$ and $R = 2$, which could be achieved by analyzing two levels in duplicate with the control rules applied across two runs. The third QC procedure uses similar rules and the same N, but operates over one run. The difference between QC procedures 2 and 3 is the control rules; the second procedure includes the use of an "8x rule" and the third does not. The 8x rule looks at the four control measurements in the current run and the previous run (that's the meaning of the $R = 2$ notation in the key), giving a total of eight when "looking back" to the previous run. A key benefit of this QC procedure is that the control rule "looks back" across control data from consecutive runs.

The fourth QC procedure is a single-rule procedure using 2.5s control limits with $N = 4$. The performance observed is nearly the same as for the multi-rule procedure immediately above—in fact, the two power curves cross near the middle. For practical purposes, the choice between these two depends on the ease of implementation in the laboratory. The advantage of the multi-rule procedure is potential improvement in performance when an 8x look-back rule is added.

QC procedures 5–8 all have $N = 2$. For a 6-sigma process, which actually would be slightly off-scale to the right on the graph, the projection of the power curve for the $1_{3.5s}$ rule shows that the desired error detection rate can be achieved with a very low level of false rejections. The y-intercept (P_{fr}) shows that the probability for false rejections

would be very low, actually 0. For a 5-sigma process, the 1_{3s} rule will provide a P_{ed} of approximately 0.87, whereas the multi-rule procedure has a P_{ed} of 0.93, and the $1_{2.5s}$ procedure has a P_{ed} of 0.95. In choosing between these rules, laboratorians should also consider their respective probabilities for false rejection, which are 0.00, 0.01, and 0.03, respectively. In practice, all three rules would be acceptable for a 5-sigma process, but if sigma were 5.2 or above, then the 3s procedure would be preferable. If sigma were a little less, for example 4.8, then the multi-rule or 2.5s single-rule procedure would be preferred.

As sigma gets lower, between 4.8 and 4.0, then the $N = 4$ QC procedures become more appropriate. When sigma is less than 4.0, providing the appropriate QC becomes much more difficult. For a sigma of 3.5, it would be productive to use the $N = 4$ multi-rule procedure with the additional 8x look-back rule. For a sigma of 3.3, the $N = 6$ multi-rule procedure provides the appropriate error detection. For a sigma of 3.0, error detection is no longer ideal since P_{ed} drops to about 0.77.

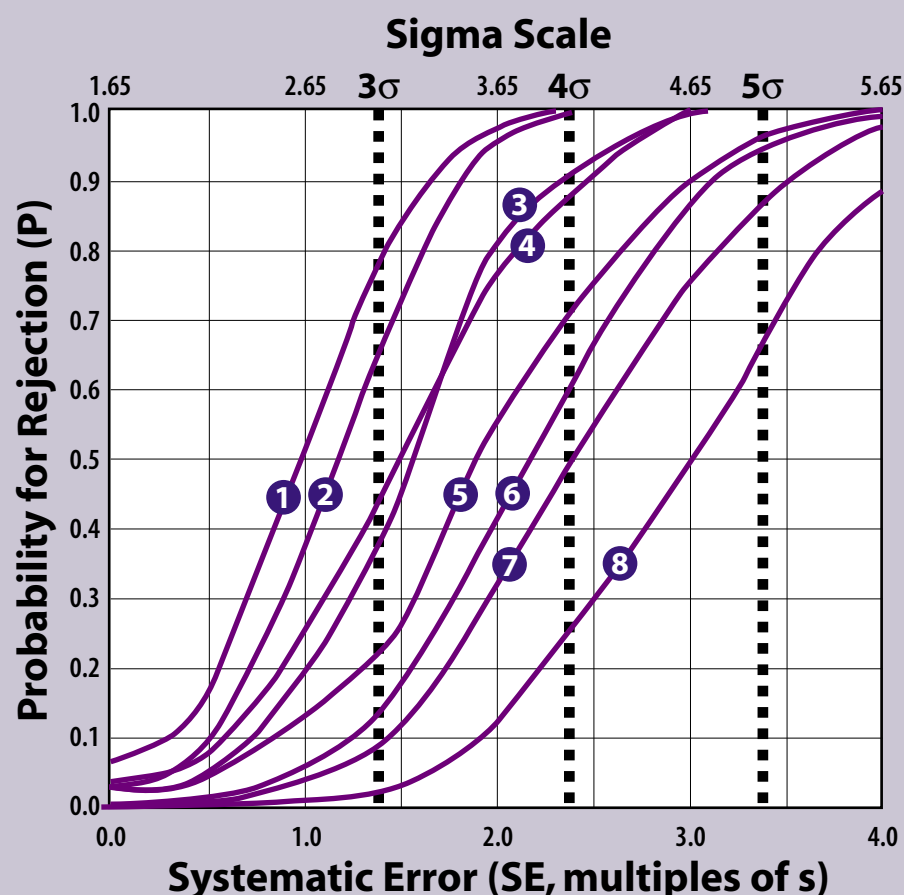
Some Specific QC Examples

Laboratorians may find some general guidelines for QC helpful (see box, p. 12); however, the following examples provide some specific recommendations for common analytes.

Potassium. Given that the TE_a according to CLIA is 0.5 mmol/L, which is 10% at a decision level of 5.0 mmol/L, if the lab method has a coefficient of variation (CV) of 1.0% and a bias of 2.0%, then the sigma metric is 8.0 $[(10-2)/1]$. Therefore, labs should set 3.5s control limits with $N=2$.

Figure 3

Planning Tool for Selecting QC Procedure



QC Procedure	QC Rule	P_{fr}	N	R
1	$1_{3s}/2$ of $3_{2s}/R_{4s}/3_{1s}/6_x$	0.07	6	1
2	$1_{3s}/2_{2s}/R_{4s}/4_{1s}/8_x$	0.03	4	2
3	$1_{3s}/2_{2s}/R_{4s}/4_{1s}$	0.03	4	1
4	$1_{2.5s}$	0.04	4	1
5	$1_{2.5s}$	0.03	2	1
6	$1_{3s}/2_{2s}/R_{4s}$	0.01	2	1
7	1_{3s}	0.00	2	1
8	$1_{3.5s}$	0.00	2	1

Probability for rejection is shown on the y-axis vs. the size of the systematic error on the x-axis (bottom scale) or the sigma-metric of the measurement procedure (top scale). The vertical lines represent measurement procedures having 3-sigma, 4-sigma, and 5-sigma performance. The key above identifies the QC procedures. P_{fr} =probability for false rejection; N=total number of control measurements; R=number of runs to which the QC procedure is applied.

Triglycerides. Given that the CLIA TE_a is 25%, if the lab method has a CV of 3.0% and a bias of 7.0%, the sigma metric is 6.0 $[(25-7)/3]$. Therefore, labs should set 3.5s control limits with $N=2$.

Glucose. Given that the CLIA TE_a is 10% at a decision level of 125 mg/dL, if the lab method has a CV of 2.0% and a bias of 0.0%, the sigma metric is 5.0 $[(10-0)/2]$. Therefore, labs should set control limits of 2.5s or 3.0s with $N=2$.

Calcium. Given that the CLIA TE_a is 1.0 mg/dL, or 10% at a decision level of 10.0 mg/dL, if the lab method has a CV of 2.0% and a bias of 2.0%, the sigma metric is 4.0 $[(10-2)/2]$. Therefore, labs should use a 2.5s single rule or multi-rule with $N=4$. Note that if the bias could be eliminated, the method would have a sigma of 5.0, which would allow the use of 2.5s or 3.0s control limits with $N=2$.

Cholesterol. Given that the CLIA TE_a is 10% at a decision level of 200 mg/dL if the lab method has a CV of 3.0% and bias of 0.0%, the sigma metric is 3.3 $[(10-0)/3]$. Therefore, labs should use a multi-rule QC procedure with $N=6$.

Sodium. Given that the CLIA TE_a is 4.0 mmol/L, which is 2.9% at a decision level of 140 mmol/L, if the lab method has a CV of 1.0% and a bias of 0.0%, the sigma metric is 2.9 $[(2.9-0)/1]$. In this case, labs should use the maximum number of QC runs that they

can afford. Note that it would be necessary to achieve a CV of about 0.6%, which would give a sigma-metric of 4.8, for $N=2$ QC procedures to assure the desired quality is being achieved.

Total QC Strategies

The analytes above were chosen to illustrate some of the difficulties laboratorians encounter when implementing various QC procedures. For example, what's the best way to deal with the fact that potassium is easy to control and sodium is almost impossible to control? A similar situation exists for cholesterol and triglycerides. And because many individual tests are often performed on chemistry instruments that measure multiple analytes, most labs will run the same number of controls for each analyte despite the different QC needs described above. Here's where we need a broader approach that recognizes statistical QC as part of the overall or Total QC strategy.

For example, for multitest systems, you will need to fix the size of N, and then try to vary the control rules to adjust the error detection capabilities. Your QC software needs to provide the flexibility to implement different rules for different tests. Next you need to pay attention to preventive maintenance, instrument function checks, and other quality checks that can be focused on the lower sigma methods. Your technical skills and knowledge

of the analytic system will be invaluable for defining and implementing the Total QC plan.

Getting Started

The QC methods discussed here are not as difficult as they may seem at first glance. To get started, labs should pick a single analytical system and make an assessment of the sigma metrics for all of the tests on that system. Begin by using CLIA acceptability criteria as the quality requirements. To get an estimate of the precision of the methods, use the daily QC data. Peer-comparison data is useful to estimate bias, but another option is simply assuming the bias to be zero and then updating the assessment when estimates of bias become important and/or are available. Identifying the high sigma methods—where minimum QC is sufficient—is important, as is identifying the worst case methods that require maximum QC and special attention. Finally, for those methods that have sigma values that are neither high nor low, optimize the QC rules and/or Ns as needed.

Some Final Words

With today's focus on patient safety and quality in health care, labs have put increased focus on QC and new management tools such as Six Sigma Quality Management. And while the CLIA EQC procedures may have created confusion about how labs should conduct QC, the QC planning tool described here is simple and easy to use and will also provide the appropriate guidance and directions that labs need in order to ensure that test results are accurate. For laboratories to reduce daily QC to weekly or even monthly QC as suggested by CMS guidelines for EQC, methods should perform at the 6 sigma level or better. The use of sigma metrics, as described here, will provide a much more objective and quantitative evaluation of the applicability of EQC than the evaluation protocols recommend by CMS in the *State Operations Manual*.

I hope that someday when the laboratory community looks back at the release of the final CLIA rules, we view them as the rebirth of efforts to do the right QC, instead of a reduction in the amount of QC that labs are required to perform. When all is said and

done, however, laboratorians must always use their technical skills and knowledge of analytic systems to define and implement appropriate total QC management. CLN

RECOMMENDED READINGS

Westgard JO. Internal quality control: planning and implementation strategies. *Ann Clin Biochem* 2003; 40: 593–611. www.westgard.com/IQCpaper2003.html

Westgard JO. CLIA Final Rules: Appropriate QC Procedures. www.westgard.com/CLIAfinalrule9.htm

Westgard JO. Six Sigma Quality Design and Control. www.westgard.com/lesson67.htm

Westgard JO. Formulating a Total Quality Control Strategy. www.westgard.com/lesson55.htm

Westgard JO. EZ Rules3: Solutions for the taxing issue of QC. www.westgard.com/essay80.htm



James O. Westgard, PhD, is a Professor in the Department of Pathology and Laboratory Medicine at the University of Wisconsin Medical School

(Madison, Wisc.) and Faculty Director of Quality Management Services for the Clinical Laboratories at the University of Wisconsin Hospital and Clinics. He teaches in the Clinical Laboratory Science Program and is co-director of a capstone program that offers a Graduate Certificate in Laboratory Quality Management via the Internet. He is also President of Westgard QC, Inc., a small business that provides education and training for quality management (www.westgard.com). Email: jowestgard@mailplus.wisc.edu or james@westgard.com.

Six Sigma and Traditional QC Some Rules of Thumb

Performance	QC Recommendations
6 Sigma	Use $N=2$ with 3.0s or 3.5s control limits. It's important to minimize false rejections, so be sure to avoid 2s control limits.
5 Sigma	Use $N=2$ and 2.5s or 3.0s control limits. The preference is for 3.0s when above 5.0 sigma and 2.5s when below 5.0 sigma.
4 Sigma	Use $N=4$ with multi-rules or 2.5s single rule. Labs should also consider the advantage of multi-rule with additional look-back rule to maximize error detection.
< 4 Sigma	Use the maximum QC that is affordable. Also maximize preventive maintenance, individual instrument, and function checks, and deploy the most experienced analysts to perform these tests.

This article is available as an 8 1/2" x 11" reprint on the AACC Web site (www.aacc.org).

Click on "Clinical Laboratory News," then "Series Articles."